STEVEN J. KAMPER, PhD[1,2]

Sample Size: Linking Evidence to Practice

J Orthop Sports Phys Ther 2022;52(8):563-564. doi:10.2519/jospt.2022.0702

Most studies include data from a sample of people because it is not possible to include the entire population in the study. The way the sample is recruited into the study[6] has implications for the generalizability of the findings.[4] The sample size (number of people from the population recruited into a study) also has implications for how you interpret the results. Sample sizes that are too small can obscure important associations or differences. When sample sizes are too large, the risk is that statistically significant findings are clinically irrelevant. Perhaps most importantly, small studies are at higher risk of bias.

Statistical Power

A priori power calculations help researchers balance the risks of sample sizes that are too small or too large by identifying a sample size that will give a high probability of identifying an important effect, if one exists. In the musculoskeletal rehabilitation field, this probability is typically set at 80% or 90%, while allowing for a 5% probability of a false positive result. A priori power calculations reported with the study can help readers assess whether a study might be underpowered or overpowered.

Why Does Sample Size Matter?

A small study carries more risk that the researchers' conclusions are inaccurate (**TABLE**). Sample size directly impacts the precision of effect estimates[3] and measures of statistical significance (*P* values).[5]

High-quality treatment studies produce a point estimate, usually a mean difference or odds/risk ratio, with a confidence interval. Although technically a little more complex, you can interpret the results such that the true effect of treatment could plausibly lie anywhere within the confidence interval[3]. The smaller the sample, the wider the confidence interval, and the less certain you are about the true treatment effect. Confidence intervals from a small sample often span large and negligible effects, which means it is uncertain whether a treatment is useful or not.

Researchers commonly conduct hypothesis tests to determine treatment effectiveness based on whether *P* values are below 0.05. Limitations of this approach notwithstanding,[2] a small sample will result in a larger *P* value regardless of how effective the treatment is; a statistically significant finding ($P<0.05$) does not mean it is 95% likely that an effect is real. When sample sizes are small, *P* values can be especially unstable. This means that the same study conducted on a different sample will almost certainly produce a different *P* value, which is one reason why the confidence intervals are more informative about potential treatment effects than the *P* values.[3]

Sampling Variability and Effect Sizes Sampling variability refers to the fact that two samples drawn from the same population will not look the same; there will be differences in the characteristics. Larger sample sizes help minimize sampling variability. For example, if the population mean weight is 72 kg with a range of 45-185 kg, it is quite possible that a small sample from the population could have a mean of 65 kg and a different small sample, a mean of 80 kg. If weight was a relevant factor to the study question, then the results that came from these two studies could be quite different. Put another way, smaller samples are less likely to be representative and generalizable to the population as a whole, so study results reflect a feature of the sample only, rather than the population.

Study Inflation Effect Underpowered studies (ie, those with too small sample sizes) that find a statistically significant effect are more likely to report an inflated effect size.[1] This is called the study inflation effect, and it is more serious the smaller the study (ie, the smaller study, the more exaggerated the effect). The implications are critical: if you read a small study that reports an effect, it is likely that the real (population) effect is smaller than in the study.

In studies that investigate predictive questions, the small sample problem often leads to model "overfitting": the prediction model "fits" the sample but does not necessarily represent the population. This means that predictive models developed from small samples are not useful in practice. Studies that report associations between predictive variables and

[1]School of Health Sciences, The University of Sydney, Sydney, Australia. [2]Nepean Blue Mountains Local Health District, Penrith, Australia.

TABLE Problems With Small Studies

Problem	Implication
• Sampling variability	• The results do not generalize well to the population.
• Study inflation effect	• The study effects are likely to be exaggerated.
• Poor precision (wide confidence intervals)	• Cannot conclude whether effects are clinically meaningful.
• Low power	• Conclusions based on hypothesis testing (*P* value) are unreliable.
• Publication bias	• Skew the body of evidence toward showing an effect.

outcome are complex—sample size is but one important consideration for readers.

Publication Bias Studies that show an effect or a significant association are more likely to be published than studies that do not. This is called publication bias, and it is especially the case for small studies. For every small study published with evidence of an effect, there are likely other studies addressing the same question that show no effect but have not been published. The small study you read may only be part of the picture.

Pilot and Feasibility Studies

An underpowered study with a small sample is not the same as a pilot or feasibility study (for this article, consider pilot and feasibility studies interchangeably). Pilot studies are not designed to test or estimate the effectiveness of a treatment, nor should you interpret the results as such. They are designed to prepare for a future definitive study that addresses the research question. Objectives consistent with pilot studies include testing the feasibility and acceptability of the data collection processes, estimating recruitment rates, and checking adherence to the intervention(s). Pilot studies are not suitable for answering questions about treatment effectiveness.

