Epilepsy & Behavior 92 (2019) 61–70

Contents lists available at ScienceDirect

Epilepsy & Behavior

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

Review

Optimal cutoff score of the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E) for detecting major depressive disorder: A meta-analysis

Do-Hyung Kim, Young-Soo Kim, Tae-Won Yang, Oh-Young Kwon *

Department of Neurology and Institute of Health Science, Gyeongsang National University Hospital, Gyeongsang National University College of Medicine, Jinju, Republic of Korea

ARTICLE INFO

Article history:
Received 11 October 2018
Revised 9 December 2018
Accepted 9 December 2018
Available online 4 January 2019

Keywords:
Epilepsy
Major depressive disorder
Psychological interview
Meta-analysis

ABSTRACT

Background and purpose: The Neurological Disorders Depression Inventory for Epilepsy (NDDI-E) is a useful inventory for screening major depressive disorder (MDD) in people with epilepsy (PWE). The cutoff score for detecting MDD has been reported with the range of >11 to >16. The aim of this study was to find optimal cutoff score of the NDDI-E for MDD detection by combining the raw data from previous studies.

Methods: We searched MEDLINE, EMBASE, Cochrane Library, Web of Science, and SCOPUS to identify proper studies. Original researches, which tested the accuracy of NDDI-E for MDD detection in adult PWE, were recruited. We included the studies in which MDD was diagnosed by a gold standard structural interview, the Mini International Neuropsychiatric Interview (MINI). In addition, we included only the studies providing enough information for meta-analysis: number of PWE with MDD, number of total PWE, and sensitivity (Se) and specificity (Spe) for each cutoff score. After collecting data from included studies, we performed a diagnostic test accuracy (DTA) meta-analysis using bivariate model.

Results: We identified 13 validation studies conducted in outpatient epilepsy clinic setting. As summary estimates of test accuracy measures, the Se, Spe, and diagnostic odds ratio (DOR) of NDDI-E for MDD detection were 0.81, 0.84, and 22.48, respectively. The analysis using the multiple thresholds model showed that the NDDI-E score of 13.2 was the best fit for MDD detection. When analyzing only with the seven data sets of the cutoff score >13, the Se, Spe, and DOR were 0.87, 0.80, and 25.72, respectively.

Conclusions: The optimal NDDI-E cutoff score for MDD detection is >13. The information provided by this DTA meta-analysis will be a useful reference for applying NDDI-E in geographic areas where no NDDI-E validation studies have been conducted for their languages.

1. Introduction

The bidirectional relationship between depression and epilepsy increases prevalence of depression in people with epilepsy (PWE). Moreover, depression affects quality of life (QOL) and treatment outcome in PWE, thus, it is critical to detect depression in PWE in epilepsy clinics [1]. For detecting major depressive disorder (MDD), the gold standard tests are structured interviews such as the Mini International Neuropsychiatric Interview (MINI), the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), axis I disorders (SCID), and the Composite International Diagnostic Interview (CIDI). These structured interviews include items inquiring symptoms, which may be adverse effects of antiepileptic drugs so that the interviews have a limitation to use in PWE. For this reason, there is no epilepsy-specific gold standard test for diagnosing MDD. Fortunately, there is a useful and epilepsy-specific screening tool for MDD: the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E) has been developed and validated on the base of gold standard tests. In the NDDI-E, the items asking symptoms, which may be caused by antiepileptic drugs were removed, and only six items were remained [2].

When NDDI-E was first developed, SCID was used as the referential standard test [2]. The authors found that a cutoff score >15 had the highest sensitivity (Se), specificity (Spe), and positive predictive value (PPV) for MDD detection [2]. Another study validated the NDDI-E based on another referential standard test, the MINI. In the study, the Se and Spe were 100% and 85%, respectively, for MDD detection [3]. To date, NDDI-E has been the most frequently validated tool for screening MDD in PWE. Perhaps this is because NDDI-E has the advantage of being epilepsy-specific. The validation studies were for various language versions of NDDI-E. The cutoff score for screening MDD has been reported

* Corresponding author at: Department of Neurology, Gyeongsang National University Hospital, Gyeongsang National University College of Medicine, 816-15 Jinjudae-ro, Jinju 52727, Republic of Korea.

E-mail address: oykwon@gnu.ac.kr (O.-Y. Kwon).

https://doi.org/10.1016/j.yebeh.2018.12.006
1525-5050/

with the range of >11 to >16, and the MINI has been most frequently adopted as the referential standard test in the studies [4].

Gill et al. summarized the validation states of depression-screening tools for PWE by a systematic review [4]. They did not combine the raw data but obtained the medians for the data that the primary studies provided: Se, Spe, and area under the curve (AUC). According to the systematic review, the most frequently indicated NDDI-E cutoff score for MDD detection was >15. This cutoff score was provided by 12 studies, and the medians were 80.5%, 86.2%, and 0.900 for Se, Spe, and AUC, respectively [4]. The cutoff score >13 had the best balance between Se and Spe: medians of Se and Spe among 10 studies were 83.7% and 86.3%, respectively, in Gill's systematic review [4]. A limitation of the review was that statistical significance was not evident because the results were descriptive. The authors also mentioned that it was difficult to statistically integrate the data because the information of the primary studies was not appropriate [4].

The MINI was most frequently used as a referential standard test in the primary studies of Gill's review, but other gold standard interviews such as CIDI, SCID, and screening tools such as the Beck Depression Inventory (BDI) and the Major Depression Inventory (MDI) were also used as referential tests [4]. The heterogeneity of the referential standards may cause a deviation of the combined results. The authors also mentioned the possible bias associated with the heterogeneity as another limitation of their review. Therefore, to increase the accuracy of the results, we need to collect the studies using the same referential standard test.

Most of the studies that validated NDDI-E for MDD detection were conducted to develop various language versions of NDDI-E other than English [4]. Thus, in the geographic areas where the NDDI-E version for the common language has been developed, the language version can be a referential inventory for clinical application, and they can use the cutoff score that the validation study provided. In contrast, it is difficult which cutoff score should be adopted when using NDDI-E in the areas whose language was not validated for NDDI-E yet. Furthermore, since the cutoff scores presented by the validation studies were ranged from >11 to >16, the wide distribution of the scores makes it more difficult to select an adequate cutoff score [4]. In addition, even in the geographic areas in which the NDDI-E version for the commonly used language has been developed, it is difficult to say that the cutoff scores have yet to be clearly validated, since the validation studies were performed on a small number of institutions, usually one institution, with a small sampling number of PWE.

The aim of this meta-analysis was to find optimal cutoff score of NDDI-E for MDD detection in PWE by combining the raw data from previous studies in which the MINI was used as the referential standard test for the validation of the NDDI-E. For the purpose, we did a diagnostic test accuracy (DTA) meta-analysis using bivariate model.

2. Methods

2.1. Search strategy

Researches, which tested the accuracy of NDDI-E for MDD detection in adult PWE, were recruited for this meta-analysis. The search was performed on 21 August 2018, in the MEDLINE, EMBASE, COCHRANE, Web of Science, and SCOPUS electronic databases from inception to July 2018. We also manually examined other bibliographies, including reference lists and gray literatures, to avoid missing pertinent articles. No restrictions were placed on the publication date or language. The references, including articles and proceedings, were exported and managed using EndNote X8.2 [5].

