Diagnosis and Treatment of Clinical Alzheimer's-Type Dementia: A Systematic Review

Evidence Summary

Background

The ultimate reason for accurately diagnosing clinical Alzheimer's-type dementia (CATD) and whether Alzheimer's disease (AD) is the underlying neuropathological etiology is to inform decision making about drug and nondrug treatments to improve patient and caregiver outcomes.

In individuals with suspected cognitive impairment, comprehensive neuropsychological testing may help clinically diagnose dementia and distinguish between dementia subtypes. However, such testing is time consuming and access is limited in some clinical settings. Therefore, we need better understanding in this population with suspected cognitive impairment (case finding) which brief cognitive tests and test combinations most accurately distinguish patients with CATD from those with normal cognition or mild cognitive impairment (MCI), and whether patient characteristics affect test classification accuracy.

Additionally, many individuals clinically diagnosed with CATD do not meet neuropathologic (gold standard) criteria for AD on post-mortem brain autopsy. Therefore, we also need better understanding of how accurate pre-mortem brain imaging and cerebrospinal

Purpose of Review

To summarize evidence on cognitive test accuracy for clinical Alzheimer's-type dementia (CATD) in suspected cognitive impairment; biomarker accuracy for Alzheimer's disease (AD) in dementia; and effects of CATD drug treatment.

Key Messages

- Many brief cognitive tests were highly (>0.8) sensitive and specific distinguishing CATD from normal cognition, but less from mild cognitive impairment.

- Amyloid PET and MRI were highly sensitive and specific distinguishing autopsy-confirmed AD from non-AD dementia; FDG-PET was highly sensitive and moderately (>0.5 to <0.8) specific; CSF tests were moderately sensitive and specific. Data were limited on biomarkers added to clinical evaluation.

- Cholinesterase inhibitors (ChI) were slightly better than placebo for cognition and function, but increased withdrawals due to adverse effects; evidence was insufficient for supplements. In moderate to severe CATD, memantine plus ChI slightly improved cognition versus ChI, but not function.

- Donepezil and antidepressants appeared similar to placebo for agitation and depression, respectively; for other prescription drugs and all supplements, evidence was insufficient on behavioral and psychological symptoms.

Agency for Healthcare Research and Quality

fluid (CSF) biomarkers are for distinguishing patients whose dementia is due to AD from those with non-AD dementia, and whether classification accuracy varies depending on patient characteristics.

Finally, although only a few prescription drugs are approved by the U.S. Food and Drug Administration (FDA) for CATD, many supplements are promoted for cognition and function. In addition, many prescription drugs are used off-label for CATD-associated behavioral and psychological symptoms of dementia (BPSD), including antipsychotics, despite FDA black box warnings about their increased mortality risk in this population.[1, 2] Less is understood about the beneficial and harmful effects of supplements for CATD-associated BPSD. To guide CATD treatment decisions for cognition, function, BPSD and other outcomes, we need to clarify the benefits and harms of prescription drugs and supplements in this population.

Purpose

The target audiences of this report are primary care clinicians who diagnose and treat the vast majority of older patients with cognitive disorders, psychologists who may perform additional cognitive testing in primary care settings, and dementia specialists who are most likely to have access to biomarker testing for further diagnostic clarification.

The purpose of this report is to summarize the evidence on (1) the accuracy of brief cognitive tests for distinguishing CATD from normal cognition and MCI in individuals with suspected cognitive impairment; (2) the accuracy of brain imaging and CSF biomarkers for distinguishing autopsy-confirmed AD from non-AD in individuals with dementia; (3) the benefits and harms of prescription drugs and supplements for cognition, function, and BPSD in patients with CATD; and (4) whether the accuracy of cognitive or biomarker tests for classifying patients and the

efficacy of CATD drug treatments vary by patient characteristics.

Main Points

- **Accuracy of brief cognitive tests for distinguishing CATD from normal cognition and MCI**:
  - Multiple brief cognitive tests had high sensitivity and specificity (defined as >0.8) for distinguishing CATD from normal cognition, including those commonly used as individual stand-alone tests, brief multidomain batteries, and individual memory and verbal fluency tests typically administered as part of a larger battery in clinical practice. These tests less accurately distinguished CATD from MCI, or mild CATD from normal cognition.
  - Few cognitive tests were evaluated in multiple studies which reported the same type of test score and used comparable cut points to define abnormality, and few studies compared classification accuracy between individual tests or their combinations.
  - There was minimal evidence addressing whether accuracy of brief cognitive tests for identifying CATD varied by study participant characteristics.

- **Accuracy of biomarkers for distinguishing autopsy-confirmed AD from non-AD dementia**:
  - Amyloid positron emission tomography (PET) brain imaging was highly sensitive and specific, and fluorodeoxyglucose (FDG)-PET was highly sensitive and moderately specific (latter defined as >0.5 to <0.8); based on single studies making direct comparisons, both may increase accuracy differentiating between AD and non-AD dementia when added to a clinical evaluation.

