Valuing Alzheimer's Disease drugs

Citation for published version:
https://doi.org/10.1017/S0266462320000574

Digital Object Identifier (DOI):
10.1017/S0266462320000574

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
International journal of technology assessment in health care

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Valuing Alzheimer's Disease drugs: A health technology assessment perspective on outcomes

Objectives
Due to the nature of the condition, health technology assessment agencies might face considerable challenges in choosing appropriate outcome measures for Alzheimer's disease drugs. To analyse which outcome measures were used in past health technology assessments in three European countries: England, Germany, and The Netherlands; to explore possible reasons for prioritisations, and derive potential implications for future assessments of Alzheimer's disease drugs.

Method
We conducted a literature review of studies that analysed decisions made in health technology assessments (across disease areas) in the three European countries. We then conducted case studies of technology assessments conducted for Alzheimer's disease drugs in these countries.

Results
Overall, outcomes measured using clinical scales dominated decisions or recommendations about whether to fund Alzheimer's disease drugs, or price negotiations. Health technology assessment processes did not always allow the inclusion of outcomes relevant to people with Alzheimer's disease, their carers and families. Processes did not include...
| early discussion and agreement on what would constitute appropriate outcome measures and cut-off points for effects.  
Conclusions  
To facilitate consistent and timely decisions about the value of new Alzheimer’s disease drugs early agreement with various stakeholders about outcomes, outcome measures and cut-offs is important to ensure that future AD drugs are appropriately valued. |
Title

Valuing Alzheimer's Disease drugs: A health technology assessment perspective on outcomes

Running title

Valuing Alzheimer's Disease drugs

Author details

Annette Bauer, MSc, Care Policy and Evaluation Centre, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom.

Raphael Wittenberg, MSc, Care Policy and Evaluation Centre, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom.

Amanda Ly, MSc, Centre for Medical Informatics, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Nine Edinburgh BioQuarter, 9 Little France Road, Edinburgh, EH16 4UX, United Kingdom.

Anders Gustavsson, PhD, Quantify Research, Hantverkargatan 8, 112 21 Stockholm, Sweden; Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Alfred Nobels allé 23, 141 83 Stockholm, Sweden.

Christin Bexelius, PhD, ROCHE. F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, 4070 Basel, Switzerland.

Claire Tochel, PhD, Centre for Medical Informatics, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Nine Edinburgh BioQuarter, 9 Little France Road, Edinburgh, EH16 4UX, United Kingdom.
Martin Knapp, Professor in Health and Care Policy, Care Policy and Evaluation Centre, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom.

Mia Nelson, PhD, Centre for Medical Informatics, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, 9 Little France Road, Edinburgh EH16 4UX, United Kingdom.

Catherine Sudlow, Professor of Neurology and Clinical Epidemiology, Centre for Medical Informatics, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, 9 Little France Road, Edinburgh EH16 4UX, United Kingdom.

CORRESPONDING AUTHOR:

Annette Bauer, Care Policy and Evaluation Centre, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom; a.bauer@lse.ac.uk; Tel: +44 207 852 3784
Objectives

Due to the nature of Alzheimer’s disease, health technology assessment agencies might face considerable challenges in choosing appropriate outcomes and outcome measures for drugs that treat the condition. This study sought to understand which outcomes informed previous health technology assessments, to explore possible reasons for prioritisations, and derive potential implications for future assessments of Alzheimer’s disease drugs.

Method

We conducted a literature review of studies that analysed decisions made in health technology assessments (across disease areas) in three European countries: England, Germany, and The Netherlands. We then conducted case studies of technology assessments conducted for Alzheimer’s disease drugs in these countries.

Results

Overall, outcomes measured using clinical scales dominated decisions or recommendations about whether to fund Alzheimer’s disease drugs, or price negotiations. Health technology assessment processes did not always allow the inclusion of outcomes relevant to people with Alzheimer’s disease, their carers and families. Processes did not include early discussion and agreement on what would constitute appropriate outcome measures and cut-off points for effects.

Conclusions

We conclude that in order to ensure that future Alzheimer’s disease drugs are valued appropriately and timely, early agreement with various stakeholders about outcomes, outcome measures and cut-offs is important.
Key words

Dementia; cholinesterase inhibitors; Alzheimer’s disease treatment; technology assessment, health; priorities, health

Acknowledgements

We would like to thank members of the ROADMAP health technology assessment and regulatory bodies expert advisory group (EXAG) as well as the following individuals, who provided advice on the scope of the work and on the interpretation of findings: Dr Amr Makady from the Zorginstituut Nederland (ZIN) and Dr Joshua Pink from the National Institute for Health and Care Excellence (NICE). We performed the study as part of the Innovative Medicines Initiative / Horizon 2020 ROADMAP (Real world Outcomes across the Alzheimer’s Disease spectrum for better care: Multi-modal data Access Platform) project. This project received funding from the Innovative Medicines Initiative 2 Joint Undertaking [grant no 116020 “ROADMAP”]. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA).

Main text

Introduction

Many countries face the prospect of rapid increases in expenditure related to Alzheimer’s disease (AD). Governments are faced with the task of making decisions about which drugs and interventions should be funded. Health technology assessment agencies or other...
decision-making bodies are responsible for such decisions based on reviews of clinical and effectiveness or cost-effectiveness evidence through a process called health technology assessment (HTA). Assessing the effectiveness or cost-effectiveness for approved AD drugs has been difficult because AD drugs have historically promised only very small or no effects in functional improvement or modifying disease progression (1,2). Whilst HTA processes vary by country, they have in common that evidence on effectiveness or cost-effectiveness is reviewed by a technical team and interpreted by a group of stakeholders, who present different perspectives such as those of clinicians, drug companies, patient representatives and researchers. Which outcomes and outcomes measures influence final decisions is likely to be based on various criteria including: whether they reflect meaningful changes in a person’s life (which is important from the perspective of people living with the condition, their families and carers); whether they are measurable in study designs (which is important from a developer and manufacturer perspective); and whether they are clinically and economically relevant (which is important from a payer perspective). Processes leading to decisions are complex, and are likely to vary between countries. The aims of our study were to understand: (1) which outcomes and outcomes measures are likely to be prioritized in HTAs for AD drugs in different countries, and (2) which processes influence these priorities. This study complemented other work which sought information on outcome prioritization from the perspective of patients, carers and practitioners (3, 4).

**Methods**

Overall approach and selection of countries
We employed two methods: a literature review and case studies. Findings of the literature review informed the design of the case studies. Both methods are explained in more detail below. Researchers with methodological expertise in systematic reviews (CT, AL) and qualitative research interviews (MN) as well as researchers specialized in medicine and neurology (CS), with knowledge of drug reimbursements and of HTA processes (CB, AG), and of dementia policies and economics (MK, RW) were involved in reviewing the methods throughout the research. In addition, the advisory group of the larger research programme of which this study was a part and which consisted of members from HTA or regulatory agencies across the world commented formally on initial findings. Three European countries were selected: England, Germany and The Netherlands. The respective HTA agencies are the National Institute for Health and Care Excellence (NICE) in England, the Institute for Quality and Efficiency in Health Care (IQWiG) in Germany and the Dutch Zorginstituut Nederland (ZIN). The choice was influenced by the size of the economy and roles and responsibilities of HTA agencies with the aim to have multiple perspectives: England and Germany present two large economies in Europe, in which HTA agencies have taken on different roles and responsibilities. For example, whereas in England drugs need to be cost-effective in order to be publicly funded (5,6), in Germany decisions about whether drugs are funded and at what price are primarily based on their added therapeutic benefit (7). The Netherlands, as a relatively small economy in Europe, has taken a middle ground approach in this regard: the cost-effectiveness of drugs needs to be proven if their cost is above a certain threshold (8). An overview of the main features of the HTA agencies in the three different countries is shown in Table 1.
Data collection and analysis

First, we conducted a literature review of studies which analyzed how outcomes are prioritized during HTA processes in the three countries. For the purpose of the literature review, we pragmatically defined prioritized outcomes (and their measures) as those that informed the final decision about whether the drug gets funded, or about its price. We made this decision based on initial searches, which showed how the issue has been investigated in the literature of HTAs. Since we expected that there would be limited evidence from the AD field, we searched for studies across disease areas. Details on search strategies, review and data extraction methods can be found in Supplement 1. Second, we gathered data on how outcomes and outcome measures had been prioritized in past HTAs of AD drugs. We conducted case studies based on information available on HTA websites, which documented the decision processes from the beginning to final recommendation. Here, we conceptualized ‘prioritization’ as the process of deriving decisions about which outcomes and measures should inform the value of AD drugs. We therefore considered any evidence of how decisions were made including views and opinions expressed by stakeholders about the importance of certain outcomes and measures, and how they thought they should inform the decision about the value of AD drugs. Information was extracted on topics relevant to outcomes and outcome measures considered in the appraisal. The framework for case studies and the data extraction form can be found in Supplement 2. The analysis was a thematic one, in which we used a mix of inductive and deductive methods for deriving themes.
About the data sources

Our literature review identified a total of 32 studies for the three countries. Thirteen studies referred to England; fourteen to Germany (this included one study which also referred to England); and six to The Netherlands. Studies used the following types of methods: quantitative analysis using statistical methods (n=10); qualitative or mixed methods (n=16); literature reviews (n=3); opinion papers or editorials (n=4). The main data sources were HTA reports and documentation of the decision processes from HTA agencies’ websites and interviews. Details of studies can be found in Supplement 3.

The case studies referred to publicly available documentation of HTAs for AD drugs (cholinesterase inhibitors and memantine) carried out between 2006 and 2010 in each of the three countries. This included altogether 6 HTAs: England (n=1; covering cholinesterase inhibitors and memantine together); Germany (n=2; one for cholinesterase inhibitors and one for memantine); Netherlands (n=3; memantine; donzepil; rivastigmine for Parkinson’s disease). What was documented varied widely between HTAs and countries but covered at a minimum:

- draft and final scope (including an agreed set of outcomes and outcome measures);
- draft and final appraisal of the reviewed evidence (including decisions or recommendations);
- consultation comments by stakeholders to draft scope and appraisal.
The list of documents that were identified as well as a list of stakeholders involved in the HTA processes can be found in Supplements 4 and 5; no documentation was available for stakeholder consultation in the Netherlands’ HTAs.

Results

A range of evidence related to relevant outcomes and outcome measures was collated under eight themes. The purpose of the collation was to have distinguishable themes that reflected the different aspects covered in the case studies and the literature. The themes are related and to a certain degree overlapping. The findings for each will be described in turn, and we refer in brackets to the numbered data source, which can be found in the Supplements 3 and 4.