Regardless whether a study shows an effect or not, the results are less reliable if the sample is small. Readers should have lower confidence in results from small studies because of sampling variability, study inflation effects, poor precision, and low power. Small studies are at risk of publication bias and often incorrectly labelled as pilot or feasibility studies. ◉

STUDY DETAILS

AUTHOR CONTRIBUTIONS: Steven J. Kamper drafted and revised the manuscript.
DATA SHARING: There are no data associated with this article.
PATIENT AND PUBLIC INVOLVEMENT: There was no patient consultation involved in this article.

REFERENCES

1. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of *p*-values. *Royal Soc Open Sci*. 2014;1:140216. https://doi.org/10.1098/rsos.140216
2. Herbert R. Research Note: Significance testing and hypothesis testing: meaningless, misleading and mostly unnecessary. *J Physiother*. 2019;65:178-181. https://doi.org/10.1016/j.jphys.2019.05.001
3. Kamper SJ. Confidence intervals: linking evidence to practice. *J Orthop Sports Phys Ther*. 2019;49:763-764. https://doi.org/10.2519/jospt.2019.0706
4. Kamper SJ. Generalizability: linking evidence to practice. *J Orthop Sports Phys Ther*. 2020;50:45-46. https://doi.org/10.2519/jospt.2020.0701
5. Kamper SJ. Interpreting outcome 2–statistical significance and clinical meaningfulness: linking evidence to practice. *J Orthop Sports Phys Ther*. 2019;49:559-560. https://doi.org/10.2519/jospt.2019.0704
6. Kamper SJ. Sampling: linking evidence to practice. *J Orthop Sports Phys Ther*. 2020;50:725-726. https://doi.org/10.2519/jospt.2020.0704

SERGIO NÚÑEZ DE ARENAS-ARROYO, MSc[1] • IVÁN CAVERO-REDONDO, PhD[1,2] • ANA TORRES-COSTOSO, PhD[3]
SARA REINA-GUTIÉRREZ, MSc[1] • MARÍA JOSÉ GUZMÁN-PAVÓN, PhD[3] • VICENTE MARTÍNEZ-VIZCAÍNO, PhD[1,4]

Accuracy of the Most Common Provocation Tests for Diagnosing Carpal Tunnel Syndrome: A Systematic Review With Meta-analysis

Carpal tunnel syndrome (CTS) is the most common entrapment mononeuropathy caused by compression of the median nerve at the wrist as it passes through a space-limited osteofibrous canal.[41] The prevalence of CTS ranges from 5% (in the general population) to 21% (often in the working population) and costs in excess of $2 billion annually for medical care. The nonmedical cost can be substantially greater.[12]

OBJECTIVE: To estimate the screening performances of the most important provocation tests for diagnosing carpal tunnel syndrome (CTS).

DESIGN: Diagnostic test accuracy systematic review with meta-analysis.

LITERATURE SEARCH: We systematically searched the MEDLINE, Scopus, Web of Science, and Cochrane databases from inception to November 2020.

STUDY SELECTION CRITERIA: Observational studies comparing the accuracies of the Durkan test (DT), the hand elevation test (HET), the Phalen test (PT), the Tinel test (TT), and the upper-limb neurodynamic test specific to the median nerve (ULNT1) with electrodiagnosis for screening for CTS.

DATA SYNTHESIS: Random-effects models for the diagnostic odds ratio (dOR) values computed by Moses' constant for a linear model and 95% confidence intervals (CIs) were used to calculate the accuracy of these tests. Hierarchical summary receiver operating characteristic curve analyses were used to summarize the overall test performance.

RESULTS: Thirty-seven studies were included in the meta-analysis, with a total sample of 2662 wrists for DT, 864 wrists for HET, 6361 wrists for PT, 6094 wrists for TT, and 571 wrists for ULNT1. The pooled dORs for screening for CTS were 15.84 (95% CI: 3.78, 66.38) for DT, 128.63 (95% CI: 40.64, 407.12) for HET, 7.23 (95% CI: 4.06, 12.86) for PT, 5.31 (95% CI: 3.49, 8.09) for TT, and 1.78 (95% CI: 0.61, 5.19) for ULNT1.

