

Text: 2884 words

Linking the Clinical Dementia Rating Scale-Sum of Boxes, the Clinician's Interview-Based

Impression Plus Caregiver Input and the Clinical Global Impression Scale:

Evidence based on Individual Participant Data from Five Randomized Clinical Trials of

Donepezil

Global Ratings of Alzheimer's Disease

Myrto **Samara**, Stephen Z **Levine**, Kazufumi **Yoshida**, Yair **Goldberg**, Andrea **Cipriani**, Orestis **Efthimiou**, Takeshi **Iwatsubo**, Stefan **Leucht**, Toshiaki A. **Furawaka**

Dr. Myrto Samara, MD, Ph.D. (Department of Psychiatry and Psychotherapy, Technical University of Munich, School of Medicine, Munich, Germany) AND (3<sup>rd</sup> Department of Psychiatry, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece).

[myrtosamara@auth.gr](mailto:myrtosamara@auth.gr) 0000-0002-6943-5091

Professor Stephen Z Levine, Ph.D., (Department of Community Mental Health, University of Haifa, Israel). [slevine@univ.haifa.ac.il](mailto:slevine@univ.haifa.ac.il) 0000-0002-5544-0420

Dr. Kazufumi Yoshida, MD, MSc. (Department of Health Promotion and Human Behaviour, Kyoto University Graduate School of Medicine/ School of Public Health, Kyoto, Japan).

[yoshida.kazufumi.77s@st.kyoto-u.ac.jp](mailto:yoshida.kazufumi.77s@st.kyoto-u.ac.jp) 0000-0003-4475-8043

Professor Yair Goldberg, Ph.D. (Faculty of Industrial Engineering and Management, Haifa Technion, Israel). [yairgo@technion.ac.il](mailto:yairgo@technion.ac.il) 0000-0002-4544-0910

Professor Andrea Cipriani, MD, Ph.D., (Department of Psychiatry, University of Oxford, Oxford, UK; Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK). [andrea.cipriani@psych.ox.ac.uk](mailto:andrea.cipriani@psych.ox.ac.uk) 0000-0001-5179-8321

Dr. Orestis Efthimiou, Ph.D. (Institute of Social and Preventive Medicine (ISPM); University of Bern, Switzerland; Department of Psychiatry, University of Oxford, Oxford, UK); [oremiou@gmail.com](mailto:oremiou@gmail.com) 0000-0002-0955-7572

Professor Takeshi Iwatsubo, MD, Ph.D. (University of Tokyo, Bunkyo-ku, Tokyo, Japan) [iwatsubo@m.u-tokyo.ac.jp](mailto:iwatsubo@m.u-tokyo.ac.jp) 0000-0003-1160-8129

Professor Stefan Leucht, MD, Ph.D. (Department of Psychiatry and Psychotherapy, Technical University of Munich, School of Medicine, Munich, Germany) [stefan.leucht@tum.de](mailto:stefan.leucht@tum.de) 0000-0002-4934-4352

Professor Toshiaki A. Furukawa MD, Ph.D. (Department of Health Promotion and Human Behaviour, Kyoto University Graduate School of Medicine/ School of Public Health, Kyoto, Japan) [furukawa@kuhp.kyoto-u.ac.jp](mailto:furukawa@kuhp.kyoto-u.ac.jp) 0000-0003-2159-3776

**Correspondence.** Myrto Samara, MD, Ph.D. (Aristotle University of Thessaloniki, Thessaloniki, Greece; Technische Universität München, Munich, Germany) <[samaramyrto@gmail.com](mailto:samaramyrto@gmail.com)>, tel: +306980107450

## **Abstract**

**Background:** In patients with Alzheimer's disease, global assessment scales, such as the Clinical Dementia Rating-Sum of Boxes (CDR-SB), the Clinician's Interview-Based Impression Plus Caregiver Input (CIBI plus), and the Clinical Global Impression (CGI) are commonly used.

**Objective:** To clinically understand and interpret the associations between these scales, we examined the linkages for the total and change scores of CDR-SB, CIBI plus and CGI.

**Methods:** Individual participant data (N= 2,198) from five pivotal randomized placebo-controlled trials of donepezil were included. Data were collected at baseline and scheduled visits for up to 6 months. Spearman's correlation coefficients  $\rho$  were examined between corresponding total and change scores of simultaneous CDR-SB, CIBI plus, and CGI ratings. To link between the simultaneous ratings, equipercentile linking was used.

**Results:** We found strong evidence that the Spearman's correlation coefficients between the CDR-SB and CGI, and CDR-SB and CIBI plus total scores were at least adequately correlated ( $\rho=0.50$  to  $0.71$ , with  $P<0.01$  ). The correlation coefficients between the change scores of CDR-SB and CGI were deemed adequate for weeks 6 to 24 ( $\rho=0.44$  to  $0.65$ ); the remaining correlations were smaller in magnitude ( $\rho=0.09$  to  $0.35$ ). Overall, the linkages were in-line with expectations e.g., CDR-SB range score of 3-4 (=very mild dementia) was linked to a CGI score of 3 (=mildly ill), and an increase of CDR-SB of 1 was linked to a change of 5 (=minimal worsening) in both CGI and CIBI plus.

**Conclusion:** The study findings can be useful for clinicians wishing to compare scores of different scales across patients. They can also help researchers understand results of studies using different scales and can facilitate meta-analyses, to increase statistical power.