Cost-effectiveness

In England, decisions about whether to fund AD drugs were based on cost-effectiveness, which in turn was based on health-related quality of life (in the form of quality adjusted life years measured with the EQ-5D) and institutionalization (Suppl. 4: 4.5). No other economic consequences (e.g. for hospital care) were included or discussed. Both, health-related quality of life and institutionalization, were in additional analysis extrapolated from clinical scales for cognition and functioning (Suppl. 4: 4.1; 4.2). In Germany and in The Netherlands, no additional economic analysis and no review of economic evidence was conducted, and there was no mention of cost-effectiveness in the scoping documents (Suppl. 4: 4.12; 4.13; 4.17-1.19). This partly reflects the different approaches in the three countries towards including cost-effectiveness evidence in HTAs (Suppl. Table 1: 3.5): Germany does not include cost-effectiveness in their HTAs. In The Netherlands, the prices of the drugs were considered ‘too low’ to justify the need for cost-effectiveness considerations i.e. as long as
they had additional value and no adverse consequences they would be funded (personal communication with ZIN representative).

Quality of Life (QoL)

There were differences in the ways HTA agencies responded to challenges of measuring QoL for people with AD: NICE allowed the prediction of QoL in the form of economic modelling based on surrogate outcomes measured with clinical scales. This approach was in contrast to the one taken by IQWiG, which does not accept the use of QoL measures like the EQ-5D and which has been consistently found to rarely accept QoL evidence (Suppl. Table 3: 3.14 3.21; 3.23; 3.25). Methodological requirements (such as a minimum follow up rate of 70%) frequently lead to the exclusion of evidence. Based on this and other methodological requirements not met by studies, IQWiG concluded that the evidence of an impact of AD drugs on QoL was insufficient (Suppl. Table 4: 4.14-4.16; 4.23-4.25). The resulting exclusion of QoL outcomes in the appraisal of AD drugs was criticized by some of the stakeholders (Suppl. Table 4: 4.12; 4.15; 4.16; 4.18; 4.19; 4.21). ZIN, whilst generally accepting and prioritizing QoL evidence including when measured through the EQ-5D [8], did not review QoL evidence in their HTAs of AD drugs (Suppl. Table 4: 4.26-4.28). We were unable to find an explanation.

Outcomes measured with clinical scales (O-CS)

A wide range of outcomes were measured with clinical scales. Table 2 presents an overview of the scales used in studies reviewed for the technology assessments. Not all scales, however, informed the advice or decisions about the value of AD drugs equally.
For Peer Review
countries O-CS such as cognition (measured, for example, with the Alzheimer’s Disease Assessment Scale-cognitive subscale; ADAS-cog) or functioning (measured, for example, with Activities of Daily Living scales; ADL) had an important influence on final decisions (Suppl. 4: 4.4; 4.15; 4.24). In the HTAs in England there was less evidence of stakeholder discussion about their relevance to people with AD (Suppl. 4: 4.7-4.10). The surrogate nature of those outcomes was made explicit in NICE’s documentation (Suppl. 4: 4.1-4.4).

The debate about the relevance of O-CS for people with AD was strongest in German HTAs (Suppl. 4: 4.20; 4.21). Whilst manufacturers argued the importance of clinical outcomes - in particular cognition - as reliable indicators of QoL with good psychometric properties, some stakeholders doubted whether clinical scales measured something that was meaningful to individuals (Suppl. 4: 4.20; 4.21). Both, IQWiG and the Federal Joint Committee (Gemeinsamer Bundesausschuss; G-BA), the body that makes the final and legally-binding decision about which drugs are funded, appeared to treat all O-CS as final health outcomes (Suppl. 4: 4.12-4.14), which meant that they bypassed some of their stricter methodological requirements that would have applied if they had been treated as surrogate outcomes (Suppl. Table 3: 3.24; 3.27).

In terms of specific measures, IQWiG did not accept the use of global assessment outcomes (measured for example with the Clinician Interview-Based Impression of Change; CIBIC), which were seen as reflecting the clinician’s perspective rather than the perspective of the person with AD (case studies). Instead, they expressed a preference for measures which evaluated personal goal attainment such as the Goal Attainment Scale (Suppl. 4: 4.14). ZIN pragmatically accepted those O-CS that had been accepted by the European Medicines Agency as validated outcome measures (Suppl. Table 3: 3.22; Suppl. 4: 4.26-4.28). This excluded the Mini-Mental State Examination (MMSE) as a measure for cognition, which was in contrast to NICE, which accepted its use as a main
outcome measure for modelling final QoL endpoints (Suppl. 4: 4.3;4.4; 4.26-4.28). ZIN noted that the wide range of outcome measures across different domains made the comparison of findings from studies difficult (Suppl. 4: 4.26-4.28).

--- Table 2 about here ---

During HTAs of AD drugs in all three countries, stakeholders raised concerns about how to interpret the identified (often very small) changes on clinical scales (Suppl. 4: 4.7-4.11; 4.16; 4.20). Several stakeholders argued that there was need for greater clarity, from the beginning, on cut-off points on various scales (Suppl. 4: 4.7-4.11; 4.16; 4.20). They should: be based on evidence, reflect disease severities, and be relevant to people with dementia. NICE tried to address the challenge of low effect sizes by giving particular weight to multi-domain changes (i.e., a simultaneous change in different scales). IQWiG considered every single outcome separately and as a result came to more conservative conclusions about the value of drugs (i.e. they concluded more uncertainty about their effectiveness), resulting in criticism from the drug manufacturers (Suppl. 4: 4.16; 4.20). In the Netherlands, ZIN expected manufacturers to set out and justify relevant cut-offs before conducting studies (Suppl. 4: 4.26-4.28). Responding to the uncertainty over clinical relevance and relevance to people with AD, it decided to make the introduction of the drugs subject to start and stop criteria and delegated the application of those to clinicians (Suppl. 4: 4.26-4.28).

Adverse effects

In England, benefit-harm considerations were not given much weight during HTAs possibly because safety concerns were addressed already as part of market authorization and aspects of adverse effects were thought to be captured in QoL outcomes (Suppl. Table 3: 4.18-4.19; 4.21).
Some stakeholders felt that the better tolerability of AD drugs, when compared to alternative treatments (such as antipsychotics), was undervalued in this approach (Suppl. 4: 4.11). In contrast, adverse effects were regarded as important outcomes from the perspective of people living with AD in Germany and the Netherlands (Suppl. 4: 4.14; 4.16; 4.26-4.28). Stakeholders of the HTAs carried out by IQWiG criticized the lack of long-term safety data on AD drugs and raised concerns about whether adverse effects had been underestimated (Suppl. 4: 4.11). In The Netherlands, ZIN sometimes left benefit-harm decisions to clinicians, as it concluded that the evidence did not support general conclusions (Suppl. 4: 4.26-4.28).

Outcomes relevant to people with AD

In both England and Germany, stakeholders (mainly patient representatives but also researchers and commissioners) argued that many outcomes relevant to people with AD were not being picked up by the clinical scales (Suppl. 4: 4.2; 4.8; 4.11; 4.21; 4.23; 4.25). They advocated for including more tangible outcomes (e.g. ability for someone to pick up the phone) as well as long-term outcomes (e.g. institutionalization). An ability to maintain aspects of personal identity was seen as another important outcome. Stakeholders highlighted an urgent need for appropriate outcome measures in early stages of AD (Suppl. 4: 4.8; 4.11).

In both countries, stakeholders thought that this required more flexible approaches towards including evidence (Suppl. 4: 4.8; 4.21; 4.23; 4.25). Whilst this need for different and more flexible processes was to a large extent shared by NICE, in Germany IQWiG and
G-BA believed that such changes would contradict legislation and reduce the necessary methodological robustness (Suppl. 4: 4.21; 4.23; 4.25).

In the Dutch HTAs for AD drugs, the challenges of considering outcomes that mattered to people with AD, their carers and families were not documented but had been - according to a ZIN representative - discussed at several stages of the process (personal communication).

Carers’ outcomes

In NICE’s HTAs of AD drugs, carers’ QoL was a primary outcome or endpoint (Suppl. Table 4: 4.1; 4.3), reflecting the priority given by NICE to this group. However, final decisions were based on an economic model, which did not include carers’ outcomes, an omission which was criticised by some stakeholders (Suppl. 4: 4.7; 4.11). In Germany, carers’ outcomes were not viewed as the responsibility of the healthcare system and were given lower priority relative to outcomes for people with AD (Suppl. 4: 4.21; 4.25). Whilst some stakeholders argued for including carers’ outcomes in its own right, there seemed to be an overall consensus that carers’ outcomes were important mainly because of their impact on the person with AD (Suppl. 4: 4.21; 4.25). In addition, IQWiG was skeptical about carer-reported outcomes for the person with dementia, which they argued reflected the needs of the carer rather than the needs of the person with dementia (Suppl. 4: 4.21; 4.25). Dutch HTAs for AD drugs did not include carers’ outcomes (Suppl. 4: 4.26-4.28).

Institutionalization
Institutionalization for someone with AD was a stated outcome in HTA agencies’ documentation and discussed as important by stakeholders in Germany and England (Suppl. 4: 4.1; 4.3; 4.12; 4.13; 4.17-4.19). Only English HTAs included institutionalization as an outcome in the economic modelling although stakeholders discussed whether it was possible to accurately predict this outcome since there were many other correlated factors such as the carer’s situation and availability of care in the community (Suppl. 4: 4.8) . In German HTAs institutionalization was viewed by some stakeholders as an outcome that was primarily important from an economic perspective (Suppl. 4: 4.21; 4.25). Some stakeholders thought that ‘institutionalization’ could not be measured separately from ‘time spent caring’, and that instead ‘hours of care provided’ should be measured independently of whether they were provided by a professional or by an unpaid carer (Suppl. 4: 4.21; 4.25). Similar to the discussion in English HTAs, stakeholders discussed the lack of evidence on these outcomes and methodological challenges of including them. In The Netherlands there was no recorded information on these outcomes.

Table 3 presents an overview of the findings. We applied categories indicating if an outcome or group of outcomes was prioritized or not prioritized. ‘Prioritized’ outcomes were those that informed decisions and ‘Not prioritized’ outcomes were those that did not inform decisions.