CONCLUSION: HET has the best clinical performance for detecting CTS and should be considered the first screening test of choice during the physical examination. The most common tests (DT, PT, and TT) have good accuracies for screening for CTS.
J Orthop Sports Phys Ther 2022;52(8):522-531. Epub: 19 June 2022. doi:10.2519/jospt.2022.10828

KEY WORDS: *carpal tunnel syndrome, diagnostic accuracy studies, meta-analysis, nerve compression syndromes*

CTS is characterized by positive and negative symptoms and signs. Positive signs (gain of function) include neuropathic pain (pain caused by a disease or an injury to the somatosensory nervous system[48,50]), nocturnal paresthesia, and dysesthesia. Negative signs of CTS included loss of sensation, weakness, and thenar muscle atrophy.[41,48] Severe CTS cases could produce serious physical, psychological, and economic consequences.[41]

The gold standard for confirming a diagnosis of CTS is electrodiagnosis.[40] However, clinical assessment has been used as the initial diagnosis step in the American Academy of Orthopaedic Surgeons recommendations.[25] Clinical assessment includes personal characteristics, observation of deformities and range of motion, muscle strength evaluation, hand diagram, sensory examination, and provocation tests. In patients in which clinical or provocation tests are positive, the diagnosis should be confirmed by electrodiagnostic procedures.[25]

The Durkan test (DT), the hand elevation test (HET), the Phalen test (PT), the Tinel test (TT), and the upper-limb

[1]Health and Social Research Center, Universidad de Castilla-La Mancha, Cuenca, Spain. [2]Rehabilitation and Health Research Center (CIRES), Universidad de las Américas, Santiago, Chile. [3]Facultad de Fisioterapia y Enfermería, Universidad de Castilla-La Mancha, Toledo, Spain. [4]Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Talca, Chile. This review was registered with PROSPERO (registration number CRD42021237602). This work was supported by the European Regional Development Fund. Sergio Núñez de Arenas-Arroyo (2020-PREDUCLM-16704) and Sara Reina-Gutiérrez (2020-PREDUCLM-15596) are supported by a grant from the University of Castilla-La Mancha. The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Dr Iván Cavero-Redondo, Universidad de Castilla-La Mancha, Edificio Melchor Cano, Centro de Estudios Sociosanitarios, Santa Teresa Jornet s/n, 16071 Cuenca, Spain. E-mail: Ivan.Cavero@uclm.es

neurodynamic test specific to the median nerve (ULNT1) are examples of provocation tests used in the physical examination when diagnosing CTS. Provocation tests detect nerve mechanosensitivity, axonal hyperactivity, axonal regeneration, or vascular compromise by increasing internal pressure in the carpal tunnel by manual mechanical compression, reduced blood supply nerve, or movement or tension on neural structures.[1,2,15,48] The tests are easy to perform, quick, reproducible, inexpensive, and noninvasive and require less skilled examiners than invasive options.

DT is a manual compression test of the carpal tunnel held for up to 30 seconds and recorded as positive if associated with the reproduction of distal paresthesias in the median nerve distribution.[15,65] HET requires the patient to raise both hands above the head for 1 to 2 minutes; if symptoms reproduce, the test is considered positive.[1,31] PT is a maximum wrist flexion held for up to 60 seconds and recorded as positive if associated with the reproduction of distal paresthesias in the median nerve distribution.[43,65] TT is a percussion over the median nerve just proximal to the wrist crease and recorded as positive if associated with distally radiated paresthesias in the median nerve distribution.[65] Finally, ULNT1 is a neurodynamic test that begins with 90° abduction and 90° external rotation of the shoulder, 90° flexion of the elbow, supination of the forearm, maximum extension of the wrist and fingers, and abduction of the thumb. Next, 1 of the physical therapist's hands prevents scapular elevation, and the other hand maintains finger abduction while the elbow is slowly extended to the point of tolerance. The test is positive if the patient has at least 1 of the following items: (1) feels reproduction of symptoms, (2) range of motion is limited at least 10° in elbow extension, or (3) contralateral neck side-bending resulted in an increase of symptoms, or ipsilateral side-bending resulted in a decrease of symptoms.[55,58]

There is no consensus for the best provocation test for diagnosing CTS. Although electrodiagnosis has traditionally been proposed as the gold standard, a clinical provocation test is needed to improve the diagnostic process and reduce medical costs and waiting lists. Previous systematic reviews[26,34,39] have tried to clarify the sensitivity and specificity of provocation tests compared to the electrodiagnosis in CTS screening. No previous meta-analysis has estimated the diagnostic performances (diagnostic odds ratio [dOR], sensitivity, specificity, positive likelihood ratio [PLR], and negative likelihood ratio [NLR]) of the most important provocation tests using electrodiagnostic methods by calculating hierarchical summary receiver operating characteristic (HSROC) models. These models are currently considered the most rigorous multivariate meta-analysis approach.[20]

The aim of this systematic review with meta-analysis was to estimate the diagnostic performance of the most important provocation tests for the diagnosis of CTS using HSROC analysis.

