Supplemental Table 1. Summary articles describing meta-analyses of cognitive measures

Authors # and type of Studies Included	Intervention/ Exposure	Inclusion Criteria	Types of Cognitive Measures	Cognitive Domains	How Data Were Combined
Angevaren, M. 2008 [16] 11 RCTs	Physical activity programs	• Adults ≥55 years old • Compared aerobic physical activity programs with another intervention or no intervention	Simple RT, Choice RT, TMT- A and B. DSST, Rand memory test story recall, RIPA, WAIS, BVRT, DS (forward and backward), RAVLT, WMS, Word comparison, Task switching paradigm, Verbal fluency, Face recognition, Stroop, Stopping task, Digit vigilance, Tracking, Letter search, Finger tapping, Visual search, Pursuit rotor task	• Cognitive speed • Verbal and Visual memory functions • Working memory • Executive functions • Perception • Cognitive inhibition • Visual attention • Auditory attention • Motor function	• WMD when same test used • SMD when different tests used to measure the same construct • Fixed and random effects models • Subgroup and sensitivity analyses
Eilander, A. 2010 [17] 19 RCTs	Micronutrient supplementation	• Healthy children aged 0-18 • Supplemented with ≥ 3 micronutrients with a placebo group • Treated for ≥4 weeks	BAS, DG, CPAS-R, CTBS, GMT, MISIC, NAR, OOHMT, PGI, PMAT-FC, NEPSY, RAVLT, SDMT, WAIS, WIAT, WISC-III/R	• Fluid intelligence • Crystallized intelligence • Short term memory • Visual perception • Retrieval ability • Cognitive processing speed • Sustained attention • Motor skills • Academic performance	• SMD (Cohen's d) • Random effects model • Assessed heterogeneity and publication bias • Subgroup and sensitivity analyses • Meta-regression (model not specified)
Falkingham, M. 2010 [18] 14 RCTs	Oral iron supplement	• Anemic and non-anemic people at least 6 years old • treated at least 4 weeks • Objective measure of cognitive performance	RAVENS, RAVLT, WDS, DS, Mazes test, HVLT, Bourden-Wisconsin concentration	• Attention/concentration • IQ • Memory • Psychomotor • Scholastic achievement	• SMDs • Random effects model • Assessed heterogeneity and publication bias • Subgroup and sensitivity analyses

Supplemental Table 1. Summary articles describing meta-analyses of cognitive measures

Authors # and type of Studies Included	Intervention/ Exposure	Inclusion Criteria	Types of Cognitive Measures	Cognitive Domains	How Data Were Combined
Guilera, G. 2009 [19] 18 RCTs	Antipsychotic medications	• Adults aged 16-65 years • Diagnosed with schizophrenia, schizo-affective or schizophreniform disorder • Used a standardized neurological test listed in the Lezak manual	Stroop, CPT, TMT-A and B, CANTAB, Rapid Visual Information Processing Test, WCST, WISC-R, WAIS, HVOT, WMS-R, CVLT, RCAVLT, HVLT, NART, Peabody Picture Vocabulary Test, FT, GPT, Benton Judgment of Lines	• Attention and vigilance • Automaticity and procedural learning • General intellectual functioning • Verbal comprehension • Perceptual processing • Psychomotricity • Reasoning • Speed of processing • Verbal learning/memory • Visual learning/memory • Working memory	• SMD (Hedges' g) were combined to create a weighted mean estimate for each cognitive domains and a global index • Random effects model • Assessed heterogeneity and publication bias • Subgroup and sensitivity analyses • Meta-regression
Hogervorst, E. 2010 [20] 38 RCTs	Hormone replacement therapy	• Postmenopausal women without dementia • Placebo group • Included cognitive measures	paragraph recall, story recall, COWAT, FAS, FR, BVRT, RAVLT, SRT, DS, Stroop, TMT-B, MMSE	• Verbal memory/fluency • Visual memory • Concentration • Executive function • Visuospatial	• Categorized studies as having a predominant positive (+1), neutral (0) or negative (-1) effect on cognitive outcomes • Number of positive, neutral and negative effects on tests • Used Chi-Square and Spearman's rank correlations
Hogervorst, E. 2009 [21] 7 RCTs	Estrogen alone or combined with Progestrogen	• Postmenopausal women with Alzheimer's disease or other types of dementia • Treated at least 2 weeks • Double-blind RCT	MMSE, BIMC, ADAS-Cog, HSD, WMS, BSRT, CERAD, DS, VRT, visual span, FR, ROVMT, BNT, Token test, TMT-A and B, DSST, Stroop	• General cognitive function • Verbal memory • Visual memory • Language • Processing Speed	• WMD when same test used • SMD when different tests used to measure the same construct • Fixed and random effects models • Assessed heterogeneity • Subgroup and sensitivity analyses

Supplemental Table 1. Summary articles describing meta-analyses of cognitive measures

Authors # and type of Studies Included	Intervention/ Exposure	Inclusion Criteria	Types of Cognitive Measures	Cognitive Domains	How Data Were Combined
Li, H 2011 [22] 17 RCTs	Cognitive stimulation/ training or cognitive rehabilitation	• Individuals with mild cognitive impairment • pre- and post-test intervention measures • Study reported means, standard deviations, t test or F test and sample size	Episodic and Semantic memory tests, TMT-A and B, WCST, Figure Rey-copy, pattern and picture reproduction, FR, MMSE	• Memory • Executive functioning • Attention/processing speed • Visuospatial ability • General cognitive function	• SMD (Cohen's d) • Effects sizes averaged within each domain • Random effects model • Assessed heterogeneity and publication bias • Subgroup and sensitivity analyses • Meta-regression
Karsdorp, PA. 2007 [23] 11 RCTs	Surgery or interventional catheterization	• Children and adolescents with congenital heart disease (2 to19 years old) including a control group • Reported data required to calculate ESs • Published in English or German	BAS, BSID, DAS, HAWIE, HAWIK, HAWIVA, KABC, LIS, MSCA, SB, WISC, WPPSI	Specific domains not reported	• SMD (Cohen's d) • Weighted mean ES (Hunter and Schmidt method) • Assessed heterogeneity and publication bias • Subgroup and sensitivity analyses
Lethaby, A. 2008 [24] 16 RCTs	Hormone replacement therapy	• Healthy women who had undergone natural or surgical menopause • Treated at least two weeks • Double blind RCT	CAMCOG, MMSE, WMS, CVLT, BNT, VRT, BVRT, VMT, TMT-A, DSST, FT, GPT, Stroop, Letter cancellation tests, WCST, WAIS, COWAT, DS (backward)	• Global cognitive function • Verbal memory and language • Visuospatial • Speed tests • Attention • Semantic Memory • Executive function	• WMD when same outcomes • SMD otherwise • Odds ratio when outcome was binary (cognitive impairment) • Fixed effects model • Assessed heterogeneity • Subgroup and sensitivity analyses
Marasco, SF. 2008 [25] 8 RCTs	Off-pump (beating heart) coronary artery bypass grafting vs. on-pump	• Patients receiving coronary artery bypass grafting • All patient populations • All language publications considered	RAVLT, GPT, TMT-A and B, WAIS III, DSST	• Verbal memory • Motor capacity • Divided attention • Executive function • Information processing	• WMD for each outcome • Fixed and random effects models • Assessed heterogeneity and publication bias • Subgroup and sensitivity analyses