-
- Magnetic resonance imaging (MRI) medial temporal atrophy was highly sensitive and specific and single-photon emission computed tomography (SPECT) cerebral blood flow had variable accuracy; SPECT plus clinical evaluation had lower sensitivity and higher specificity than clinical evaluation alone, but no studies directly compared MRI plus clinical evaluation versus clinical evaluation alone.
- Individual CSF tests and ratios were moderately sensitive and specific; in the few direct comparisons, beta amyloid 42 (Aß42)/p-tau ratio, t-tau/Aß42 ratio and p-tau appeared more accurate and Aß42 and t-tau appeared least accurate.
- Combinations of CSF tests may have the highest mix of sensitivity and specificity and may increase accuracy for distinguishing AD from frontotemporal lobar degeneration (FTLD) when added to clinical evaluation.
- There was minimal evidence addressing whether the accuracy of biomarker testing for identifying AD varied by study participant characteristics.
- No studies reported data on the accuracy of blood tests for identifying autopsy-confirmed AD.

- **Efficacy and harms of prescription drug treatment for CATD**:
  - In adults with mild to moderate CATD—
    - Cholinesterase inhibitors compared with placebo produced small improvements in cognition, function, staging, and clinical impression of change, but standard doses may increase serious adverse events and withdrawals due to adverse events.
    - In patients not receiving cholinesterase inhibitors, memantine may improve clinical impression of change, did not improve function, and evidence was insufficient about cognition, other efficacy outcomes, and harms.
    - In patients receiving cholinesterase inhibitors, memantine did not improve clinical impression of change, and evidence was insufficient for cognition, function, staging, and harms.
  - In adults with moderate to severe CATD—
    - Cholinesterase inhibitors produced small improvements in cognition, function, and clinical impression of change, but standard doses may increase serious adverse events and withdrawals due to adverse events.
    - In patients receiving cholinesterase inhibitors, memantine improved scores on brief multidomain cognitive batteries and clinical impression of change, did not improve function, and evidence was insufficient for brief cognitive tests commonly used as individual stand-alone tests, staging, and harms.
  - In adults with CATD and BPSD—
    - Evidence was insufficient to draw conclusions about the efficacy of antipsychotics, antidepressants, anticonvulsants, or memantine compared with placebo for agitation, psychosis, aggression or disinhibited sexual behavior (or of estrogen for disinhibited sexual behavior).
    - Cholinesterase inhibitors did not improve agitation more than placebo, and evidence was insufficient to draw conclusions about their effects on other BPSD or harms.
  - There was minimal evidence addressing whether efficacy of prescription drugs for CATD treatment varied by study participant characteristics.

- **Efficacy and harms of supplements for CATD**:
  - In adults with CATD—
    - Omega-3 fatty acids did not improve cognition, and the nutritional drink Souvenaid® did not improve function; evidence for both was insufficient for other outcomes.
    - Evidence was insufficient to draw conclusions about differences in efficacy and harms of gingko biloba versus donepezil, or saffron extract versus memantine, for cognition, function, or quality of life.
    - Evidence was insufficient about efficacy and harms of other supplements, including antioxidants, gingko biloba, ginseng, curcumin, and vitamin E, for cognition, function, BPSD, or other efficacy outcomes.
    - There was minimal data addressing whether efficacy of supplements for CATD treatment varied by study participant characteristics.

Methods

We used methods consistent with those outlined in the Agency for Healthcare Research and Quality Evidence-based Practice Center Program Methods Guidance (https://effectivehealthcare.ahrq.gov/topics/cer-methods-guide/overview), and we describe these in the full report. Our cognitive testing and drug treatment searches covered from database inception to March 2019, and the biomarker testing search covered from 2012 to March 2019.

Results

Brief cognitive tests for distinguishing CATD from normal cognition or MCI in adults with suspected cognitive impairment. Fifty-six unique, low or medium risk-of-bias (ROB) studies evaluated the accuracy of one or more brief cognitive tests for distinguishing CATD from normal cognition or MCI, including 26 of individual tests commonly used as stand-alone tests (n=6,953); ten of brief multidomain batteries (n=2,676); 17 of individual memory tests (n=4,061), five of individual executive function tests (n=1,167), and ten of individual language tests (n=1,676), all typically administered as part of a larger battery in clinical practice; and nine of test combinations (n=1,688). Some epidemiological cohorts were frequently used, and the extent of participant overlap across studies was unknown. Results for the most commonly evaluated brief cognitive tests are presented in Tables A and B.

Table A. Summary of findings* for CATD versus normal cognition for selected† commonly reported brief cognitive tests and metrics