METHODS

This systematic review was reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement,[42] the PRISMA diagnostic test accuracy (PRISMA-DTA) extension,[36] and the recommendations of the *Cochrane Handbook for Systematic Reviews of Diagnostic*

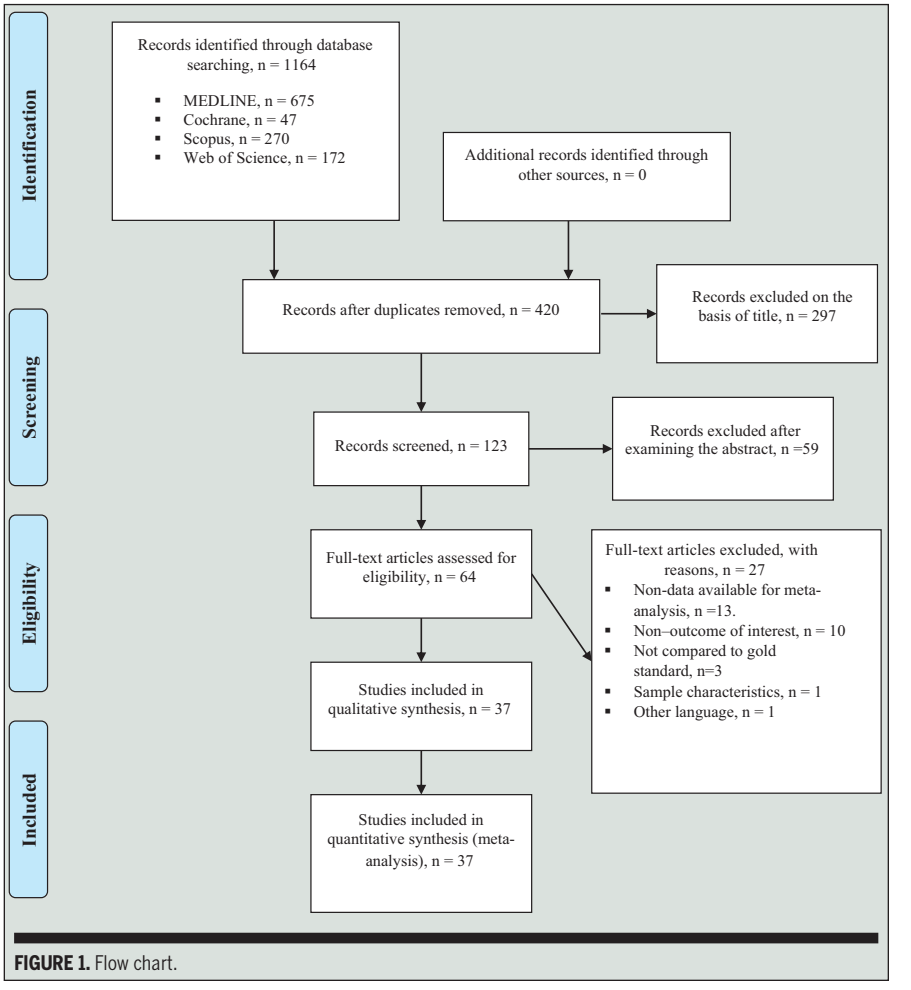

FIGURE 1. Flow chart.

TABLE 1 Characteristics of the Studies Included in the Meta-analysis[a]

