SLEEP, 2022, 1–12

https://doi.org/10.1093/sleep/zsac150
Advance Access Publication Date: 26 July 2022
Original Article

ORIGINAL ARTICLE

Comparison of rapid eye movement without atonia quantification methods to diagnose rapid eye movement sleep behavior disorder: a systematic review

Jung-Ick Byun[1], Tae-Won Yang[2,3,4], Jun-Sang Sunwoo[5,], Won Chul Shin[1], Oh-Young Kwon[2,4,6,*,†] and Ki-Young Jung[7,*,†,]

[1]Department of Neurology, Kyung Hee University Hospital at Gangdong, Kyung Hee University School of Medicine, Seoul, Republic of Korea, [2]Department of Neurology, Gyeongsang National University College of Medicine, Jinju, Korea, [3]Department of Neurology, Gyeongsang National University Changwon Hospital, Changwon, Korea, [4]Institute of Health Science, Gyeongsang National University College of Medicine, Jinju, Korea, [5]Department of Neurology, Kangbuk Samsung Hospital, Seoul, Korea, [6]Department of Neurology, Gyeongsang National University Hospital, Jinju, Korea and [7]Department of Neurology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[†]These authors contributed equally to this work.

*Corresponding author. Oh-Young Kwon, Department of Neurology, Gyeongsang National University College of Medicine, 15, Jinjudae-ro 816beon-gil, Jinju 52727, Korea. Email: oykwon@gnu.ac.kr; Ki-Young Jung, Department of Neurology, Seoul National University Hospital, Neuroscience Research Institute, Seoul National University College of Medicine, 101 Daehak-Ro, Jongno-Gu, Seoul 110-744, Republic of Korea. Email: jungky@snu.ac.kr.

Abstract

Study Objectives: Rapid eye movement (REM) sleep without atonia (RWA) is essential for diagnosing REM sleep behavior disorder (RBD). Manual and automatic quantifications of RWA that use different criteria have been validated. This study compared the RWA quantification methods for diagnosing RBD.

Methods: The PubMed, EMBASE, Web of Science, and Cochrane Library databases were systemically searched for studies published from inception to December 2021. The inclusion criteria were cohort, cross-sectional, and case-control studies assessing the sensitivity and specificity of RWA quantification methods. Pooled estimates of the sensitivity, specificity, diagnostic odds ratio (DOR), and area under the curve (AUC) were determined. Risk of bias and certainty of evidence was assessed using the Quality Assessment of Diagnostic Accuracy Studies tool and the Grading of Recommendations, Assessment, Development, and Evaluations framework, respectively.

Results: Fourteen articles including 402 patients with RBD met the inclusion criteria. Manual methods evaluating any chin and phasic flexor digitorum superficialis (FDS) activity had the highest DOR (138.8, 95% CI = 21.8% to 881.7%) and AUC (0.9686). The automatic REM atonia index (RAI) showed similar or higher sensitivity (89.1%, 95% CI = 84.6% to 92.7%) but a lower specificity (73.5%), DOR (43.1), and AUC (0.9369) than the manual techniques.

Conclusions: In this meta-analysis, manual RWA quantification that employed chin or phasic FDS activity had the best RBD diagnostic performance. The automatic RAI method may be useful for screening patients with RBD. The results should be interpreted carefully because of the high risk of bias in patient selection and significant heterogeneity among the studies.

PROSPERO Registration number: CRD42021276445.

Statement of Significance

Various rapid eye movement (REM) sleep without atonia (RWA) scoring methods have been validated for diagnosing REM sleep behavior disorder (RBD). We performed a systematic review to investigate and compare the diagnostic performance of the RWA quantification methods. The results of our meta-analysis showed outstanding diagnostic performance of all RWA quantification methods with area under the curve (AUC) values over 0.9. The manual scoring method employing any chin muscle and phasic flexor digitorum superficialis was the best method for diagnosing RBD, followed by the manual method adopting any chin and phasic tibialis anterior activity and applying any chin activity only. The automated REM atonia index had an acceptable sensitivity but had a lower specificity and AUC than the manual methods.

Submitted: 23 March, 2022; **Revised:** 23 June, 2022

Key words: diagnostic accuracy; REM sleep behavior disorder; REM sleep without atonia; REM atonia index; meta-analysis

Introduction

Rapid eye movement (REM) sleep without atonia (RWA) is an abnormal electromyography (EMG) activity during REM sleep recorded by polysomnography. Measuring RWA is crucial for the diagnosis of REM sleep behavior disorder (RBD) [1]. Because qualitative examination of RWA by visual interpretation alone is insufficient to reliably diagnose RBD, quantitative assessment of RWA is recommended for RBD diagnosis [2, 3]. Studies have validated various manual or automated methods for the quantification of RWA.

Several manual scoring methods have utilized phasic, tonic, and "any" muscle activity in chin muscle with or without limb muscle, such as flexor digitorum superficialis (FDS) or tibialis anterior (TA), using various epoch length and EMG amplitude criteria for RWA quantification [4]. The Montreal method by Lapierre and Montplaisir uses tonic and phasic RWA employing only chin muscle [5]. The Sleep Innsbruck Barcelona (SINBAR) method uses tonic and phasic RWA employing both chin and limb muscles, specifically FDS, and uses an additional "any RWA" criterion [6]. McCarter *et al.* [7] also utilized chin and limb muscles for tonic and phasic RWA but used the TA muscle instead of the FDS for the "any RWA." Although criteria may differ between the scoring methods, overall, they quantify the proportion of RWA activity during REM sleep.

Acceptable sensitivity and specificity of RWA for RBD diagnosis were reported regardless of the RWA measurement techniques [4, 8]. Although any chin EMG activity was recommended as an acceptable method [4], adding upper limb EMG to any chin EMG activity is suggested for better RWA assessment [9]. Recently, the International RBD Study Group (IRBDSG) recommended employing both chin and FDS muscle activity for RWA measurements according to the SINBAR scoring method.

Because manual RWA measurement is time-consuming, requires advanced expertise, and can be biased [8], automatic RWA quantification may be advantageous. The REM atonia index (RAI) is the most commonly used automatic method and has been validated in numerous cohorts [7, 10–12]. The RAI subdivided the chin EMG signal into 1-second mini-epochs and calculated the RWA as (amplitude ≤ 1 μV)/(100- (1 μV < amplitude ≤ 2 μV)) [13]. Even though the RAI may be a fast and practical method for busy clinical practice, the American Academy of Sleep Medicine scoring method and IRBDSG guidelines do not recommend the automatic scoring method [4].

The diagnostic performance of the manual and automatic approaches for RWA quantification has not been systematically analyzed. We performed a systematic review to investigate RWA quantification methods for the diagnosis of RBD. The aim of our study was to compare the diagnostic performance of various RWA quantification methods, including manual procedures and an automated RAI.

Methods

This study was performed in accordance with the recommendations of the Preferred Reporting Items for Systematic reviews and Meta-Analyses 2020' (PRISMA 2020) guidelines [14]. The protocol of this study was specified in advance and was registered in PROSPERO (Registration No. CRD42021276445).

Eligibility criteria

Inclusion criteria. We included studies that provided information for true and false positive (TP and FP) and true and false negative (TN and FN) values by evaluating the performance of manual or automated RWA quantification methods for RBD diagnosis. Selected studies included a group of participants with RBD (isolated or secondary) according to International Classification of Sleep Disorders (ICSD) criteria [1] (reference standards) and a group of controls without RBD. We considered prospective and retrospective cohort, cross-sectional and case-control studies. The results of baseline evaluations in longitudinal studies were also considered.

Exclusion criteria. We excluded duplicates or subcohorts of already published cohorts and studies investigating animals or pediatric populations. We did not exclude studies based on publication date or language.