Cognitive Test Category	Test Metric	Studies, N (Patients Analyzed)	CATD Prevalence, Median (Range)	SN, Median (Range)	SP, Median (Range)	TP per 1000 Patients, Median (Range)	TP per 1000 Patients, Median (Range)	TP per 1000 Patients, Median (Range)	TP per 1000 Patients, Median (Range)
Individual Stand-Alone Tests	Clock drawing totals	8 (n=1,022)	0.50 (0.15-0.64)	0.79 (0.36-0.93)	0.88 (0.42-1.00)	355 (122-581)	480 (205-612)	60 (0-314)	105 (18-320)
	MMSE total	7 (n=1,724)	0.50 (0.15-0.71)	0.88 (0.56-1.00)	0.94 (0.59-1.00)	414 (113-669)	474 (251-745)	27 (0-255)	43 (0-182)
	MoCA total	2 (n=864)	0.71 (0.60-0.71)	0.94 (0.93-0.96)	0.94 (0.91-1.00)	669 (557-683)	289 (263-378)	23 (0-26)	41 (28-43)
Brief Multidomain Batteries	DRS total	2 (n=507)	0.60 (0.50-0.71)	0.97 (0.96-0.97)	0.96 (0.92-0.99)	583 (480-686)	375 (290-460)	21 (3-40)	21 (20-21)
Memory‡	List learning, trials & totals	6 (n=1,784)	0.21 (0.11-0.50)	0.82 (0.35-0.96)	0.96 (0.73-1.00)	178 (73-480)	650 (470-837)	24 (0-240)	33 (10-295)
	List delayed recall & retention	5 (n=937)	0.50 (0.16-0.50)	0.89 (0.62-0.96)	0.94 (0.76-0.98)	430 (140-480)	480 (400-706)	30 (13-151)	35 (19-190)
	Prose recall & retention	3 (n=895)	0.40 (0.11-0.54)	0.77 (0.71-0.87)	0.87 (0.81-0.89)	334 (78-435)	524 (369-739)	78 (55-151)	65 (32-125)
Executive‡	TMT B, completion time	2 (n=457)	0.33 (0.16-0.50)	0.86 (0.85-0.87)	0.86 (0.83-0.88)	282 (415-740)	578 (415-740)	93 (85-101)	48 (21-75)
Language‡	Semantic (category) fluency	9 (n=1,586)	0.50 (0.15-0.68)	0.92 (0.35-1.0)	0.89 (0.81-1.0)	470 (142-627)	430 (307-751)	42 (0-143)	35 (0-325)
	Phonemic (letter) fluency	4 (n=830)	0.63 (0.15-0.68)	0.77 (0.72-0.89)	0.86 (0.69-0.93)	505 (111-557)	333 (236-659)	53 (28-261)	83 (38-176)
	BNT total	2 (n=479)	0.50 (0.16-0.50)	0.65 (0.53-0.84)	0.92 (0.85-0.92)	270 (119-420)	460 (460-715)	40 (40-126)	153 (40-235)

Table A. Summary of findings* for CATD versus normal cognition for selected† commonly reported brief cognitive tests and metrics (continued)

Cognitive Test Category	Test Metric	Studies, N (Patients Analyzed)	CATD Prevalence, Median (Range)	SN, Median (Range)	SP, Median (Range)	TP per 1000 Patients, Median (Range)	TP per 1000 Patients, Median (Range)	TP per 1000 Patients, Median (Range)	TP per 1000 Patients, Median (Range)
Combinations	WMS LM, WAIS DSy, BNT60	2 (n=302)	0.47 (0.44-0.50)	0.82 (0.68-0.95)	0.87 (0.74-1.00)	382 (342-421)	462 (368-557)	65 (0-129)	92 (22-161)

BNT=Boston Naming Test; CATD=clinical Alzheimer's-type dementia; DRS=Dementia Rating Scale; DSy=Digit Symbol; FN=false negative; FP=false positive; LM=Logical Memory; MMSE=Mini-Mental State Examination; MoCA=Montreal Cognitive Assessment; SN=sensitivity; SP=specificity; TMT B=Trail Making Test part B; TN=true negative; TP=true positive; WAIS=Wechsler Adult Intelligence Scale; WMS=Wechsler Memory Scale.

*No studies reported data on harms of brief cognitive testing for identifying CATD.

†Data presented for brief cognitive tests evaluated in two or more eligible studies and rated low or moderate risk of bias.

‡Individual tests that typically are administered as part of a larger battery in clinical practice.

Table B. Summary of findings* for CATD versus MCI for selected† commonly reported cognitive tests and metrics

Cognitive Test Category	Test Metric	Studies, N (Patients Analyzed)	CATD Prevalence, Median (Range)	SN, Median (Range)	SP, Median (Range)	TP per 1000 Patients, Median (Range)	TN per 1000 Patients, Median (Range)	FP per 1000 Patients, Median (Range)	FN per 1000 Patients, Median (Range)
Individual Stand-Alone Tests	CLOX 1 (draw)	2 (n=150)	0.58 (0.50-0.65)	0.67 (0.58-0.76)	0.86 (0.72-1.00)	393 (288-498)	374 (249-500)	48 (0-97)	185 (157-212)
	MMSE total	2 (n=604)	0.69 (0.61-0.76)	0.84 (0.79-0.88)	0.81 (0.79-0.83)	570 (542-598)	257 (192-322)	58 (51-65)	115 (71-159)
	MoCA total	3 (n=1189)	0.72 (0.25-0.76)	0.79 (0.67-0.97)	0.79 (0.78-0.88)	541 (205-655)	223 (189-659)	60 (51-90)	114 (22-250)
Memory‡	List learning, trials & totals	2 (n=139)	0.47 (0.47-0.65)	0.65 (0.35-0.91)	0.72 (0.66-0.90)	307 (165-596)	380 (238-475)	107 (53-179)	165 (59-307)
Memory‡	List delayed recall & retention	3 (n=327)	0.47 (0.47-0.62)	0.83 (0.73-0.90)	0.65 (0.52-0.83)	411 (345-564)	274 (262-438)	163 (90-253)	80 (52-128)

CATD=clinical Alzheimer's-type dementia; FN=false negative; FP=false positive; MMSE=Mini-Mental State Examination; MoCA=Montreal Cognitive Assessment; MCI=mild cognitive impairment; SN=sensitivity; SP=specificity; TN=true negative; TP=true positive.