---- Table 3 about here ----
Discussion

This study assessed the outcomes and outcome measures that dominated HTAs of AD drugs in three European countries, and the processes that influenced those priorities. This is to our knowledge the first study, which examines how outcomes and measures for AD drugs are currently prioritised in technology assessments. This study contributes to an increased transparency about reasons for and challenges of including certain outcomes when assessing the value of AD drugs. Overall, we identified some challenges in the process of how outcomes, outcome measures and cut-off points were defined in technology assessments of AD drugs. This included a lack of early involvement of stakeholders in discussions of appropriate outcomes and outcome measures as well as of cut-off points for appropriate effect sizes. In addition, a narrow focus on evidence from certain types of studies, namely randomised controlled trials, led to a strong focus on outcomes measured with clinical scales to the potential exclusion of (long-term) outcomes relevant for people with AD.

Our study was exploratory in nature, and we chose to conduct two methods to address the gap in evidence about the role of outcomes and outcome measure in HTAs of AD drugs. We first reviewed studies that analysed the influence of outcomes and measures on decisions of the value of drugs in HTAs. Whilst this provided useful knowledge about common decision making patterns in HTAs (and reasons for those), it provided only limited information about the process, by which decisions were made about outcomes and measures, and the process by which they influenced decisions. Whilst we have no affirmative knowledge of the reason...
for this missing focus of studies, it is plausible that decisions about outcomes and measures are regarded objective or neutral. It is also likely to reflect a wide acceptance of outcomes measured with clinical scales as patient-relevant. As a result, designers of studies and manufacturers have to make decisions about outcomes and measures without certainty whether those will be accepted by HTAs. In the case of HTAs for AD drugs, this is likely to have contributed to the use of a wide range of measures. With the second method, the case studies, we therefore sought to address the gap in evidence about the process by which outcomes and measures are influencing decisions through in-depth analysis of reports produced for technology assessments. This kind of analysis allowed us to understand the nature of decision processes, and stakeholder viewpoints. Although this study was explorative in nature, we were able to shed new light on the important, currently under-investigated role of outcomes and outcome measures in influencing the value of AD drugs.

In terms of methodological robustness, the literature review, although pragmatic, applied systematic search strategies and involved detailed data extraction. Researchers with a high and diverse level of methodological and clinical expertise were involved in and contributed to the robustness of the research process. Approval of the research methods and interpretation of the findings was provided by experts in the field. In terms of limitations, for the case studies, we were reliant on publicly available information, which was limited, especially for the Dutch case studies. Furthermore, by focusing on HTAs with the most comprehensive information and those that were most comparable between countries, we might have missed some aspects of more recent updates of HTAs. Overall, our findings need to be interpreted in the context of a rapidly evolving field. Considerations that decision makers need to take into account today may very well change in the future, for example in light of new evidence and new technologies.
The findings from this study suggest that there are substantial challenges in including outcomes relevant to people with AD when assessing the value and cost-effectiveness of AD drugs. Those challenges are not only relevant to existing AD drugs but to other types of treatment and interventions, which seek to prevent or alter the progression of AD. Unless there is an agreed set of outcomes, outcome measure and cut-offs that define a meaningful diversion from the path without intervention, it will be challenging to assess the value of a drug or an intervention (in particular in relation to other interventions). In the future, this is likely to be relevant to pricing or investment decisions for disease-modifying treatments, which may need to be offered at pre-dementia stages, and which would require measuring surrogate outcomes such as imaging or other biomarkers (9). Without outcome measures that are acceptable to relevant stakeholders – including patients, carers and the wider public - and agreed before HTAs are conducted or preferably even before studies are being developed, there is a risk of delays in the appropriate evaluation of, and access to, new treatments (10). Clear methodological guidance on accepted outcome measures in fields such as prevention and diagnostics is therefore needed (11).

This includes the need to consider patient-relevant outcomes in HTAs in addition to clinical outcomes (12). Whilst in early stages, innovative methods have been developed (and tested) that allow HTA agencies to consider patient preferences over different outcomes when developing methodological guidance (12). Knowledge is also becoming increasingly available about how to best include patient and carers’ perspectives in HTAs (13). Decisions about the value of drugs in HTAs in some countries (including England) have shown to be substantially influenced by aspects of value not captured by clinical and economic evidence (14). Whilst this is a reflection of including patient and stakeholders perspectives it also raises questions about transparency and consistency of decisions (14). Therefore, including
outcomes, measures and cut-offs that are more patient-relevant (and agreeing on those in advance) is likely to contribute to more consistent decision making as it reduces the need for additional considerations that in effect address the issue of evidence not being sufficiently relevant to what matters to patients, carers and the wider public.

Furthermore, the challenges we identified suggest a need for collaborative approaches between multiple stakeholders to enable decisions on outcomes and measures to be made early in the process. Some of the required processes are already in place, to varying degrees in different countries, whilst others still need to be developed.

Such multi-stakeholder approaches should go hand in hand with including wider sets of evidence, often referred to as real-world evidence (15). This requires an investment in data that can be used to demonstrate long-term impact on costs and outcomes (1). This might include data on the costs associated with different rates of disease progression so that cost savings linked to a delay in disease progression can be estimated. Findings from a study that modelled the likely cost-effectiveness of disease-modifying treatments (should they become available) showed that in England the benefit from deferring onset by one year would be substantial at about £28,000 (in 2012/13 prices) (16). This highlights the importance of including such data in decision making. Unless the impacts on disease progression, QoL, need for care and costs over time are considered, there is a risk that that future AD drugs and interventions are not valued in line with patient, carers and wider public interests.

Conclusions

This study investigated the role of outcomes and outcome measures in HTAs of AD drugs in three European countries. The findings highlight the strong priority placed on outcomes measured with clinical scales as well as the challenges of considering measures that capture
changes in disease progression that are potentially relevant from the perspective of people living with the condition, their families and carers. We conclude that there is an urgent need to reform HTA processes to appropriately assess the value of AD drugs.

Conflicts of interest

CS was lead (jointly with CB) of work package 2 of the ROADMAP project; CS is lead of dementia outcomes work package for the UK MRC-funded Dementias Platform UK. CB is employed by F. Hoffman-La Roche. AG is a partner of Quantify Research, providing consultancy services to pharmaceutical companies and other private and public organisations and institutions. AGs contribution to ROADMAP was on behalf of Roche Pharmaceuticals. AB, AL, CT, MK, MN, RW have no conflict of interests to declare.

Acknowledgements

We would like to thank members of the ROADMAP HTA and regulatory bodies expert advisory group (EXAG) as well as the following individuals, who provided advice on the scope of the work and on the interpretation of findings: Dr Amr Makady from the Zorginstituut Nederland (ZIN), Joshua Pink and Dr Jacoline Bouvy from the National Institute for Health and Care Excellence (NICE).

Funding: This study was part of the Real World outcomes across the AD spectrum for better care (ROADMAP) project. This project received funding from the Innovative Medicines Initiative 2 Joint Undertaking [grant agreement no 116020 (“ROADMAP”)]. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA).
References


Legend

Table 1: Features of health technology assessment (HTA) agencies in England, Germany, and The Netherlands

Table 2: Clinical scales used in studies identified in health technology assessments (HTAs) of Alzheimer’s disease drugs in England, Germany, and The Netherlands

Table 3: Prioritization of outcomes and outcome measures in health technology assessments (HTAs) of Alzheimer’s disease (AD) drugs
Supplementary file

Supplement 1: Details on search strategies, inclusion and exclusion criteria, data extracted and quality appraisal

The following databases were searched for articles published between 2007 and November 2017: CINAHL, MEDLINE, SocScience, EconLit, Elsevier Science Direct. The following subject headings and keywords were used: outcome-related term (i.e. outcome OR benefit OR effect OR endpoint) AND country-related term (i.e. Germany OR Netherlands OR England) AND a technology assessment-related term (i.e. benefit assessment OR technology assessment). If the number of results was particularly large, we added an additional search term for decision making process (i.e. process OR decision making).

In addition, the smart search (CINAHL), recommendations based on previously read articles (Elsevier Direct) and articles frequently cited together (PubMed) functions of databases were used. Additional searches in journals of particular relevance such as ‘Value in Health’ and ‘Medical Decision Making’ were also carried out. A few reference searches for key articles were carried out to test if all relevant articles were captured in the searches.

Articles were included that referred to decisions, processes and standards of health technology assessments if they made reference to the role of outcomes. Excluded were articles, which were critical discussions about the use of specific methods - such as: the quality-adjusted life years (QALY) measure; social discount rates in economic evaluation methods; multi-criteria decision making – or which related to personalised medicine, described the influence of HTA processes on market access to drugs or focused on price setting mechanisms and negotiations.

The following information was extracted for each study: study ID; setting; purpose; design; type of data and analysis method; further details about methods (where required); results; conclusions and limitations stated by author(s). For each study a rating was generated to reflect its relevance for our research questions.

In a next step, information was summarised for each country using the following headings (which were identified during the initial analysis of information):

- Responsibilities of HTA and other relevant agencies in regards to HTA or reimbursement process;
- HTA process and requirements;
- Decision making process and criteria;
- Price negotiations and status of listing decision;
- Stakeholder involvement in process;
- Surrogate and composite outcomes;
- Quality of life (and quality-adjusted life years);
- Cost-effectiveness;
- Sub groups.
Supplement 2: Case study framework and data extraction form

Case study framework

Case studies were carried out for health technology assessments / appraisals (HTA) carried out in the dementia/ AD field in three countries: England, Germany and Netherlands.

For England the case study referred to one Multiple Technology Appraisal for donezepil, galantamine, rivastigmine and memantine for the treatment of Alzheimer’s disease (AD)\(^1\), which was published 23 March 2011 with last update 11 May 2016. Relevant publicly available documents were included to inform the case studies, such as:

- Guidance and appendices;
- Research recommendations information;
- Documents produced as part of the guidance development such as:
  - Background information (includes review decision documents, press releases);
  - Assessment report documents;
  - Draft and final protocol documents;
  - Draft and final matrix documents;
  - Draft and final scope documents (including consultation comments);
  - Appraisal consultation documents (e.g. Assessment reports; Consultee and commentator comments on the assessment report; Manufacturer and Non-manufacturer Submissions; Expert written personal statements);
  - Final appraisal determination documents (including comments on appraisal consultation)

In Germany case studies referred to the following 3 single drug benefit assessments: Memantine in AD; Cholinesterase inhibitors in AD; Ginkgo compounds. There were no technology appraisals in form of early benefit assessments carried out for dementia/ AD drugs since introduction of the new legislation (AMNOG) in 2011. Instead, all appraisals refer to drug assessments carried out before AMNOG. Relevant information included the following documents from the IQWiG website:

- Final and preliminary reports
- Documentation and appraisal of comments on the preliminary report
- Report plan (different versions) and amendments
- Documentation and appraisal of comments on the report plan
- Executive summary of the working paper ’Memantine in Alzheimer’s disease: Results of the unpublished studies IE2101 and MEM-MD-22 as well as unpublished responder analyses’
- Press releases

---

\(^1\) [https://www.nice.org.uk/guidance/ta217](https://www.nice.org.uk/guidance/ta217)
We also looked at the G-BA website for manufacturers’ value decisions, G-BA value decisions (Tragende Gruende). Further information about decisions and the role of clinical endpoints in those decisions were also available online\(^2\) \(^3\) \(^4\) \(^5\).