Search strategy

Referring to key articles that evaluated the sensitivity and specificity of RWA for RBD diagnosis, we established a search strategy after selecting the following words and phrases: "REM sleep behavior disorder," "parasomnia," "polysomnography," "sleep monitoring," "electromyography," "muscle," "quantitative," "atonia," "RWA," "activity," "tonic," and "phasic." The initial search was performed on November 9, 2021, through multiple databases for relevant studies: MEDLINE, EMBASE, Web of Science, and the Cochrane Library. After reviewing the initial search results, the authors (J.-I.B., T.-W.Y., O.-Y.K.) discussed and amended the search strategy and performed the final search on December 10, 2021. Databases were searched from their inception to December 2021 for articles using a combination of keywords and Medical Subject Heading terms (Supplementary Table S1). Moreover, we manually searched the reference lists of the identified publications for additional studies.

Selection process

First, each of the two researchers (J.-I.B. and T.-W.Y.) independently screened the relevant studies according to the titles and abstracts included in the search results. Extensive screening was performed to avoid omitting any relevant studies. Subsequently, the same researchers independently underwent the selection process after evaluating the full texts of the included articles to assess their eligibility for inclusion in this meta-analysis. All authors discussed and came to consensus regarding the selection criteria, and any disagreement was resolved by discussion and by the participation of a third arbitrator (K.-Y.J.) when necessary.

Data collection process

Data extraction was first performed by one researcher (J.-I.B.) and subsequently verified by the other two researchers (T.-W.Y. and O.-Y.K.). For the meta-analysis, absolute numbers of TP, FP, TN, and FN values were derived from the data provided in the articles and were summarized in 2 × 2 tables. In some cases, we

contacted the authors to request the data that was included in a study but not shown. For each study, the following additional data were collected: patient demographics; RBD duration; proportion of REM sleep; apnea-hypopnea index; muscles used; definition of muscle contraction and epoch duration; and cutoff criteria for RWA.

Synthesis methods

Based on the TP, TN, FP, and FN values, we calculated the pooled sensitivity, specificity, diagnostic odds ratio (DOR), and area under the curve (AUC) using random-effects models (using the DerSimonian–Laird method) by using Meta-DiSc 1.4 [15]. The DOR is a single overall indicator of diagnostic performance; a higher DOR indicates the better diagnostic performance of the RWA scoring method [16, 17]. A summary receiver operating characteristic (SROC) curve was constructed with the Moses-Shapiro-Littenberg method [18].

Heterogeneity among studies was evaluated by means of Cochrane Q statistics and the Higgins I^2 index [19]. We considered significant Cochran's Q statistics as an indicator for heterogeneity beyond sampling variation, and $I^2 > 75\%$ was considered to be highly heterogeneous. When heterogeneity in DOR was present, meta-regression analysis was performed to identify the source of variability by the following predefined influencing factors: age, the proportion of males, RBD type (isolated, secondary, or unspecified), the proportion of REM sleep, and exclusion or arousal or respiratory events. Scoring criteria were also included as an influencing factor for the manual scoring methods. The analysis was also performed using Meta-DiSc 1.4 [15].

Quality assessment of the literature

The quality of each article was independently assessed by the two researchers (J.-I.B. and O.-Y.K.) using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [20] using Review Manager, version 5.4 (Cochrane Collaboration) [21]. The QUADAS evaluates the following four domains: patient selection, reference standard, index test, and flow and timing of patients through the study. Discordance between the authors was resolved by discussion with the third arbitrator (T.-W.Y.) when necessary. We summarized the quality of the evidence and its applicability by describing the number of studies with a high/low/unclear risk of bias.

Certainty of evidence for the pooled estimates

Quality of evidence was assessed using the GRADEpro Guideline Development Tool (GRADEpro GDT) online software [22]. The GRADE criteria for the diagnostic test evaluates the following five domains (risk of bias, indirectness, inconsistency, imprecision, and publication bias) by classifying each domain into the following four grades: high indicates that further study is unlikely to change confidence in the effect size and direction; moderate indicates that further study may change confidence, low indicates that further study is very likely to change confidence, and very low indicates that the effect cannot be estimated accurately [23]. Two researchers (J.-I.B. and O.-Y.K.) independently assessed the quality, and any discordance was resolved by discussion with the third researcher (T.-W.Y.).

Publication bias

When there were more than 10 studies with the same RWA quantification method, tests for publication bias were performed using the meta package in R [24]. Publication bias was assessed by visual inspection of the symmetry of funnel plots and Egger's regression intercept test [25].

Results

We initially identified 1700 articles, and 1084 of them remained after the removal of duplicates. A total of 34 articles were retrieved, and the full text of these articles was read after screening the titles and abstracts. The detailed steps of the article selection process are shown in a PRISMA flow diagram (Figure 1). Finally, a total of 14 studies met the study criteria and were included in the meta-analysis (Figure 1) [6, 7, 10–12, 26–34]. The reasons for excluding the other studies are listed in Supplementary Table S2.

Characteristics of included studies

The features and data of the included studies are shown in Tables 1 and 2. These tables summarize the information of the studies collected from the manual and automatic RWA quantification methods. The data came from six studies that used manual scoring, three that used the automatic RAI method, and five that used both manual and automatic RWA quantification.

Manual RWA quantification was performed in eleven studies. The chin muscle was mainly used, and the FDS or TA was occasionally used. Muscle activity was measured in tonic, phasic, or both phases. The quantifying protocols in which muscle type and activity were combined and the studies that employed each protocol were as follows. Three studies quantified any chin and phasic FDS activity (n = 87 patients; mean age: 65 years [range: 63–67]; male/female: 62/25), and the cutoff rate was 32% [6, 10, 29]. Four studies used any chin and phasic TA activity (n =79 patients; mean age: 65 years (range: 58–69); male/female: 60/19), and the mean applied cutoff rate was 34.3% (range: 8–46.4) [6, 7, 12, 31]. Any chin activity was measured in six studies (n =162 patients; mean age range: 63–69 years; male/female: 113/49; mean RBD duration range: 5.7–9.4 years), and the average cutoff rate was 16.9% (range: 6.1–21.6) [6, 7, 10, 12, 29, 32]. Eleven studies recorded phasic chin activity (n = 340 patients; mean age range: 54–69 years; male/female: 233/107; mean RBD duration range: 4.4–9.4 years) with a mean cutoff rate of 14.8% (range: 3.2–47.5) [6, 7, 10–12, 29–34]. Nine studies assessed tonic chin activity (n = 305 patients; mean age range: 54–67 years; male/female: 200/105; mean RBD duration range: 4.4–7.9 years) with a mean cutoff rate of 7.9% (range: 0.99–30) [6, 10, 11, 29–34]. Eight studies [7, 10–12, 26–28, 32] evaluated the performance of the automatic RAI method (n = 248 patients; mean age range: 58–69 years; male/female: 176/72; mean RBD duration range: 5.7–9.4 years) for diagnosing RBD. The cutoff criterion used for the RAI was a mean of 0.87 (range: 0.8–0.9).