*No studies reported data on harms of brief cognitive testing for identifying CATD.

†Data presented for brief cognitive tests evaluated in two or more eligible studies and rated low or moderate risk of bias.

‡Individual tests that typically are administered as part of a larger battery in clinical practice.

- **Biomarkers for distinguishing AD from non-AD in adults with CATD**. Twenty-four unique, low or medium ROB studies (n=2,152) evaluated the accuracy of biomarkers for distinguishing autopsy-confirmed AD from non-AD dementia, including 15 of brain imaging (n=1,225), and nine of CSF biomarkers (n=927). No studies examined the accuracy of blood testing. Results for the most commonly evaluated biomarker tests are presented in Table C.

Table C. Summary of findings of brain imaging and CSF biomarker tests for autopsy-confirmed AD versus non-AD dementias*

Biomarker Category	Test	Studies, N (Patients Analyzed)	AD Prevalence, Median (Range)	SN, Median (Range)	SP, Median (Range)	TP per 1000 Patients, Median (Range)	TN per 1000 Patients, Median (Range)	FP per 1000 Patients, Median (Range)	FN per 1000 Patients, Median (Range)
Brain Imaging	Amyloid PET	4 (n=426)	0.64 (0.33-0.79)	0.91 (0.79-0.98)	0.92 (0.76-1.0)	599 (261-682)	324 (208-555)	28 (0-119)	65 (13-111)
	FDG PET	2 (n=182)	0.64 (0.57-0.70)	0.89 (0.84-0.94)	0.74 (0.73-0.74)	568 (477-659)	268 (217-318)	97 (80-114)	67 (43-91)
	SPECT cerebral perfusion	3 (n=205)	0.56 (0.48-0.64)	0.64 (0.57-0.94)	0.83 (0.76-0.92)	437 (271-603)	390 (301-479)	48 (42-55)	125 (41-208)
	MRI MTA	2 (n=161)	0.33 (0.24-0.42)	0.91 (0.91-0.91)	0.89 (0.84-0.94)	300 (217-383)	602 (487-717)	70 (43-96)	28 (22-35)
CSF	Aß42	2 (n=362)	0.65 (0.65-0.66)	0.77 (0.74-0.79)	0.58 (0.53-0.62)	425 (373-477)	262 (220-303)	88 (41-135)	225 (168-282)
	t-tau	3 (n=449)	0.66 (0.65-0.85)	0.65 (0.62-0.72)	0.66 (0.64-0.69)	428 (401-609)	226 (103-228)	117 (46-129)	241 (228-244)
	p-tau	2 (n=362)	0.85 (0.63-0.94)	0.79 (0.78-0.80)	0.61 (0.60-0.61)	597 (451-823)	94 (59-293)	48 (0-138)	170 (81-489)
	Aß42/p-tau ratio	1 (n=217)	0.65	0.83	0.60	535	213	142	110
	Aß42/t-tau ratio	1 (n=217)	0.65	0.75	0.57	484	203	152	161

Table C. Summary of findings of brain imaging and CSF biomarker tests for autopsy-confirmed AD versus non-AD dementias* (continued)

AD=Alzheimer's disease; CSF=cerebrospinal fluid FDG=fluorodeoxyglucose; FN=false negative; FP=false positive; MRI= magnetic resonance imaging; MTA=medial temporal atrophy; PET=positron emission tomography; SN=sensitivity; SP=specificity; SPECT=single-photon emission computerized tomography; TN=true negative; TP=true positive.

*Two included amyloid PET studies reported data on physical harms, with one reporting that florbetapir was well tolerated and associated with no serious adverse events,[3] and the other that five percent of participants who received flutemetamol experienced mild to moderate treatment-related adverse events and that two deaths and two nonfatal serious adverse events were considered unrelated to flutemetamol.[4] No other brain imaging studies reported data on harms.

Drugs for cognition, function, and harms in adults with CATD. Fifty-four unique, low or medium ROB trials evaluated the efficacy and harms of prescription drugs or supplements for cognition, function, and harms in CATD. These included 25 of cholinesterase inhibitors versus placebo (n=9,476), 11 that compared different cholinesterase doses with each other (n=5,893) (7 of which also included a placebo comparison), six of memantine versus placebo (n=2,227), 11 of supplements versus placebo (n=2,004), three that compared different prescription drugs (n=454), and five that compared prescription drugs with supplements (n=258). The main findings of these studies for cognition, function, and harms are summarized in Table D.