Supplemental Table 1. Summary articles describing meta-analyses of cognitive measures

Authors # and type of Studies Included	Intervention/ Exposure	Inclusion Criteria	Types of Cognitive Measures	Cognitive Domains	How Data Were Combined
Martin, M. 2011 [26] 36 RCTs	Mental training, problem solving training, speed training, cognitive restructuration technique	• Healthy people 60 years or older people with mild cognitive impairment (without dementia) • Described cognitive training and specific domains of cognitive function • At least 2 measurements • Published in English or German	Luria task, TMT, visuomanual coordination, abstraction proverbs, phonematic fluency, GUMT, subjective memory tests, alpha span, Brown-Peterson test, DS, UFOV, Road Sign Test, letter comparison, auditory memory, letter series test, word series test, letter sets test, HVLT, Rivermead Behavioural Memory Test, Hopkins Prospective Memory Task, Memory controllability Inventory, RAVENS, RAVLT, FNT, BVRT, WAIS	• Memory • Attention • Speed	• Selected one variable from each study to represent the outcome measure • WMD when same outcomes • SMD otherwise • Odds ratio when outcome was binary (improvement) • Fixed and random effects model • Assessed heterogeneity • Sensitivity analyses
Metternich, B. 2010 [27] 14 RCTs	Non-pharmacological (e.g., Mental Training, Psycho-educational programs)	• Patients reporting subjective memory complaints • Published in English, Dutch, German or French • Reported sufficient data to conduct the meta-analyses	MIA, MCI, MFQ, FNT, CVLT, HVLT, Visual Verbal Learning Test, WMS, BSRT, GUMT	• Subjective memory • Objective memory	• SMD (Hedges' g) • Random effects models • Assessed heterogeneity
Repantis, D. 2010 [28] 91 RCTs	Modafinil and Methylphenidate use	• Participants showed no signs of psychiatric disorder, cognitive decline or other diseases • Published single- or double-blind or quasi-randomized controlled trials	Not reported	• Attention and vigilance • Memory and learning • Executive functions and information processing	• SMD (Cohen's d) • Assessed heterogeneity • Sensitivity analyses • Meta-regression using linear mixed model

Supplemental Table 1. Summary articles describing meta-analyses of cognitive measures

Authors # and type of Studies Included	Intervention/ Exposure	Inclusion Criteria	Types of Cognitive Measures	Cognitive Domains	How Data Were Combined
Woodward, ND. 2007 [29] 16 RCTs	Atypical antipsychotic drugs	• Participants diagnosed with schizophrenia or schizoaffective disorder • Prospective double blind design • Treated at least one week • No other antipsychotic medications administered • Sample size at least 10 • At least one of the identified tests reported	TMT-A and B, Continuous performance test, DSST, SDMT, WCST, CVLT, RVLT, BVLT, COWAT, Category Instance Generation Test, FT/Oscillation Test, GPT	• Attention • Processing speed • Executive function • Verbal learning • Delayed verbal recall • Verbal fluency • Motor skill	• SMD (difference in mean change divided by pooled baseline and retest SD) • ES for Global Cognitive Index or average ES across all tests • Fixed and random effects model • Assessed heterogeneity and publication bias • Sensitivity analyses
Campbell, LK. 2007 [30] 28 Observational studies	Treatment for Acute Lymphocytic Leukemia	• Post-treatment cognitive data for childhood ALL patients in 1st remission and control group • Neurocognitive measures with adequate psychometric properties and published normative data • Published in English	WPPSI, WISC, WAIS, SB, MSCA, K-BIT, KABC, WRAT, WJ, DS, TMT, Stroop, WISC, FT, GPT, Purdue Pegboard, WRAML, CVLT, RAVLT, BSRT, VMI, RAVLT, WAIS-R BD, Judgment of Line Orientation, BVRT	• Overall Cognitive functioning • Academic achievement • Attention • Executive functioning • Processing speed • Psychomotor skill • Verbal memory • Visuospatial skill • Visuospatial memory	• SMD (Hedges' g) • ESs averaged if multiple measures for the same neurocognitive domain • Random effects model • Assessed heterogeneity and publication bias • Subgroup and sensitivity analyses
Goodman, M. 2002 [31] 22 Observational studies	Lead in the workplace	• Central tendency for lead exposure was <70µg/dl • Totals of exposed and unexposed workers were reported • Test score means and measures of dispersion provided	WAIS-R BD, LM, DSST, visual interference, BVRT, Paired associates, Visual reproduction, Flicker fusion, SDMT, DS (forward and backward), TMT-A and B, Simple RT, Picture completion, GPT, FT	• Specific domains not reported	• SMD (Hedges' and Olkin's d_i) • Fixed and random effects models • Assessed heterogeneity and publication bias • Subgroup and sensitivity analyses

Supplemental Table 1. Summary articles describing meta-analyses of cognitive measures