In The Netherlands, case studies referred to short pharmacotherapeutic reports for donezepil (for the indication and symptomatic treatment of mild to moderately severe Alzheimer’s dementia)\(^6\); rivastigmin for people with Parkison’s disease and memantine.

Across case studies, the following information were extracted with respect to the following questions:

<table>
<thead>
<tr>
<th>Study endpoints (mortality/morbidity/quality of life)</th>
<th>What endpoints were set out during scoping?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Which primary endpoints were used in studies that supported the recommendation? (This might include information about categorised clinical endpoints and clinical scales)</td>
</tr>
<tr>
<td></td>
<td>Which surrogate endpoints were used that supported recommendations, which methods of validation were used? Did the Committee discuss the appropriateness of surrogate endpoint as validated indicators of endpoints?</td>
</tr>
<tr>
<td></td>
<td>How were patient preferences (satisfaction, adherence, complaints) and patient reported outcomes considered?</td>
</tr>
<tr>
<td></td>
<td>Which endpoints were considered in cost-effectiveness analysis that supported recommendations? How was clinical evidence mapped to final endpoint quality of life (in cost-effectiveness analysis)?</td>
</tr>
<tr>
<td></td>
<td>Were aspects of meaningful delay and disease progression considered in endpoints?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stakeholder views and influence</th>
<th>Which suggestions were made in regards to clinical endpoints?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Which challenges around including relevant clinical endpoints were discussed by stakeholders?</td>
</tr>
<tr>
<td></td>
<td>Were any clinical endpoints were considered differently as a result of stakeholder involvement?</td>
</tr>
<tr>
<td></td>
<td>Which clinical endpoints were identified as relevant for future research?</td>
</tr>
</tbody>
</table>

| Uncertainty | How did uncertainty in data influence discussions about outcome? Were there any criteria or rationales that made an uncertain outcomes more acceptable? |

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Which thresholds were applied in regards to clinical measures and/or cost-effectiveness?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>How were benefit harm ratios considered?</td>
</tr>
</tbody>
</table>

---


\(^5\) [https://www.ispor.org/News/articles/Oct06/german_policy.asp](https://www.ispor.org/News/articles/Oct06/german_policy.asp)

For each HTA, information was extracted from publicly available documentation relating to the HTA using a range of categories that were derived from headings used in analysis of data from the literature review and from an initial analysis of the information. The categories were as follows:

Outcomes included:

- Outcomes set out during scoping
- Outcomes considered during review
- Outcomes considered differently as a result of stakeholder involvement (Suggestions made by stakeholders in regards to outcomes)
- Outcomes identified as relevant for future research
- Outcomes used in studies that supported the recommendation
- Outcomes considered in cost-effectiveness analysis

Challenges around including outcomes:

- Types of evidence considered
- Surrogate outcomes and methods of validation
- Patient preferences and patient-reported outcomes
- Aspects of meaningful delay and disease progression
- Influence of data uncertainty on outcomes
- Thresholds in regards to outcomes measures or cost-effectiveness

Supplement 3: Details of studies identified in literature review
<table>
<thead>
<tr>
<th>Study ID</th>
<th>Purpose</th>
<th>Setting</th>
<th>Method</th>
<th>Data sources</th>
<th>Details</th>
<th>Results</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Allen et al 2017 (Low)</td>
<td>To compare initial Canadian national HTA recommendations with the initial decisions of the other HTA agencies, and to identify factors for differing national HTA recommendations between the four HTA agencies.</td>
<td>Australia, Canada, England, Scotland</td>
<td>Medicines that were reviewed by all four agencies and received a negative recommendation from only one agency were selected as case studies. Statistical analysis of HTA recommendations classified as positive or negative (numerically coded); percentage agreement was calculated.</td>
<td>Information from websites of HTA and bodies responsible for final reimbursement decision</td>
<td>Process map using a previously developed mapping methodology; this enabled identification and relationship between HTA agencies and responsible body for reimbursement decision</td>
<td>HTA bodies considered clinical efficacy; adverse effects; cost-effectiveness; all have implicit or explicit quality-adjusted life-year threshold; factors influencing decisions were: uncertainties surrounding a range of factors including: cost-effectiveness; comparator choice; clinical benefit; safety; trial design; submission timing</td>
<td>Use of publicly available sources; inclusion criteria limited to products listed on Controlled Drug Regulation, which resulted in exclusion of cancer medicines</td>
</tr>
<tr>
<td>3.2 Carroll et al 2017 (Medium)</td>
<td>To explore the type of additional exploratory analyses conducted by Evidence Review Groups and their impact on the recommendations made by NICE</td>
<td>England</td>
<td>A content analysis of relevant documents was undertaken to identify and extract relevant data, and narrative synthesis was used to rationalize and present these data.</td>
<td>100 most recently completed single technology appraisals since 2009 with published guidance were selected for inclusion.</td>
<td>Categories for exploratory analyses developed with research team; this was used to inform coding; all data extraction were double checked by two researchers.</td>
<td>The additional analyses undertaken by Evidence Review Groups in the appraisal of company submissions are highly influential in the policy-making and decision-making process; clear influence on 47% of final appraisal determinations</td>
<td>No limitations stated by author(s)</td>
</tr>
<tr>
<td>3.3 Cerri et al 2013 (High)</td>
<td>This study examined the impact of evidence, process and context factors on NICE decisions; to assess which of factors best explains the pattern of NICE decisions</td>
<td>England</td>
<td>With multinomial logistic regression, the relative contribution of explanatory variables on NICE decisions was assessed</td>
<td>A data set of NICE decisions 2004-2009 in HTAs was created, including 32 variables extracted from published information. A three-category decision outcome variable was created.</td>
<td>A total of 65 technology appraisals (118 technologies) were analysed</td>
<td>Results showed significant associations (p&lt;0.10) between NICE decision outcome and four variables: (i) demonstration of statistical superiority of the primary endpoint in clinical trials by the appraised technology; (ii) the incremental cost-effectiveness ratio (ICER); (iii) the number of pharmaceuticals appraised within the same appraisal; and (iv) the appraisal year.</td>
<td>No limitations stated by author(s)</td>
</tr>
<tr>
<td>3.4 Clement et al 2009 (High)</td>
<td>To assess how committees use evidence on effectiveness and cost-effectiveness (including any barrier to such use) and what additional factors have influenced decisions</td>
<td>Australia, Canada, England</td>
<td>Descriptive analysis of retrospective data from HTA bodies; 3 case studies: diabetes mellitus, ranibizumab for age-related macular degeneration, and teriparatide for osteoporosis</td>
<td>All publicly available documents as of 31st December 2008</td>
<td>Factors that influenced decisions: The differences in listing decisions often appeared less about the interpretation of the clinical or economic evidence and more about differences in agency processes in terms of outcomes: More than 50% of submissions reviewed by NICE used clinical end points (rather than clinical scales or surrogates), and if surrogate outcome were used they were more likely to be judged valid by committee</td>
<td>Use of publicly available sources; there may be subtle issues that were not captured, particularly in the deliberation process; surprisingly few common drugs across the 3 systems, making comparisons across committees less conclusive</td>
<td></td>
</tr>
<tr>
<td>3.5 Dakin et al 2014 (Medium)</td>
<td>To investigate the influence of cost-effectiveness and other factors on NICE decisions and whether NICE’s decision-making has changed over time</td>
<td>England</td>
<td>Logistic regression to predict whether a technology was recommended or not; NICE’s decisions as binary choices for/ against a technology in a specific patient group</td>
<td>Data on all NICE decisions published by December 2011 were obtained from HTAinSite [<a href="http://www.htainsite.com">www.htainsite.com</a>].</td>
<td>Independent variables comprised of the following: clinical and economic evidence; characteristics of patients, disease or treatment; and contextual factors potentially affecting decision-making.</td>
<td>Cost-effectiveness was main driver for NICE decisions; past decisions appear to have been based on a higher threshold than £20 000–£30 000/QALY; this may reflect consideration of other factors that cannot be easily quantified.</td>
<td></td>
</tr>
<tr>
<td>3.6 Drummond &amp; Sorenson 2009 (Medium)</td>
<td>Opinion paper that explains NICE activities, achievements, challenges and lessons learnt</td>
<td>England</td>
<td>Opinion paper</td>
<td>-</td>
<td>-</td>
<td>No direct conclusions; issues discussed around QALY, ICER, sub group analysis and stakeholder involvement</td>
<td></td>
</tr>
<tr>
<td>3.7 Drummond et al 2013 (Medium)</td>
<td>Opinion paper that explores HTA approaches, in both methods and policy, to help bring about reconciliation between</td>
<td>Europe</td>
<td>Opinion paper</td>
<td>-</td>
<td>-</td>
<td>HTA initiatives are likely to give manufacturers an incentive to more closely align their research and development with social objectives; adequate stakeholder involvement is needed to ensure</td>
<td>N/A</td>
</tr>
<tr>
<td>3.8 Fischer 2012 (High)</td>
<td>To structure empirical evidence of coverage decisions made in practice based on the components ‘methods and evidence’, ‘criteria and standards’, ‘decision outcome’ and ‘processes’</td>
<td>Focus on England, scope international</td>
<td>Literature review</td>
<td>Electronic databases, journals and HTA websites were searched for publications between 1993 and June 2011. Included were analysis of past decisions and application of quantitative methods.</td>
<td>Each study was categorized by the scope of decision-making and the components covered by the variables used in quantitative analysis.</td>
<td>Important influence of therapeutic value where decision makers did not explicitly account for cost-effectiveness; the ICER had significant influence on decisions in Canada, Australia and the UK, but usually in combination with other aspects such as burden of disease or health condition. Budget considerations were significant influences in Australian and Dutch decision-making.</td>
<td>No limitations stated by author(s)</td>
</tr>
<tr>
<td>3.9 Kreis and Schmidt 2013 (Low)</td>
<td>This article explores operational processes and underlying rationales of public engagement at HTA agencies</td>
<td>France, Germany, United Kingdom</td>
<td>Authors explored qualitatively public engagement processes and underlying rationales</td>
<td>The analysis is based on website information, legal framework documents, published and grey literature, and semi structured, in-depth interviews with top officials at these agencies</td>
<td>Authors used the term public as the broadest generic term to include engagement of individual citizens, patients, consumers (or users), laypeople, or formal or informal representatives of groups of these</td>
<td>Engagement processes differed across agencies, particularly regarding the areas in which the public is involved, which groups of the public are involved, what weight they have in influencing decisions, how they are recruited and supported, and how potential conflicts of interests are addressed.</td>
<td>No limitations stated by author(s)</td>
</tr>
<tr>
<td>3.10 Nicod and Kanavos 2012 (Medium)</td>
<td>To identify diverging HTA recommendations across five countries, understand the rationale for decision-making, and suggest ways forward to minimize inter-country differences</td>
<td>England, Scotland, Sweden, Canada, and Australia</td>
<td>Comparative statistical analysis of HTA recommendations for 287 drug-indication pairs appraised by countries between 2007 and 2009, including an in-depth analysis of two case studies</td>
<td>Appraisal reports from each agency</td>
<td>Agreement levels were measured using kappa scores. Associations between the HTA recommendations and the HTA body issuing the recommendation were explored</td>
<td>Substantial disparities in recommendations for/against drugs; HTA processes potentially influenced by: different priorities in different settings; different perception of benefit and value, and use of different tools of addressing uncertainty; patient preferences and characteristics</td>
<td>No limitations stated by author(s)</td>
</tr>
</tbody>
</table>
### 3.11 Nicod et al 2017 (High)