2.2. Study selection

For the initial screening process, two reviewers (OY Kwon and DH Kim) individually examined the titles and abstracts of the studies included in the search results to identify studies suitable for this meta-analysis. We performed the initial screening extensively not to omit any study proper to this meta-analysis. In the subsequent selection process, two reviewers independently rescreened the full-length articles or abstracts identified in the initial screening. Studies with disagreements between the two reviewers were reappraised by discussion or consultation with the third author (YS Kim).

Studies were included if they fulfilled all the following criteria: adult subjects, original researches, diagnostic accuracy studies, studies using the MINI as the referential gold standard test, and studies which provide adequate data for combined analysis. The reason why we included the studies only when using the MINI was to eliminate the impact of the variability of the reference standard tests on the diagnostic accuracy of the NDDI-E for MDD detection. The reason for choosing the MINI was that we had known, when we planned this meta-analysis, that the MINI had been the most frequently adopted referential standard test in the previous studies validating the NDDI-E [4]. For the sake of the adequate data, we included only the studies in which we can obtain the number of PWE with MDD, number of total PWE, and Spe and Se for each cutoff score.

2.3. Data synthesis and analysis

2.3.1. Data extraction

Data extracted from the selected studies included study information (author, year, area in which the study was performed, language for NDDI-E, number of clinics included in each study), population demographics of PWE (age, sex, number), and outcomes associated with MDD (point prevalence, cutoff score of NDDI-E for detecting MDD). Disparities of extracted data among the reviewers were discussed until a consensus was reached. The extracted data are summarized in Table 1.

We collected studies to evaluate the DTA of NDDI-E. These studies used the MINI as a gold standard reference to diagnose MDD. We extracted the number of PWE with MDD, number of total PWE, and Spe and Se for each cutoff score in all included studies. We used the extracted data to calculate the following numbers of PWE in terms of NDDI-E test: true positive (tp), false positive (fp), false negative (fn), and true negative (tn). Therefore, tp meant the number of PWE screened as MDD by the NDDI-E and diagnosed as MDD by the MINI.

2.3.2. Main analysis

When we performed DTA meta-analysis, we used the bivariate model. R (Version 3.3.3, The R Foundation for Statistical Computing, Vienna, Austria) was used for an analysis with multiple thresholds model. SAS (Version 9.4, Statistical Analysis System (SAS) Institute Inc., Cary, NC, USA) was used for an analysis getting summary estimates of test accuracy measures and a subgroup analysis using meta-regression. Stata (Version 14.2, StataCorp LP, College Station, Texas, USA) was used for the summary receiver operating characteristic (SROC) curve analysis.

In the DTA meta-analysis, the multiple thresholds model is appropriate when the included studies provided data for multiple cutoff scores. This model derives the optimal cutoff score using Youden index [6]. We obtained the optimal NDDI-E cutoff score using the analysis of multiple thresholds model.

Summary estimates of test accuracy measures were obtained by an analysis of the data summing all the data on cutoff scores provided in the studies together. They included Se, Spe, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR). For the cutoff scores having data from two or more studies, Se, Spe, PLR, NLR, and DOR for each cutoff score were obtained by subgroup analysis using meta-regression. The SROC curve and its AUC estimate were obtained by the SROC curve analysis.

2.3.3. Heterogeneity

The Cochrane handbook recommends graphical representation when evaluating heterogeneity in the DTA meta-analysis [7]. We evaluated the heterogeneity using SROC curve as a graphic representation. The heterogeneity can be assessed by observing the degree of closeness

Table 1
Characteristics and demographics of the 13 studies included in the diagnostic test accuracy (DTA) meta-analysis.

Study identification	Continent	Country	Language for NDDI-E	Number of clinics	Number of PWE	Age						Female (%)	MDD determined by MINI (%)	Cutoff score of NDDI-E for MDD
						Total		No MDD		MDD				
						Mean	SD	Mean	SD	Mean	SD			
Tong 2015	East Asia	China	Chinese	1	202	30.8	11.0	31.0	11.7	30.2	8.8	43.1	26.7	>12
Micoulaud 2015	Europe	France	French	2	116	40.4	13.8	40.4	14.8	40.5	11.1	58.6	28.5	>15
Hansen 2015	Europe	Denmark	Danish	1	124	NC	NC	NC	NC	NC	NC	NC	11.7	>13
Guo 2015	East Asia	China	Chinese	1	248	NC	NC	31.1	10.3	30.5	10.4	48.0	16.5	>14
Thomson 2014	South America	Argentina	Spanish	1	155	47.0	18.0	46.0	19.0	50.0	16.0	61.9	16.1	>15
de Oliveira 2014	South America	Brazil	Portuguese	5	126	39.3	10.2	38.9	10.3	40.4	9.9	54.0	27.8	>15
Zis 2013	Europe	Greece	Greek	1	101	36.8	13.2	34.7	11.7	37.4	13.6	46.5	21.8	>15
Tadokoro 2012	East Asia	Japan	Japanese	2	159	NC	NC	38.5	13.1	34.3	8.3	50.0	24.2	>16
Mula 2012	Europe	Italy	Italian	4	120	44.9	14.4	NC	NC	NC	NC	35.5	21.5	>13
Ko 2012	East Asia	Korea	Korean	1	121	NC	NC	39.9	14.4	42.9	12.5	59.2	24.2	>11
Di Capua 2012	Europe	Spain	Spanish	1	121	42.0[a]	NC	39.0[a]	NC	50.0[a]	NC	44.7	26.7	>13
de Oliveira 2011	South America	Brazil	Portuguese	1	142	39.1	10.1	NC	NC	NC	NC	50.7	26.4	>15
de Oliveira 2010	South America	Brazil	Portuguese	1	98	NC	NC	40.0	10.0	40.0	10.0	56.1	27.6	>15

PWE: people with epilepsy, SD: standard deviation, MDD: major depressive disorder, MINI: Mini International Neuropsychiatric Interview, NDDI-E: Neurological Disorders Depression Inventory for Epilepsy, Tx: treatment, NC: no comment or not clear.
[a] Median age.

between 95% prediction region and 95% confidence region in the SROC curve.

2.3.4. Publication bias

Publication bias was assessed by Deeks' funnel plot asymmetry test [8]. For the assessment of publication bias, we included all individual data sets of "tp, fp, fn, and tn" at all cutoff scores provided in all 13 included studies. If the p value for the bias estimate analyzed from the Deeks' funnel plot is less than 0.1, it is regarded as having a likelihood of publication bias.

2.4. Quality assessment of literatures

We used Quality Assessment of Diagnostic Accuracy Studies, version 2 [QUADAS-2 (www.quadas.org)] for the quality assessment of literatures because this is a DTA meta-analysis [9]. In this tool, overall assessment of bias is evaluated in four domains: patient selection, index test, reference standard, and flow and timing. In each domain, signaling questions can be used as the guides of assessment. When using the QUADAS-2, if one of the four areas is assessed as having high or unclear risk of bias, the overall classification is considered to be at high risk of bias. The level of applicability of the included studies can also be evaluated using the QUADAS-2. For the evaluation of applicability, three of the 4 domains except flow and timing are used.

2.5. Meta-analysis guidelines

There are useful guidelines for the minimum requirements for systematic reviews and meta-analysis. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is one of the guidelines. It is evidence-based and can be used to improve the quality of reviews [10]. For this meta-analysis, we referred the PRISMA guideline.