Table D. Efficacy and harms of prescription drugs versus placebo in participants with CATD*

Drug Intervention	ChEI -Co-Use†	CATD Severity	Cognition	Function	SAE
Ch-I	NA	Mild to moderate	Favors Ch-I (Low SOE) Mean Δ**: SMD 0.24-0.52 Likelihood >4-pt ADAS-Cog improvement: NNTB 5-11	Favors Ch-I (Low SOE) Mean Δ††: SMD -0.02 to 0.22	Ch-i may increase risk (Low SOE) 10.7% vs. 8.7%
		Moderate to severe	Favors AChEI (Low SOE) Mean Δ‡‡: SMD 0.29	Favors Ch-I (Low SOE) Mean Δ**: SMD -0.10 to 0.18	No difference (Low SOE)
Memantine	No	Mild to moderate‡	(Insufficient SOE) Mean ADAS-Cog Δ: SMD -0.13	No difference (Low SOE) Mean ADCS-ADL Δ: SMD 0.00	(Insufficient SOE) 10% vs. 10%
	Yes	Mild to moderate	(Insufficient SOE) Mean Δ**: SMD -0.11 to 0.04	No difference (Low SOE) Mean ADCS-ADL Δ: SMD 0.12	(Insufficient SOE) 12.4% vs. 13.9%
		Moderate to severe	Brief individual stand-alone tests: (Insufficient SOE) Mean MMSE Δ: SMD 0.47 Brief multidomain batteries: Favors memantine (Low SOE) Mean SIB Δ: SMD 0.27	No difference (Low SOE) Mean Δ††: SMD -0.26 to 0.14	(Insufficient SOE) 8.2% vs. 6.3%

Δ=change; ADAS-Cog=Alzheimer's Disease Assessment Scale-Cognitive Subscale; BADLS=Bristol Activities of Daily Living Scale; CATD=clinical Alzheimer's-type dementia; Ch-I=cholinesterase inhibitors; DAD=Disability Assessment for Dementia; MDS-ADL=7-item Minimum Data Set Activities of Daily Living Self-Performance scale; MMSE=Mini-Mental State Examination; NA=not applicable; PDS=Progressive Deterioration Scale; SAE=serious adverse events; SIB=Severe Impairment Battery; SMD=standardized mean difference SOE=strength of evidence.

*Strength of evidence grades for a treatment comparison and outcome reflected our confidence about the direction of the treatment effect (i.e., whether or not the outcome differs between treatments). Definitions for strength of evidence grades are detailed in the Methods section of the main report.

Table D. Efficacy and harms of prescription drugs versus placebo in participants with CATD* (continued)

†Indicates whether all trial participants received a cholinesterase inhibitor as a co-intervention.

‡No eligible studies compared memantine with placebo in patients with moderate to severe CATD who were not receiving a cholinesterase inhibitor.

**MMSE or ADAS-Cog

††DAD, ADCS-ADL or PDS

‡‡MMSE or SIB

***ADCS-ADL or MDS-ADL

†††ADCS-ADL or BADLS

Drugs for BPSD in adults with CATD and BPSD. Eleven unique, low or medium ROB trials evaluated the efficacy and harms of drug treatment compared with placebo on BPSD in patients with CATD, including four of antipsychotics (n=522), four of antidepressants (n=836), one of cholinesterase inhibitors (n=272), one of anticonvulsants (n=153), and one of supplements (the Japanese herbal medicine, Yokukansan) (n=145). Two trials compared different prescription drugs (n=414). The main findings of these studies for BPSD and harms are summarized in Table E.

Table E. Efficacy and harms of prescription drug treatments versus placebo for BPSD in patients with CATD and agitation and/or aggression†

Drug Treatment	Agitation	Aggression	Psychosis	SAE
Anti-psychotic	Insufficient SOE (standard- and low-dose haloperidol, pimavanserin, quetiapine)	Insufficient SOE (standard- and low-dose haloperidol, pimavanserin)	Insufficient SOE (aripiprazole, standard- and low-dose haloperidol, pimavanserin)	Insufficient SOE (aripiprazole, pimavanserin, quetiapine)
Anti-depressant	Insufficient SOE (citalopram,* sertraline)	NR	Insufficient SOE (citalopram*)	Insufficient SOE (citalopram,* mirtazapine, sertraline)
Cholinesterase inhibitor	No difference (low SOE) Likelihood 30% CMAI improvement: RR, 0.96 (0.56-1.62) for donepezil	NR	NR	NR
Anti-convulsant	Insufficient SOE (valproic acid)	NR	NR	NR

CATD=clinical Alzheimer's-type dementia; CMAI=Cohen-Mansfield Agitation Inventory; NR=not reported; NS=no statistically significant difference; RR=relative risk; SAE=serious adverse event; SOE=strength of evidence

*Citalopram was dosed up to 30 mg/day, which exceeds the currently recommended maximum dose of 20 mg/day.

†Strength of evidence grades for a treatment comparison and outcome reflected our confidence about the direction of the treatment effect (i.e., whether or not the outcome differs between treatments). Definitions for strength of evidence grades are detailed in the Methods section of the main report.