To better understand the reasons for differences in reimbursement decisions for orphan drugs in four European countries.

**Methods**
- **England, Scotland, Sweden, France**
- Semi structured interviews with representatives of HTA bodies
- Semi-structured interviews; eight representatives from the four HTA bodies were interviewed between March and June 2015
- An interview topic guide was developed on the basis of findings from a systematic comparison of HTA decisions for 10 orphan drugs.
- Qualitative thematic data analysis using the framework approach

**Results**
- Decisions regarding orphan drugs made in context of lower quality evidence; threshold of acceptable uncertainty varied by country; NICE more likely to accept surrogate endpoints for orphan drugs; NICE always prefers overall survival to progression-free survival; HRQOL data were considered as a hard end point by NICE.
- Safety only implicitly considered because already part of marketing authorisation.

**No limitations stated by author(s)**

### 3.12 Oyebode et al 2016 (Low)

To determine the aspects of expert advice that decision-makers find most useful in the development of evidence-based guidance and to identify the characteristics of experts providing the most useful advice.

**Methods**
- **England**
- (1) Interviews examined the usefulness of expert advice during guidance development.
- (2) Associations between usefulness score and characteristics of the expert advisor were investigated using univariate and multivariate analyses
- (1) Semi-structured interviews with 17 members of the Interventional Procedures Advisory Committee of NICE.
- (2) Data were extracted from 211 experts’ questionnaires for 41 consecutive procedures.
- (1) Transcripts were analysed inductively to identify themes;
- (2) Usefulness of advice was scored using an index developed through the qualitative work.

**Results**
- Values and challenges of using expert opinion in HTA processes are analysed

**Authors reflect on their own potential bias due the researchers' previous experience at NICE and working in public health and medical roles; concept of 'usefulness' was potentially problematic**

### 3.13 Spinner et al 2013 (Medium)

To assess whether different clinical evidence bases may have influenced listing recommendations.

**Methods**
- **Australia, Canada, England and Wales**
- Authors reviewed the evidence considered for each listing recommendation, identified the similarities and differences, and evaluated the extent to which different clinical appraisal reports between 2007 and 2010 (including manufacturers’ submissions) for nine drugs for which the three agencies had Not provided

**Results**
- Decisions across HTA bodies associated with differences in the clinical evidence base considered. NICE considered indirect and/or mixed-treatment comparisons; in some cases, NICE excluded trials from review if the drug and/or the comparator

**Small number of case studies; only publicly available documents were considered**
<table>
<thead>
<tr>
<th>Germany (n=14)</th>
<th>3.14 Biome et al 2017 (High)</th>
<th>To determine methodological requirements for QoL measurement and data presentation in early benefit assessment (EBA)</th>
<th>Qualitative content analysis based on documents of all EBAs completed by 2014; analysis included information extraction, coding, critical discussion and consensus building. Documents publicly available on the G-BA website including: manufacturer dossier; dossier evaluation and benefit assessment by IQWiG or Federal Joint Committee (G-BA); protocol of the oral hearing; rationale of the G-BA decision (&quot;Tragende Gruende&quot;=main justifications)</th>
<th>No association between the inclusion of QoL data in benefit dossiers and the G-BA's rating decision might be explained by non-compliance with the various methodological requirements found in our analysis, so that in most cases, the mere inclusion of QoL data in the dossier did not lead to a positive evaluation of QoL benefit. In addition, many EBAs did include QoL outcomes, but there were no statistically or clinically significant effects.</th>
<th>No limitations stated by author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.15 Fischer and Stargardt 2014 (Medium)</td>
<td>To explain the decisions made in early benefit assessments (EBAs), clarify the roles of manufacturers, IQWiG, German Federal Joint Committee (G-BA), and guide manufacturers in developing future submissions</td>
<td>Authors evaluated differences in rating decisions by manufacturers, the IQWiG, and the G-BA with regard to each pharmaceutical’s added benefit. Authors used Cohen’s kappa to analyze agreement between rating decisions; chi-square test and bivariate regression were used to identify associations between components of the EBA. Data extracted for EBAs for which the G-BA made a rating decision between 2011 and 2013. Authors developed a variable list including: rating decisions of manufacturers, IQWiG, G-BA; characteristics of the process; types of evidence submitted; methods used to Two independent reviewers extracted data. Once completed, the worksheets were compared to identify any deviations. Interrater reliability was good, with an average Cohen’s kappa coefficient of 0.63 (range, 0.28 to 1.00) for categorical variables and an average Pearson’s correlation coefficient of 0.80</td>
<td>While the G-BA tended to disagree with the rating of benefit by manufacturers, it softened IQWiG’s decisions, potentially to make the final outcome more acceptable. Concerns voiced that the G-BA might be exceeding its statutory authority by taking cost or procedural considerations into account appear to be unfounded. Choosing appropriate evidence to submit for each endpoint remained a challenge, as submission of</td>
<td>No limitations stated by author(s)</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Focus</td>
<td>Methods</td>
<td>Results</td>
<td>Limitations</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>---------</td>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>3.16 Fischer et al 2016 (Medium)</td>
<td>Germany, compared with England, Scotland, Australia</td>
<td>Authors analysed statistically decisions made for comparable patient subgroups by the four agencies between 2011 and 2014. First, decisions were compared (a) by their final outcome, i.e. whether a health benefit was identified, and (b) by the agencies’ judgement on comparative effectiveness. Subsequently, they partially explored reasons for differences between HTA agencies.</td>
<td>For each comparison, authors analysed the agreement between G-BA and each of the other HTA agencies. Agreement was quantified by calculating Cohen’s Kappa, to determine whether agreement between two raters was by chance.</td>
<td>No limitations stated by author(s)</td>
<td></td>
</tr>
<tr>
<td>3.17 Ivandic et al 2014 (High)</td>
<td>Germany, England</td>
<td>The following aspects were examined: guidance texts on methodology and information sources for the assessment; clinical study design and methodology; statistical analysis, quality of evidence base, extrapolation of results (modelling), and generalisability of study</td>
<td>The findings are presented separately for the two HTA systems and thus may serve as stand-alone references. A concise, integrated comparison follows to highlight the main similarities and Methodological requirements differed mainly in the acceptance of low-level evidence, surrogate endpoints, and data modeling. Some of the discrepancies may be explained, at least in part, by differences in the health care system and procedural aspects (e.g. timing of assessment).</td>
<td>No limitations stated by author(s)</td>
<td></td>
</tr>
<tr>
<td>Section</td>
<td>Study</td>
<td>Methods</td>
<td>Results</td>
<td>Conclusion</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>---------</td>
<td>---------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>3.18</td>
<td>Griffith and Griffith 2015 (Low)</td>
<td>Analysis of past decisions of German HTA to inform future submissions</td>
<td>All IQWiG decisions from January 2011 to May 2015 were assessed, and the effect of the clinical evidence base on the submission outcome was examined.</td>
<td>Completed single drug appraisals from Jan 2011 to May 2015: Recommendation ('added benefit' or 'no added benefit'), indication, rationale, and evidence base were extracted.</td>
<td>Over half of drugs appraised by IQWiG since 2011 have been given 'no added benefit' status, and direct evidence against an appropriate comparator remains a priority for a favourable decision.</td>
</tr>
<tr>
<td>3.19</td>
<td>Kohler et al 2015 (Medium)</td>
<td>To determine the information gain from AMNOG documents compared with non-AMNOG documents for methods and results of studies available at market entry of new drugs.</td>
<td>Authors assessed reporting quality for each study and each available document for eight methods and 11 results items. For each document type they calculated the proportion of items with complete reporting for methods and results, for each item and overall, and compared the findings.</td>
<td>Dossier assessments conducted by IQWiG between 1 Jan 2011 and 28 Feb 2013; European public assessment reports, journal publications, and registry reports.</td>
<td>Concludes that AMNOG documents provide a considerably higher proportion of complete information than European public assessment reports; this includes information on methods, results and patient relevant outcomes. The information gap was most striking when the drug was approved only in a certain subpopulation.</td>
</tr>
<tr>
<td>3.20</td>
<td>Kvitkina et al 2014 (Low)</td>
<td>To describe the feasibility of the early benefit assessment on the basis of patient-relevant outcomes by systematically characterising the outcomes available in manufacturers' dossiers and comparing the companies' and IQWiG's evaluations regarding patient relevance and surrogate validity</td>
<td>Dossier assessments were used for data extraction; the outcomes available and the respective evaluations were extracted and compared. 12 out of 22 submitted dossiers contained sufficient data to assess outcomes; all 12 assessable dossiers provided data on patient-relevant outcomes.</td>
<td>Publicly available manufacturers' dossiers; published between October 2011 and June 2012.</td>
<td>Data on mortality and adverse events were available in almost all dossiers; data on morbidity and health-related quality of life available in 8 and 7 dossiers, respectively. Of a total of 214 outcomes extracted by IQWiG, 124 patient-relevant and 3 surrogate outcomes were included in IQWiG's assessment (companies: a total of 183 outcomes included, of which 172 were patient-relevant and 11 were surrogates outcomes partly deviated from each other.</td>
</tr>
</tbody>
</table>
### 3.21 Lauenroth and Stargardt 2017 (High)

To analyze how value is determined within the scope of the German Pharmaceutical Restructuring Act

**Germany**

Generalized linear model regression to analyze impact of added benefit on difference between negotiated prices and prices of comparators

All pharmaceuticals that had undergone assessment, appraisal, and price negotiations in Germany before June 30, 2016

Data were extracted from G-BA databases; added benefit was defined in various ways; in all models, they controlled for additional criteria such as size of patient population, European price levels, and whether the comparators were generic.