**Keywords:** Alzheimer; global ratings; CIBIC; CIBIS; CDR-SB; CGI; minimal clinically important difference.

## **Introduction**

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by progressive impairments in cognition, social and functioning activities, and ultimately death [1]. Multiple rating scales have been developed to measure functional impairment. These scales ascertain patient and caregiver information in addition to the clinician's global impressions of the patient's functional and cognitive capacities [2]. The scales are widely used in research and clinical settings to support a diagnosis, assess disease severity or progression, and quantify treatment efficacy. In 1990, the US Food and Drug Administration (FDA) highlighted that a global clinical assessment is essential and should be a co-primary outcome measure, in addition to those of cognitive functioning, when testing antidementia drugs based on the premise that a change in the patient's global condition is clinically meaningful [3]. Recently, the FDA [4], the European Medicines Agency (EMA) [5] and the Japanese Pharmaceuticals and Medical Devices Agency (PMDA) [6] suggested that, at least for prodromal AD/mild cognitive impairment due to AD, a single primary endpoint examining both cognition and function could also be considered appropriate, a shift from the generally applied co-primary approach; nevertheless, this approach remains a requirement according to all three regulatory agencies for established AD. In response to the relevant guidelines, various global assessment scales are used.

The Clinical Dementia Rating (CDR) scale is a frequently applied global assessment instrument used to appraise the cognitive and functional aspects of dementia [7]. While the instrument is subjective as it relies on clinical impression after examining the patient and a caregiver/informant, evidence indicates that this measure has high validity and reliability [7-12]. A more expanded alternative to the CDR is the Clinician's Interview-Based Impression Plus Caregiver Input (CIBI plus). The CIBI plus requires a semi-structured interview with the patient

without other clinical information or test scores but with input from a responsible caregiver and so aims to capture ‘global’ changes across domains of cognition, function and behavior [13, 14]. Finally, the Clinical Global Impression (CGI) scales are used to measure overall symptom severity (CGI-S) or change (CGI-C) in a variety of neurological and psychiatric conditions such as anxiety, bipolar disorder, depression, and schizophrenia [15]. Thus, the CGI is often disorder- and even study-specific. Its administration, scale wording, or scoring instructions may differ, capturing a variety of changes depending on the version included in the study. Nevertheless, the CGI is extremely useful in everyday clinical practice as it is quick and easy to complete, and its interpretation is straightforward even by clinicians who are not experts in a specific condition. Evidence demonstrates that the CGI links well to standardized detailed efficacy scales in anxiety [16-18], depression [19, 20], and schizophrenia [21, 22].

Linking different scales has been undertaken to examine measures of cognition in dementia research [23-26]. However, to date, no study has linked global assessment rating total and changes in AD. Many situations may require the conversion of scores obtained from one scale to equivalent scores on another scale. Score conversions are clinically useful, for example, when comparing different efficacy outcomes across patients. In research settings, linking measures is required when aiming to meta-analyse studies using different scales, and in postmortem research where information may be available on one but not another measure. Therefore, we hereby aimed to link CDR-SB, CIBI plus, and CGI using data from five pivotal double-blind, randomized placebo clinical trials of AD.

## Materials and Methods

### Individual patient data

We requested individual participant data from double-blind, randomized placebo-controlled trials of donepezil conducted by Eisai Co. Ltd. in AD. Data access was provided following the submission of an a priori analytic plan and analyzed via a secure Internet cloud-based platform (<http://www.clinicalstudydatarequest.com>). We included only trials in which patients with AD were assessed using at least two of the following three scales: Clinical Dementia Rating-Sum of Boxes (CDR-SB), Clinician's Interview-Based Impression Plus Caregiver Input (CIBI plus), and Clinical Global Impression (CGI). Institutional review boards had approved all studies, and all participants had given written informed consent.

### Measures

*Clinical Dementia Rating (CDR)*: The CDR examines the patient's performance in six domains. These are memory, orientation, judgment and problem solving, community affairs, home and hobbies performance, and personal care [7]. It involves a semi-structured interview with the patient and an informant or caregiver. It is clinician-rated on a 5-point scale with a rating of 0 meaning 'no dementia', 0.5 'questionable dementia', and 1, 2, and 3 showing 'mild, moderate and severe cognitive impairment' respectively. The six domains are summed to create a 0-18 Sum of Boxes (SB) score, which is considered a more detailed index than the CDR global score (0-3) in cases of mild dementia [27], allowing a reliable differentiation between mild cognitive impairment and AD [28, 29]. According to these interpretive guidelines, a CDR-SB range score of 0.5-4.0 corresponds to 'questionable cognitive impairment', of 4.5-9.0 corresponds to 'mild dementia', of 9.5-15.5 to 'moderate dementia' and of 16-18 to 'severe dementia' [28, 29].



*Clinician's Interview-Based Impression Plus Caregiver Input (CIBI plus)*: The CIBI plus is a measure of severity or change in four domains. Namely, general condition, cognitive function, behavior, and activities of daily living [13, 14]. Ratings are based on a semi-structured interview with the patient and a caregiver, and range from 1 to 7 points; for *CIBI of Severity plus (CIBIS plus)*, 1 corresponds to 'no symptoms' and 7 to 'extremely severe', whereas for *CIBI of Change plus (CIBIC plus)*, 1 corresponds to 'markedly improved' and 7 to 'markedly worse'.

*Clinical Global Impression (CGI)*: The CGI is rated on a 7-point scale, ranging from 1 to 7, and reflects a clinician's global judgment about illness severity or progression. For *CGI-Severity (CGI-S)* scale, 1 corresponds to 'normal', 2 to 'borderline mentally ill', 3 to 'mildly ill', 4 to 'moderately ill', 5 to 'markedly ill', 6 to 'severely ill' and 7 to 'among the most extremely ill patients', whereas for *CGI-Change (CGI-C)*, 1 corresponds to 'very much improved', 2 to 'much improved', 3 to 'minimally improved', 4 to 'no change', 5 to 'minimally worse', 6 to 'much worse', and 7 to 'very much worse'.