3. Results

3.1. Identification of relevant studies

We probed five electronic databases to get relevant studies: MEDLINE, EMBASE, Cochrane Library, Web of Science, and SCOPUS. A flowchart of the process used for identification of studies is shown in Fig. 1. A total of 1891 studies were identified by searching the five databases: 814 from MEDLINE, 403 from EMBASE, 53 from Cochrane Library, 37 from Web of Science, and 584 from SCOPUS. First, 450 duplicated studies were excluded. An additional 1404 of the remaining 1441 studies were also excluded because they did not meet the selection criteria on the initial screen. The full texts of the remaining 37 studies were evaluated in the subsequent selection process. By selection, 24 studies were excluded for various reasons: no or different gold standard (n = 11), insufficient data (n = 4), duplicated data (n = 3), not diagnostic study (n = 2), validation with a different purpose (n = 2), including dysthymia in validation (n = 1), and prescreening before NDDI-E (n = 1). Finally, 13 studies were included in this meta-analysis.

3.2. Features of the included studies

The characteristics and demographics of the studies included in the meta-analysis are summarized in Table 1. The publication years of the studies ranged from 2010 to 2015. All the 13 studies were articles. In each of East Asia [11–14] and South America [15–18], four studies were performed, respectively. Five studies were performed in Europe [19–23]. Three studies were conducted in Brazil [15–17]. Two studies were conducted in China [11,14]. One study was performed in each of Argentina [18], Denmark [20], France [21], Greece [23], Italy [22], Japan [13], Korea [12], and Spain [19].

All the 13 studies included in this meta-analysis were conducted to validate various language versions of NDDI-E other than English and used the MINI as a gold standard tool for diagnosing MDD. The validation was for Portuguese in three studies [15–17], Spanish in two studies [18,19], and Chinese in two studies [11,14]. Danish [20], French [21], Greek [23], Italian [22], Japanese [13], and Korean [12] were each validated in one study. Nine of the 13 studies were conducted in a single clinic [11,12,14,15,17–20,23]. Two studies recruited PWE from two clinics [13,21], and the remaining two studies recruited PWE from four [22] and five clinics [16], respectively.

The total number of PWE in the 13 studies included in the meta-analysis was 1833. The sampling number of studies ranged from 98 to 248, with a median of 124. Eleven of the 13 studies provided the mean age of PWE included in the study populations in different ways. Five studies reported the mean age of the total PWE population, and the mean ages of PWE with MDD and PWE without MDD, separately [14,16,18,21,23]. Two studies reported only the mean age of the total PWE [17,22]. Four studies reported the mean ages of PWE with MDD and PWE without MDD, separately, but not that of total PWE [11–13, 15]. The mean age of the total PWE ranged from 30.8 to 47.0 years, with the median of 39.3 years. The mean age of PWE without MDD ranged from 31.0 to 46.0 years, with the median of 38.9 years. The mean age of PWE with MDD ranged from 30.2 to 50.0 years, with the median of 40.0 years.

Another of the 13 studies provided the median ages of the total PWE, PWE with MDD, and PWE without MDD [19]. The median ages were

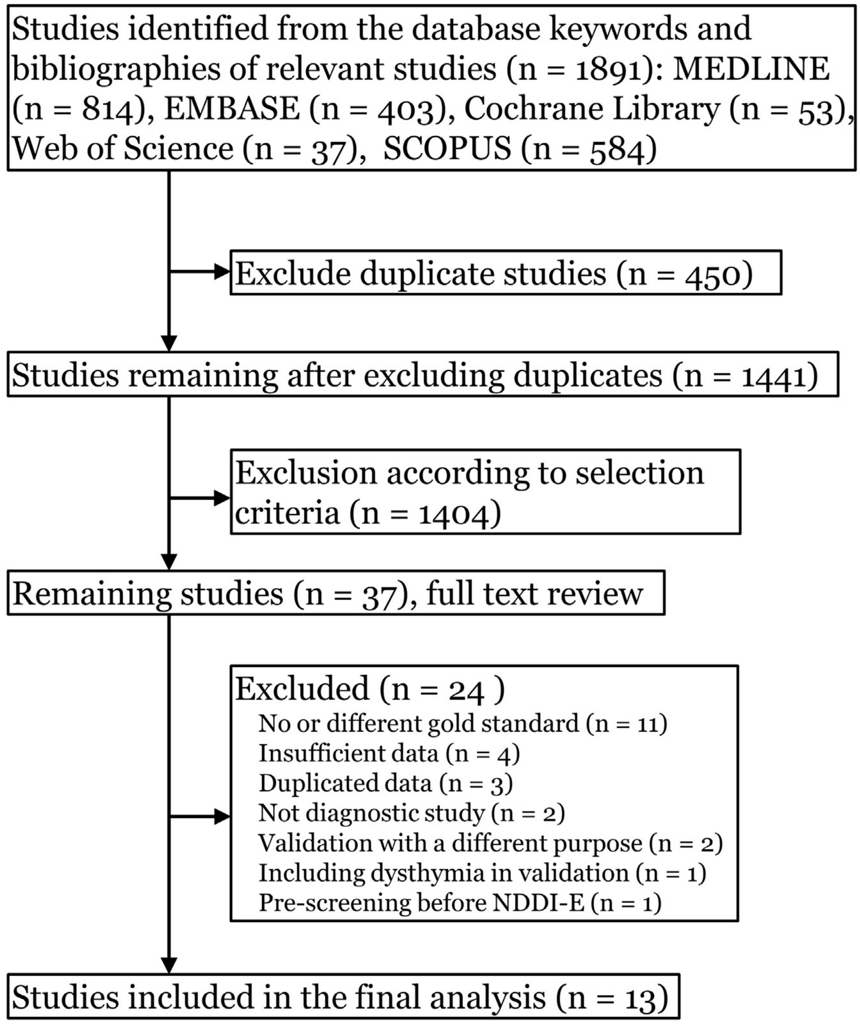

Fig. 1. Flow diagram for identification of relevant studies. A total of 1891 studies were identified by searching five databases. We excluded 450 duplicated studies and an additional 1404 of the remaining 1441 studies that did not satisfy the selection criteria. We reviewed full texts of the remaining 37 studies and excluded 24 studies because of various reasons. Finally, 13 studies were included in this meta-analysis.

42.0, 39.0, and 50.0 years for the total PWE, PWE without MDD, and PWE with MDD, respectively. In the other remaining study, we could not find the age information of the total PWE, PWE with MDD, and PWE without MDD, because the authors presented the ages of PWE based on current psychiatric comorbidities [20]. Twelve articles reported the gender ratio. The percentage of female in the PWE ranged from 35.5% to 61.9% with a median of 50.4%. In the 13 included studies, the point prevalence of MDD determined by the MINI ranged from 11.7% to 28.5%, with a median of 24.2%. The NDDI-E cutoff scores for MDD detection, of the 13 studies, ranged from >11 to >16, with a median of >15.

3.3. Summary measures

The included studies of this meta-analysis provided raw data of cutoff scores from >9 to >20. Among them, the raw data of cutoff scores >11, >12, >13, >14, >15, >16, and >18 were provided by more than one study (Table 2). The number of studies for the cutoff scores were 4, 6, 7, 4, 10, 2, and 4, respectively. The largest number of studies that provided the raw data for a single cutoff score was 10, and the cutoff score was >15. The sum of the PWE numbers of the 10 studies was 1386. The second largest number of studies was seven for the cutoff score >13. The sum of PWE numbers of the seven studies was 1078.