Authors # and type of Studies Included	Intervention/ Exposure	Inclusion Criteria	Types of Cognitive Measures	Cognitive Domains	How Data Were Combined
Grant, I. 2003 [32] 11 Observational studies	Cannabis use	• Included a "cannabis only" users and control group • Cannabis group drug-free on day of testing • Information on other substance use • Information on history of neurological or psychiatric problems • Data on length of abstinence from cannabis before testing	WAIS-R DS and DVIG, WAIS-R Vocabulary, Verbal Fluency, WCST, RAVENS, WAIS-R BD, Object Assembly, GPT, FT, CVLT, RAVLT	• Simple reaction time • Attention • Verbal/language • Abstraction/executive functioning • Perceptual motor • Simple motor • Learning • Forgetting/retrieval	• SMD (Hedges' and Olkin's d_i) • Weighted average of ESs calculated if multiple measures for the same neurocognitive domain • Fixed effects model • Assessed heterogeneity • Sensitivity analyses • Meta-regression
Valentini, E. 2010 [33] 24 Single or double-blind experimental study design	Mobile phone electromagnetic fields	• Human provocation/ laboratory studies on modulated RF-EMF • Experimental design with real and sham EMF • Reported speed measures • Used comparable tasks • Published in English	TMT-B, two and ten Choice RT, Simple RT, SUB, Sentence verification, VIG	• Attention and speed of processing • Divided and sustained attention • Working memory • Semantic memory	• SMD (Hedges' and Olkin's d_i) • Fixed and random effects models • Assessed heterogeneity and publication bias • Sensitivity analyses • Meta-regression
Wheaton, P. 2009 [34] 22 Repeated Measure Observational studies	Drug treatment for Traumatic Brain Injury (TBI)	• Adults (≥16 years) with TBI with age and severity matched control group • Treated ≤7 days after TBI • no history of TBI, mental health problems or substance abuse or pre-existing impairments • Not recently treated with pharmaceuticals to enhance cognition • Published in English	PASAT, Story memory	• Attention • Memory • General cognition	• SMD (Cohen's *d)* • ESs averaged if multiple measures for the same neurocognitive domain • Fixed and effects model • Assessed publication bias • Sensitivity analyses

Supplemental Table 1. Summary articles describing meta-analyses of cognitive measures

Authors # and type of Studies Included	Intervention/ Exposure	Inclusion Criteria	Types of Cognitive Measures	Cognitive Domains	How Data Were Combined
Barth, A. 2008 [35] 10 Single or double-blind experimental study design	electromagnetic fields emitted by GSM mobile phones	• Treatment group (phone switched on) and control group (phone switched off) or repeated measurements with alternate switching • Range of GSM phone 900 - 1800 MHz • Healthy participants • At least one cognitive test that is used in another study	SRT, CRT, VIG, SUB, VER, N-Back test, TMT, DS (forward and backward), Spatial span forward and backward	• Information processing • Reaction time • Attention • Memory • Executive functions	• ES (not specified) • Fixed and random effects models • Assessed heterogeneity
Brands, A. 2005 [36] 33 Observational studies	Type 1 Diabetes	• Adults (18 years and older) with type 1 diabetes with a defined control group • Cognitive performance measured using standard neuropsychological or reliable experimental testing method at normal glucose values • Published in English between 1980-2004	Measures not reported	• Overall intelligence • Working memory • Immediate memory • Delayed memory • Psychomotor efficiency • Processing speed • Motor speed • Attention • Cognitive flexibility • Visual perception	• SMD (Cohen's d) • ESs averaged if multiple measures for the same neurocognitive domain • Fixed effects model • Assessed heterogeneity and publication bias • Sensitivity analyses
Sibley, B. 2003 [37] 44 Quasi-experimental and cross-sectional studies	Physical activity	• Elementary school-aged children • English studies conducted that could be obtained	Measures not reported	• Perceptual skills • IQ • Achievement • Verbal tests • Math tests • Memory • Developmental level/academic readiness	• SMD (Hedges' g) • Fixed effects model • Assessed heterogeneity • Subgroup analyses
Balint, S. 2009 [38] 25 Observational studies	No intervention	• Adults (>18 years) with ADHD (diagnosed using DSM-III-R or DSM-IV criteria) and normal controls • raw data for ES calculation available • Published in English	Stroop, TMT, WAIS-R DS and DSST, CPT	• Attention (simple, focused and sustained)	• SMD (Cohen's d) • Random effects model • Assessed heterogeneity • Sensitivity analyses • Meta-regression using linear mixed model

Supplemental Table 1. Summary articles describing meta-analyses of cognitive measures

Authors # and type of Studies Included	Intervention/ Exposure	Inclusion Criteria	Types of Cognitive Measures	Cognitive Domains	How Data Were Combined
Bhutta, A. 2002 [39] 15 Case-control studies	Born preterm	• School-aged children (≥5 years) who were born preterm and controls • Report cognitive data, behavioral data or both • Studies had an attrition rate < 30% • Published in 1980 or later	BAS, MIQS, WPPSI, KABC, WISC	• IQ • Other domains not reported	• WMD (because normative scores available for all tests) • Fixed and random effects models • Assessed heterogeneity and publication bias • Sensitivity analyses • Meta-regression using linear model
Bora, E. 2009 [40] 62 Observational studies	Bipolar disorder (BD)	• Neuropsychological data on remitted adults with BD or first-degree relatives of patients with BD and healthy control group • Used at least one cognitive measure that was used in ≥3 studies • Published in English between 1995-Oct 2007	RAVLT, CVLT, VLT, WMS-R, CPT, TMT-A and B, FAS, WCST, CANTAB, WAIS-R DS, Stroop, ROCF, NART	• Verbal learning/memory • Visual memory • Sustained attention • Processing speed • Verbal fluency • Set shifting • Working memory • Response inhibition • Visuospatial abilities • General intelligence	• SMD (Hedges' g) • Random effects model • Assessed heterogeneity and publication bias • Meta-regression using random effects model using restricted-information maximum likelihood method
Jansen, C. 2005 [41] 16 Observational studies	Chemotherapy	• Adults receiving chemotherapy and a control group or normative data • neuropsychological testing when had or were presently receiving chemotherapy • Used reliable, valid, and standardized tests • Reported sufficient data to estimate ES	CPT, DRS, HRNB, HSCS, RBANS, RCFT, TMT, WAIS, WMS	• Attention or concentration • Executive function • Speed of information processing • Language • Motor function • Visuospatial skill • Verbal memory • Visual memory	• SMD • ESs averaged if multiple measures for the same neurocognitive domain • Model not reported
Krabbendam, L. 2005 [42] 31 Observational studies	Bipolar disorder or schizophrenia	• Compared adult participants with schizophrenia and with bipolar disorder • Used standardized neuropsychological testing procedures • Published in English	DS, Letter-Number Span, DSST, TMT-A and B, Stroop, WCST	• Verbal working memory • Verbal fluency • Mental speed • Executive control • Concept formation and shifting	• SMD (Cohen's d) • ESs averaged if multiple measures for the same neurocognitive domain • Random effects model • Assessed heterogeneity • Subgroup analyses

Supplemental Table 1. Summary articles describing meta-analyses of cognitive measures