### Statistical analysis

Initially, we computed descriptive statistics of the trial measures. For each study, donepezil and placebo arms were found to be very similar in terms of patient characteristics (see eTable 1 for descriptive statistics of the studies) and were pooled to increase statistical power. We analyzed all trials collectively as a unique population rather than by trial to maximize the sample size.

Nevertheless, results on linking were presented separately for drug and placebo arms as well. We used the Spearman correlation coefficient ( $\rho$ ), and the nominal level of statistical significance was adopted ( $P < 0.05$ ). Spearman correlation coefficient was considered poor for values  $< 0.40$ , adequate for values between 0.40 and 0.70, and strong for  $> 0.70$  [30]. Then, equipercenile

linking was computed. This technique identifies scores on each scale that have the same percentile ranks and allows for a conversion between them. Linking assumes a nonparametric association between any two measures, aims to concord between those measures, allows for possible measurement errors on both scales compared and has no independent and dependent variables; thus, linked scores are interchangeable [31]. This method has been used in previous studies extensively [16, 20-22, 32]. In instances of multiple assessments over time, the median value across the different measurement points was used to define the corresponding scores between scales [26]. Equipercetile linking was conducted using the equate library [31] in R 3.6.2.[33].

## Results

We included five pivotal randomized placebo-controlled trials of donepezil [34-38] that employed at least one simultaneous scheduled administration of the CDR-SB, CIBI plus, and CGI in the present analysis. Table 1 details the studies. Data from 2,198 trial participants with AD based on the Diagnostic and Statistical Manual of Mental Disorders, Revised Third or Fourth Edition (DSM-III-R/-IV) [39, 40], were included. None of the trials required biomarker confirmation of AD pathology. There were 854 men and 1,344 women, with a mean age (standard deviation [SD]) of 72.4 (7.5) years. Total and change rating scale score descriptive statistics are shown in Table 2. The CDR-SB was administered in all five trials, whereas the CIBI plus in three trials and the CGI in two trials with no overlap between them (Table 1). Independent raters scored CDR-SB and the global measurement in three of the studies [36-38]. Details on the specific versions of CGI used in the selected studies are provided in the online Supplement A.

### CDR-SB, CGI-S and CIBIS plus

The Spearman correlation coefficients between CDR-SB and CGI-S scores were strong at both available timepoints ( $\rho=0.68$  at baseline, and  $\rho=0.71$  at week 12; Table 2), with strong evidence against the null hypothesis of zero correlation ( $p\text{-value}<0.01$ ) in all instances. Equipercentile linking of CDR-SB scores with CGI-S scores was calculated. An estimated CDR-SB score of 2 corresponded to CGI-S score of 2, CDR-SB of 4 to CGI-S of 3, CDR-SB of 6.5 to CGI-S of 4, and CDR-SB of 11.5 to CGI-S of 5 (Figure A, Table 3).

For both CDR-SB and CIBIS plus scales, only baseline assessments were available. The Spearman correlation coefficient between their scores was deemed adequate ( $\rho=0.50$  at baseline;

Table 2),  $P < 0.01$ . Equipercetile linking of CDR-SB scores with CIBIS plus is presented in Figure B and corresponding scores in Table 3. Approximately, CDR-SB score of 2.5 corresponded to CIBIS plus score of 1, CDR-SB of 3.5 to CIBIS plus of 2, CDR-SB of 5.5 to CIBIS plus of 3, CDR-SB of 7.0 to CIBIS plus of 4, CDR-SB of 11.0 to CIBIS plus of 5, and CDR-SB of 13.0 to CIBIS plus of 6.

#### CDR-SB change, CGI-C and CIBIC plus

The Spearman correlation coefficients between CDR-SB change and CGI-C scores were poor at the first 3 weeks ( $\rho = 0.16$  to  $0.34$ ), and adequate at weeks 4 to 24 ( $\rho = 0.44$  to  $0.65$ ), with  $P < 0.01$  at all time points apart from week 1 (Table 2). Equipercetile linking of CDR-SB change with CGI-C scores was calculated. Based on the median values across different time points, a CDR-SB decrease of 3.5 corresponded to CGI-C score of 1, a CDR-SB decrease of 2.0 corresponded to CGI-C score of 2, a CDR-SB decrease of 0.5 corresponded to CGI-C score of 3, no change on CDR-SB scores corresponded to CGI-C score of 4, a CDR-SB increase of 1.0 to CGI-C score of 5, a CDR-SB increase of 2.5 corresponded to CGI-C score of 6, and a CDR-SB increase of 5.5 corresponded to CGI-C score of 7 (See Figure C and Table 4).

The Spearman correlation coefficients between CDR-SB change and CIBIC plus scores were poor over time (range  $\rho = 0.09$  to  $0.35$ ; Table 2). Equipercetile linking of CDR-SB change with CIBIC plus scores was calculated. Based on the median values across different timepoints, a CDR-SB decrease of 5.0 corresponded to CIBIC plus score of 1, a CDR-SB decrease of 3.0 corresponded to CGI-C score of 2, a CDR-SB decrease of 1.0 corresponded to CGI-C score of 3, no change on CDR-SB corresponded to CGI-C score of 4, a CDR-SB increase of 1.0 corresponded to CGI-C score of 5, a CDR-SB increase of 3.0 corresponded to CGI-C score of 6,

and a CDR-SB increase of 6 corresponded to CGI-C score of 7 (See Figure D and Table 4). In the online Supplement B, results on linking were presented separately for drug and placebo arms (see eTables 2-5).