Limitations

Evidence on the accuracy of brief cognitive tests for distinguishing CATD from normal cognition and MCI in adults with suspected cognitive impairment had several limitations. We found few eligible studies for most individual cognitive tests, fewer for test combinations, none for several common tests (e.g., Mini-Cog, Saint Louis University Mental Status test [SLUMS], Telephone Interview for Cognitive Status [TICS]), and minimal data on web-based tests. Similarly, few studies evaluated the accuracy of brain imaging and CSF tests compared with autopsy-confirmed diagnoses, and none examined blood tests. Studies for both cognitive and biomarker tests were limited by small sample sizes. Cognitive test studies were heterogeneous in several ways. They varied in their definitions of normal cognition and in test scoring metrics. Further, they rarely used normative or other prespecified cut points to distinguish normal from abnormal. Rather, cut points most often were selected to maximize classification accuracy within the study cohort. Brain imaging and CSF studies were also methodologically heterogeneous. These studies varied in composition of non-AD comparison groups, interval between imaging or CSF collection and autopsy, methods of image acquisition or CSF assay and analysis, neuropathologic reference standards, and use of test cut points unique to their individual study cohorts. Biomarker studies were limited because many study participants with biomarker measures did not complete autopsy and weren't included in analyses. No studies using an autopsy-confirmed AD reference group evaluated the classification accuracy of MRI hippocampal atrophy, computed tomography (CT), tau (positron emission tomography) PET, or (functional magnetic resonance imaging) fMRI brain imaging; beta amyloid (Aβ)42/Aβ40 ratio, or neurofilament light protein CSF tests; or any blood tests. Further, few studies examined the classification accuracy of test combinations. Because cognitive test and biomarker study populations were predominately white and relatively young (mean age 73 to 74 years for cognitive studies and mean dementia symptom onset in participants' early 60s to early 70s for biomarker studies), we could not determine generalizability of results to other racial/ethnic groups or older populations. Further, there was little evidence about whether accuracy varied by study participant characteristics and no cognitive testing studies and few brain imaging or CSF testing studies reported on harms.

Evidence on the efficacy and harms of CATD drugs had several limitations. We found few trials for individual drug treatments, especially for supplements and BPSD drugs, and most study sample sizes were small. This resulted in low statistical power for even somewhat common events and large mean differences between groups that could be clinically meaningful if real. We also found few trials that stratified results by CATD severity. Because we analyzed studies grouped by participant CATD severity and graded SOE for treatment effects within these severity categories, it is possible that SOE grades would have been different in cases when lumping studies regardless of baseline CATD severity may have been clinically reasonable (e.g., for harms). This review limited prescription drug classes evaluated for cognition and function to cholinesterase inhibitors and memantine, and prescription drug classes evaluated for BPSD to cholinesterase inhibitors, memantine, antipsychotics, antidepressants, anxiolytics, antiepileptics/mood stabilizers, hormonal agents and cannabinoids. This review required studies of cognition and function to be at least 24 weeks long, and studies of agitation, aggression, and psychosis to be at least 2 weeks long; trials reporting only on acute and shorter-term treatment effects were excluded. In addition, few included trials were longer than 26 weeks, so longer-term drug effects were unclear. Because

trial populations were predominately white, we could not determine generalizability to other racial/ethnic groups. Few trials directly compared different drug treatments. Few trials reported results for CATD staging, individual cognitive domains, quality of life, or caregiver outcomes, and no eligible studies without high risk of bias reported results for disinhibited sexual behavior. Harms reporting was poor. Many eligible trials were excluded from analyses due to high risk of bias, often because of high attrition, especially trials longer than 26 weeks and those that compared two active treatments. Many trials analyzed results using methods of accounting for missing data that may overestimate treatment benefit. It was difficult to interpret the relevance of small between-group differences in continuous outcomes and most trials did not report on between-group differences for the likelihood of experiencing clinically important treatment effects (i.e., responder analyses). Lastly, few trials evaluated whether treatment efficacy and harms varied by study participant characteristics.

Implications and Conclusions

Cognitive test studies showed that among individuals with suspected cognitive impairment (case finding), selected brief cognitive tests, including those commonly used as individual stand-alone tests, brief multidomain batteries, and memory verbal fluency tests typically administered as part of a larger battery in clinical practice are accurate for distinguishing between CATD and normal cognition, but somewhat less accurate distinguishing between smaller differences in cognitive function (e.g., distinguishing mild CATD from normal cognition, or CATD from MCI). However, because few studies directly compared the accuracy of different tests, different test scoring metrics, different cut-points for defining tests as abnormal, or combinations of tests, we could not definitively determine which test or combination of tests is most accurate and which

cut-point is best for each test and test metric. We found even less information about whether test accuracy varied by patient characteristics. So, brief cognitive tests may help identify which patients with suspected cognitive impairment are more likely to have CATD but are not considered sufficient alone to make the clinical diagnosis. Brief cognitive test results may help clinicians decide who warrants further diagnostic evaluation, including a detailed history of cognitive symptoms, focused neurological exam, and possible neuropsychological testing and specialty referral. These brief cognitive test results also may be sufficient for objectively documenting cognitive impairment in more impaired patients with a recognized history of cognitive and functional decline typical for CATD.