Authors # and type of Studies Included	Intervention/ Exposure	Inclusion Criteria	Types of Cognitive Measures	Cognitive Domains	How Data Were Combined
McDermott, LM. 2009 [43] 14 Observational studies	Depression	• Participants diagnosed with major or minor depression • Reported sufficient data to estimate ES • Related severity of depression to neuropsychological test performance	WCST, word fluency, TMT-A and B, COWAT, Hayling test B, Stroop, MCST, DS (forward and backward), CANTAB, semantic fluency test, Verbal fluency test, MFFT-20, ZVT, TAP, DSST, GPT, RAVLT, AVLT, signal detection, recognition test	• EM • Executive function • Processing speed • Semantic memory • Visuospatial memory	• Pearson correlation coefficient • If more than one test in a cognitive domain, then one representative test was chosen • Random effects model • Assessed heterogeneity and publication bias
Naguib, JM. 2009 [44] 24 Case-control studies	Type 1 diabetes	• Children (≤19 years old) with type 1 diabetes and defined control group • At least 3 subjects per group • Used standardized neuropsychological tests	Not reported	• Intelligence • Visuospatial • Language and education • Memory and learning • Psychomotor activity • Attention • Executive function	• SMD (Cohen's d) • Fixed and random effects models • Assessed heterogeneity and publication bias • Sensitivity and subgroup analyses
Nieto, RG. 2011 [45] 12 Observational studies	Early onset schizophrenia or pediatric bipolar disorder	• Included children (≤18 years old) with early onset schizophrenia or pediatric bipolar disorder and a healthy control group • Data available to calculate ESs • Published in English	Not reported	• Attention • Working memory • Executive control • Visual memory • Verbal learning/memory • Visuospatial skills • Verbal fluency • Processing speed	• SMD (Hedges and Olkin) • Random effects model • Assessed heterogeneity and publication bias • Subgroup and sensitivity analyses
Quinn, TJ. 2011 [46] 21 Observational studies (case-control, cross-sectional or longitudinal)	Circulating hemostatic measures	• Adult patient population (≥18 years) • Analysis included measures of at least one circulating blood biomarker that was pertinent to hemostasis	DST, LM, RAVENS, VFT	• Speed of processing • Verbal declarative memory • Non-verbal reasoning • Executive function • General composite cognitive function	• SMD • Fixed and random effect models • Assessed heterogeneity and publication bias • Sensitivity analyses

Supplemental Table 1. Summary articles describing meta-analyses of cognitive measures

Authors # and type of Studies Included	Intervention/ Exposure	Inclusion Criteria	Types of Cognitive Measures	Cognitive Domains	How Data Were Combined
Voss, MW. 2010 [47] 20 Observational studies	Sports expertise	• Used a controlled laboratory examination of cognitive skills • Compared expert athletes with matched control group of non-expert athletes • Published in English	Not reported	• Attentional cuing • Processing speed	• SMD (Hedges' g) • ESs averaged if multiple measures for the same domain • Random effects model • Assessed heterogeneity and publication bias • Subgroup analyses
Zhang, JP. 2010 [48] 8 Observational studies	Genetic variation in human dystrobrevin binding protein 1 (DTNBP1)	• Participants with DTNBP1 polymorphisms and healthy control • Reported the full-scale IQ score	CANTAB, COWAT CPT-I/P, CVLT, MWT-B, WAIS-III, WAIS-R, WMS-III, WRAT-3	• General cognitive ability	• SMD (Hedges' g) • Random effect model • Assessed heterogeneity and publication bias • Sensitivity analyses • Meta-regression

ADAS-Cog = Alzheimer's Disease Assessment Scale-Cognitive; ADHD = attention deficit/hyperactivity disorder; ALL = acute lymphocytic leukemia; AVLT = Auditory Verbal Learning Test; BAS = British Ability Scale; BD = Block design; BIMC = Blessed Information-Memory-Concentration; BNT = Boston Naming Test; BSID = Bayley Scales of Infant Development; BSRT = Buschke Selective Reminding Test; BVLT = Buschke Verbal Learning Test; BVRT = Benton Visual Retention Test; CAMCOG = Cambridge Cognition Examination; CANTAB = Cambridge Neuropsychological Test Automated Battery; CBCL = Child Behaviour Checklist; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; CHD = congenital heart disease; Choice RT = Choice Reaction Time; COWAT = Controlled Oral Word Association Test; CPAS-R = Cognitive Psychomotor Assessment System-Revised; CPT = Continuous Performance Test; CPT-I/P = Continuous Performance Test-Identical Pairs Version; CRT = Choice Reaction Task; CTBS = Comprehensive Test of Basic Skills; CVLT = California Verbal Learning Test; DAS = Differential Ability Scale; DG = Differentiele Geschiktheidsbatterij; DRS = Dementia Rating Scale; DS = Digit span; DSM (-III-R, -IV) = Diagnostic and Statistical Manual of Mental Disorders (3rd Edition-Revised, 4th Edition); DSST = Digit symbol substitution test; DTNBP1 =; EMF = electromagnetic fields; ERT = Estrogen replacement therapy; ES = effect size; FAS = FAS, letter fluency test; FNT = Face name task; FR = Face Recognition; FT = Finger Tapping; GCI = Global Cognitive Index; GMT = Group Mathematics Test; GPT = Grooved Pegboard Test; GSM/UMTS = Global System for Mobile Communications (originally Groupe Spécial Mobile)/Universal Mobile Telecommunications System; GSM = Global System for Mobile Communications (originally Groupe Spécial Mobile); GUMT = Guild Memory Test; HAWIE = Hamburger Wechsler Intelligence Test for Adults; HAWIK = Hamburger Wechsler Intelligence Test for Children; HAWIVA = Hamburger Wechsler for Children in Pre-school Age; HRNB = Halstein-Reitan Neuropsychologic Battery; HRT = hormone replacement therapy; HSCS = High Sensitivity Cognitive Screen; HSD = Hasegawa dementia scale; HVLT = Hopkins Verbal Learning Test; HVOT = Hooper Visual Organization Test; IQ = intelligence quotient; KABC = Kaufman Assessment Battery of Childhood; K-BIT = Kaufman Brief Intelligence Test; LIS = Leiter International Scale; LM = logical memory; MCI = Memory Controllability Inventory; MCST = Modified Card Sorting Test; MFFT-20 = Matching Familiar Figures Test-20; MFQ = Memory Functioning Questionnaire; MHz = megahertz; MIA = Metamemory in Adulthood Questionnaire; MIQS = McCarthy IQ Scale; MISIC = Malin's Intelligence Scale for Children; MMSE = Mini Mental State Examination; MPH = methylphenidate; MSCA = McCarthy Scales of Children's Abilities; MWT-B = Mehrfachwahl-Wortschatz-Intelligenztest-Version B; NAR = Neale Analysis of Reading; NART = National Adult Reading Test; NEPSY = Developmental Neuropsychological Assessment; OOHMT = Otis Ottawa d'Habitele Mentale Test; PASAT = Paced Auditory Serial Addition Test; PGI = Post Graduate Institute (India); PMAT-FC = Primary Mental Abilities Test for Filipino Children; RAVENS = Raven's progressive matrices; RAVLT = Rey Auditory Verbal Learning Test; RBANS = Repeatable Battery for the Assessment of Neuropsychologic Status; RCAVLT = Rey and Crawford Auditory Verbal Learning Test; RCFT = Rey-Osterrieth Complex Figure Test; RCT = randomized controlled trial; RIPA = Ross Information Processing Assessment; ROCF = Rey Osterreich Complex Figure; ROVMT = Rey-Osterrieth Visual Memory Test; RVLT = Rey Visual Learning Test; SB = Stanford Binet Scale; SD = standard deviation; SDMT = Symbol Digit Modalities Test; Simple RT = Simple reaction time; SMC = subjective memory complaints; SMD = standardized mean difference; SRT = Simple Reaction Task; SUB = Subtraction; TAP = Test of Attentional Performance; TBI = traumatic brain injury; TMT = Trail Making Test; VER = Sentence Verification; VFT = verbal fluency test; VIG = Vigilance; VLT = verbal learning test; VMI = visual-motor integration; VMT = Visuospatial Memory Test; VRT = Visual Retention Test; vs. = versus; WAIS = Wechsler Adult Intelligence Scale; WAIS-R = Wechsler Adult Intelligence Scale-Revised; WCST = Wisconsin Card Sorting Test; WIAT = Wechsler Individual Achievement Test Screener; WISC (-III, -R) = Wechsler Intelligence Scale for Children (3rd Edition, Revised); WMD = weighted mean difference; WMS = Wechsler Memory Scale; WMS-III = Wechsler Memory Scale-Third Edition; WMS-III = Wechsler Memory Scale-Third Edition; WMS-R = Wechsler Memory Scale-Third Edition-Revised; WPPSI = Wechsler Preschool and Primary Scale of Intelligence; WRAML = Wide Range Assessment of Memory and Learning; WRAT = Wide Range Achievement Test; WRAT-3 = Wide Range Achievement Test-Third Edition; ZVT = Zahlen-Verbindungs-Test