## Discussion

The current study was the first to examine the linkage among global rating scales in patients with mild to moderate AD (i.e., CDR-SB , CIBI Plus and CGI total and change scores). Based on individual participant data derived from five pivotal randomized controlled trials that compared donepezil and placebo, our results offered researchers and clinicians a concordance table of total and change scores between these global rating scales. There are several notable features to the current study. First, these broadly used scales have never been linked before. Second, the results are from individual-level data on a large sample of participants (n=2,198), strengthening the robustness of our findings. Third, the concordance tables may solve practical problems in everyday clinical practice and research. Fourth, placebo and donepezil treated participants were included in the analysis, which results in a greater spread of post-baseline scores.

The CDR-SB had adequate correlations with CGI-S and CIBIS plus total scores, increasing our confidence in the linking functions. Results suggested that a CDR-SB score of 2-2.5 (=questionable cognitive impairment) was linked to a CGI-S score of 2 (=borderline mentally ill) and to a CIBIS plus score of 1 (=normal). Similarly, a CDR-SB score of 3.5-4 (=very mild dementia) was linked to a CGI-S score of 3 (=mildly ill) and to a CIBIS plus score of 2 (=borderline mentally ill); a CDR-SB range score of 5.5-7 to a CGI-S score of 4 (=moderately ill) and a CIBIS plus score of 3 (=mildly ill) or 4 (=moderately ill); and a CDR-SB range score of 11-13 (=moderate dementia) to a CGI-S score of 5 (=markedly ill) and to a CIBIS plus score of 5 (=markedly ill) or 6 (=severely ill).

In contrast, the correlation of change scores between the rating scales examined was lower, especially for the linking of CDR-SB and CIBIC plus ratings, suggesting that conversion of change scores may be less reliable than total scores, a finding replicated in many other studies

[41, 42]. These lower correlations may partly reflect floor and ceiling effects in extreme change scores (i.e., scales that lack sensitivity to differentiate moderate from marked change and vice-versa) as it is evident in Table 4, with similar scores indicating moderate and marked improvement. In addition, extreme values often result in weakened linkages owing to few available data. Therefore, more caution is necessary when interpreting the results in the very low or very high end of the score range. As for the minimal clinically important difference, our results suggested that a 1-point CDR-SB increase corresponds to minimal worsening (Table 4) whereas Andrews et al. reported that, for patients with mild, moderate or severe AD, a 2 points CDR-SB increase is indicative of a minimal clinically important decline using anchor-based methods [43]; nevertheless, our study was not designed to identify minimal clinically important differences. Moreover, correlations of change scores between CDR-SB and CGI-C were adequate for weeks 16 to 24. Several potential reasons could explain why the change score correlations increased over time (Table 2). One explanation is that clinicians expect a delayed treatment effect of antideementia drugs [44]; thus, as the trial progressed, the variability of scale scores increased, as did the correlation coefficients between them. Another explanation is that the natural progression of the disease induced an increase in functional impairment and its variability.

One should consider additional general drawbacks when analyzing change scores in AD. First of all, clinicians appear to find it easier to identify progressive deterioration than to identify amelioration [45, 46] given the understanding that dementia progression is inevitable, even under treatment. In other words, the model of progressive decline is better defined and identified than the model of improvement [47, 48]. Furthermore, in AD, any given patient's picture may remain stable for an extended time interval or fluctuate within or by day [49]. Thus, it is evident that, in

order for a fixed change of symptom severity to accumulate and become clinically detectable, RCTs with long durations may be needed [50]. Indeed, a European task force consensus recommended 18-month-long trials be used in AD [51], but our analysis was restricted to a maximum of 6 months. Offsetting this limitation, we present information by period for researchers and clinicians where required.

Participants in our analysis were highly selected since they came from double-blind, randomized placebo-controlled trials with strict inclusion criteria (Table 1). Based on trial inclusion criteria, no patient had very early or prodromal AD (stages 1, 2 and 3) or severe AD (stage 6) since all patients were required to have an initial CDR severity of 1 (mild AD, stage 4) or 2 points (moderate AD, stage 5)- an entry cut-off for all studies (Table 1). This inclusion selection cut-off limited the variability of patient severity needed in linking studies and no linking results were available for total scores corresponding to prodromal or severe dementia. Also, research has shown that such selection criteria allow only limited generalizations from clinical trial data to the general population [52, 53]. Nonetheless, our findings may apply to persons with mild to moderate symptoms of AD. Epidemiological studies with longer follow-up periods may be needed to inform clinical practice better and replicate the current results. Furthermore, there may be order-of-assessment effects in the participant ratings.

Notwithstanding the limitations of our study, our results can be useful in clinical and research settings. In clinical settings, they provide clinicians an accessible means to interpret different ratings, and so for instance, compare patients. In research settings, they are useful for meta-analysis (e.g., when the conversion of a covariate or efficacy measure is needed), to pool and/or harmonize data across studies and so increase statistical power, and when postmortem data is examined since information may be limited and linking could be a sensible alternative.



In conclusion, the provided linkages between CDR-SB and CGI as well as CDR-SB and CIBI plus offer the first evidence-based insight into the associations between these global rating scales, which could contribute to a better understanding and improved interpretation of clinical data and trial results in Alzheimer's disease.

## **Author Disclosure**

### Acknowledgements

We acknowledge Eisai Co. Ltd for providing us with the study data. Eisai Co. Ltd did not provide study design, critical input, or manuscript review for the study. The views expressed in this manuscript may or may not express the views of Eisai Co. Ltd. We acknowledge

<http://www.clinicalstudydatarequest.com> for hosting the study data.