As summary estimates of test accuracy measures, we obtained Se, Spe, PLR, NLR, and DOR. Sensitivity was 0.81, 95% confidence interval (CI) [0.76, 0.85], and Spe was 0.84, 95% CI [0.82, 0.86]. Positive likelihood ratio was 5.04, 95% CI [4.24, 5.98], and NLR was 0.22, 95% CI [0.17, 0.29]. Diagnostic odds ratio was 22.48, 95% CI [15.02, 33.66].

With consideration of the cutoff score as a continuous variable, we performed the DTA meta-analysis using a multiple thresholds model to find the optimal NDDI-E cutoff score for MDD detection (Fig. 2). In the meta-analysis, the estimated distribution functions were transformed back from the regression lines of negative test result for the PWE with MDD and those without MDD. Using the graph showing the estimated distribution functions of Spe and "1 − Spe", we visualized the cutoff line at which Youden index was maximum (Fig. 2-B). According to the graph, the optimal NDDI-E cutoff score for MDD detection was 13.2. The Spe and "1 − Spe" were 0.80 and 0.14 (Se = 0.86), respectively.

Table 2
Summary estimates of test accuracy measures from the DTA meta-analysis after combining the 13 included studies validating the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E) for detecting major depressive disorder (MDD).

Parameter	Estimate	Lower	Upper
Sensitivity	0.81	0.76	0.85
Specificity	0.84	0.82	0.86
Diagnostic odds ratio	22.48	15.02	33.66
Positive likelihood ratio	5.04	4.24	5.98
Negative likelihood ratio	0.22	0.17	0.29

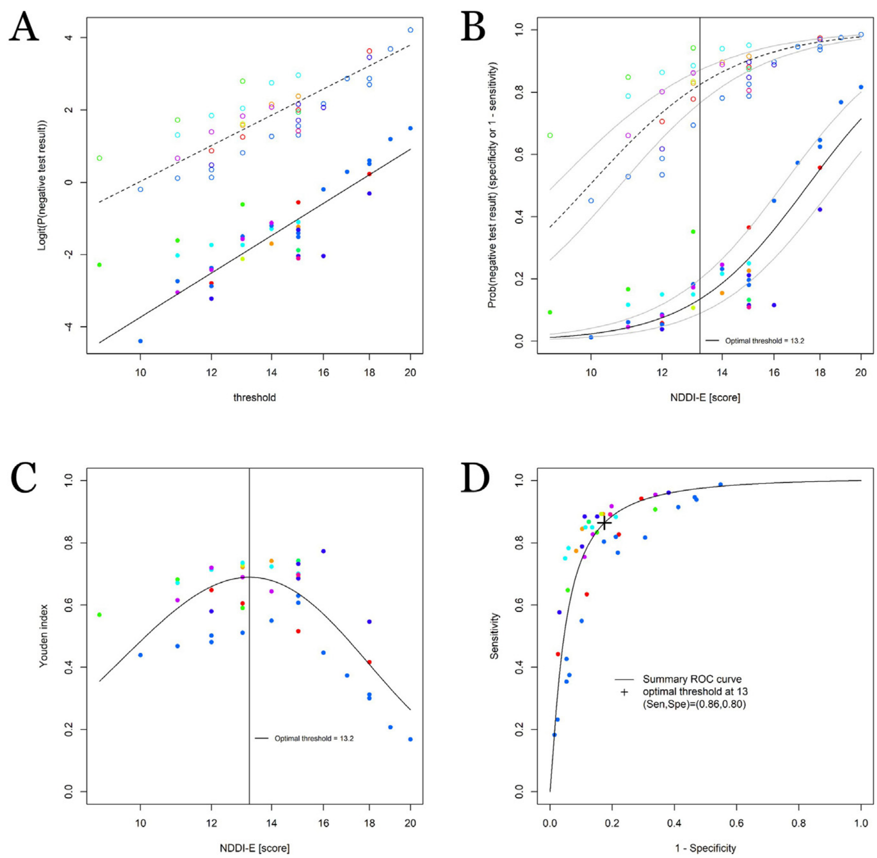

Fig. 2. Diagnostic test accuracy meta-analysis using multiple thresholds model for optimal cutoff score of the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E) for detecting major depressive disorder (MDD) in people with epilepsy (PWE). (A) Regression lines of negative test results of NDDI-E for the PWE with MDD (filled circles, solid line) and those without MDD (open circles, dashed line) had their slopes. The slopes of the regression lines meant that the Se and Spe change with cutoff scores. (B) Estimated distribution functions of negative test result for the PWE with MDD (filled circles, solid line) and those without MDD (open circles, dashed line) were transformed back from the regression lines. The gray lines marked the confidence regions. The cutoff line with maximum Youden index was visualized at the cutoff score 13.2. Thus, the optimal NDDI-E cutoff score for MDD detection was >13. The Spe and "1 − Spe" were 0.80 and 0.14 (Se = 0.86), respectively, at the cutoff score. (C) The optimal cutoff score of NDDI-E for MDD detection could be confirmed in the Youden index curve as 13.2. The optimal cutoff score was depicted as a solid vertical line. (D) The optimal cutoff score 13.2 was marked as a cross in the estimated summary receiver operating characteristics curve.

In the cutoff scores whose data were provided by two or more studies, the estimates of parameters including Se, Spe, PLR, NLR, and DOR were obtained for each cutoff score (Fig. 3). The estimates were calculated by the subgroup analysis using meta-regression and presented in Tables 3 and 4. At the cutoff score >13, Se was 0.87, 95% CI [0.80, 0.91], Spe was 0.80, 95% CI [0.73, 0.86], PLR was 6.02, 95% CI [3.21, 5.82], NLR was 0.17, 95% CI [0.11, 0.25], and DOR was 25.77 with 95% CI [15.36, 43.25].

All data presented by cutoff score were extracted from all 13 studies and included in the SROC curve analysis to calculate the integrated AUC estimate. The AUC estimate was 0.91, 95% CI [0.88, 0.93] (Fig. 4). In the graph of SROC curve obtained from the analysis, we could assess the heterogeneity by observing the closeness between the confidence region and the prediction region. Because the area of 95% prediction region was markedly wider than the area of 95% confidence region in the graph, we could assess that the likelihood of heterogeneity was high. The heterogeneity was also confirmed by quantitative statistics. The correlation coefficient between log-transformed Se and Spe was −0.99 and it was smaller than zero. Therefore, we could interpret that there was a heterogeneity between the objects of this analysis.

The SROC curve analysis was performed for each of the six cutoff scores whose raw data were provided from four or more studies. The AUC estimates are summarized in Table 4. At NDDI-E score >13, the AUC estimate was 0.90, 95% CI [0.87, 0.92]. In the graph of SROC curve at the NDDI-E cutoff score >11, we could assess that the likelihood of heterogeneity was high because the area of 95% prediction region was wider than the area of 95% confidence region (Fig. 5).

3.4. Publication bias

In Deeks' funnel plot asymmetry test [8], the *p* value of the bias estimate was 0.29. Because this value was not less than 0.1, the likelihood of publication bias was low. We could also observe that there was no asymmetry among the data points of the funnel plot (Fig. 6).