Literature quality of the included studies

The methodological quality was assessed using QUADAS-2 scores. All included studies were graded high or unclear in

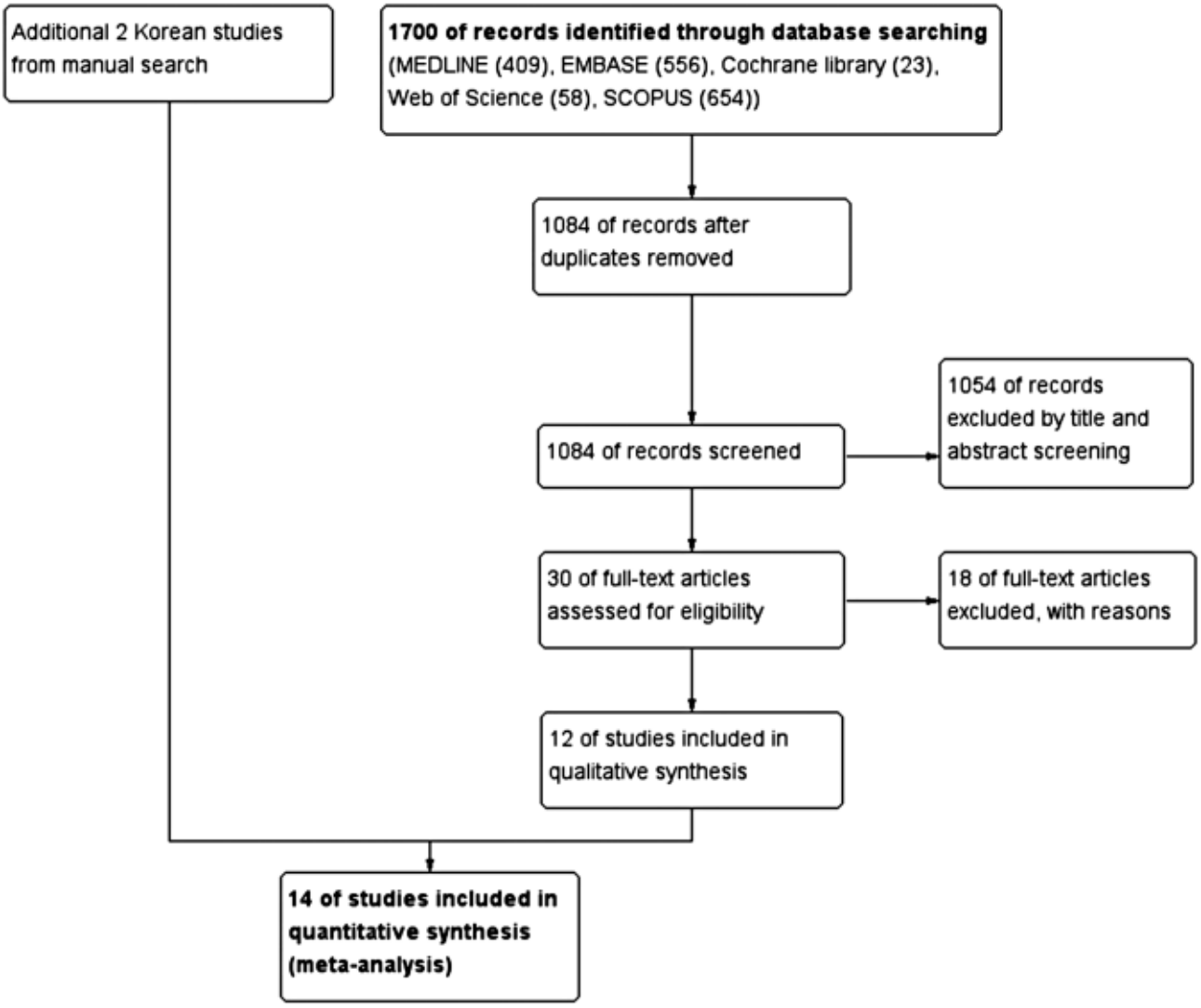

Figure 1. Flow diagram of the selection of the relevant studies. A total of 1700 research articles were identified by searching five databases. Of these articles, 616 duplicate studies were removed, and an additional 1050 studies that did not satisfy the selection criteria were excluded from the analyses. The full texts of the remaining 34 studies were reviewed, and 20 of these were excluded for various reasons. Thus, a total of 14 studies were ultimately selected for the meta-analysis.

1 or more domains of the QUADAS-2, so they were at risk of bias. In the patient selection domain, bias risk in 10 studies and applicability in 13 studies were graded as a considerable risk because they had a case-control design or had disease controls, not healthy controls. Ten studies that did not prespecify a cutoff threshold were scored as "unclear" in the index domain. All except one study used ICSD criteria to diagnose RBD, and that one was rated "low" in the reference standard domain. In the flow and timing domain, more than half (n = 8, 57.1%) of the studies were evaluated as "high risk" because they did not specify RBD duration (Figure 2 and Supplementary Figure S1).

Results of syntheses

Pooled estimates of sensitivity, specificity, DOR, and AUC values are summarized in Table 3. Forest plots for the sensitivity and specificity with each method are shown in Figure 3 and Supplementary Figure S2. Forest plots for the DOR and SROC curves are presented in Figure 4 and Supplementary Figure S3. The highest sensitivity was seen with the manual procedure using any chin and phasic FDS activity (93.1%, 95% CI = 85.6% to 97.4%). The manual method adopting phasic chin activity had the lowest sensitivity (60.0%, 95% CI = 54.6% to 65.2%). The automatic RAI technique showed pooled sensitivity (89.1%, 95% CI = 84.6% to 92.7%) similar to the manual method employing any chin and phasic TA activity (87.3%, 95% CI = 78.0% to 93.8%) or the manual method evaluating any chin activity (90.1%, 95% CI = 84.5% to 94.2%).

The specificity was highest in the manual method adopting any chin and phasic TA activity (95.6%, 95% CI = 90.1% to 98.6%). This was followed by high specificity values with the manual technique adopting tonic chin activity (93.7%, 95% CI = 90.0% to 96.3%) and the manual method employing any chin and phasic FDS activity (91.3%, 95% CI = 84.6% to 95.8%). The automatic RAI had the lowest specificity (73.5%, 95% CI = 67.8% to 78.7%) (Figure 3 and Supplementary Figure S2).

The DOR was highest for the manual method employing any chin and phasic FDS activity (138.8, 95% CI = 21.8% to 881.7%). This was followed by high DOR values with the

Table 1. Characteristics of studies that provided information about manual methods for quantifying REM sleep disorder without atonia among all studies included in the current meta-analysis.

Study	Design	Groups (N)	Age (years)	Sex (M/F)	RBD Dx/ Exclusion criteria (OSA, REM duration)	RBD duration (years)	REM sleep (% or min)	AHI (per h)	Artifact correction	Muscle investigated	Analysis description	Cut-off (%)
Ferri 2014	CC	HC (75)	61.0 ± 12.1	57/ 18			19.6 ± 5.5%	-				
		RBD (74)	62.1 ± 9.7	56/ 18	ICSD-1/ -	7.9 ± 7.5	19.8 ± 8.3%	-	-	Chin	Tonic (30 s, amp×2) Phasic (2 sec, amp×4)	PC (15), TC (30)
Figorilli 2017	CS	PD-RBD (25)	62.7 ± 10.1	11/ 14			13.7 ± 8.2%	2.9 ± 3.9				
		PD+ RBD (37)	66.0 ± 7.5	24/ 13	ICSD-3/ REM time <5 min	-	10.5 ± 5.5%	5.5 ± 9.2	Arousal, respiratory events	Chin, FDS, TA	Tonic (30 s, amp×2) Phasic (3 s, amp×4) Any	PC (16), TC (30), AC (18), AC + PF (32)
Frauscher 2012	CS	OSA (30)	66.9 ± 8.6	25/5			20.1 ± 6.2	3.2 ± 2.4				
		RBD (30)	67.0 ± 7.7	20/ 10	ICSD-2/ AHI>10, REM AHI>10	8.5 ± 8.2	17.3 ± 7.5%	2.6 ± 2.2	Arousal	Chin, FDS, TA, EDB, Biceps, SCM	Tonic (30 s, amp×2) Phasic (3 s, amp×2) Any	PC (16.3), TC (8.7), AC (18.2), AC + PF (31.9), AC + PT (46.4)
Frauscher 2014	CS	HC (60)	50.9 ± 13.8	18/ 42			12.2 ± 4.6%	3.1 ± 4.2				
		RBD (20)	65.1 ± 11.6	18/2	ICSD-2/ AHI>10	-	15.8 ± 5.9%	2.9 ± 2.3	Arousal	Chin, FDS	Tonic (30 s, amp×2) Phasic (3 s, amp×2) Any	PC (16), TC (9.6), AC (18), AC + PF (32)
Khalil 2013	CC	HC (10)	35.6 ± 18.6	5/5			-	-				
		RBD (16)	54.3 ± 16.2	11/5	- / REM time <5 min				Arousal	Chin	Tonic (20 s, amp×2) Phasic (2 sec, amp×4)	PC (3.17), TC (1.22)
Kim 2020	CC	PSG (14)	59.0 ± 9.6	7/7			15.9 ± 9.3%	15.4 ± 12				
		IRBD (14)	58.5 ± 6.5	7/7	ICSD-3		15.7 ± 9.9%	11.9 ± 9.5		Chin, TA	Tonic (30 s, amp×4) Phasic (3 sec, amp×4) Any	TC (4.9) PC (4.2) AC + PT (8.0)
Lee 2014	CC	PSG (10)	67.4 ± 6.8	6/4			18.6 ± 5.2%	4.8 ± 4.9				
		IRBD (40)	65.8 ± 10.9	18/ 22	ICSD-2		17.4 ± 7.2%	7.0 ± 13.2		Chin	Tonic (30 s, amp×4) Phasic (3 s, amp×4) Any	TC (1.2) PC (3.8) AC (6.1)
Lee 2015	CC	OSA (15)	59.9 ± 9.0	9/6			20.1 ± 6.0%	5.1 ± 3.1				
		IRBD (17)	64.5 ± 7.2	10/7	ICSD-3/ -	4.4	20.8 ± 7.9%	3.0 ± 3.5		Chin	Tonic (30 s, amp×2) Phasic (3 s, amp×4)	TC (6.5) PC (9.5)