<http://www.clinicalstudydatarequest.com> did not provide study design, critical input, or manuscript review for the study. The views expressed in this manuscript may or may not express the views of <http://www.clinicalstudydatarequest.com>. Cipriani is supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility, by an NIHR Research Professorship (grant RP-2017-08-ST2-006), by the NIHR Oxford and Thames Valley Applied Research Collaboration and by the NIHR Oxford Health Biomedical Research Centre (grant BRC-1215-20005). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health. Efthimiou is supported by project grant No. 180083 from the Swiss National Science Foundation (SNSF).

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Conflict of Interest

Samara, Levine, Yoshida, Goldberg, Efthimiou: The authors report no conflicts with any product mentioned or concept discussed in this article.

Iwatsubo: has served as a consultant of Eisai, Roche and Biogen in the last 3 years.

Cipriani: has received research and consultancy fees from INCiPiT (Italian Network for Paediatric Trials), CARIPLO Foundation and Angelini Pharma.

Leucht: In the past 3 years Leucht has received honoraria as a consultant or for lectures for LB Pharma, Otsuka, Lundbeck, Boehringer Ingelheim, LTS Lohmann, Janssen, Johnson&Johnson, TEVA, MSD, Sandoz, SanofiAventis, Angelini, Sunovion, Recordati and Geodon Richter.

Furukawa: reports personal fees from Mitsubishi-Tanabe, MSD and Shionogi, and a grant from Mitsubishi-Tanabe, outside the submitted work; TAF has a patent 2018-177688 concerning smartphone CBT apps pending, and intellectual properties for Kokoro-app licensed to Tanabe-Mitsubishi.

#### Data Availability Statement

Data access made available upon request to <http://www.clinicalstudydatarequest.com>

## References

- [1] Whitehouse PJ, Price DL, Struble RG, Clark AW, Coyle JT, Delon MR (1982) Alzheimer's disease and senile dementia: loss of neurons in the basal forebrain. *Science* **215**, 1237-1239.
- [2] Robert P, Ferris S, Gauthier S, Ihl R, Winblad B, Tennigkeit F (2010) Review of Alzheimer's disease scales: is there a need for a new multi-domain scale for therapy evaluation in medical practice? *Alzheimers Res Ther* **2**, 24.
- [3] Leber P (1990) *Guidelines for the clinical evaluation of antidementia drugs*, US Food and Drug Administration, Rockville, MD, USA.
- [4] U.S. Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation, Early Alzheimer's Disease: Developing Drugs For Treatment, Guidelines for Industry, <https://www.fda.gov/media/110903/download>, Accessed 27/2/2021.
- [5] European Medicines Agency, Guideline on the clinical investigation of medicines for the treatment of Alzheimer's disease, [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicines-treatment-alzheimers-disease-revision-2\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicines-treatment-alzheimers-disease-revision-2_en.pdf), Accessed 27/2/2021.
- [6] Ministry of Health Labour and Welfare (MHLW) and The University of Tokyo Hospital, Project to Promote the Development of Innovative Pharmaceuticals, Medical Devices, and Regenerative Medical Products: Issues to Consider in the Clinical Evaluation and Development of Drugs for Alzheimer's Disease, <https://www.pmda.go.jp/files/000221585.pdf>, Accessed 27/2/2021.
- [7] Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL (1982) A new clinical scale for the staging of dementia. *The British Journal of Psychiatry* **140**, 566-572.

- [8] McCulla MM, Coats M, Van Fleet N, Duchek J, Grant E, Morris JC (1989) Reliability of clinical nurse specialists in the staging of dementia. *Arch Neurol* **46**, 1210-1211.
- [9] Morris JC, Ernesto C, Schafer K, Coats M, Leon S, Sano M, Thal LJ, Woodbury P (1997) Clinical dementia rating training and reliability in multicenter studies: the Alzheimer's Disease Cooperative Study experience. *Neurology* **48**, 1508-1510.
- [10] Nguyen VT, Quach THT, Pham AG, Tran TC (2019) Feasibility, Reliability, and Validity of the Vietnamese Version of the Clinical Dementia Rating. *Dement Geriatr Cogn Disord* **48**, 308-316.
- [11] Nyunt MS, Chong MS, Lim WS, Lee TS, Yap P, Ng TP (2013) Reliability and Validity of the Clinical Dementia Rating for Community-Living Elderly Subjects without an Informant. *Dement Geriatr Cogn Dis Extra* **3**, 407-416.
- [12] Karam GE, Khandakji MN, Sahakian NS, Dandan JC, Karam EG (2018) Validation into Arabic versions of Dementia Rating Scales, Dementia Caregivers Scales, and Dementia Research Instruments. *Alzheimers Dement (Amst)* **10**, 796-801.
- [13] Knopman DS, Knapp MJ, Gracon SI, Davis CS (1994) The Clinician Interview-Based Impression (CIBI): a clinician's global change rating scale in Alzheimer's disease. *Neurology* **44**, 2315-2321.
- [14] Reisberg B, Ferris SH (1994) Clinician's interview-based impression of change-plus. In *Guide to Assessment Scales in Dementia.*, Kelly C, Newton-Howes G, eds. Science Press, London.
- [15] Guy W (1976) *ECDEU assessment manual for psychopharmacology-Revised*, National Institute of Mental Health, Rockville, Maryland, USA.
- [16] Furukawa TA, Katherine Shear M, Barlow DH, Gorman JM, Woods SW, Money R, Etschel E, Engel RR, Leucht S (2009) Evidence-based guidelines for interpretation of the Panic Disorder Severity Scale. *Depress Anxiety* **26**, 922-929.