Authors conclude that price premiums were driven by health gain, the proportion of people benefitting from a pharmaceutical, European price levels, and whether the comparator was generic. QoL did not play a role in current decision making

No limitations stated by author(s)

### 3.22 Leverkus and Chuang-Stein 2016 (Medium)

To investigate requirements of benefit assessment with special attention on: choice of the comparator, patient relevant endpoints, subgroup analyses, extent of benefit, determination of net benefit, primary and secondary endpoints, and uncertainty of the additional benefit.

**Germany**

Authors state they contrast the approaches taken by the G-BA and IQWiG with those of the European Medicines Agency (EMA).


For principles underlying regulatory decisions, they reference primarily the International Conference on Harmonization (ICH) E9 (Statistical Principles for Clinical Trials, 1998) document.

Provides comprehensive overview and opinion on methodological requirements and issues in German HTA process, with particular focus on the role of outcomes and evidence types

No limitations stated by author(s)

### 3.23 Lohberg et al 2016 (High)

To analyse how QoL is defined in early benefit assessment (EBA) and which role does it play

**Germany**

Qualitative analysis all benefit assessments completed by the end of 2013 were processed. Additionally, data on the decision outcomes were collected and analysed

Publicly available dossiers (summaries), dossier evaluations, protocols of the oral hearings, the final resolutions of the Federal Joint Committee (G-BA) and main

Documents were imported to software and searched for QoL terms; resulting paragraphs were reduced and summarized by two researchers; coding was performed on the basis of summaries

QoL has not been well defined in HTA processes and does not inform final decisions; they identified the absence or the inappropriate presentation of QoL data; at the same the stakeholders saw the value and importance of including QoL in EBA

No limitations stated by author(s)
<table>
<thead>
<tr>
<th>Study</th>
<th>Title</th>
<th>Methods</th>
<th>Findings</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.24 Riedel et al 2014 (Low)</td>
<td>To explain some fundamental concepts in Health Economic Evaluations (HEE) and how these concepts are adapted in different countries, notably in Germany</td>
<td>Bibliographic search to identify existing methods of health economic evaluation of new drugs used by HTA agencies in 11 countries and comparison with German HTA agency</td>
<td>Although the core principles of HEE are very similar worldwide, there is a lack of harmonization. Overcoming the fourth hurdle (the reimbursement hurdle) is likely to be increasingly challenging for new drugs.</td>
<td>No limitations stated by author(s)</td>
</tr>
<tr>
<td>3.25 Ruof et al 2014 (a) (High)</td>
<td>To analyse the outcomes 18 months after introduction of the new AMNOG legislation on early benefits assessments (EBA)</td>
<td>All EBAs commenced prior to June 2012 were included and analysed (proportions were calculated; no statistical analysis was carried out)</td>
<td>Considerable variance was observed in additional benefit reported by manufacturers, IQWiG and G-BA. Areas of disagreement included comparator selection, definition of subgroups and patient-relevant endpoints, and classification and balancing of adverse events.</td>
<td>No limitations stated by author(s)</td>
</tr>
<tr>
<td>3.26 Ruof et al 2014 (b) (High)</td>
<td>To compare endpoints and related benefit categories used in marketing authorisation to those considered by G-BA in the field of oncology</td>
<td>Evaluation of early benefit assessments (EBAs) in oncology commencing prior to 31 December 2013</td>
<td>Inconsistencies in acceptance of morbidity and QoL outcomes between G-BA and EMA; EMA accepted well established and clinically relevant morbidity endpoints (e.g. progression-free survival and response rate), which were mostly excluded by G-BA; final decisions by G-BA mostly driven by mortality outcomes</td>
<td>No limitations stated by author(s)</td>
</tr>
<tr>
<td>3.27 Staab et al 2016 (High)</td>
<td>To evaluate the acceptance of clinically medicines for oncological, metabolic and infectious diseases with EBAs</td>
<td>Manufacturer’s dossiers, regulatory documents were analysed to determine patient</td>
<td>Inconsistencies were identified in patient relevance of morbidity-related PEPs as well as in</td>
<td>No limitations stated by author(s)</td>
</tr>
</tbody>
</table>
For Peer Review

acknowledged primary endpoints (PEPs) from regulatory trials in early benefit assessments (EBAs) conducted by the Federal Joint Committee (G-BA) finalised before 25 January 2016 were evaluated. Assessments, G-BA appraisals and oral hearing minutes were reviewed, and PEPs relevance of outcomes from G-BA perspective; acceptance of symptomatic vs. asymptomatic outcomes were also analysed. Acceptance of asymptomatic endpoints by the G-BA

Netherlands (N=6)

| 3.28 Angelis et al 2017 (High) | To study the practices, processes and policies of value-assessment for new medicines across eight European countries and the role of HTA beyond economic evaluation and clinical benefit assessment | French, Germany, England, Sweden, Italy, Netherlands, Poland and Spain | A systematic (peer review and grey) literature review was conducted using an analytical framework examining: (1) ‘Responsibilities and structure of HTA agencies’; (2) ‘Evidence and evaluation criteria considered in HTAs’; (3) ‘Methods and techniques applied in HTAs’; and (4) ‘Outcomes and implementation of HTAs’ | Two electronic databases (MEDLINE—through PubMed resource—and the Social Science Citation Index—through the Web of Science portal) were searched up to January 2014; with article searches taking place in February 2013 in the first instance and update taking place at the end of January 2014 | Systematic literature review method based on the Centre for Reviews and Dissemination (CRD) guidance. Feedback from the Advance-HTA consortium partners was provided in August 2014. Additional input, including the most recent updates on national HTA processes, was collected from HTA experts and national competent authorities between March and May 2016. | Debates about health utilities/preferred health gain; for example, while NICE favours the use of the QALY, IQWiG strongly opposes its use on the grounds that it does not reflect patient-level utilities. Increasing use of incorporating real world data; considerable subjectivity in the criteria selection used to interpret evidence and determine product value; increasing realisation by many HTA agencies that value is multi-dimension; move away from only relying on ‘scientific value judgments’ (safety/efficacy/effectiveness); need for methodological approaches that encompass multiple evaluation criteria explicitly. | No limitations stated by author(s) |

| 3.29 Cerri et al 2014 (Medium) | To examine the factors that influence decisions made by the Dutch HTA agency (CVZ) to recommend, restrict or not recommend pharmaceutical | Netherlands | Descriptive statistics for each variable, stratified by outcome group (recommended, restricted or not recommended); chi-squared test for categorical variables; ANOVA test for continuous variables; CVZ decisions in 2004–2009. A data set of CVZ decisions pertaining to pharmaceutical technologies was created, including 29 variables | Technologies included in list 1A/1B or on the expensive drug list considered recommended; those included in list 29 variables | The multinomial model showed significant associations (p < 0.10) between CVZ outcome and several variables, including: (1) use of an active comparator and demonstration of statistical superiority of the primary endpoint in clinical trials, (2) pharmaceutical budget impact. | Reliance on publicly available data sources; data extraction performed by single researcher (under supervision from senior researchers) |
| **3.30 Franken et al 2013**  
**Medium** | **3.31 Le Polain et al 2010**  
**Medium** |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>technologies for use in the Netherlands</strong></td>
<td><strong>To describe and critically evaluate drug reimbursement decision processes, to identify their strengths and weaknesses and to formulate general policy recommendations.</strong></td>
</tr>
</tbody>
</table>
| **Kruskal-Wallis for not normally distributed indicators.**  
A multinomial logit regression was used in the analysis to model the probabilities associated with the three types of technology appraisal outcome. | **Comparative study (1) for the description of drug reimbursement decision processes, authors used the Hutton framework; (2) systems were evaluated using accountability for reasonableness framework by Daniels and Sabin.** |
| **extracted from published information.** | **Literature, policy documents and interviews with stakeholders** |
| **2 were considered restricted;** | **The paper provides a wide range of information on assessment and appraisal processes of Dutch HTA, and draws conclusions about criteria: For example, although there is no formal hierarchy in assessment criteria, most interviewees stated that effectiveness, efficacy and side effects were often the most important criteria determining the therapeutic value. Interviewees also acknowledged that the majority of time in a meeting of the Dutch HTA is devoted to determining the therapeutic value, less time is spent on** |
| **associated with introduction of the technology, (3) therapeutic indication and (4) prevalence of the target population. Results confirm the value of a comprehensive and multivariate approach to understanding CVZ decision-making.** | **Analysis took place in supply-driven context; it was beyond the scope of this study to explore opportunities to move towards demand-driven system, where the societal needs drive the industry’s strategic plan.** |

Authors investigated all reimbursement dossiers published in the period January 2005 to July 2011.  
Data sources included all Dutch and Swedish drug reimbursement information published in the period January 2005 to July 2011.  
The analysis started in 2005 because that was the first year in which pharmacoeconomic evidence was required for reimbursement decision making in The Netherlands.  
Therapeutic value appeared to be the most decisive criterion; the relative importance of full economic evaluations is more modest than would generally be expected, especially in The Netherlands; both countries could make the appraisal process more transparent by more explicitly showing the role of different criteria.  
Reliance on publicly available data sources.

To investigate the role of pharmacoeconomic evidence in drug reimbursement decision making; and (ii) to determine the extent to which appraising the importance of full economic evaluations relative to other evidence is a transparent process.  

Austria, Belgium, France, the Netherlands and Sweden  

Analysis took place in supply-driven context; it was beyond the scope of this study to explore opportunities to move towards demand-driven system, where the societal needs drive the industry’s strategic plan.
To review the current approach to HTA used in The Netherlands in medical specialist care; the authors seek to provide a basic understanding of the strengths and weaknesses of the specific practices and processes.

Authors explore trends in future of (Dutch) HTA: What can be expected is a growing incentive for all parties to generate HTA data; increasing trend for conditional reimbursement linked to requirements for data collection and further study; further work is needed to understand how assessments and procedures jointly affect decision-making and to develop best practice guidelines; broader appraisals might be needed where the assessment will also cover optimal positioning of a service amongst the variety of services available to patients.