Author	Country	Test	Age	n (% Female)	n (Wrists)	Prevalence %	Sens (%)	Spec (%)	dOR
Amirfeyz et al, 2011	UK	DT	58.5 (26-91)[b]	103 (63.2)	163	57.0	84.3	78.6	19.07
Boland & Kiernan, 2009	Australia	DT	49.5 (NR)	64 (74.4)	86	86.0	14.0	96.0	1.72
Durkan, 1991	USA	DT	45.0 (NR)	23 (74.2)	96	47.9	87.0	90.0	60.00
El Miedany et al, 2008	Egypt	DT	20-91	284 (68.6)	414	56.0	46.0	25.0	0.29
Fertl et al, 1998	Austria	DT	48.75 (21-78)[b]	39 (82.9)	103	61.2	83.0	92.0	184.36
Kuhlman & Hennessey, 1997	USA	DT	NR	NR	228	62.3	28.0	74.0	1.14
Küçükakkaş & Yurdakul, 2019	Turkey	DT	46.7 (12.7)	367 (79.3)	512	44.1	96.0	67.0	49.15
Ma & Kim, 2012	Korea	DT	56.8 (48-64)[b]	76 (84.5)	90	50.0	84.4	82.2	25.11
Mondelli et al, 2001	Italy	DT	57.9 (15.3)	237 (72.7)	326	54.9	42.0	99.0	105.29
Richter & Brüser, 1999	Germany	DT	52.0	77 (71.3)	108	54.0	87.0	96.0	174.57
Sadanandan & Rijesh, 2017	India	DT	44.99 (28-64)[b]	71 (87.3)	73	97.3	42.0	99.0	1.46
Szabo et al, 1999	USA	DT	18-73	71 (71)	187	46.5	89.0	91.0	77.86
Tetro et al, 1998	USA	DT	48.1 (21.5-83.9)[b]	36 (31.6)	191	49.7	75.0	93.0	37.61
Zhang et al, 2020	USA	DT	59 (13)	39 (70.9)	85	89.4	71.0	22.0	0.70
Ahn, 2001	Korea	HET	55 (26-87)[b]	200 (100)	400	50.0	75.5	98.5	202.36
Amirfeyz et al, 2005	UK	HET	56.5 (23-94)[b]	63 (66.3)	95	50.5	88.0	98.0	322.00
Amirfeyz et al, 2011	UK	HET	58.5 (26-91)[b]	103 (63.2)	163	57.0	98.6	91.4	981.33
Kasundra et al, 2015	India	HET	43.9 (14)	50 (86.2)	116	80	84.9	82.6	29.23
Ma & Kim, 2012	Korea	HET	56.8 (48-64)[b]	76 (84.5)	90	50.0	86.7	88.9	52.00
Ahn, 2001	Korea	PT	55 (26, 87)[b]	200 (100)	400	50	68.0	91.0	21.00
Amirfeyz et al, 2005	UK	PT	56.5 (23, 94)[b]	63 (66.3)	95	50.5	83.0	98.0	230.00
Amirfeyz et al, 2011	UK	PT	58.5 (26-91)[b]	103 (63.2)	163	57.0	87.1	84.3	36.20
Boland & Kiernan, 2009	Australia	PT	49.5 (NR)	64 (74.4)	86	86.0	64.0	75.0	5.22
Brüske et al, 2002	Poland	PT	53 (21-82)[b]	89 (80)	247	59.5	85.0	89.0	48.55
Buch-Jaeger & Foucher, 1994	France	PT	52 (29-81)[b]	90 (80.3)	172	61.0	58.0	54.0	1.61
Campos-Serna et al, 2020	Mexico	PT	57.15 (45.2-65)[b]	546 (84.0)	650	66.61	65.82	49.77	1.91
Chiquete et al, 2011	Mexico	PT	45.4 (NR)	63 (87.5)	72	55.5	66.7	73.3	5.31
Durkan, 1991	USA	PT	45.0 (NR)	23 (74.2)	96	47.9	70.0	84.0	12.00
El Miedany et al, 2008	Egypt	PT	20-91	284 (68.6)	414	56.0	47.0	17.0	0.18
Fertl et al, 1998	Austria	PT	48.75 (21-78)[b]	39 (82.9)	103	61.2	79.0	92.0	47.44
González-Roig et al, 2008	Cuba	PT	43.38 (13.2)	226 (93.0)	243	75.3	68.2	83.5	0.43
Hansen et al, 2004	USA	PT	46.6 (17-75)[b]	82 (58)	142	67.0	34.0	74.0	1.48
Hegmann et al, 2018	USA	PT	42.2 (11.4)	790 (66.2)	918	11.5	52.8	87.7	7.97
Heller et al, 2008	Israel	PT	55 (29-78)[b]	49 (81.6)	80	72.5	67.0	59.0	2.96
Kuhlman & Hennessey, 1997	USA	PT	NR	NR	228	62.3	51.0	76.0	3.18
Küçükakkaş & Yurdakul, 2019	Turkey	PT	46.7 (12.7)	367 (79.3)	512	44.1	86.0	57.0	8.16
Ma & Kim, 2012	Korea	PT	56.8 (48-64)[b]	76 (84.5)	90	50.0	84.4	86.7	35.29
MacDermid et al, 1997	Canada	PT	42 (NR)	NR	84	42.8	86.5	88.0	43.40
Mondelli et al, 2001	Italy	PT	57.9 (15.3)	237 (72.7)	326	54.9	59.0	93.0	19.44
Naranjo et al, 2007	Spain	PT	47 (11)	56 (82.3)	105	76.2	76.7	30.4	1.51
Richter & Brüser, 1999	Germany	PT	52.0	77 (71.3)	108	54.0	85.0	98.0	304.75
Sadanandan & Rijesh, 2017	India	PT	44.99 (28-64)[b]	71 (87.3)	73	97.3	59.0	86.0	2.90
Sawaya & Sakr, 2009	USA	PT	30-78	22 (82)	54	50.0	52.0	52.0	1.16
Kasundra et al, 2015	India	PT	43.9 (14)	50 (86.2)	116	80	84.9	73.9	15.99
Szabo et al, 1999	USA	PT	18-73	71 (71)	187	46.5	75.0	95.0	56.14