- [17] Bandelow B, Baldwin DS, Dolberg OT, Andersen HF, Stein DJ (2006) What is the threshold for symptomatic response and remission for major depressive disorder, panic disorder, social anxiety disorder, and generalized anxiety disorder? *J Clin Psychiatry* **67**, 1428-1434.
- [18] Zaider TI, Heimberg RG, Fresco DM, Schneier FR, Liebowitz MR (2003) Evaluation of the clinical global impression scale among individuals with social anxiety disorder. *Psychol Med* **33**, 611-622.
- [19] Bobo WV, Anglero GC, Jenkins G, Hall-Flavin DK, Weinshilboum R, Biernacka JM (2016) Validation of the 17-item Hamilton Depression Rating Scale definition of response for adults with major depressive disorder using equipercntile linking to Clinical Global Impression scale ratings: analysis of Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS) data. *Hum Psychopharmacol* **31**, 185-192.
- [20] Leucht S, Fennema H, Engel RR, Kaspers-Janssen M, Lepping P, Szegedi A (2017) What does the MADRS mean? Equipercntile linking with the CGI using a company database of mirtazapine studies. *J Affect Disord* **210**, 287-293.
- [21] Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR (2006) Linking the PANSS, BPRS, and CGI: clinical implications. *Neuropsychopharmacology* **31**, 2318-2325.
- [22] Levine SZ, Rabinowitz J, Engel R, Etschel E, Leucht S (2008) Extrapolation between measures of symptom severity and change: an examination of the PANSS and CGI. *Schizophr Res* **98**, 318-322.
- [23] Balsis S, Benge JF, Lowe DA, Geraci L, Doody RS (2015) How Do Scores on the ADAS-Cog, MMSE, and CDR-SOB Correspond? *Clin Neuropsychol* **29**, 1002-1009.
- [24] Gross AL, Kueider-Paisley AM, Sullivan C, Schretlen D, International Neuropsychological Normative Database I (2019) Comparison of Approaches for Equating Different Versions of the Mini-Mental State Examination Administered in 22 Studies. *Am J Epidemiol* **188**, 2202-2212.

- [25] Roalf DR, Moore TM, Mechanic-Hamilton D, Wolk DA, Arnold SE, Weintraub DA, Moberg PJ (2017) Bridging cognitive screening tests in neurologic disorders: A crosswalk between the short Montreal Cognitive Assessment and Mini-Mental State Examination. *Alzheimers Dement* **13**, 947-952.
- [26] Levine SZ, Kazufumi Y, Goldberg Y, Samara M, Cipriani A, Efthimiou O, Iwatsubo T, Leucht S, Furukawa TA (2020) Linking the Mini-Mental State Examination, the Alzheimer's Disease Assessment Scale-Cognitive Subscale, and the Severe Impairment Battery: Evidence from Individual Participant Data from Five Randomized Clinical Trials of Donepezil. *Evidence-Based Mental Health* **Epub ahead of print**.
- [27] Lynch CA, Walsh C, Blanco A, Moran M, Coen RF, Walsh JB, Lawlor BA (2006) The clinical dementia rating sum of box score in mild dementia. *Dement Geriatr Cogn Disord* **21**, 40-43.
- [28] O'Bryant SE, Lacritz LH, Hall J, Waring SC, Chan W, Khodr ZG, Massman PJ, Hobson V, Cullum CM (2010) Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. *Arch Neurol* **67**, 746-749.
- [29] O'Bryant SE, Waring SC, Cullum CM, Hall J, Lacritz L, Massman PJ, Lupo PJ, Reisch JS, Doody R, Texas Alzheimer's Research C (2008) Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's research consortium study. *Arch Neurol* **65**, 1091-1095.
- [30] Post MW (2016) What to Do With "Moderate" Reliability and Validity Coefficients? *Arch Phys Med Rehabil* **97**, 1051-1052.
- [31] Albano AD (2016) equate: An R Package for Observed-Score Linking and Equating. *Journal of Statistical Software; Vol 1, Issue 8 (2016)* **74**.
- [32] Samara MT, Engel RR, Millier A, Kandenwein J, Toumi M, Leucht S (2014) Equipercentile linking of scales measuring functioning and symptoms: examining the GAF, SOFAS, CGI-S, and PANSS. *Eur Neuropsychopharmacol* **24**, 1767-1772.
- [33] R Core Team (2013) R Foundation for Statistical Computing, Vienna, Austria. .

- [34] Homma A, Takeda M, Imai Y, Udaka F, Hasegawa K, Kameyama M, Nishimura T (2000) Clinical efficacy and safety of donepezil on cognitive and global function in patients with Alzheimer's disease. A 24-week, multicenter, double-blind, placebo-controlled study in Japan. E2020 Study Group. *Dement Geriatr Cogn Disord* **11**, 299-313.
- [35] Burns A, Rossor M, Hecker J, Gauthier S, Petit H, Moller HJ, Rogers SL, Friedhoff LT (1999) The effects of donepezil in Alzheimer's disease - results from a multinational trial. *Dement Geriatr Cogn Disord* **10**, 237-244.
- [36] Rogers SL, Friedhoff LT (1996) The efficacy and safety of donepezil in patients with Alzheimer's disease: results of a US Multicentre, Randomized, Double-Blind, Placebo-Controlled Trial. The Donepezil Study Group. *Dementia* **7**, 293-303.
- [37] Rogers SL, Doody RS, Mohs RC, Friedhoff LT (1998) Donepezil improves cognition and global function in Alzheimer disease: a 15-week, double-blind, placebo-controlled study. Donepezil Study Group. *Arch Intern Med* **158**, 1021-1031.
- [38] Rogers SL, Farlow MR, Doody RS, Mohs R, Friedhoff LT (1998) A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Donepezil Study Group. *Neurology* **50**, 136-145.
- [39] American Psychiatric Association (1987) *Diagnostic and statistical manual of mental disorders.*, American Psychiatric Association, Washington, DC.
- [40] American Psychiatric Association (1994) *Diagnostic and statistical manual of mental disorders.*, American Psychiatric Association, Washington, DC.
- [41] Cronbach LJ, Furby L (1970) How we should measure "change": Or should we? *Psychol Bull* **74**, 68-80.
- [42] Gu Z, Emons WHM, Sijtsma K (2018) Review of Issues About Classical Change Scores: A Multilevel Modeling Perspective on Some Enduring Beliefs. *Psychometrika* **83**, 674-695.
- [43] Andrews JS, Desai U, Kirson NY, Zichlin ML, Ball DE, Matthews BR (2019) Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. *Alzheimers Dement (N Y)* **5**, 354-363.