In this editorial, the authors highlight the distinguishing features of the new Dutch guidelines for economic evaluation; and highlight which developments, in their opinion, are desirable in coming updates, but are still in development or controversial.

New guidelines set preference for QALYs measured with the EQ-5D if appropriate but also offer alternative approaches for areas in which QoL might not be appropriate such as: prevention; diagnostics; medical devices; long-term care; forensics; reference is also made to multi-criteria decision making.
References


Griffiths EA. The German NICE or the German nasty? An analysis of IQWIG decisions and requirements for an 'added benefit'. Value Health. 2015;18:A335.


## Supplement 4: HTA documents analysed for case studies

### National Institute for Health and care Excellence (NICE), England

**Draft documents for consultation**

- **4.1** Health Technology Appraisal Donepezil, galantamine, rivastigmine and memantine for the treatment of mild to moderate Alzheimer’s disease (Part review of TA 111) Draft scope

**Final documents**

- **4.2** Alzheimer’s disease - donepezil, galantamine, rivastigmine and memantine (review): appraisal consultation document (online)

**Reports by the Assessment group**

- **4.3** Health Technology Appraisal Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer’s disease (Review of TA 111) Final Scope

- **4.4** Final Appraisal Determination Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer’s disease (review of NICE technology appraisal guidance 111)

**Comments to Technology Assessment Report (TAR)**

- **4.5** The effectiveness and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer’s disease (review of TA111): a systematic review and economic model, Produced by: Peninsula Technology Assessment Group (PenTAG), University of Exeter [Note that this includes a revised section on results]

- **4.6** Overview Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer’s disease (Review of NICE technology appraisal guidance 111)

**Responses by Assessment Group**

- **4.7** Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer’s disease (Review of TA 111), Responses by various stakeholders including: Eisai/Pfizer; NHS Quality Improvement Scotland; NHS West Kent and NHS Islington; Novartis; Shire Pharmaceuticals; Alzheimer’s Society; RICE (The Research Institute for the Care of Older People); Lundbeck

**Submissions**

- **4.11** Various submissions including by manufacturers and other stakeholders e.g. Alzheimer’s Society Report; British Geriatrics Society; Royal College of Psychiatrists (Faculty of old age psychiatry); NHS Quality Improvement Scotland

### Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Germany

**Cholinesterase Inhibitors: Donepezil, Galantamin, Rivastigmin**


- **4.13** Amendment 1 zum Berichtsplan „Cholinesterasehemmer bei Alzheimer Demenz“, [Auftrag A05/19A], 12.06.2006; Amendment 1 to Report Plan version 1.0. Last accessed 10th January 2018


- **4.16** IQWiG. Cholinesterasehemmer bei Alzheimer Demenz. Abschlussbericht A05-19A. Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG); Februar 2007. Last accessed 10th January 2018 [⇒ this is the German version of 4.15; in addition to 4.15 it includes the documented stakeholder involvement through meeting and written consultation]
Memantine


4.18 Amendment 1 zum Berichtsplan „Memantin bei Alzheimer Demenz“ [Auftrag A05/19C], 12.06.2006; Amendment 1 to the report plan version. Last accessed 10th January 2018

4.19 Amendment 2 zum Berichtsplan Memantin bei Alzheimer Demenz, Auftrag A05-19C Version 1.0 Stand: 06.08.2007; Amendment 2 to the report plan version. Last accessed 10th January 2018

4.20 Memantin bei Alzheimer Demenz, Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan, Auftrag A05-19C Version 1.0 Stand: 11.02.2008; documentation and appraisal of comments on the report plan version 1.0. Last accessed 10th January 2018


4.22 Memantin bei Alzheimer Demenz Vorbericht (vorläufige Nutzenbewertung), Auftrag A05-19C Version 1.0 Stand: 01.08.2008; Preliminary report. Last accessed 10th January 2018


Zorginstituut Nederland, previously: College voor Zorgverzekeringen (CVZ), Netherlands

4.26 CFH rapport 07/11 memantine (Ebixa®), (2e)herbeoordeling, Op 2 april 2007 uitgebracht aan de minister van Volksgezondheid, Welzijn en Sport

4.27 GVS-rapport 13/11 donepezil (hydrochloride) Aspen® Vastgesteld op 24 juni 2013, College voor zorgverzekeringen, Diemen.

4.28 Farmacotherapeutisch rapport rivastigmine (Exelon*) bij Parkinsondementie, 2006

Supplement 5: List of stakeholders involved in HTAs as identified in case studies

**ENGLAND**

<table>
<thead>
<tr>
<th>Stakeholder group: Manufacturers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accord Healthcare (donepezil)</td>
<td>Novartis (rivastigmine)</td>
</tr>
<tr>
<td>Aspire Pharma (galantamine, rivastigmine)</td>
<td>Pfizer (donepezil)</td>
</tr>
<tr>
<td>Actavis UK (all four drugs)</td>
<td>Ranbaxy (donepezil)</td>
</tr>
<tr>
<td>Consilient Healthcare (galantamine, memantine)</td>
<td>Sandoz (all four drugs)</td>
</tr>
<tr>
<td>Dr Reddy’s Laboratories (all but galatamine)</td>
<td>Shire (galantamine)</td>
</tr>
<tr>
<td>Eisai (donepezil)</td>
<td>Teva UK (all four drugs)</td>
</tr>
<tr>
<td>Lundbeck Ltd (memantine)</td>
<td>Wockhard UK (donezepil)</td>
</tr>
<tr>
<td>Mylan (galantamine, memantine)</td>
<td>Zentiva UK (all but rivastigmine)</td>
</tr>
</tbody>
</table>

**Stakeholder group: Patient/ carer groups**
<table>
<thead>
<tr>
<th>Stakeholder Group: Professional Associations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Association of British Neurologists</td>
<td>Royal College of General Practitioners</td>
</tr>
<tr>
<td>Association of Directors of Adult Social Services</td>
<td>Royal College of Nursing</td>
</tr>
<tr>
<td>British Geriatrics Society</td>
<td>Royal College of Pathologists</td>
</tr>
<tr>
<td>British Neuropathological Society</td>
<td>Royal College of Physicians</td>
</tr>
<tr>
<td>British Neuropsychiatry Association</td>
<td>Royal College of Psychiatrists</td>
</tr>
<tr>
<td>College of Mental Health Pharmacy</td>
<td>Royal Pharmaceutical Society</td>
</tr>
<tr>
<td>Dementia Action Alliance</td>
<td>Royal Society of Medicine</td>
</tr>
<tr>
<td>Institute of Neurology</td>
<td>United Kingdom Clinical Pharmacy Association</td>
</tr>
<tr>
<td>Primary Care Neurology Society</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Others</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Department of Health</td>
<td>NHS South Eastern Hampshire CCG</td>
</tr>
<tr>
<td>NHS England</td>
<td>Welsh Government</td>
</tr>
<tr>
<td>NHS Somerset CCG</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>English Translation or Description</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bundesverband für Gesundheitsinformation und Verbraucherschutz e. V.</td>
<td>Association for health information for the public and consumer protection (charity)</td>
<td></td>
</tr>
<tr>
<td>Deutsche Alzheimer Gesellschaft e. V.</td>
<td>German charity for Alzheimer</td>
<td></td>
</tr>
<tr>
<td>Deutsche Gesellschaft für Neurologie; Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde</td>
<td>Professional association for psychiatry, psychotherapy and neurology</td>
<td></td>
</tr>
<tr>
<td>Hirnliga e.V.</td>
<td>Charity for the brain; refers to Alzheimer</td>
<td></td>
</tr>
<tr>
<td>Institut für angewandte Statistik</td>
<td>Institute for applied statistics</td>
<td></td>
</tr>
<tr>
<td>Institut für Arzneimittelsicherheit in der Psychiatrie</td>
<td>Institute for safety of psychiatric drugs</td>
<td></td>
</tr>
<tr>
<td>Karolinska Institutet</td>
<td>Swedish medical university</td>
<td></td>
</tr>
<tr>
<td>Kompetenznetz Demenz</td>
<td>Network for researchers, clinicians, people living with Alzheimer and their families</td>
<td></td>
</tr>
<tr>
<td>Lundbeck GmbH</td>
<td>Pharma company</td>
<td></td>
</tr>
<tr>
<td>Merz Pharmaceuticals GmbH</td>
<td>Pharma company</td>
<td></td>
</tr>
<tr>
<td>Novartis Pharma GmbH</td>
<td>Pharma company</td>
<td></td>
</tr>
<tr>
<td>The Research Institute for the Care of Older People (RICE)</td>
<td>/</td>
<td></td>
</tr>
<tr>
<td>Verband Forschender Arzneimittelhersteller e. V. (VFA)</td>
<td>Association of pharma companies involved in research</td>
<td></td>
</tr>
<tr>
<td><strong>Verein zur Förderung der Forschung auf dem Gebiet der experimentellen Neurologie</strong></td>
<td>Association to promote research in neurology</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>Technology Assessment for Cholinesterase inhibitors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eisai GmbH</td>
<td>Pharma</td>
<td></td>
</tr>
<tr>
<td>Novartis GmbH</td>
<td>Pharma</td>
<td></td>
</tr>
<tr>
<td>Pfizer GmbH</td>
<td>Pharma</td>
<td></td>
</tr>
<tr>
<td>Janssen-Cilag GmbH</td>
<td>Pharma</td>
<td></td>
</tr>
<tr>
<td>Merz Pharmaceuticals GmbH</td>
<td>Pharma</td>
<td></td>
</tr>
<tr>
<td>Verband Forschender Arzneimittelhersteller e.V.</td>
<td>Association of Pharma Companies involved in Research</td>
<td></td>
</tr>
<tr>
<td>University of Manchester</td>
<td>University, England (UK)</td>
<td></td>
</tr>
<tr>
<td>Alzheimer-Ethik e.V.</td>
<td>Charity for Alzheimer, founded by carers</td>
<td></td>
</tr>
<tr>
<td>Universitätsparklinikum Freiburg</td>
<td>University</td>
<td></td>
</tr>
<tr>
<td>Deutsche Gesellschaft für Gerontologie und Geriatrie</td>
<td>German Society of Gerontology and Geriatrics</td>
<td></td>
</tr>
<tr>
<td>Arznei-Telegramm</td>
<td>News magazine about drugs</td>
<td></td>
</tr>
<tr>
<td>Deutsche Gesellschaft f. Gerontopsychiatrie und –psychotherapie (DGGPP) e. V.</td>
<td>German Psychogeriatric Association</td>
<td></td>
</tr>
<tr>
<td>Deutsche Alzheimer Gesellschaft e. V.</td>
<td>German Alzheimer Association (charity)</td>
<td></td>
</tr>
<tr>
<td>Universitätsparklinikum Hamburg-Eppendorf</td>
<td>Medical University in Hamburg, Germany</td>
<td></td>
</tr>
<tr>
<td>Kompetenznetz Demenzen</td>
<td>Network for dementia</td>
<td></td>
</tr>
<tr>
<td>Hirnliga e.V.</td>
<td>Charity for the Brain, specifically Dementia</td>
<td></td>
</tr>
<tr>
<td>Bezirkskrankenhaus Günzburg</td>
<td>Hospital</td>
<td></td>
</tr>
<tr>
<td>Institut für Klinische Pharmakologie, Klinikum Bremen-Mitte</td>
<td>Pharmacological Institute, Medical university</td>
<td></td>
</tr>
</tbody>
</table>
Table 1: Features of health technology assessment (HTA) agencies in England, Germany, and The Netherlands