Supplemental-Table 2. Summary of supplemental articles on individual participant data meta-analysis and methods to support statistical harmonization

Topic	Citation	Summary
General articles on conducting IPD	Blettner, M. 1999 [3]	• Described the strengths and limitations of four methods of summarizing data: qualitative summary, meta-analysis of published data, re-analysis of IPD, and prospectively planned pooled analyses • Harmonization of data not mentioned
	Cooper, H. 2009 [57]	• Discussed the relative merits of conducting an IPD vs. aggregated data (AD) analysis • IPD permits subgroup analysis and quality assurance of the original analysis reported in the literature, but is more costly
	Curran, PJ. 2009 [61]	• Discussed issues around conducting integrative data analysis (IDA) as defined as the statistical analysis of a single data set that consists of two or more separate samples that have been pooled into one • Identified two possible methods to deal with heterogeneity due to measurement: nonlinear factor analysis (NFA) and item response theory (IRT) • They recognize the issue of combining complex constructs, such as depression, but they do not provide detailed information on how to handle this issue in pooling data from different studies
	Friedenreich, CM. 1993 [67]	• Presented a methodology for the pooling and analysis of epidemiologic studies using individual subject level data • Discussed random and fixed effect models, examining homogeneity of effects, explaining any heterogeneity, sensitivity analyses and quality assessment
	Ioannidis, JPA. 2002 [73]	• Discussed advantages and disadvantages of IPD meta-analysis of time to event data in genetic epidemiology • Standardization of information across studies using *a priori* definitions was listed as an advantage as a standardized set of variables was available for all studies • Other issues around harmonization were not mentioned
	Riley, RD. 2010 [2]	• Discussed the rationale, conduct and reporting of IPD meta-analyses • Did not discuss the issue of different variables/measures being available among datasets
	Schmid, CH. 2003 [87]	• Discussed issues around conducting an IPD analysis using data from multiple international RCTs evaluating the effect of ACE inhibitors for treatment of nondiabetic renal disease
	Simmonds, MC. 2005 [89]	• Reviewed methods used to conduct IPD meta-analyses conducted during 1999-2001. • Harmonization of data not mentioned
	Van der Steen, JT. 2008 [96]	• Discuss benefits and pitfalls of pooling databases from comparable observational studies of lower respiratory infection in nursing home residents in the U.S. (Missouri) and the Netherlands (Amsterdam) • Identified issues in comparability in measurements in terms of: 1) question wording and response options, 2) clinical meaning, 3) response distributions • If response distributions to the same question differed by population, they tried to do qualitative interviews with physicians to determine whether the variable had a different meaning between the countries • Did not discuss specific methodology for constructing new variables when differences there were differences between the variables in the two databases
	van Walraven, C. 2010 [97]	• Discussed reward and challenges of IPD meta-analysis. • Reward: outcome and analytical harmonization • Challenge: getting and harmonizing data

Supplemental Table 2. Summary of supplemental articles on individual participant data meta-analysis and methods to support statistical harmonization (cont'd)