- [44] Knopman DS (2008) Clinical trial design issues in mild to moderate Alzheimer disease. *Cogn Behav Neurol* **21**, 197-201.
- [45] Quinn J, Moore M, Benson DF, Clark CM, Doody R, Jagust W, Knopman D, Kaye JA (2002) A videotaped CIBIC for dementia patients: validity and reliability in a simulated clinical trial. *Neurology* **58**, 433-437.
- [46] Horikoshi S, Kunii Y, Matsumoto J, Gotoh D, Miura I, Yabe H (2020) Does Treatment Response With Antidementia Drugs After 6 Months in Alzheimer's Disease Predict Long-term Treatment Outcome? *J Clin Psychopharmacol* **40**, 195-197.
- [47] Rockwood K, Wallack M, Tallis R (2003) The treatment of Alzheimer's disease: success short of cure. *Lancet Neurol* **2**, 630-633.
- [48] Rockwood K, Morris JC (2006) Global assessment measures In *Clinical Diagnosis and Management of Alzheimer's Disease*, Gauthier S, ed. Informa Healthcare, London.
- [49] Alzheimer's A (2015) 2015 Alzheimer's disease facts and figures. *Alzheimers Dement* **11**, 332-384.
- [50] Schneider LS, Sano M (2009) Current Alzheimer's disease clinical trials: methods and placebo outcomes. *Alzheimers Dement* **5**, 388-397.
- [51] Vellas B, Andrieu S, Sampaio C, Wilcock G, European Task Force g (2007) Disease-modifying trials in Alzheimer's disease: a European task force consensus. *Lancet Neurol* **6**, 56-62.
- [52] Malmivaara A (2019) Generalizability of findings from randomized controlled trials is limited in the leading general medical journals. *J Clin Epidemiol* **107**, 36-41.
- [53] Canevelli M, Bruno G, Vanacore N, de Lena C, Cesari M (2016) Are we really tackling the "evidence-based medicine issue" in Alzheimer's disease? *Eur J Intern Med* **35**, e29-e30.

**Table 1 Trial characteristics of the analytic dataset**

Authors	N	Sex	Mean age (SD)	Assessments available for linkage by visit	Trial inclusion criteria
Homma et al., 2000 [34]	268	Male: 89 Female: 179	70.5 (7.2)	CGI-C, CDR-SB change 4, 8, 12, 16, 20, 24	CDR: 1 or 2 MMSE: 10-26 ADAS cog: at least 15
Burns et al., 1999 [35]	815	Male: 345 Female: 470	71.6 (7.4)	CIBIS plus, CDR-SB 0 CIBIC plus, CDR-SB change 6, 12, 18, 24	CDR: 1 or 2 MMSE: 10-26
Rogers and Friedhoff, 1996 [36]	161	Male: 64 Female: 97	72 (7.5)	CGI-S, CDR-SB 0, 12 CGI-C, CDR-SB change 1, 3, 6, 9, 12	CDR: 1 or 2 MMSE: 10-26
Rogers et al., 1998 [37]	481	Male: 176 Female: 305	74.0 (7.6)	CIBIS plus, CDR-SB 0 CIBIC plus, CDR-SB change 3, 6, 9, 12	CDR: 1 or 2 MMSE: 10-26
Rogers et al. 1998 [38]	473	Male: 180 Female: 293	73.5 (7.2)	CIBIS plus, CDR-SB 0 CIBIC plus, CDR-SB change 6, 12, 18, 24	CDR: 1 or 2 MMSE: 10-26

Note. Abbreviations. CDR: Clinical Dementia Rating global score, MMSE: Mini-mental State Examination, ADAS cog: Alzheimer's Disease Assessment Scale-Cognitive Subscale.

**Table 2 Sample characteristics**

Visit		CGI-S and CDR-SB scores			CIBIS plus and CDR-SB scores			
Week	N	CGI-S	CDR-SB	$\rho$	N	CIBIS plus	CDR-SB	$\rho$
Baseline	161	3.88 (0.68)	6.85 (2.52)	<b>0.68</b>	1743	3.71 (0.78)	6.79 (2.23)	<b>0.50</b>
Week								
12	144	3.88 (0.69)	6.81 (2.71)	<b>0.71</b>				
Visit		CGI-C and CDR-SB change scores			CIBIC plus and CDR-SB change scores			
Week	N	CGI-C	CDR-SB change	$\rho$	N	CIBIC plus	CDR-SB change	$\rho$
Week 1	159	3.91 (0.38)	-0.08 (0.39)	0.16				
Week 3	151	3.83 (0.57)	-0.10 (0.55)	<b>0.27</b>	441	3.94 (0.68)	-0.13 (1.01)	0.09
Week 4	261	3.81 (0.63)	-0.01 (0.45)	<b>0.34</b>				
Week 6	146	3.81 (0.65)	-0.03 (0.98)	<b>0.48</b>	1576	3.92 (0.80)	-0.09 (1.04)	<b>0.18</b>
Week 8	252	3.71 (0.87)	-0.01 (0.67)	<b>0.44</b>				
Week 9	145	3.87 (0.72)	-0.08 (1.06)	<b>0.45</b>	423	3.88 (0.89)	-0.24 (1.33)	<b>0.12</b>
Week								
12	386	3.75 (0.91)	0.01 (0.98)	<b>0.49</b>	1495	4.02 (0.95)	-0.11 (1.29)	<b>0.24</b>
Week								
16	235	3.85 (1.17)	0.13 (1.03)	<b>0.57</b>				
Week								
18					1036	4.17 (1.03)	-0.01 (1.49)	<b>0.28</b>
Week								
20	226	3.88 (1.27)	0.19 (1.36)	<b>0.59</b>				
Week								
24	227	3.92 (1.32)	0.30 (1.52)	<b>0.65</b>	1008	4.28 (1.07)	0.13 (1.65)	<b>0.35</b>