Table 1. Continued

Study	Design	Groups (N)	Age (years)	Sex (M/F)	RBD Dx/ Exclusion criteria (OSA, REM duration)	RBD duration (years)	REM sleep (% or min)	AHI (per h)	Artifact correction	Muscle investigated	Analysis description	Cut-off (%)
McCarter 2014	CC	OSA (40)	67.8 ± 8.8	33/7			70.9 ± 22.2min	8.6 ± 8.6				
		PD+ RBD (20)	69.2 ± 7.5	20/0	ICSD-2/ REM AHI>30	9.4 ± 13.2	70.5 ± 36 min	10.7 ± 13.2	Arousal, respiratory events	Chin, TA	Tonic (30 s, amp×2) Phasic (3 s, amp×4) Any	PC (15.5), AC (21.6), AC + PT (43.4)
McCarter 2017	CC	OSA (30)	66.1 ± 13.5	25/5			20.4 ± 6.2%	4.5 ± 5.1				
		IRBD (15)	66.8 ± 5.1	13/2	ICSD-2/ AHI>25, REM time <5 min	5.7 ± 3.1	20.6 ± 7.2%	5.5 ± 6.1	Arousal, respiratory events	Chin, TA	Tonic (30 s, amp×2) Phasic (3 s, amp×4) Any	PC (15.8), AC (19.7), AC + PT (39.5)
Yang 2020	CC	PSG (29)	67.2 ± 6.1	11/ 18			19.7 ± 6.4%	5.9 ± 3.9				
		IRBD (57)	67.1 ± 6.6	21/ 36	ICSD-3/ AHI>15, PLMI>15		19.4 ± 7.5%	6.6 ± 4.1		Chin, TA	Tonic (30 s, 10 μV) Phasic (3 s, 10 μV)	TC (0.99), PC (0.46)

Abbreviations: REM, rapid eye movement; RBD, REM sleep behavior disorder; IRBD, isolated RBD; M, male; F, female; AHI, apnea-hypopnea index; IRBD, isolated RBD; PD+RBD, Parkinson's disease patient with RBD; PD-RBD, Parkinson's disease patient without RBD; CC, case-control; CS, cross-sectional; HC, healthy controls; PSG, sleep disorder controls; OSA, obstructive sleep apnea control; ICSD, international classification of sleep disorders; FDS, flexor digitorum superficialis; TA, tibialis anterior; EDB, extensor digitorum brevis; SCM, sternocleidomastoid; amp, amplitude; PC, phasic chin; TC, tonic chin; AC, any chin; PF, phasic FDS; PT, phasic TA.

RWA quantification techniques with the method using any chin and phasic TA activity (137.5, 95% CI = 27.5% to 687.3%) and the method applying any chin activity (130.4, 95% CI = 30.7% to 553.0%). The highest AUC was also seen with the manual approach employing any chin and phasic FDS activity (0.9686). This was followed by high AUC values with the techniques assessing any chin activity (0.9657) or using any chin and phasic TA activity (0.9642). Both DOR and AUC were lowest with the manual method using phasic chin activity (32.5 and 0.9348, respectively), followed by the automated RAI method (43.1 and 0.9369, respectively) (Figure 4 and Supplementary Figure S3).

Assessment of study heterogeneity

Significant heterogeneity in sensitivity was observed in meta-analyses of performances of the automatic RAI and manual methods employing phasic and tonic chin activity. Significant heterogeneity in specificity was found for all analysis groups except for the manual method adopting any chin and phasic TA activity. Regarding DOR, only the automatic RAI method and the manual method applying phasic chin activity showed significant heterogeneity among the studies (Table 4).

Meta-regression to identify factors affecting heterogeneity with DOR was performed for the two RWA quantification methods that showed heterogeneity with DOR. A multivariate meta-regression model showed no independent predictor for either the automatic RAI method or the manual method applying phasic chin activity. Univariate meta-regression found that the heterogeneity in the RAI method was associated with the age of patients with RBD (ratio of DOR [rDOR] = 1.39, 95% CI = 1.15% to1.67%, p = .006). No significant predictor was found for manual phasic chin activity.

Certainty of evidence for the pooled estimates

Most of the certainty of the evidence for the pooled estimates was graded as low because of the serious risk of bias due to patient selection or inconsistency between the included studies. The manual method employing any chin and phasic FDS activity or adopting any chin and phasic TA activity was suspected to have a risk of bias because of the small sample size (Supplementary Table S3).

Publication bias

Because the meta-analysis of manual methods applying phasic chin activity contained 11 primary studies, the effect of small-sized studies was assessed. Visual inspection of the funnel plot of the manual methods that evaluated phasic chin activity revealed no evidence of publication bias (Supplementary Figure S4). The Egger regression intercept test confirmed that there was no bias (intercept: 2.2922, p-value: .3115).

Discussion

We performed a meta-analysis to compare diagnostic performance between the RWA quantification methods for RBD.