<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>ENGLAND</th>
<th>GERMANY</th>
<th>NETHERLANDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTA Agency’s name</td>
<td>National Institute for Health and Care Excellence (NICE)</td>
<td>Institute for Quality and Efficiency in Health Care (IQWiG)</td>
<td>Zorginstituut Netherlands (ZIN)</td>
</tr>
<tr>
<td>Roles &amp; responsibilities</td>
<td>Advice about which drugs get funded based on cost-effectiveness (=fourth hurdle)</td>
<td>Advice on added value used to inform price negotiations</td>
<td>Advice informs whether drug gets on positive list (which influences price)</td>
</tr>
<tr>
<td>Value assessment</td>
<td>Cost-effectiveness</td>
<td>Clinical benefit</td>
<td>Clinical benefit; cost-effectiveness for drugs above certain threshold</td>
</tr>
<tr>
<td>Enforcement of methods</td>
<td>Guidance (manual) and reference case</td>
<td>Legislation and guidance</td>
<td>Guidance and reference case</td>
</tr>
<tr>
<td>Stakeholder involvement</td>
<td>Public consultations, Committee of experts (including patient representatives) co-develops guideline</td>
<td>Consultation with experts (including patient representatives)</td>
<td>Consultation with experts (including patient representatives)</td>
</tr>
</tbody>
</table>
Table 2: Clinical scales used in studies identified in health technology assessments (HTAs) of Alzheimer’s disease drugs in England, Germany and The Netherlands

<table>
<thead>
<tr>
<th>Drug covered in HTAs</th>
<th>Memantine</th>
<th>Donezepil</th>
<th>Galantamine</th>
<th>Rivastigmine</th>
<th>Different cholinesterase inhibitors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>England</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognition</td>
<td>SIB, ADAS-cog</td>
<td>ADAS-cog, MMSE, SIB scales</td>
<td>ADAS-cog</td>
<td>ADAS-cog, MMSE</td>
<td>MMSE, SIB</td>
</tr>
<tr>
<td>Functioning</td>
<td>ADCS-ADL, FAST</td>
<td>ADL</td>
<td>ADCS-ADL, DAD, GAS</td>
<td>PDS, ADCS-ADL, ADL</td>
<td>ADL</td>
</tr>
<tr>
<td>Behavioural functioning and psychological symptoms</td>
<td>NPI, NPI nursing home version, BGP</td>
<td>NPI, in particular aggression and agitation</td>
<td>NPI</td>
<td>NPI, Hamilton DS</td>
<td>NPI</td>
</tr>
<tr>
<td>Global assessment</td>
<td>CIBIC-plus</td>
<td>CDR, CDR-SB, CIBIC-plus, GBS</td>
<td>CIBIC-plus</td>
<td>CIBIC-plus, GDS, ADCS-CGIC</td>
<td>GDS</td>
</tr>
<tr>
<td><strong>Germany</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognition</td>
<td>SIB, MMSE, BGP-C</td>
<td>ADAS-cog, MMSE, SIP (for people with Down syndrome)</td>
<td>ADAS-cog</td>
<td>ADAS-cog</td>
<td>ADAS-cog, SIB, MMSE</td>
</tr>
<tr>
<td>Functioning</td>
<td>ADCS-ADL-sev, BGP, BGP-D</td>
<td>IDDD, PSMS-plus, IADL-plus, DAD, CMCS, PDS</td>
<td>DAD, ADCS-ADL, GAS</td>
<td>PDS</td>
<td>BADLS, ADCS-AL, PMS, Blessed-Roth Dementia Scale</td>
</tr>
<tr>
<td>Behavioural functioning and psychological symptoms</td>
<td>NPI</td>
<td>NPI, NPI nursing home</td>
<td>NPI</td>
<td>NOSGER, BEHAVE-AD</td>
<td>NPI, NPI-D, BEHAVE-AD</td>
</tr>
<tr>
<td>Global assessment</td>
<td>CIBIC-plus, ADCS-CGIC, GDS, FAST</td>
<td>CIBIC-plus, J-CGIC, GBS, CDR, CDR-SB</td>
<td>CIBIC-plus</td>
<td>CIBIC-plus</td>
<td>GDS</td>
</tr>
<tr>
<td>(information provided in supplement)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>The Netherlands</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognition</td>
<td>ADAS-cog, SPI</td>
<td>SIB, MMSE</td>
<td>ADAS-cog, MMSE</td>
<td>ADAS-cog, ADCS-CGIC MMSE</td>
<td>ADAS-cog, ADCS-CGIC MMSE</td>
</tr>
</tbody>
</table>
### Functioning

<table>
<thead>
<tr>
<th>Behavioural functioning and psychological symptoms</th>
<th>ADCS-ADL</th>
<th>ADCS-ADL</th>
<th>BrADL</th>
<th>ADCS-ADL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPI</td>
<td>NPI</td>
<td>NPI-10</td>
<td>NPI</td>
<td>NPI-10</td>
</tr>
</tbody>
</table>

### Global Assessment

| CIBIC-PLUS | GDS | / | / |

Index of abbreviations used in Table: Activities of daily living (ADL); Alzheimer’s Disease Assessment Scale – cognitive subscale (ADAS-cog); Alzheimer’s Disease Cooperative Study – Activities of Daily Living (ADCS-ADL); the mild cognitive impairment ADL scale (ADCS-MCIADL); Allocation of caregiver time burden (ACTS); Behavioural Rating Scale for Geriatric Patients (BGP); Behavioral Pathology in Alzheimer's Disease Rating Scale (BEHAVE-AD) and the Behavioral Rating Scale for Dementias (BRSD); Behavioral Rating Scale for Geriatric Patients – Cognitive Subscale (BGP-C); Bristol Activities of daily Living Scale (BrADL); Clinical dementia rating (CDR); CDR sum of boxes (CDR-SB); Clinician’s Interview-based Impression of Change (CIBIC)-plus; Caregiver Activity Survey (CAS); Caregiving burden scale (CBS); Caregiver stress Scale (CSS); Caregiver-rated Modified Crichton Scale (CMCS); Disability Assessment for Dementia (DAD); Functional Assessment Staging (FAST); Goal Attainment Scale (GAS); Global Deterioration Scale (GDS); Gottfried, Brine and Steen scale (GBS); Interview for Deterioration in Daily Living Activities in Dementia (IDDD), subscale; Instrumental Activities of Daily Living (IADL)-plus; Japanese-Clinical Global Impression of Change (J-CGIC); Mini Mental State Examination (MMSE), Neuropsychiatric Inventory (NPI); NOSGER (Nurses’ Observation Scale for Geriatric Patients); Neuropsychiatric Inventory Caregiver Distress Scale (NPI-D); Neuropsychiatric inventory (NPI); Progressive Deterioration Scale (PDS); Severe Impairment Battery (SIB) scales; Physical Self-Maintenance Scale (PSMS)-plus; Severe Impairment Battery (SIB)
Table 3 Prioritization of outcomes and outcome measures in health technology assessments (HTAs) of Alzheimer’s disease (AD) drugs

<table>
<thead>
<tr>
<th>Outcome</th>
<th>ENGLAND</th>
<th>GERMANY</th>
<th>THE NETHERLANDS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost-effectiveness</strong></td>
<td>Prioritized Economic modelling conducted</td>
<td>Not prioritized No economic modelling conducted</td>
<td>Not prioritized No economic modelling conducted</td>
<td></td>
</tr>
<tr>
<td><strong>Quality of life</strong></td>
<td>Prioritized Derived from outcomes measured with clinical scales to inform the economic modelling</td>
<td>Not prioritized Method requirements (e.g. 70% follow up rate) prevented influence of outcome on decisions</td>
<td>Not prioritized No assessment</td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes measured with clinical scales</strong></td>
<td>Prioritized Cognition (including Mini-Mental State Examination; MMSE), Activities of daily living (ADL) and multi-domain change prioritized</td>
<td>Prioritized Global assessment outcomes less prioritized</td>
<td>Prioritized Mini-Mental State Examination (MMSE) less prioritized / not accepted</td>
<td></td>
</tr>
<tr>
<td><strong>Adverse effects</strong></td>
<td>Not prioritized Formally included but not referred to in decision</td>
<td>Prioritized Included and referred to in decision</td>
<td>Prioritized Included and referred to in decision</td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes relevant to people with Alzheimer's Disease</strong></td>
<td>Not prioritized Focus on evidence from randomised controlled trials that measure outcomes with clinical scales</td>
<td>Not prioritized Focus on evidence from randomised controlled trials that measure outcomes with clinical scale</td>
<td>Not prioritized Focus on evidence from randomised controlled trials that measure outcomes with clinical scale</td>
<td></td>
</tr>
<tr>
<td><strong>Carers' outcomes</strong></td>
<td>Prioritized Included as primary endpoint but lower priority than outcomes for person with AD; methodological challenges prevent influence of outcome on decisions</td>
<td>Not prioritized Not considered responsibility of health care system; relevant only if impact on patient outcomes</td>
<td>Not prioritized Potential role in economic modelling but not carried out so far for AD drugs</td>
<td></td>
</tr>
<tr>
<td><strong>Institutionalization</strong></td>
<td>Prioritized Included in economic modelling (by linking outcomes with clinical scales to non-AD data sets)</td>
<td>Not prioritized Absence of trial evidence; method requirements also prevent influence of outcome on decisions</td>
<td>Not prioritized Not included due to absence of trial evidence</td>
<td></td>
</tr>
</tbody>
</table>