Biomarker studies showed that several types of brain imaging and CSF tests are highly sensitive and specific for distinguishing autopsy-confirmed AD from non-AD dementia. Based on few studies, amyloid PET and FDG-PET imaging but not SPECT appear to increase classification accuracy added to a clinical evaluation when directly compared to accuracy of a clinical evaluation alone. One study (reporting data for one of two assays evaluated) suggested that a model incorporating results from multiple CSF biomarkers may improve categorization of patients between AD and FTLD when added to clinical evaluation alone. We found no analogous data for MRI or other CSF tests. One study reported that the combination of CT and amyloid PET was not more accurate than amyloid PET alone. Data were unclear for which combination of tests and which test cut points are best for distinguishing between autopsy-confirmed AD and non-AD in individuals with CATD. However, even if future research confirms that biomarkers and their combinations improve classification accuracy when added to clinical evaluation, applicability is likely to be limited as long as access to such testing is limited in many clinical settings and there are no disease-

modifying drug treatments for AD and non-AD dementias.

Trials of about 6 months showed benefits for cholinesterase inhibitors compared with placebo regardless of baseline CATD severity. However, average differences for cognition and function between treatment groups were small, with standardized mean differences mostly between 0.20 to 0.40 for cognition and about 0.20 for function. Responder analyses showed that compared with placebo, for approximately every 5 to 13 participants assigned cholinesterase inhibitors, one additional individual was improved at 6 months on a cognitive battery (>4-point improvement in Alzheimer's Disease Assessment Scale-Cognition [ADAS-Cog]) or a global change measure (Clinician's Interview-Based Impression of Change with caregiver input [CIBIC-Plus] or Clinical Global Impression of Improvement [CGIC]). Whether these 6 month improvements are clinically meaningful is unclear. Data on moderate or marked improvement for cognition or function were not reported and moderate or marked improvement on the global change measures was rare and no more likely with cholinesterase inhibitor treatment than placebo. We could not determine the likelihood of longer-term benefits, because no eligible cholinesterase inhibitor trials with low or medium risk of bias reported efficacy outcomes beyond 6 months. For memantine, one trial suggested that for every 6 additional participants assigned memantine compared with placebo, one additional individual was improved on the CIBIC-Plus global change measure, but other eligible memantine trials did not report responder analyses and mean differences between memantine and placebo for measures of cognition and function were not statistically significant or small, particularly in patients with mild to moderate CATD.

On the whole, evidence to guide treatment decisions about prescription drugs for BPSD in patients with CATD was lacking. Few eligible trials with low or medium risk of bias examined treatment efficacy and harms of prescription drugs in patients with CATD and BPSD and were at least 2 weeks in duration. While a few trials reported some findings suggesting possible treatment benefit, the evidence was insufficient to draw conclusions. This was largely due to small sample sizes and inconsistent results within and between trials. For example, one trial of aripiprazole showed statistically significant improvement compared with placebo in two of three psychosis scores, a small trial of standard-dose haloperidol versus placebo showed numerically higher likelihood of improvement in agitation and psychosis (likelihood of absolute risk differences >25% and/or standardized mean differences (SMDs) >0.4 that were not statistically significant), and a trial of quetiapine showed no difference in mean change for agitation compared with placebo. No eligible trials of antipsychotics reported data on stroke and just three (n=451) reported data on deaths (4.4% for the antipsychotic group vs. 1.8% for placebo), too few to draw conclusions but not inconsistent with FDA warnings.[1, 2] For antidepressants, one trial of citalopram up to 30 mg/day reported statistically significant improvement compared with placebo for a minority of agitation and psychosis outcomes. However, this dose exceeds the current maximum recommended dose of 20 mg/day. A trial of sertraline showed no statistically significant difference in agitation compared with placebo. Strength of evidence for all efficacy outcomes for both these antidepressant trials was considered insufficient to draw conclusions and insufficient to guide treatment decisions. We found no evidence from qualifying trials for other antidepressants, low-strength evidence that donepezil and placebo did not differ for agitation in one trial, and only insufficient evidence about the efficacy of antiseizure drugs, memantine, and estrogen. Only one trial compared different prescription drugs for agitation and evidence was insufficient to draw conclusions about differences in BPSD or harms for continued antipsychotics compared with switching to memantine. No trials compared antipsychotics with antidepressants.

Few eligible trials with low or medium risk of bias examined efficacy and harms of supplements on the outcomes of cognition, function, and BPSD in patients with CATD. Two trials each for the nutritional drink Souvenaid® and omega-3 fatty acids showed no benefit compared with placebo for function and cognition, respectively. Several other trials showed statistically significant benefits compared with placebo for one or more outcomes. However, due to small sample sizes, few trials for each intervention, and study limitations, evidence was insufficient to draw conclusions about the efficacy and safety of these supplements for CATD treatment for all these outcomes.