Table 2. Characteristics of studies that provided information about the automatic RAI method for quantifying REM sleep without atonia among the studies included in the current meta-analysis

Study	Study design	Groups (N)	Age (years)	Sex (M/F)	Exclusion criteria (OSA, REM duration)	RBD duration (years)	REM sleep (% or min)	AHI(perh)	Artifact correction	Cut-off
Cesari (2019)	CC	IRBD (29)	57.7 ± 17.2	21/8	ICSD-3/ —	—	14.1 ± 7.9%	11.2 ± 16.7	Arousal, respiratory events	—
		HC (27)	56.6 ± 9.2	13/14	—	—	20.1 ± 5.9%	3.3 ± 4.9		
Ferri (2012)	CC	PD+RBD (16)	68.9 ± 8.2	10/6	ICSD-2/ AHI > 10/h	—	10.1 ± 5.0%	—	Noise reduction	0.9
		PD-RBD (11)	66.5 ± 6.4	7/4		—	10.6 ± 6.3%	—		
Ferri (2013)	CC	IRBD (17)	66.0 ± 4.9	14/3	ICSD-2/ —	9.1 ± 12.9	17.6 ± 7.9%	—	Noise reduction	0.9
		Aged HC (14)	58.2 ± 9.6	—		—	16.0 ± 5.0%	—		
Ferri (2014)	CC	RBD (74)	62.1 ± 9.7	56/18	ICSD-1/ —	7.9 ± 7.5	19.8 ± 8.3%	—	Noise reduction	0.8
		HC (75)	61.0 ± 12.1	57/18		—	19.6 ± 5.5%	—		
Figorilli (2017)	CC	PD+RBD (37)	66 ± 7.5	24/13	ICSD-3/ REM time <5 min	—	10.5 ± 5.5%	5.5 ± 9.2	Noise reduction	0.8
		PD-RBD (25)	62.7 ± 10.1	11/14			13.7 ± 8.2%	2.9 ± 3.9		
Lee (2014)	CC	IRBD (40)	65.8 ± 10.9	18/22	ICSD-2		17.4 ± 7.2%	7.0 ± 13.2	N/A	
		PSG (10)	67.4 ± 6.8	6/4			18.6 ± 5.2%	4.8 ± 4.9		
McCarter (2014)	CC	PD+RBD (20)	69.2 ± 7.5	20/0	ICSD-2/ REM AHI > 30	9.4 ± 13.2	70.5 ± 36 min	10.7 ± 13.2	Arousal, respiratory events	0.88
		OSA (40)	67.8 ± 8.8	33/7			70.9 ± 22.2 min	8.6 ± 8.6		
McCarter (2017)	CC	IRBD (15)	66.8 ± 5.1	13/2	ICSD-2/ AHI > 25, REM time < 5 min	5.7 ± 3.1	20.6 ± 7.2%	5.5 ± 6.1	Arousal, respiratory events	0.86
		OSA (30)	66.1 ± 13.5	25/5			20.4 ± 6.2%	4.5 ± 5.1		

REM, rapid eye movement; RBD, REM sleep behavior disorder; M, male; F, female; AHI, apnea-hypopnea index; IRBD, isolated RBD; PD + RBD, Parkinson's disease patient with RBD; PD-RBD, Parkinson's disease patient without RBD; CC, case-control; HC, healthy controls; PSG, sleep disorder controls; OSA, obstructive sleep apnea; ICSD, international classification of sleep disorders.

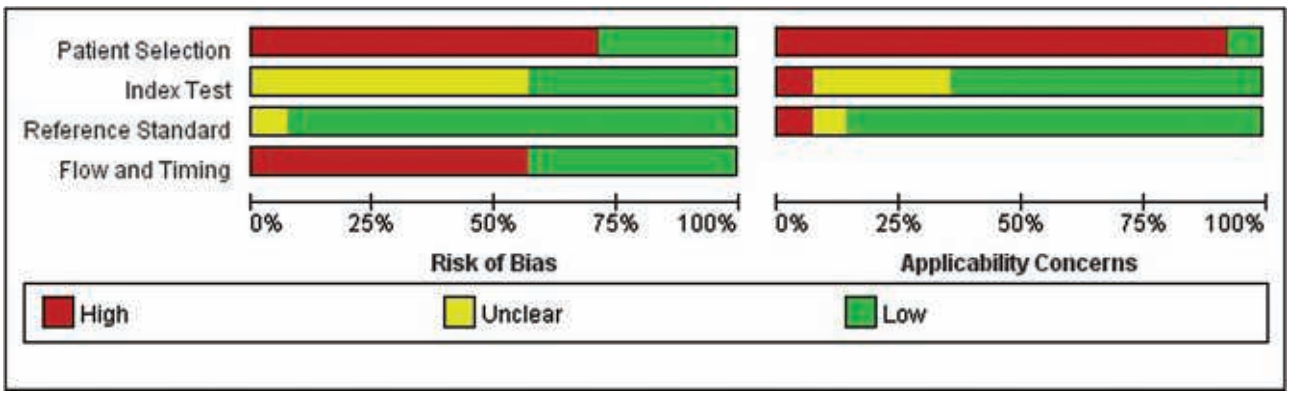

Figure 2. Summary of the risk of bias and applicability concerns of the included papers using the Quality Assessment of Diagnostic Accuracy Studies 2 tool. Most studies were rated as having "high risk" in the patient selection domain due to a case–control design. More than half of the studies were evaluated as "high risk" in the flow and timing domain because they did not specify RBD duration.

The results of this meta-analysis showed outstanding diagnostic performance of all RWA quantification methods, with AUC values above 0.9 but the certainty of the evidence was generally low. Overall, manual scoring of any chin and phasic FDS activity was the best for RBD diagnosis, with the highest DOR followed by any chin and phasic TA activity or any chin activity. Phasic or tonic chin EMG activity alone showed low diagnostic performance. Automated RAI had acceptable sensitivity but had lower specificity, DOR, and AUC values compared with manual methods using tonic and/or phasic activity of chin with or without the phasic activity of the FDS or TA muscle.

Employing phasic FDS EMG activity in addition to any chin muscle activity for RWA quantification is currently recommended for RBD diagnosis in the guidelines of the SINBAR group [6] and IRBDSG [4]. Chin EMG activity can easily be affected by breathing or snoring artifacts; in contrast, FDS activity is minimally affected by these artifacts [6, 35]. In addition, interrater variability was low when utilizing FDS activity for RWA quantification [35]. In the present meta-analysis, the additional measurement of a limb muscle to the chin muscle slightly increased diagnostic accuracy for RBD. Applying any chin activity alone for manual RWA quantification showed an overall convincing sensitivity and specificity achieving 90%. The additional use of phasic

Table 3. Pooled estimates from the current meta-analysis according to the RWA quantification methods for diagnosing REM sleep behavior disorder

Index test	Number of studies	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	Diagnostic OR (95% CI)	AUC
Manual (any chin) + phasic FDS	3	0.931 (0.856–0.974)	0.913 (0.846–0.958)	138.80 (21.849–881.72)	0.9686
Manual (any chin) + phasic TA	4	0.873 (0.780–0.938)	0.956 (0.901–0.986)	137.48 (27.501–687.33)	0.9642
Manual (any chin)	6	0.901 (0.845–0.942)	0.892 (0.840–0.932)	130.37 (30.735–552.97)	0.9657
Manual (phasic chin)	11	0.600 (0.546–0.652)	0.896 (0.859–0.927)	32.450 (12.173–86.502)	0.9348
Manual (tonic chin)	9	0.823 (0.775–0.864)	0.937 (0.900–0.963)	66.966 (25.644–174.87)	0.9546
Automatic RAI	7	0.891 (0.846–0.927)	0.735 (0.678–0.787)	43.061 (13.302–139.40)	0.9369

REM, rapid eye movement; RWA, REM sleep without atonia; FDS, flexor digitorum superficialis; TA, tibialis anterior; RAI, REM atonia index; CI, confidence interval; OR, odds ratio; AUC, area under the curve.