Topic	Citation	Summary
IPD analysis Methods	Bennett, DA. 2003 [54]	• Reviewed analytic methods for prospective cohort studies using time to event data for single studies and IPD meta-analyses • Discussed issues around missing data (event times and covariates) for individual studies as well as for IPD meta-analyses • Suggested running a simulation sensitivity analysis to determine the extent of biasing and underestimation of standard errors using different methods for imputation of event times • In their example, the authors did not employ imputation methods for covariates due to the size of the data set • The authors reported the number and nature of the missing covariate values according to key variables such as cohort, censoring status, age at recruitment.
	Granda, P., Blasczyk, E. 2010 [8]	• Defined general approaches to harmonization • Input harmonization aims to achieve standardizes measurement processes and methods in all national or regional populations • Output harmonization uses different national or regional measurements possibly derived from non-standard measurement tools • An ex-post strategy to output harmonization (i.e., surveys made comparable after the fact, retrospective harmonization) requires a conversion process • This conversion process should be transparent , well documented ,and reversible • Focus should be given to both variable level and survey level harmonization • Need to develop criteria to assess the quality of harmonization
	Granda, P., Wolf, C., Hadorn, R. 2010 [15]	• Discussed strategies and issues around harmonization of survey data • Provided methods for assessing the quality of harmonization or the degree to which the original information is preserved in the harmonized data • This is most applicable to direct harmonization (i.e., when a single harmonized variable is created directly a single questionnaire item)
	Hofer, SM. 2009 [71]	• Discussed the challenges of meta-analytic and pooled data approaches using cognitive aging literature as an example • Discussed concurrent calibration (cocalibration) of data using IRT models or with latent variable approaches based on item- or scale-level data across studies • Feasibility of pooling variable is limited when variables are not operationally defined in the same way • Using standardized variables (T scores) or proportion correct requires assuming the measurement properties of the variables are relatively comparable and linear • Also need to consider population characteristics such as age, birth cohort, education ranges • Proposed a coordinated analysis approach to enhance communication and collaboration among researchers, facilitate reproducible research, archive analysis and measurement alignment process, to provide a stronger basis for cumulative science, and to permit quick entry into completed analyses
	Jones, AP. 2009 [74]	• Discussed methods used to combine longitudinal clinical trial data across studies using IPD and aggregate data methods • Did not discuss the issue of different variables/measures being available among datasets
	Mathew, T. 2010 [76]	• Compared One-step (linear function of the mean obtained from a linear model of IPD) vs. Two-step (linear function of the mean obtained from linear model of summary data) meta-analysis models using IPD • It provides a nice overview of IPD meta-analysis • Did not discuss the issue of different variables/measures being available among datasets

Supplemental Table 2. Summary of supplemental articles on individual participant data meta-analysis and methods to support statistical harmonization (cont'd)

Topic	Citation	Summary
Comparison of imputation methods	Burgess, S. 2011 [55]	• Described four Bayesian methods for imputing missing data based on a missing at random (MAR) assumption in the context of genetic epidemiology: multiple imputations, single nucleotide polymorphism (SNP) imputation, latent variables, and haplotype imputation • Results of a simulation study and application to the British Women's Heart and Health Study were presented • Method analogous to the 2-stage least-squares method except it accounts for the observational correlation between phenotype and outcome. This analysis was done using WinBUGS • Precision was improved using four imputation methods – equivalent to 25% increase in sample size • All imputation methods give similar results
	Donegan, S. 2010 [63]	• Reviewed the reporting and methodological quality of indirect comparisons (which could be considered an extreme missing data situation) • Authors conducted a systemic review including 43 reviews in which clinical effectiveness of two interventions were indirectly compared • In general, the underlying assumptions of conducting an indirect comparison analysis were not routinely described or tested
	Peyre, H. 2011 [83]	• Compared imputation method for data Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR) within one study • Methods compared: personal mean score (PMS), multiple imputation (MI), hot deck (HD) imputation and full information maximum likelihood (FIML) • MI and FIML superior to PMS and HD in terms of accuracy and precision • HD tended to underestimate and PMD associated with insignificant bias
	Siddique, J. 2011 [88]	• The authors used an imputation approach to calibrate rater bias in the diagnostic assessment of posttraumatic stress disorder (PTSD) • Nurse practitioners were twice as likely to diagnose PTSD than a clinical psychologist – and each patient was randomly assigned to be rated by only one rater • A Bayesian random effects censored ordinal probit model was used to identify a latent moderate class of patients • A Markov chain Monte Carlo algorithm was used to estimate the posterior distribution of the model parameters and generate multiple imputations for the recalibrated diagnosis variable
	Spratt, M. 2010 [91]	• Examined how the choice of imputation model and the number of imputations affected estimates of prevalence and associations in a study of wheezing among 81-month-old children in Avon Longitudinal Study of Parents and Children • Preliminary analyses of the association of measured variables with missingness and outcome variables are required to determine the plausibility of the assumptions underlying both complete-case and multiple-imputation-based analyses. • They applied a covariate (socioeconomic status) which was available on all subjects as an intermediate variable to generate multiple imputations procedures • Analyses of MI should often be based on 25 or more imputed values in order to reduce the impact of random sampling inherent in the MI process
	Sterne, JAC. 2009 [92]	• Reviewed the reasons why missing data may lead to bias and loss of information • Discussed situations in which multiple imputation may help reduce bias and increase precision as well as the potential pitfalls • Proposed guidelines for reporting analyses using multiple imputation

Supplemental Table 2. Summary of supplemental articles on individual participant data meta-analysis and methods to support statistical harmonization (cont'd)

Topic	Citation	Summary
Methods for evaluating equivalence	Crane, PK. 2008 [59]	• Compared item- and scale-level strategies for handling demographic heterogeneity when measuring executive function • Examined the extent to which item-level and scale-level adjustment for demographic variables influenced the relationships with various composite executive function scores with an external criterion (MRI) • The authors created composite scores for executive function using classical test theory and item response theory in which demographic differences were ignored or taken into account • Candidate scores were compared using 3 linear regression models; model A included demographic terms as independent variables, model B include MRI variables, and model C included both • R^2 was used to estimate effect sizes
	Teresi, JA. 2007 [93]	• Discussed methods based on IRT that can be used to examine differential item functioning (DIF) within study subgroups • The method used was the item response theory log-likelihood ratio (IRTLR) approach • This method could also be extended to testing DIF among study populations

Abbreviations: ACE = angiotensin-converting-enzyme; DIF = differential item functioning; FIML = fill information and maximum likelihood; HD = hot deck; IDA = integrative data analysis; IPD = individual patient data; IRT = item response theory; IRTLR = item response theory log-likelihood ratio; MAR = missing at random; MCAR = missing completely at random; MI = multiple imputation; MNAR = missing not at random; MRI = magnetic resonance imaging; NFA = nonlinear factor analysis; PMS = personal mean score; PTSD = post traumatic stress disorder; RCT = randomized controlled trial; SNP = single nucleopeptide polymorphisms; U.S. = United States of America