3.5. Study quality assessment

All the recruited 13 studies included in this meta-analysis had low risk of bias according to the assessment using QUADAS-2. The results are shown in Supplementary Table 1. We assessed the primary studies with careful consideration of the following points. Although all the studies described the MINI and the NDDI-E in detail, they did not explain how these tests were conducted and interpreted. Two studies described that the MINI was performed before the NDDI-E. In one study, after performing the MINI at the clinic, PWE took the NDDI-E questionnaire home, completed it, and brought it back to the clinic [20]. The authors of the other study described that the MINI and the NDDI-E were performed for each PWE on the same day, and the MINI was performed prior to NDDI-E [12]. However, even in these two studies, it is unclear whether the NDDI-E test was performed without knowing the results of the MINI test. Taken together, by the signaling questions, the studies were required to be inspected for the "index test" and "reference standard" items of QUADAS-2. However, since NDDI-E is a self-report questionnaire and the MINI is a structural interview with defined

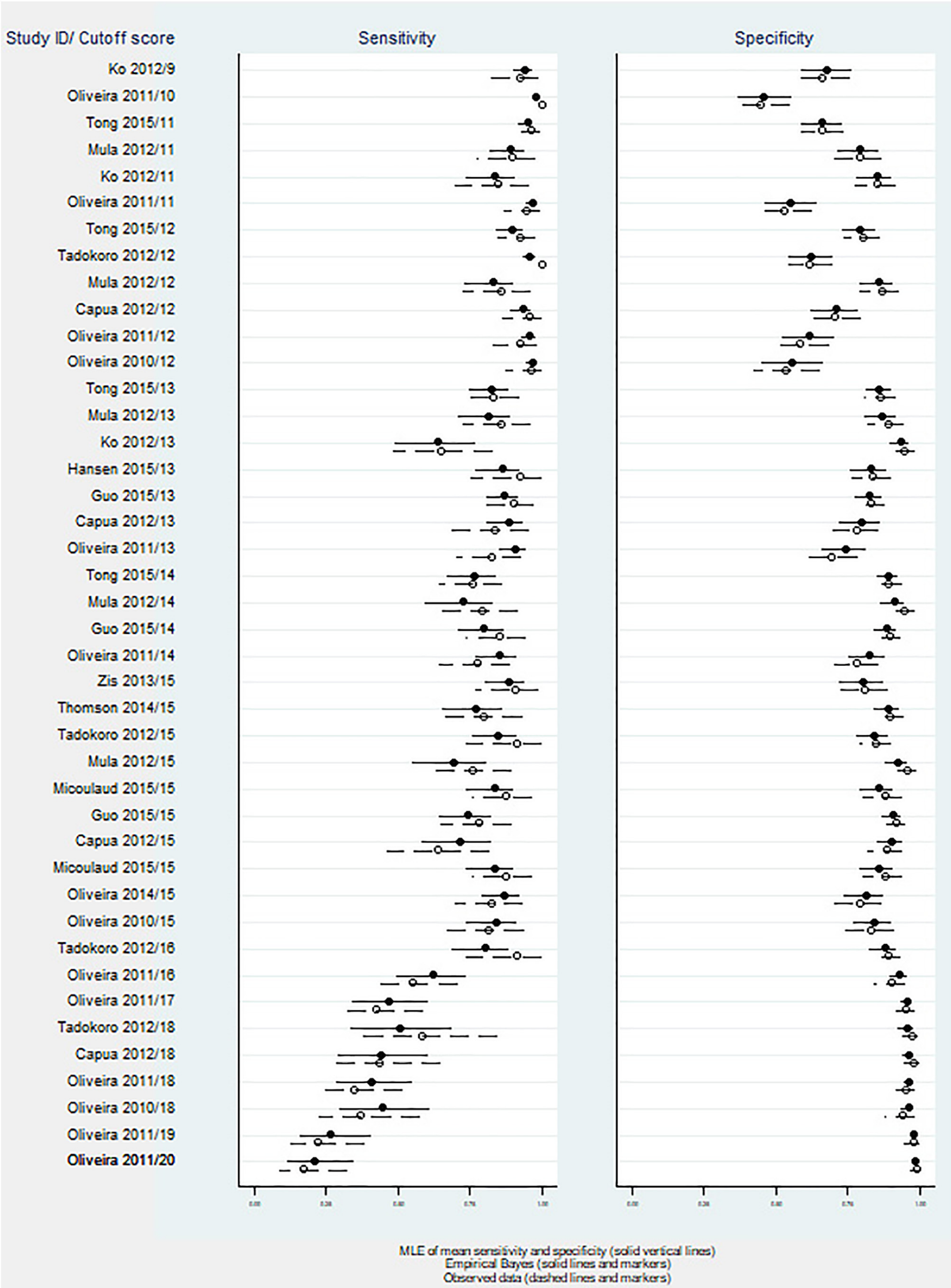

Fig. 3. Paired forest plot depiction of empirical Bayes predicted versus observed sensitivity and specificity by cutoff score. The sensitivity and specificity were well balanced at the optimal NDDI-E cutoff score >13.

Table 3
Parameter estimates of the NDDI-E for detecting MDD after DTA meta-analysis of each cutoff score.

Cutoff score	Number of study	Sensitivity (%)			Specificity (%)			Positive likelihood ratio			Negative likelihood ratio		
		Estimate	95% CI		Estimate	95% CI		Estimate	95% CI		Estimate	95% CI	
			Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
>11	4	0.95	0.90	0.97	0.61	0.51	0.71	2.45	1.90	3.16	0.09	0.04	0.17
>12	6	0.94	0.89	0.97	0.71	0.63	0.79	3.28	2.52	4.29	0.09	0.05	0.16
>13[a]	7	0.87	0.80	0.91	0.80	0.73	0.86	4.32	3.21	5.82	0.17	0.11	0.25
>14	4	0.82	0.73	0.88	0.86	0.80	0.91	6.02	4.18	8.69	0.21	0.14	0.31
>15	10	0.80	0.73	0.86	0.89	0.85	0.93	7.60	5.52	10.46	0.22	0.16	0.30
>16	2	0.63	0.46	0.77	0.93	0.88	0.96	8.98	5.27	15.29	0.40	0.26	0.62
>18	4	0.39	0.28	0.52	0.98	0.96	0.99	15.85	8.43	29.80	0.63	0.51	0.76

CI: confidence interval.
[a] Optimal cutoff score.

items, the loose description may not cause risks of bias for those two QUADAS-2 items.

In the recruited studies, the authors also did not mention whether there was an appropriate interval between the MINI and the NDDI-E test. However, because both tests were a self-report questionnaire or an item-specific test, we interpreted that there was no possibility of causing risk of bias to the QUADAS-2 item "flow and timing". Three of the included 13 studies did not address the consecutiveness of patient collection and did not even describe the period of recruitment [15–17]. Nevertheless, we interpreted that there was no risk of bias to the QUADAS-2 item "patient selection". This is because the three studies were performed in the epilepsy clinics as other studies.

4. Discussion

We performed a DTA meta-analysis with 13 studies that validated the NDDI-E for MDD detection in PWE based on the MINI as a referential standard test. In the studies, the NDDI-E cutoff score for MDD detection ranged from >11 to >16. The cutoff scores that we could obtain the raw data ranged from >9 to >20 from the included studies. Summary estimates of test accuracy measures presented that the Se, Spe, DOR, and AUC estimate of NDDI-E for MDD detection were 0.81, 0.84, 22.48, and 0.91 respectively. The most appropriate cutoff score was >13 when assessed by a DTA meta-analysis using multiple thresholds model. In the separate analysis for each cutoff score, the Se, Spe, DOR, and AUC estimate at cutoff score >13 were 0.87, 0.80, 25.77, and 0.90, respectively. This information will be helpful for the clinical applications of the NDDI-E when a specific language version has not been validated yet. Furthermore, we collected only studies using the MINI as a reference standard for this review to target a more homogeneous group. This may have increased the reliability of the data and meta-analysis results.