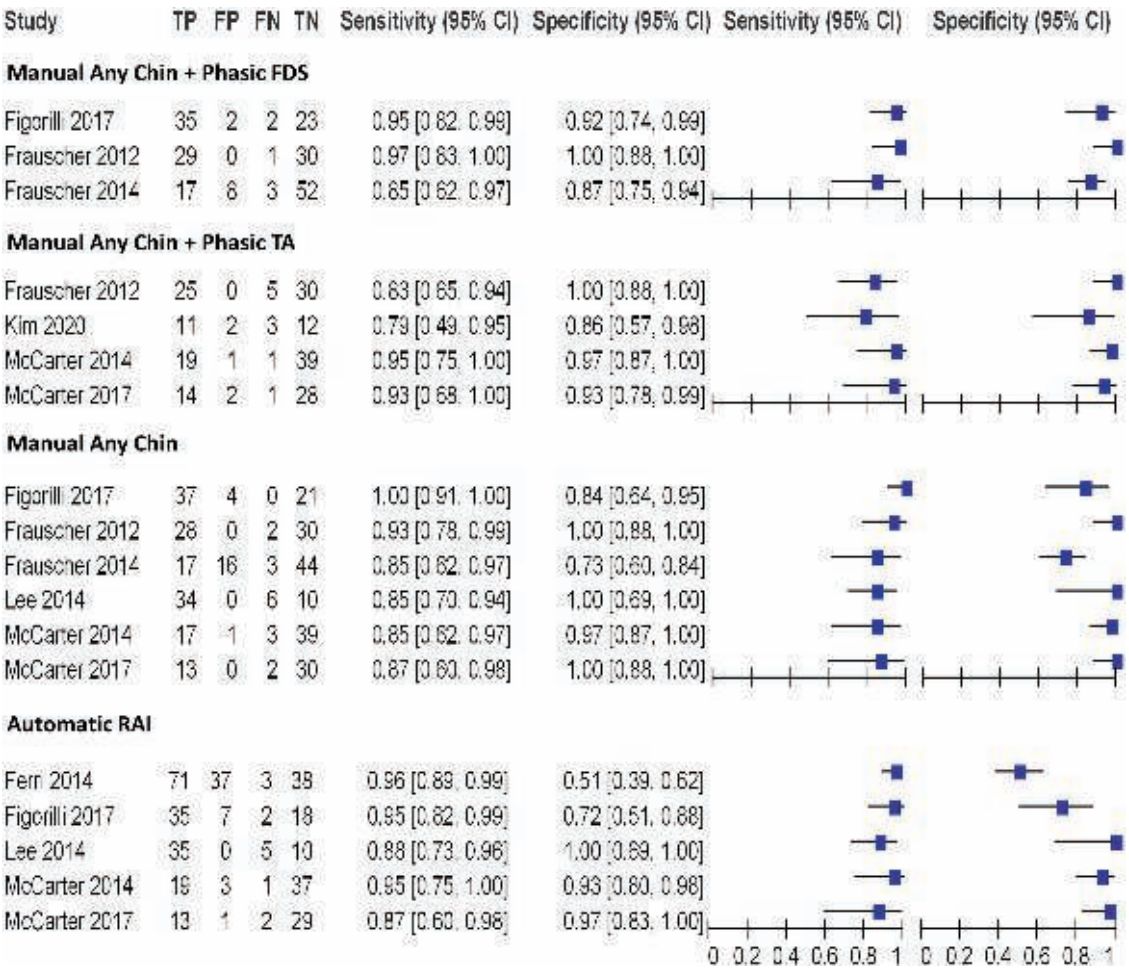

Figure 3. Comparison of pooled sensitivity and specificity for the three most accurate manual methods and an automatic process. Among the four manual or automatic techniques used to diagnose rapid eye movement sleep behavior disorder, the manual approach that employed any chin and phasic FDS activity had the highest sensitivity, and the method using any chin and phasic TA activity had the highest specificity. The automatic RAI technique showed similar sensitivity to the manual procedures but lower specificity. FN, false negative; FP, false positive; TN, true negative; TP, true positive; FDS, flexor digitorum superficialis; TA, tibialis anterior; RAI, REM atonia index.

FDS activity increased the sensitivity from 90.1% to 93.1%, specificity from 89.2% to 91.3%, DOR from 130.37 to 138.80, and AUC from 0.9657 to 0.9686.

Based on our results, adding TA activity instead of FDS activity to the manual method employing chin activity increased DOR but decreased AUC. Because TA activity in the RWA quantification procedure can contain other activities such as periodic limb movement during sleep, the addition of TA activity to the chin EMG for the quantification may have no benefit in discerning patients with RBD from the controls in older age [6, 7]. Our review showed that the manual method employing any chin and phasic TA activity had lower sensitivity and AUC but higher specificity and similar DOR than any chin and phasic FDS activity. Thus, employing an additional phasic limb muscle, regardless of whether it is the FDS or TA, may increase the diagnostic performance of RBD compared with assessing chin activity alone and reduce the likelihood of false negatives when diagnosing RBD. Moreover, adopting upper limb muscles, such as the FDS,

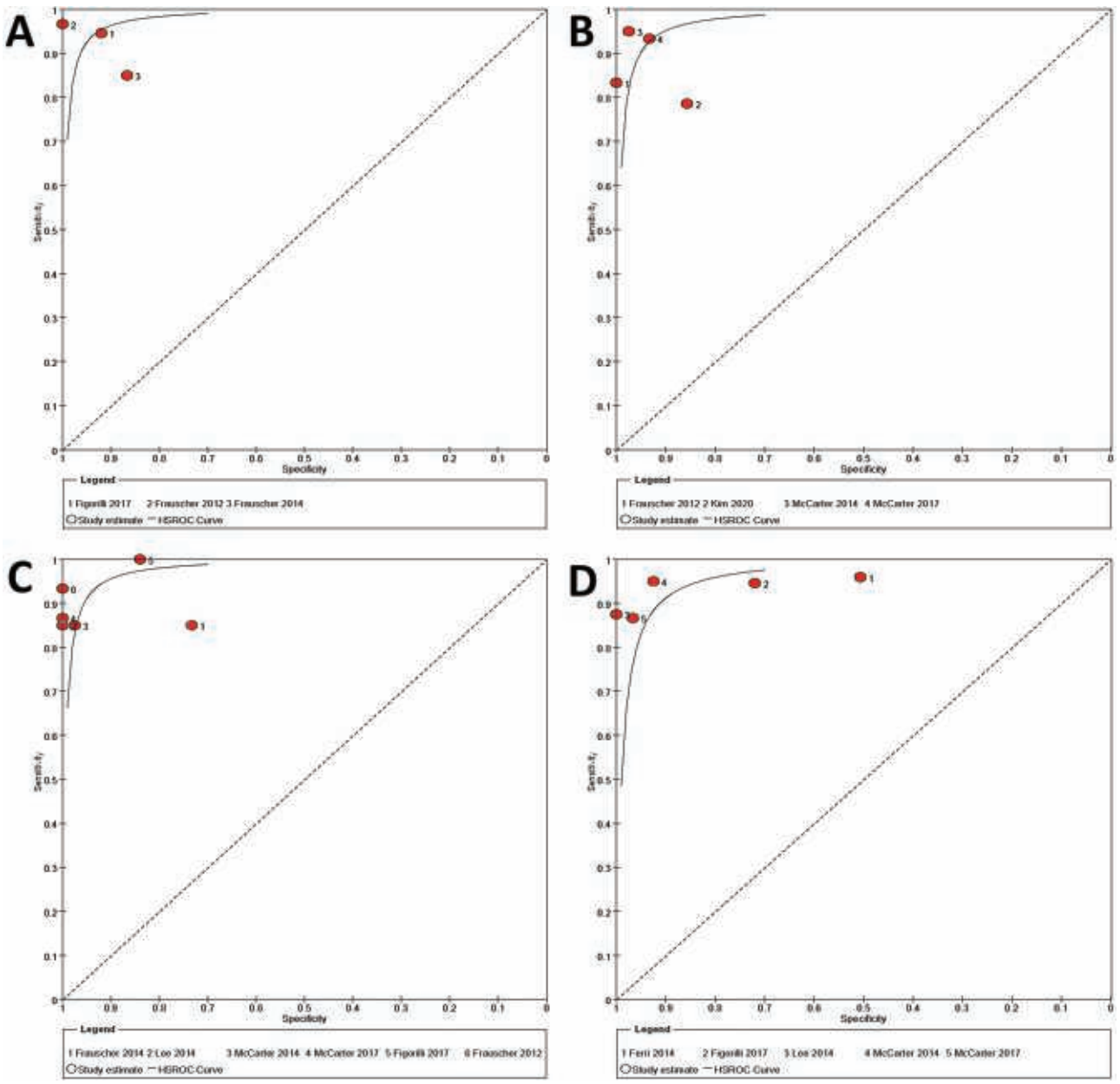

Figure 4. Comparison of the summary receiver operating characteristic curves for the three most accurate three manual methods and the automatic rapid eye movement atonia index technique. The highest AUC was observed in the manual approach that employed any chin and phasic FDS activity (A), followed by the manual method that assessed any chin activity (C), and the manual process that quantified any chin and phasic TA activity (B). The AUC was lowest in the automated RAI method (D) among the four techniques. AUC, area under the curve, SROC, summary receiver operating characteristic; FDS, flexor digitorum superficialis; TA, tibialis anterior; RAI, rapid eye movement atonia index.