Supplemental Table 3. Examples of studies presenting harmonized data

Topic	Citation	Summary
Examples of analyses of harmonized data	Anstey, KJ. 2010 [49]	• Harmonized data from the Dynamic Analyses to Optimizing Ageing (DYNOPTA) project • Harmonized data [including cognitive measures] from 9 Australian cohorts using response conversion (see van Buuren above) • Did not give details on the analysis
	Bath, PA. 2010 [50]	• Harmonized data from the Longitudinal Aging Study Amsterdam (LASA) and the Nottingham Longitudinal Study on Activity and Ageing (NLSAA) [including cognitive measures] • LASA used the Mini Mental State Exam (MMSE: 30 point scale) and the NLSAA used the Clifton Assessment Procedures for the Elderly (CAPE: 12 point scale) • The derived variables were simply MMSE/30 and CAPE/12
	Beer-Boorst S. 2000a [52] Beer-Boorst, S. 2000b [53]	• Developed a common surveillance system to allow for the comparison of lifestyle and biological risk factors from different populations across Europe including seven collaborating centers [European Alimentation (EURALIM)] • Common variables included: diet, health, lifestyle and demographic variables • Did not discuss method of harmonization
	Crane, PK. 2008 [58]	• Used IRT to cocalibrate cognitive scales from three large community-based studies (the Cardiovascular Health Study [CHS], the Adult Changes in Thought Study [ACT] and the Indianapolis site from the Indianapolis-Ibadan Dementia Project • The primary objective was to cocalibrate the Mini Mental State Examination (MMSE), Modified Mini Mental State Exam (3MS), Cognitive Abilities Screening Instrument (CASI), and The Community Screening Instrument for Dementia (CSI 'D') • Used McDonald's bifactor model to evaluate whether the scales were unidimensional • Identified anchor items that were comparable across tests – only included identical items (e.g., interlocking pentagons) • Used Samejima's graded response model to estimate the probability of each response category for each item for any level of cognitive functioning. This formula was used to determine the most likely response for every cognitive functioning level
	Curran PJ. 2008 [60]	• Used IRT to fit a series of growth curve models to a single pooled sample that consists of data drawn from three separate studies of developmental internalizing symptomology • The studies examined children with and without alcoholic parents (The Michigan Longitudinal Study [MLS], the Adolescent/Adult Family Development Project [AFDP], and the Alcohol and Health Behavior Project [AHBP]) • There were 21 unique dichotomous self-report items to define internalizing symptomology; four items were present in all studies • Dimensionality Step: Factor analysis was used to examine the dimensionality of the 21 items by conducting an exploratory factor analysis in each study to assess unidimensionality based on traditional measures including eigenvalues, scree plots and estimates of incremental variance • Calibration Step: Fitted a standard 2PL IRT model to the 21 dichotomous items from a single randomly selected assessment for each participant in the pooled sample • DIF Step: Estimated a series of multiple group IRT models as a function of developmental status, gender and study group membership • Scoring Step: calculated individual time-specific scale scores for every participant at every time point at which they were assessed using a modal a posteriori method.

Supplemental Table 3. Examples of studies presenting harmonized data (cont'd)

Topic	Citation	Summary
Examples of analyses of harmonized data (cont'd)	Darby S. 2006 [62]	• Authors used data from 13 studies of residential radon and lung cancer carried out in Europe (Austria, the Czech Republic, Finland [2], France, Germany [2], Italy, Spain, Sweden [3], and the UK) • Data were assembled according to a common format and uniform definitions were used except for study-specific definitions for social status [did not detail how this was put into a common metric] • Data were analyzed using a linear odds model; models were fit using conditional maximum likelihood (similar to conditional logistic regression)
	The Fibrinogen Studies Collaboration. 2009 [66]	• The authors combine data on the association between plasma fibrinogen and coronary heart disease in 31 cohort studies using proportional hazards (Cox) model, stratified by cohort, sex and (for the two RCTs) trial arm • All studies provided data on fibrinogen level, age, smoking status, total cholesterol, SBP and BMI • Some studies also provided data on HDL and LDL cholesterol, alcohol consumption, triglycerides and history of diabetes • The authors use a two-stage process. At the first stage partially and (where possible) fully adjusted estimates are obtained from each study, together with their standard errors (a key issue is estimating the within study correlation of the two estimates) • At the second stage, the results are combined in a bivariate meta-analysis • This study addresses the issue of when studies included in an IPD meta-analysis include some, but not all, important confounding variables • The proposed bivariate model, with estimates of the parameter of interest either fully or partially adjusted for confounding factors may be useful also for more difficult constructs. Some studies measuring the construct fully and other studies measuring the construct only partially could possibly analyzed with this bivariate approach
	Grimm KJ. 2010 [69] (see also Duncan, GJ. 2007 [64]	• The authors examined the associations between early behavioral and cognitive skills with later achievement using data from 3 longitudinal studies (the NICHD Study of Early Child Care and Youth Development [SECCYD], the National Longitudinal Survey of Youth–Children and Young Adults [NLSY-CYA], and the Early Childhood Longitudinal Study Kindergarten Cohort [ECLS-K]) • Behavior scales differed among the studies and categorization of children into "normative", "problematic" and "clinical" groups was done using set cut-offs or based on the observed distribution of the data (e.g., T-scores) • The authors used a combined item-response and growth curve model to account for differential reliability
	Khachaturian, AS. 2010 [75]	• The authors describe the challenges and opportunities for developing a national database for successful aging • One of the main challenges is defining a "case" for population-based prevention studies • Clinical assessment conducted by experts produce accurate diagnoses, but are very costly, labor intensive and require highly trained personnel • Population studies, because of a lower yield, necessitates greater efficiency and lower-cost less-highly trained personnel • Need for multi-stage assessment among subject, other informants (e.g., family members), as well as clinical assessments by clinicians and non-clinicians (including surveys by mail and telephone cognitive assessments) • The critical questions identified were: 1) can these assessment approaches be refined in order to detect or predict individuals who may develop future impairments, or declination, in cognition or behavior, or even scaled down for high volume throughput; 2) can technologies be developed to allow the most passive, non-intrusive assessment of the individual's cognitive and behavioral function; and 3) will the collected longitudinal data afford the possibility to measure intra-person change, vis-a-vis Bayesian-modeling approaches • The issue of measuring within-person change over time was also highlighted as the ultimate aim is to predict the trajectory of an individual's cognitive-behavioral-functional health, the rate of decline, and the point at which one crosses the threshold from an asymptomatic stage to a phenotype resembling pre-MCI, then to MCI, and then to AD

Supplemental Table 3. Examples of studies presenting harmonized data (cont'd)