Note. Abbreviations. N: Number of sample size, CGI-S: Clinical Global Impression-Severity, CDR-SB: Clinical Dementia Rating-Sum of Boxes, CIBIS plus: Clinician’s Interview-Based Impression of Severity Plus Caregiver Input, CIBIC plus: Clinician’s Interview-Based Impression of Change Plus Caregiver Input,  $\rho$ : Spearman’s correlation coefficient. Values of  $\rho$  in bold are statistically significant ( $P < .05$ ) Sample size varies owing to different visit schedules across trials.

**Table 3 Conversion table between CGI-S, CDR-SB and CIBIS plus scores**

CGI-S interpretation and ratings		Conversion between CGI-S and CDR-SB scores		CDR-SB ratings and interpretation		Conversion between CDR-SB and CIBIS plus scores		CIBIS plus ratings and interpretation	
1	Normal, not at all ill			0	Normal				
2	Borderline mentally ill	2	2	0.5-2.5	Questionable impairment	2.5	1	1	Normal, not at all ill
3	Mildly ill	3	4 (3.0-4.5)	3.0-4.0	Very mild dementia	3.5	2	2	Borderline mentally ill
4	Moderately ill	4	6.5 (5.0-9.0)	4.5-9.0	Mild dementia	5.5	3	3	Mildly ill
						7.0	4	4	Moderately ill
5	Markedly ill	5	11.5 (7.5-14)	9.5-15.5	Moderate dementia	11.0	5	5	Markedly ill
						13.0	6	6	Severely ill
6	Severely ill								
7	Among the most extremely ill patients			16.0-18.0	Severe dementia			7	Among the most extremely ill patients

Note. Abbreviations. CGI-S: Clinical Global Impression-Severity, CDR-SB: Clinical Dementia Rating-Sum of Boxes, CIBIS plus: Clinician’s Interview-Based Impression of Severity Plus Caregiver Input.

**Table 4 Conversion table between CGI-C, CDR-SB change and CIBIC plus scores**

CGI-C interpretation and ratings		Conversion between CGI-C and CDR-SB change scores		Change in CDR-SB scores	Conversion between the CDR-SB change and CIBIC plus scores		CIBIC plus ratings and interpretation	
1	Marked improvement	<b>1</b>	<b>-3.5</b> (-5 to -2.5)			<b>-5</b> (-6.0 to -4.5)	<b>1</b>	1
2	Moderate improvement	<b>2</b>	<b>-2</b> (-2.5 to -1.1)	<b>-3</b> (-3.5 to -2.8)		<b>2</b>	2	Moderate improvement
3	Minimal improvement	<b>3</b>	<b>-0.5</b> (-1.1 to -0.4)	<b>-1</b> (-1.6 to -0.7)		<b>3</b>	3	Minimal improvement
4	No change	<b>4</b>	<b>0</b> (0 to 0.1)	<b>0</b> (-0.3 to 0)		<b>4</b>	4	No change
5	Minimal worsening	<b>5</b>	<b>1</b> (0.6 to 1.4)	<b>1</b> (0.5 to 1.0)		<b>5</b>	5	Minimal worsening
6	Moderate worsening	<b>6</b>	<b>2.5</b> (1.9 to 7)	<b>3</b> (2.8 to 4.2)		<b>6</b>	6	Moderate worsening
7	Marked worsening	<b>7</b>	<b>5.5</b> (4 to 5.5)	<b>6</b> (5.4 to 6.1)		<b>7</b>	7	Marked worsening

Note. Abbreviations. CGI-C: Clinical Global Impression-Change, CDR-SB: Clinical Dementia Rating-Sum of Boxes, CIBIC plus: Clinician’s Interview-Based Impression of Change Plus Caregiver Input. Conversion scores are based on median value across different measurement points to define the corresponding scores between the scales.

## Figure legends

### Figure 1

**1A.** Clinical Dementia Rating Sum of Boxes (CDR–SB) linked to Clinical Global Impression Severity (CGI-S) at baseline and week 12

**1B.** Clinical Dementia Rating Sum of Boxes (CDR–SB) linked to Clinician's Interview–Based Impression of Severity Plus Caregiver Input (CIBIS plus) at baseline

**1C** Clinical Dementia Rating Sum of Boxes (CDR–SB) Change Score linked to Clinical Dementia Rating Sum of Boxes (CDR–SB) Change Score Clinical Global Impression Change (CGI-C) from week 1 to week 24 at scheduled visits with the median score across assessment points.

**1D** Clinical Dementia Rating Sum of Boxes (CDR–SB) Change Score linked to Clinician's Interview–Based Impression of Change Plus Caregiver Input (CIBIC plus) from week 3 to week 24 at scheduled visits with the median score across assessment points