Future Research Recommendations

Future research about the accuracy of brief cognitive testing for CATD should continue to define CATD and MCI based on standard clinical criteria and should define normal cognition based on a formal cognitive evaluation rather than self-report or brief testing. Studies should evaluate the accuracy of commonly used or promoted brief cognitive tests for which we identified no eligible studies. Studies should prespecify test cut points or the methodology for defining them to enable external validation beyond single study populations. Studies should compare the accuracy of different individual and combined brief cognitive tests in the same study population and evaluate whether cognitive test accuracy varies by study participant characteristics (e.g., age, sex, race, education). Studies should systematically collect data on potential harms of cognitive testing. Directly or through modeling, studies should evaluate whether brief cognitive testing of patients with suspected cognitive impairment modifies subsequent drug and nondrug treatment decisions and, more importantly, alters patient and caregiver outcomes.

Future research about the accuracy of biomarkers for distinguishing AD from non-AD dementias in patients with CATD should compare the accuracy of brain imaging, CSF and blood biomarkers with autopsy-confirmed AD pathology. Among participants with collected biomarkers, studies should compare characteristics between participants with and without available autopsy data to better identify potential attrition biases in studies using autopsy neuropathology as a reference standard. Research should better clarify how biomarker accuracy varies as a function of the time between biomarker collection and autopsy, to inform how changes in biomarkers and brain neuropathology over time affect test accuracy, and the strengths and limitations of using biomarkers as a surrogate for brain neuropathology. Future studies should evaluate the accuracy of biomarkers for which we identified no eligible studies (e.g., MRI hippocampal atrophy, CT, tau PET, and fMRI for brain imaging; Aß42/Aß40 ratio and neurofilament light protein for CSF; and blood biomarkers). Studies should report information about participant clinical diagnosis to make it clear how often clinical diagnoses are reclassified based on biomarker testing. Studies should standardize imaging and assay analytic methods and rating criteria that are feasible to implement in typical clinical settings. Studies should externally validate cut points for optimally distinguishing AD from non-AD dementias across populations, including in typical clinical populations. Studies should compare different individual and combined brain imaging and CSF tests in the same study population and evaluate whether test accuracy varies by study participant characteristics (e.g., age, sex, race, education). Studies should systematically collect data on potential psychological and physical harms of biomarker testing. Lastly, directly or through modeling, studies should evaluate whether biomarker testing affects drug and nondrug treatment decisions and, more importantly, alters patient and caregiver outcomes.

Future trials investigating drug treatment for CATD should be large enough to detect the likelihood of treatment response as defined for clinically important cognitive, functional, and global outcome measures. Trials should routinely report on patient quality of life and caregiver

outcomes. Future trials should investigate treatment efficacy and harms beyond 6 months to increase applicability to clinic populations who may be treated for years. Trials should enroll more diverse participants, including nonwhites and older patients, and pre-specify analyses with sufficient statistical power to examine whether treatment effects are modified by patient characteristics, including age, sex, race/ethnicity, baseline CATD severity, baseline BPSD severity, and living setting. Additional trials should examine drug treatments and doses for which data suggest signals of possible benefits, but for which the strength of evidence is insufficient, such as antipsychotics and antidepressants for agitation and psychosis. Antipsychotics and antidepressants should be directly compared for treatment of BPSD. Future BPSD trials also should directly compare drug and nondrug treatment strategies, and drug trials should specify whether participants receive a concomitant psychosocial intervention. Future BPSD drug trials should be longer to better establish the evidence for long-term efficacy and safety. Supplements should be subjected to rigorous trial examination, both for efficacy and safety compared with placebo, and for comparative effectiveness and safety compared with FDA approved prescription drugs. Future drug trials for BPSD, which likely will continue to target agitation, aggression, and psychosis, should also prespecify disinhibited sexual behavior, depression, and anxiety as secondary efficacy outcomes.

References

1. U.S. Food & Drug Administration. U.S. Food and Drug Administration (FDA) Public Health Advisory: Deaths with antipsychotics in elderly patients with behavioral disturbances. Department of Health and Human Services, U.S. Government; 2005. https://wayback. archive-it.org/7993/20170113112252/http://www.fda. gov/DrugsDrugSafetyPostmarketDrugSafety InformationforPatientsandProviders/ucm053171.htm. Accessed on July 10 2019.

2. U.S. Food & Drug Administration. Information for Healthcare Professionals: Conventional Antipsychotics. https://wayback.archive-it.org/7993/20170722190727/ https://www.fda.gov/Drugs/DrugSafety/ PostmarketDrugSafetyInformationfor PatientsandProviders/ucm124830.htm. Accessed on July 10 2019.

3. Clark CM, Pontecorvo MJ, Beach TG, et al. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: a prospective cohort study.[Erratum appears in Lancet Neurol. 2012 Aug;11(8):658]. Lancet Neurol. 2012 Aug;11(8):669-78. DOI: 10.1016/S1474-4422(12) 70142-4. PMID: 22749065.

4. Curtis C, Gamez JE, Singh U, et al. Phase 3 trial of flutemetamol labeled with radioactive fluorine 18 imaging and neuritic plaque density. JAMA Neurol. 2015 Mar;72(3):287-94. doi: https://dx.doi.org/10.1001/ jamaneurol.2014.4144. PMID: 25622185.

Full Report