Table 4. Heterogeneity statistics for the pooled estimates from the current meta-analysis according to the RWA quantification methods to diagnose REM sleep behavior disorder

Index test	No.	Pooled sensitivity	Pooled specificity	Diagnostic OR
Manual any chin + FDS	3	$\chi^2 = 2.43$, $p = .297$, $I^2 = 17.6\%$	$\chi^2 = 6.89$, $p = .032$, $I^2 = 71.0\%$	$\chi^2 = 4.59$, $p = .101$, $I^2 = 56.4\%$
Manual any chin + TA	4	$\chi^2 = 3.14$, $p = .370$, $I^2 = 4.6\%$	$\chi^2 = 5.51$, $p = .138$, $I^2 = 45.6\%$	$\chi^2 = 4.93$, $p = .177$, $I^2 = 39.1\%$
Manual any chin	6	$\chi^2 = 10.34$, $p = .066$, $I^2 = 51.6\%$	$\chi^2 = 32.32$, $p < .0001$, $I^2 = 84.5\%$	$\chi^2 = 10.04$, $p = .074$, $I^2 = 50.2\%$
Manual phasic chin	11	$\chi^2 = 170.92$, $p < .0001$, $I^2 = 94.1\%$	$\chi^2 = 55.04$, $p < .0001$, $I^2 = 81.8\%$	$\chi^2 = 23.04$, $p = .011$, $I^2 = 56.6\%$
Manual tonic chin	9	$\chi^2 = 32.00$, $p < .0001$, $I^2 = 75.0\%$	$\chi^2 = 23.18$, $p = .003$, $I^2 = 65.5\%$	$\chi^2 = 14.37$, $p = .073$, $I^2 = 44.3\%$
Automatic RAI	7	$\chi^2 = 19.48$, $p = .007$, $I^2 = 64.1\%$	$\chi^2 = 49.27$, $p = .000$, $I^2 = 85.8\%$	$\chi^2 = 23.24$, $p = .002$, $I^2 = 69.9\%$

REM, rapid eye movement; RWA, REM sleep without atonia; FDS, flexor digitorum superficialis; TA, tibialis anterior; RAI, REM atonia index; CI, confidence interval; OR, odds ratio; AUC, area under the curve.

for the additional phasic limb muscle may be better than lower limb muscles, such as the TA, for the accurate diagnosis of RBD. However, because the increase in the AUC was minimal, the advantage of additional measurement of phasic limb muscle activity may not be useful in clinical practice.

The DOR of the manual scoring method employing only phasic or tonic chin activity was lower than that of the procedure adopting any chin activity. Therefore, the use of any chin activity, rather than phasic or tonic chin activity alone, is recommended for diagnosing RBD. Between the manual processes adopting phasic or tonic activity, the RWA quantification with tonic chin activity showed higher DOR than the phasic chin activity, which was consistent with previous studies [10, 34]. The reason for the discrepancy is unclear. Phasic RWA activity is associated with degeneration of the intermediate ventromedial medulla. However, tonic RWA activity is associated with that of the locus subcoeruleus [35]. Alpha-synuclein pathophysiology may be more directly associated with tonic activity because high baseline tonic activity predicts phenoconversion to Parkinson's disease (PD) [12].

The automatic RAI approach had higher sensitivity than manual methods employing phasic or tonic chin activity in the present meta-analysis. The RAI method had similar sensitivity to manual techniques adopting any chin activity or any chin and phasic TA activity. However, the RAI approach had lower specificity, DOR, and AUC values than the manual procedures. The results of our meta-analysis were consistent with previous studies that directly compared the RAI method with manual techniques. According to one study, the accuracy of the RAI approach was somewhat lower than that of the manual methods [12], which was corroborated by our results. Another study comparing the automatic RAI and two manual methods also showed that the RAI method had substantial accuracy agreement with the manual techniques. Nevertheless, its specificity was slightly lower than that of the manual methods, which is also similar to our results [10]. Another study evaluated patients with PD and found substantial agreement of diagnostic accuracy between the RAI method and manual procedures [11]. The RAI method has adequate sensitivity despite its limited specificity. The night-to-night variability in the quantification of RWA using the RAI technique was under 20%, indicating that it is reproducible [28]. Furthermore, the procedure takes less time and effort compared to manual methods.

Based on these findings, the RAI approach could be a useful screening tool for RBD diagnosis in busy clinics. Mean subject age was significantly influenced the heterogeneity in the automatic RAI data. Additionally, age is known to affect RWA. The RAI in normal individuals without RBD peaks in early adulthood and slightly decreased with age [36]. In another study, higher phasic muscle activity was independently associated with older age and male sex in sleep disorder patients without RBD [37]. Moreover, age was an independent factor even after adjusting for sex and the presence of RBD in patients with RBD comorbid with PD [7]. Therefore, age should be considered when using the RAI technique to quantify RWA for RBD diagnosis.

Several limitations should be considered when interpreting the results of our meta-analysis. Because we evaluated RBD patients who required RWA quantification for diagnosis, the sensitivity, and overall test performance may have been overestimated, as mentioned in the previous study [38]. The majority of the studies were retrospective and cross-sectional in design, leading to significant bias in patient selection. Recently, many other automated scoring methods for RWA have been introduced [39–41]. Because we analyzed the diagnostic performance of only the automatic RAI method, this study cannot represent the diagnostic accuracy of automatic scoring in general. Moreover, recent additional scoring methods for RWA, such as phasic muscle burst duration [12] or mixed chin EMG activity [42], were not included.

This meta-analysis of RWA quantification methods for RBD diagnosis found that the manual method employing any chin and phasic FDS activity had the best performance. In contrast, adopting chin activity alone and the automatic RAI technique had the worst results. The automatic RAI approach exhibited similar sensitivity but worse specificity, DOR, and AUC than the manual procedure that measured any chin and phasic FDS activity or any chin and phasic TA activity. Based on the results of our meta-analysis, the manual method using chin and phasic FDS activity is the best RWA quantification method for RBD diagnosis. The automatic RAI method can be used as a screening tool to diagnose RBD before manual methods are performed to confirm the diagnosis.

Supplementary Material

Supplementary material is available at *SLEEP* online.

Funding

This work was supported by the National Research Foundation of Korea (NRF) (No. NRF-2020R1C1C1013160), the Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2017M3C7A1029688), and a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (2017R1A2B2012280).

Acknowledgments

We thank Dr Birgit Högl, Dr Joan Santamaria, and the SINBAR Group for providing the sensitivity value for the RWA data in their article.

Disclosure Statement

Financial disclosure: The authors have no financial relationships relevant to this article to disclose.

Nonfinancial disclosure: The authors have no conflicts of interest relevant to this article to disclose

Author Contributions

J.-I.B., T.-W.Y., O.-Y.K., and K.-Y.J. conceptualized and designed the study, assisted in data collection, analyzed the data, and drafted the initial manuscript; J.-S.S. and W.C.S. critically analyzed the data and reviewed the manuscript. O.-Y.K. and K.-Y.J. reviewed the manuscript. All authors have approved the final submitted manuscript and agreed to be accountable for all aspects of the work.