Topic	Citation	Summary
Examples of analyses of harmonized data (cont'd)	McArdle, JJ. 1998 [79] [see also McArdle, JJ. 1994; [77] McArdle, JJ. 1997 [78]]	• The authors use methods based on linear structural equations models with incomplete or missing data to analyze longitudinal twin data for two cognitive variables to evaluate a biometric genetic hypothesis in the context of a developmental model of intellectual growth and change (biometric genetic analysis of intellectual abilities [BGIA]) • In this study, the same measurement scales (block design and vocabulary measures) were used over time, however the number of observations, the age at first administration, and the interval between administrations differed within the twin pairs and among the sets of twins. The raw scores were transformed into percentage-correct scales (0-100). • The authors used all available data, including participants with incomplete and possibly nonrandomly missing data • The authors incorporated a twin analysis including means and age effects; a longitudinal analyses based on latent growth components; and a biometric-genetic analyses for components of growth using linear structural equations models
	Minicuci, N. 2011 [82]	• Compared measures of Disability Free Life Expectancy (DFLE) across different surveys conducted in Bulgaria (National Health Interview Survey [NHIS]), Italy (Multidisciplinary Survey among Italian Families [IMF-S]) and Latin America (the Salud, Bienestar y Envejecimiento [SABE]) • Harmonized 5 ADL questions common to all surveys • Dichotomized responses to create a common scale
	Pluijm, SMF. 2005 [84]	• Constructed a harmonized measure of ADL using data from six countries contributing data to the Comparison of Longitudinal European Studies on Aging (CLESA) Study • There was overlap in the ADL items among countries, but only 2 of the 11 possible items were asked in all surveys • Items that were incompatible across countries because of cultural differences were excluded from the harmonization process • Harmonization focused on the four items comprising the Katz ADL index; all four items were present in four of the six country surveys; five- and six-item scales were constructed in the countries that had the additional items in common • In countries where the two items were not measured, the data for these was extrapolated from other "comparable" ADL items • Because they used different response options among the surveys all items were dichotomized to put them on a consistent scale • Subjects were excluded if 2 or more items were missing; hot deck methods were used to impute values when one of the items was missing due to nonresponse • Reliability and validity of the four item scale was assessed
	Ruggles, S. 2003 [85] Esteve, A. 2003 [65]	• The Integrated Public Use Microdata Series (IPUMS)-International involved working with census data from different time periods and institutional origins • The first stage of harmonization involved standardizing the data formats and correcting errors • The second stage of harmonization involved harmonizing the codes for all variables across datasets • Variable-level harmonization involved recoding variables to maximize comparability across datasets. • In this example, the content included among the datasets was greatly overlapping, but different numeric classification systems were used

Supplemental Table 3. Examples of studies presenting harmonized data (cont'd)

Topic	Citation	Summary
Examples of analyses of harmonized data (cont'd)	Schenker, N. 2007 [86]	• The authors describe several situations in which data from multiple surveys were used to enhance estimation of measures of health • The four projects involved: (1) combining estimates from a survey of households and a survey of nursing homes to extend coverage; (2) using information from an interview survey to bridge the transition in race reporting in the United States census; (3) combining information from an examination survey and an interview survey to improve on analyses of self-reported data; and (4) combining information from two interview surveys to enhance small-area estimation • In project 3, the authors discussed methods for combining information from two surveys conducted by the National Center for Health Statistics to improve on analyses of self-reported data on health conditions • One of the surveys, the National Health and Nutrition Examination Survey, was unusual in that it not only asked self-report questions on health conditions during face-to-face interviews, but it also obtained clinical measures based on physical examinations • The other survey, the National Health Interview Survey, was larger, and it obtained a rich set of variables for use in multivariate analyses, but it relied on self-report questions for its information on health conditions • 'Measurement error' models that predict clinical outcomes from self-reported answers and covariates were fitted to data from the National Health and Nutrition Examination Survey, and the fitted models were then applied to data from the National Health Interview Survey to adjust for possible inaccuracies due to self-reporting • Multiple imputation was used to properly reflect the sources of variability in subsequent analyses
	Slimani, N. 2002 [90]	• Harmonized data from 10 Western European countries (Denmark, France, Germany, Greece, Italy, Norway, Spain, Sweden, The Netherlands, UK) which from the European Prospective Investigation into Cancer and Nutrition (EPIC) project • Information on usual individual dietary intakes was obtained using different dietary assessment methods developed and validated in each participating country • A calibration approach was adopted to adjust for possible systematic over- or underestimation in dietary intake measurements and correct for attenuation bias in relative risk estimates • A single 24-hour dietary recall was collected from a random sample of 5-12% (1.5% in the UK) of the EPIC cohorts, weighted according to the cumulative number of cancer cases expected per fixed age and sex stratum • Standardized software (EPIC-SOFT) was developed to assess dietary intake reported across the EPIC centers
	van Buuren, S. 2003 [94] Hopman-Rock, M. 2000 [72]	• Used response conversion to harmonize international disability information from ERGOPLUS (Rotterdam) and EURIDISS (3 countries in Europe) • The first step was to create a conversion key; used Rasch modeling (a partial credit model) to estimate the parameters for the conversion key • The second step involved the conversion of the observed data onto the common scale

Abbreviations: 2PL IRT = two-parameter logistic using item response theory; 3MS = Modified Mini Mental State Exam; ACT = Adult Changes in Thought Study; AD = Alzheimer's disease; ADL = activities of daily living; AFDP = Adolescent/Adult Family Development Project; AHBP = Alcohol and Health Behavior Project; BGIA = biometric genetic analysis of intellectual abilities; BMI = body mass index; CAPE = Clifton Assessment Procedures for the Elderly; CASI = Cognitive Abilities Screening Instrument; CHS = Cardiovascular Health Study; CLESA = Comparison of Longitudinal European Studies on Aging; CSI 'D' = Community Screening Instrument for Dementia; DFLE = Disability Free Life Expectancy; DIF = differential item functioning; DYNOPTA = Dynamic Analyses to Optimizing Ageing; ECLS-K = Early Childhood Longitudinal Study Kindergarten Cohort; EPIC = European Prospective Investigation into Cancer and Nutrition; EURALIM = European Alimentation; HDL = high-density lipoprotein; IMF-S = Multidisciplinary Survey among Italian Families; IPD = independent patient data; IPUMS = Integrated Public Use Microdata Series; IRT = item response theory; LASA = Longitudinal Aging Study Amsterdam; LDL = low-density lipoprotein; MCI = mild cognitive impairment; MLS = Michigan Longitudinal Study; MMSE = Mini Mental State Examination; NHIS = National Health Interview Survey; NLSAA = Nottingham Longitudinal Study on Activity and Ageing; NLSY-CYA = National Longitudinal Survey of Youth-Children and Young Adults; RCT = randomized controlled trial; SABE = Salud, Bienestar y Envejecimiento; SBP = systolic blood pressure; SECCYD = NICHD Study of Early Child Care and Youth Development; UK = United Kingdom