A systematic review for validated screening tools for depression in PWE has been recently reported by Gill et al. [4]. According to the review, NDDI-E was the most frequently validated inventory among the depression screening tools for PWE. The authors did not combine raw data and presented only the median values of Se, Spe, and AUC of the NDDI-E cutoff scores from the included studies. They noted that they could not recommend the best cutoff score because the meta-analysis was impossible due to the difficulty of obtaining raw data from the primary studies. The clarity of the results of the systematic review has also been obscured because the authors did not limit the types of reference standard tests used in the recruited studies. On the other hand, we extracted data from the 13 included studies and calculated the numbers of tp, fp, fn, and tn using the extracted data. With the numbers, we could found that the NDDI-E cutoff score >13 was optimal for MDD detection by a DTA meta-analysis using multiple thresholds model.

Some studies reported the cutoff score for MDD detection differently even though the same language version of NDDI-E was tested in the studies. Two studies validated the Spanish version of NDDI-E, however, the continents, where these two studies were conducted, were different from each other. The cutoff score >15 was reported in Argentina of South America [18], and the cutoff score >13 was reported in Spain of Europe [19]. The Chinese version of NDDI-E was also validated in two studies, with the cutoff scores >14 [11] and >12 [14], respectively. For

Table 4
Estimates of diagnostic odds ratio and area under the curve of the NDDI-E for detecting MDD after DTA meta-analysis for each cutoff score.

Cutoff score	Number of study	Diagnostic odds ratio			Area under the curve		
		Estimate	95% CI		Estimate	95% CI	
			Lower	Upper		Lower	Upper
>11	4	28.82	13.21	62.88	0.92	0.89	0.94
>12	6	37.70	19.46	73.04	0.94	0.91	0.95
>13[a]	7	25.77	15.36	43.25	0.90	0.87	0.92
>14	4	28.73	16.44	50.21	0.83	0.8	0.86
>15	10	34.09	22.11	52.56	0.90	0.87	0.92
>16[b]	2	22.40	9.97	50.30	–	–	–
>18	4	25.29	12.18	52.51	0.88	0.85	0.90

CI: confidence interval.
[a] Optimal cutoff score.
[b] Area under the curve could not be calculated by Stata.

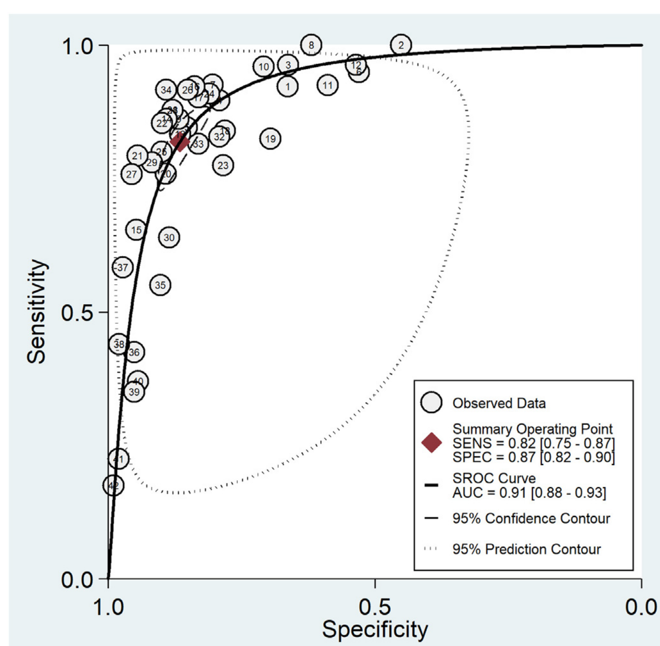

Fig. 4. Summary receiver operating characteristic (SROC) curve analysis for all data from all included studies. The likelihood of heterogeneity was high because the area of 95% prediction region was markedly wider than the area of 95% confidence region. The integrated AUC estimate was 0.91, 95% CI [0.88, 0.93].

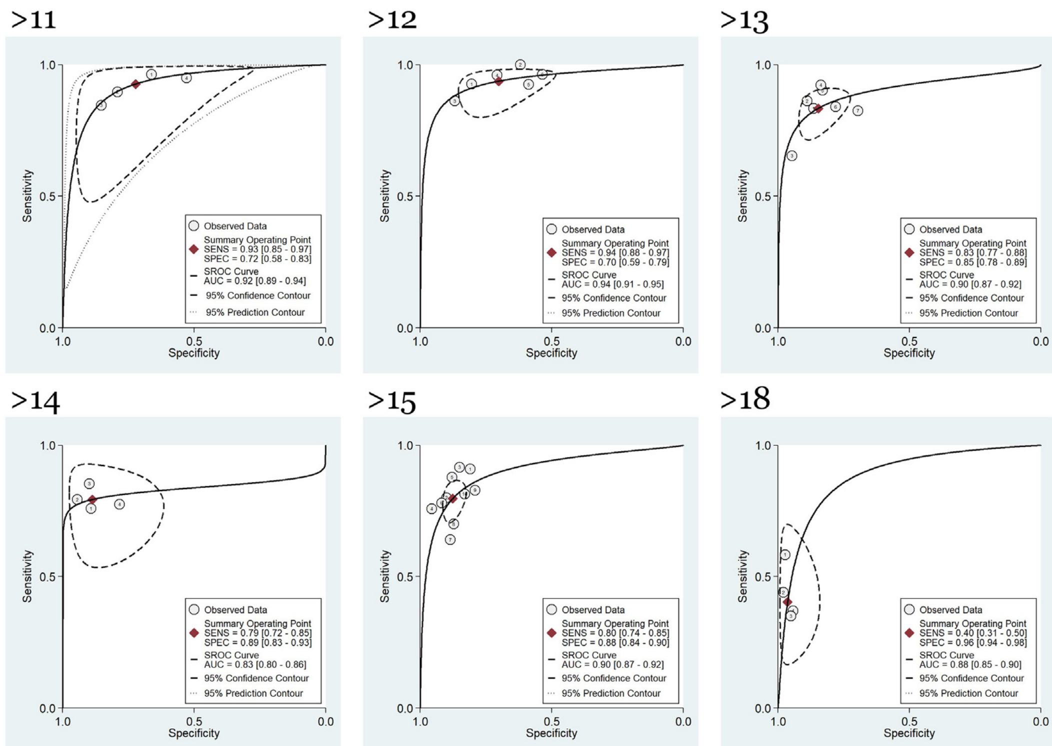

Fig. 5. SROC curve analysis for each of the six cutoff scores whose raw data were provided from four or more studies. In the graph of SROC curve at the NDDI-E cutoff score >11, the likelihood of heterogeneity was high because the area of 95% prediction region was wider than the area of 95% confidence region.

this reason, it is not yet clear whether the NDDI-E cutoff scores that have been validated so far for a particular language are appropriate for the language.

In the studies included in this meta-analysis, PWE samples may have appeared to be insufficient to represent PWE in certain populations, and the numbers of PWE sample may have seemed to be not enough. Nine studies of the included 13 studies were performed in a single clinic, and only four of the studies were performed in more than one institution. The sampling number of the included studies ranged from 98 to 248, with a median of 124. The limitation of the primary studies may have affected their cutoff scores for MDD detection.