Data Availability

Data will be available from the corresponding author upon reasonable request.

References

1. American Academy of Sleep Medicine. *The International Classification of Sleep Disorders-Third Edition (ICSD-3)*. Darien, IL: American Academy of Sleep Medicine; 2014.
2. McCarter SJ, *et al*. REM sleep atonia loss distinguishes synucleinopathy in older adults with cognitive impairment. *Neurology*. 2020; **94**(1):e15–e29.
3. McCarter SJ, *et al*. Submentalis rapid eye movement sleep muscle activity: a potential biomarker for synucleinopathy. *Ann Neurol*. 2019; **86**(6):969–974.
4. Cesari M, *et al*. Video-polysomnography procedures for diagnosis of rapid eye movement sleep behavior disorder (RBD) and the identification of its prodromal stages: guidelines from the international RBD study group. *Sleep*. 2021;**45**(3). doi:10.1093/sleep/zsab257.
5. Lapierre O, *et al*. Polysomnographic features of REM sleep behavior disorder: development of a scoring method. *Neurology*. 1992;**42**(7): 1371–1374.
6. Frauscher B, *et al*. Normative EMG values during REM sleep for the diagnosis of REM sleep behavior disorder. *Sleep*. 2012;**35**(6):835–847. doi: 10.5665/sleep.1886.
7. McCarter SJ, *et al*. Diagnostic thresholds for quantitative REM sleep phasic burst duration, phasic and tonic muscle activity, and REM atonia index in REM sleep behavior disorder with and without comorbid obstructive sleep apnea. *Sleep*. 2014;**37**(10):1649–1662. doi:10.5665/sleep.4074.
8. Dijkstra F, *et al*. REM sleep without atonia and the relation with Lewy body disease. *Parkinsonism Relat Disord*. 2019;**67**:90–98.
9. Stefani A, *et al*. Need for a consensus on definitions and on research methods in RBD and its prodromal phases. *Sleep*. 2019;**42**(5). doi:10.1093/sleep/zsz057.
10. Figorilli M, *et al*. Comparison between automatic and visual scorings of REM sleep without atonia for the diagnosis of REM sleep behavior disorder in Parkinson disease. *Sleep*. 2017;**40**(2). doi:10.1093/sleep/zsw060.
11. Ferri R, *et al* . Comparison between an automatic and a visual scoring method of the chin muscle tone during rapid eye movement sleep. *Sleep Med*. 2014;**15**(6):661–665.
12. McCarter SJ, *et al*. Diagnostic REM sleep muscle activity thresholds in patients with idiopathic REM sleep behavior disorder with and without obstructive sleep apnea. *Sleep Med*. 2017;**33**:23–29.
13. Ferri R, *et al*. A quantitative statistical analysis of the submentalis muscle EMG amplitude during sleep in normal controls and patients with REM sleep behavior disorder. *J Sleep Res*. 2008;**17**(1):89–100.
14. Page MJ, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;**372**:n71.
15. Zamora J, *et al*. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006;**6**(1):31.
16. DerSimonian R, *et al*. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;**7**(3):177–188.
17. Glas AS, *et al*. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 2003;**56**(11):1129–1135.
18. Moses LE, *et al*. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med*. 1993;**12**(14):1293–1316.
19. Higgins JP, *et al*. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;**21**(11):1539–1558.
20. Whiting PF, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;**155**(8):529–536.
21. *Review Manager (RevMan)*. *Version* 5.4. The Cochrane Collaboration; 2020.
22. GRADEpro GDT. *GRADEpro Guideline Development Tool [Software]*. McMaster University and Evidence Prime; 2021.
23. Schünemann H, *et al*. *GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations*; 2013. https://gdt.gradepro.org/app/handbook/handbook.html. Accessed August 15, 2019.
24. Balduzzi S, *et al*. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health*. 2019;**22**(4):153–160.
25. Egger M, *et al*. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;**315**(7109):629–634.
26. Cesari M, *et al*. Validation of a new data-driven automated algorithm for muscular activity detection in REM sleep behavior disorder. *J Neurosci Methods*. 2019;**312**:53–64.
27. Ferri R, *et al*. A preliminary quantitative analysis of REM sleep chin EMG in Parkinson's disease with or without REM sleep behavior disorder. *Sleep Med*. 2012;**13**(6):707–713.
28. Ferri R, *et al*. Night-to-night variability of automatic quantitative parameters of the chin EMG amplitude (Atonia Index) in REM sleep behavior disorder. *J Clin Sleep Med*. 2013;**9**(3):253–258.
29. Frauscher B, *et al*. Validation of an integrated software for the detection of rapid eye movement sleep behavior disorder. *Sleep*. 2014;**37**(10):1663–1671.
30. Khalil A, *et al*. Loss of rapid eye movement sleep atonia in patients with REM sleep behavioral disorder, narcolepsy, and isolated loss of REM atonia. *J Clin Sleep Med*. 2013;**9**(10):1039–1048.
31. Kim J, *et al*. Quantification of rapid eye movement sleep without atonia to diagnose rapid eye movement sleep behavior disorder: a retrospective and case-control study. *J Sleep Med*. 2020;**17**(1):31–36.
32. Lee J-H, , *et al*. Quantification of REM sleep without atonia in Korean REM sleep behavior disorder patients: comparison of manual and computer-assisted scoring methods. *J Korean Sleep Res Soc*. 2014;**11**(2):50–56.
33. Lee SA, Kim CS, Cho CU, Kim B, Lee GH. Quantitative EMG criteria for diagnosing idiopathic REM sleep behavior disorder. *Sleep Breath*. 2015;**19**(2):685–691.
34. Yang JH, *et al*. Automated assessment of quantitative REM sleep without atonia for diagnosis of REM sleep behavior disorder. *Chronobiol Med*. 2020;**2**(3):123–128.
35. Cesari M, *et al*. Flexor digitorum superficialis muscular activity is more reliable than mentalis muscular activity for rapid eye movement sleep without atonia quantification: a study of interrater reliability for artifact correction in the context of semiautomated scoring of rapid eye movement sleep without atonia. *Sleep*. 2021;**44**(9). doi:10.1093/sleep/zsab094.
36. Ferri R, *et al*. A quantitative analysis of the submentalis muscle electromyographic amplitude during rapid eye movement sleep across the lifespan. *J Sleep Res*. 2012;**21**(3):257–263.
37. McCarter SJ, *et al*. Greatest rapid eye movement sleep atonia loss in men and older age. *Ann Clin Transl Neurol*. 2014;**1**(9):733–738.
38. Figorilli M, *et al*. Comparison between automatic and visual scorings of REM sleep without atonia for the diagnosis of

REM sleep behavior disorder in Parkinson disease. *Sleep*. 2017;**40**(2). doi:10.1093/sleep/zsw060.

39. Burns JW, *et al*. EMG variance during polysomnography as an assessment for REM sleep behavior disorder. *Sleep*. 2007;**30**(12):1771–1778. doi:10.1093/sleep/30.12.1771.
40. Cesari M, *et al*. External validation of a data-driven algorithm for muscular activity identification during sleep. *J Sleep Res*. 2019;**28**(6):e12868.
41. Frandsen R, *et al*. Analysis of automated quantification of motor activity in REM sleep behaviour disorder. *J Sleep Res*. 2015;**24**(5):583–590.
42. Nepozitek J, *et al*. Simultaneous tonic and phasic REM sleep without atonia best predicts early phenoconversion to neurodegenerative disease in idiopathic REM sleep behavior disorder. *Sleep*. 2019;**42**(9). doi:10.1093/sleep/zsz132.