Although the validation of NDDI-E was performed in neighboring countries, the cutoff scores for MDD detection were different significantly in some cases. Even though Korea and Japan are neighboring countries, the validated cutoff scores for MDD detection were markedly different from each other. The score was >16 in the Japanese study [13]

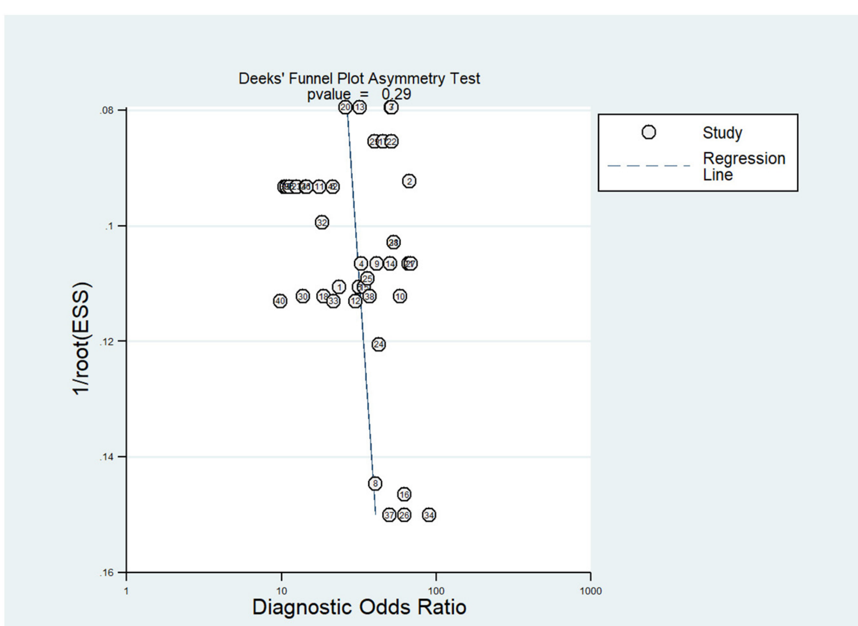

Fig. 6. Deeks' funnel plot asymmetry test for assessment of publication bias. The *p* value of the bias estimate was 0.29, and there was no asymmetry among the data points. Taken together, the likelihood of publication bias was low.

and >11 in the Korean study [12]. A similar pattern was found in studies conducted in Europe. Greece and Italy are also close countries with similar cultures, but the cutoff scores were different from each other. The score was >15 in the Greek study [23] and >13 in the Italian study [22]. This distinction was also observed between France and Spain. The two countries are adjacent, but the French study and the Spanish study provided different cutoff scores from each other. The score was >15 in the former study [21] and >13 in the latter study [19]. Thus, differences in languages and cultures among countries in the same region may affect the NDDI-E cutoff score for MDD detection.

We extracted as much of the raw data as possible from the included studies for this meta-analysis, but there were shortcomings. There were studies that only provided data for a particular NDDI-E score. Even the studies that provided raw data for multiple cutoff scores often provided raw data only for certain scores, or skipped raw data for some scores. In addition, as we mentioned above, sampling of PWE may not be appropriate, and the numbers of the sample may not be enough to validate the NDDI-E for MDD detection in the primary studies. To overcome this, more future researches with proper sample of PWE and raw data of enough consecutive cutoff scores may be needed.

There is another major limitation in this meta-analysis. We could assess that the likelihood of heterogeneity among the data was high through the SROC curve in which all data presented by cutoff scores extracted from all 13 studies were integrated. Because of the heterogeneity, care should be taken when interpreting the results of our meta-analysis. In other words, the component studies of our meta-analysis were all quite different, so the results of our meta-analysis could not be an accurate summary. Therefore, a high level of caution is needed when applying the results of this study to actual practice.

5. Conclusion

We collected 13 studies that validated NDDI-E for MDD detection based on the MDD diagnosis made by the MINI, and extracted the raw data from the studies. In the DTA meta-analysis combining the raw data, the appropriate NDDI-E cutoff score for MDD detection was >13. There are various language versions of NDDI-E that have been validated so far, and each of them presented its own cutoff score for MDD detection. However, the determination of the cutoff score may have been influenced by natures of the validation studies. Accordingly, the information of our DTA meta-analysis, which was derived from combining raw data of primary studies, may play an important reference role for practical use of NDDI-E. The information may be particularly useful when using NDDI-E in geographic areas whose current languages have not been validated for the NDDI-E or the areas that are culturally different from the countries in which the language versions were validated.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yebeh.2018.12.006.

Conflict of interest statement

The authors report no financial interests or potential conflicts of interest related to the present study.

Acknowledgments

We thank Rock Bum Kim, Clinical Associate Professor, Regional Cardiocerebrovascular Disease Center, Gyeongsang National University Hospital, for supporting the DTA meta-analysis.

Appendix A. Search strategy

1. **MEDLINE**

#1. "area under curve"[MeSH Terms] OR "ROC curve"[MeSH Terms]
#2. "area under curve"[tiab] OR "ROC curve"[tiab]
#3. #1 OR #2
#4. "interview, psychological"[MeSHTerms] OR "psychiatric status rating scales"[MeSHTerms] OR "psychological tests"[MeSH terms]
#5. "psychological interview*"[tiab] OR "psychiatric status rating scale*"[tiab] OR "psychological test*"[tiab]
#6. #4 OR #5
#7. NDDIE[tw] OR NDDI-E[tw] OR "neurological disorders depression inventory"[tw]
#8. #3 OR #6 OR #7
#9. depression[MeSH Terms] OR "depressive disorder, major"[MeSH Terms]
#10. depression*[tiab] OR "depression disorder*"[tiab] OR depressive [tiab] OR "major depressive disorder"[tiab]
#11. #9 OR #10
#12. epilepsy[MeSHTerms]
#13. epilepsy[tiab] OR epilepsies[tiab] OR "seizure disorder*"[tiab] OR epileptic[tiab]
#14. #12 OR #13
#15. #8 AND #11 AND #14

2. **EMBASE**

#1. 'area under curve'/exp OR 'ROC curve'/exp
#2. 'area under curve':ab,ti OR 'ROC curve':ab,ti
#3. #1 OR #2
#4. 'interview, psychological'/exp OR 'psychiatric status rating scales'/exp OR 'psychological tests'/exp
#5. 'psychological interview*':ab,ti OR 'psychiatric status rating scale*':ab,ti OR 'psychological test*':ab,ti
#6. #4 OR #5
#7. NDDIE OR 'NDDI-E' OR 'neurological disorders depression inventory'
#8. #3 OR #6 OR #7
#9. 'depression'/exp OR 'depressive disorder, major'/exp
#10. depression*:ab,ti OR 'depression disorder*':ab,ti OR depressive:ti,ab OR 'major depressive disorder':ti,ab
#11. #9 OR #10
#12. 'epilepsy'/exp
#13. epilepsy:ab,ti OR epilepsies:ab,ti OR 'seizure disorder*':ab,ti OR epileptic:ab,ti
#14. #12 OR #13
#15. #8 AND #11 AND #14
#16. #15 AND [embase]/lim NOT ([embase]/lim AND [medline]/lim)

3. **Cochrane Library**

#1. (MeSH descriptor: [area under curve] explode all trees) OR (MeSH descriptor: [ROC curve] explode all trees)
#2. ("area under curve" OR "ROC curve":ti,ab,kw)
#3. #1 OR #2
#4. (MeSH descriptor: [interview, psychological] explode all trees) OR (MeSH descriptor: [psychiatric status rating scales] explode all trees) OR (MeSH descriptor: [psychological tests] explode all trees)
#5. ("psychological interview*" OR "psychiatric status rating scale*" OR "psychological test*":ti,ab,kw)
#6. #4 OR #5
#7. (NDDIE OR NDDI-E OR "neurological disorders depression inventory":ti,ab,kw)
#8. #3 OR #6 OR #7
#9. (MeSH descriptor: [depression] explode all trees) OR (MeSH descriptor: [depressive disorder, major] explode all trees)
#10. (depression* OR "depression disorder*" OR depressive OR "major depressive disorder":ti,ab,kw)

#11. #9 OR #10

#12. (MeSH descriptor: [epilepsy] explode all trees)

#13. (epilepsy OR epilepsies OR "seizure disorder*" OR epileptic:ti,ab,kw)

#14. #12 OR #13

#15. #8 AND #11 AND #14

4. **Web of Science**

#1. TS=("area under curve" OR "ROC curve") OR TI=("area under curve*" OR "ROC curve")

#2. TS=("interview, psychological" OR "psychiatric status rating scales" OR "psychological tests") OR TI=("psychological interview*" OR "psychiatric status rating scale*" OR "psychological test*")

#3. TI=(NDDIE OR NDDI-E OR "neurological disorders depression inventory")

#4. #1 OR #2 OR #3

#5. TS=(depression OR "depressive disorder, major") OR TI=(depression OR"depression disorder*" OR depressive OR "major depressive disorder")

#6. TS=(epilepsy) OR TI=(epilepsy OR epilepsies OR "seizure disorder*" OR epileptic)

#7. #4 AND #5 AND #6

5. **SCOPUS**

((INDEXTERMS("area under curve" OR "ROC curve") OR TITLE-ABS-KEY("area under curve" OR "ROC curve")) OR (INDEXTERMS("interview, psychological" OR "psychiatric status rating scales" OR "psychological tests") OR TITLE-ABS-KEY("psychological interveiw*" OR "psychiatric status rating scale*" OR "psychological test*")) OR (TITLE-ABS-KEY(NDDIE OR NDDI-E OR "neurological disorders depression inventory"))) AND (INDEXTERMS(depression OR "depressive disorder, major") OR TITLE-ABS-KEY(depression* OR "depression disorder*" OR depressive OR "major depressive disorder")) AND (INDEXTERMS(epilepsy) OR TITLE-ABS-KEY(epilepsy OR epilepsies OR "seizure disorder*" OR epileptic))

References

[1] Kwon OY, Park SP. Depression and anxiety in people with epilepsy. J Clin Neurol 2014;10:175–88.
[2] Gilliam FG, Barry JJ, Hermann BP, Meador KJ, Vahle V, Kanner AM. Rapid detection of major depression in epilepsy: a multicentre study. Lancet Neurol 2006;5:399–405.
[3] Margrove K, Mensah S, Thapar A, Kerr M. Depression screening for patients with epilepsy in a primary care setting using the Patient Health Questionnaire-2 and the Neurological Disorders Depression Inventory for Epilepsy. Epilepsy Behav 2011;21:387–90.
[4] Gill SJ, Lukmanji S, Fiest KM, Patten SB, Wiebe S, Jette N. Depression screening tools in persons with epilepsy: a systematic review of validated tools. Epilepsia 2017;58:695–705.
[5] Endnote X8.2. Philadelphia: Clarivate Analytics; 2017.
[6] Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. BMC Med Res Methodol 2016;16:97.
[7] Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: analysing and presenting results. In: Deeks JJ, Bossuyt P, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy The Cochrane Collaboration; 2010 Available from: http://srdta.cochrane.org.
[8] Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58:882–93.
[9] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.
[10] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151:264–9 [w64].
[11] Guo Y, Chen ZM, Zhang YX, Ge YB, Shen CH, Ding Y, et al. Reliability and validity of the Chinese version of the Neurological Disorders Depression Inventory for Epilepsy (C-NDDI-E). Epilepsy Behav 2015;45:225–8.
[12] Ko PW, Hwang J, Lim HW, Park SP. Reliability and validity of the Korean version of the Neurological Disorders Depression Inventory for Epilepsy (K-NDDI-E). Epilepsy Behav 2012;25:539–42.
[13] Tadokoro Y, Oshima T, Fukuchi T, Kanner AM, Kanemoto K. Screening for major depressive episodes in Japanese patients with epilepsy: validation and translation of the Japanese version of Neurological Disorders Depression Inventory for Epilepsy (NDDI-E). Epilepsy Behav 2012;25:18–22.
[14] Tong X, An D, Lan L, Zhou X, Zhang Q, Xiao F, et al. Validation of the Chinese version of the Neurological Disorders Depression Inventory for Epilepsy (C-NDDI-E) in West China. Epilepsy Behav 2015;47:6–10.
[15] de Oliveira GN, Kummer A, Salgado JV, Portela EJ, Sousa-Pereira SR, David AS, et al. Brazilian version of the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E). Epilepsy Behav 2010;19:328–31.
[16] de Oliveira GN, Lessa JM, Goncalves AP, Portela EJ, Sander JW, Teixeira AL. Screening for depression in people with epilepsy: comparative study among neurological disorders depression inventory for epilepsy (NDDI-E), hospital anxiety and depression scale depression subscale (HADS-D), and Beck depression inventory (BDI). Epilepsy Behav 2014;34:50–4.
[17] de Oliveira GNM, de Araûjo Filho GM, Kummer A, Salgado JV, Portela EJ, Sousa-Pereira SR, et al. Neurological Disorders Depression Inventory for Epilepsy (NDDI-E): Brazilian version of a screening instrument. J Epilepsy Clin Neurophysiol 2011;17:49–53.
[18] Thomson AE, Calle A, Fontela ME, Yepez L, Munoz Giacomelli F, Jauregui A, et al. Screening of major depression in epilepsy: the Neurologic Depression Disorders Inventory in Epilepsy-Spanish version (Argentina). Epilepsia 2014;55:331–4.
[19] Di Capua D, Garcia-Garcia ME, Reig-Ferrer A, Fuentes-Ferrer M, Toledano R, Gil-Nagel A, et al. Validation of the Spanish version of the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E). Epilepsy Behav 2012;24:493–6.
[20] Hansen CP, Amiri M. Combined detection of depression and anxiety in epilepsy patients using the Neurological Disorders Depression Inventory for Epilepsy and the World Health Organization well-being index. Seizure 2015;33:41–5.
[21] Micoulaud-Franchi JA, Barkate G, Trebuchon-Da Fonseca A, Vaugier L, Gavaret M, Bartolomei F, et al. One step closer to a global tool for rapid screening of major depression in epilepsy: validation of the French NDDI-E. Epilepsy Behav 2015;44:11–6.
[22] Mula M, Iudice A, La Neve A, Mazza M, Bartolini E, De Caro MF, et al. Validation of the Italian version of the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E). Epilepsy Behav 2012;24:329–31.
[23] Zis P, Yfanti P, Siatouni A, Tavernarakis A, Gatzonis S. Validation of the Greek version of the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E). Epilepsy Behav 2013;29:513–5.