Table continues on next page.

TABLE 1 Characteristics of the Studies Included in the Meta-analysis[a] (continued)

Author	Country	Test	Age	n (% Female)	n (Wrists)	Prevalence %	Sens (%)	Spec (%)	dOR
Tetro et al, 1998	USA	PT	48.1 (21.5-83.9)[b]	36 (31.6)	191	49.7	61	83	7.84
Wainner et al, 2005	USA	PT	45.7 (10.6)	41 (50.0)	118	32.2	77.0	40.0	2.15
Walters & Rice, 2002	USA	PT	42.5 (24-61)[b]	54 (70.0)	77	61.03	85.0	50.0	5.71
Widodo et al, 2020	Indonesia	PT	51.7 (45-59)[b]	19 (100)	19	73.68	0.82	100	9.33
Wiesman et al, 2003	USA	PT	46.5 (10)	33 (55.9)	107	43.92	85.0	90.0	51.43
Zhang et al, 2020	USA	PT	59 (13)	39 (70.9)	85	89.4	50	33	0.50
Ahn, 2001	Korea	TT	55 (26-87)[b]	200 (100)	400	50	68.0	90.0	18.69
Amirfeyz et al, 2005	UK	TT	56.5 (23, 94)[b]	63 (66.3)	95	50.5	48.0	94.0	13.49
Amirfeyz et al, 2011	UK	TT	58.5 (26-91)[b]	103 (63.2)	163	57.0	92.9	64.3	14.48
Brüske et al, 2002	Poland	TT	53 (21-82)[b]	89 (80)	247	59.5	67.0	68.0	4.38
Buch-Jaeger & Foucher, 1994	France	TT	52 (29-81)[b]	90 (80.3)	172	61.0	42.0	63.0	1.21
Campos-Serna et al, 2020	Mexico	TT	57.15 (45.2-65)[b]	546 (84.0)	650	66.61	71.6	44.2	2.00
Chiquete et al, 2011	Mexico	TT	45.4 (NR)	63 (87.5)	72	55.5	72.2	66.7	5.03
Durkan, 1991		TT	45.0 (NR)	23 (74.2)	96	47.9	56.0	80.0	5.20
El Miedany et al, 2008	Egypt	TT	20-91	284 (68.6)	414	56.0	30.0	65.0	0.80
González-Roig et al, 2008	Cuba	TT	43.38 (13.2)	226 (93.0)	243	75.3	52.3	88.6	8.20
Hansen et al, 2004	USA	TT	46.6 (17-75)[b]	82 (58)	142	67	27.0	91.0	4.05
Hegmann et al, 2018	USA	TT	42.2 (11.4)	790 (66.2)	918	11.5	38.0	90.0	5.47
Heller et al, 2008	Israel	TT	55 (29-78)[b]	49 (81.6)	80	72.5	60.0	77.0	5.17
Kuhlman & Hennessey, 1997	USA	TT	NR	NR	228	62.3	23.0	87.0	2.06
Küçükakkaş & Yurdakul, 2019	Turkey	TT	46.7 (12.7)	367 (79.3)	512	44.1	89.0	41.0	5.57
Ma & Kim, 2012	Korea	TT	56.8 (48-64)[b]	76 (84.5)	90	50.0	82.2	88.9	37.00
MacDermid et al, 1997	Canada	TT	42 (NR)	NR	84	42.8	50.0	93.0	15.00
Mondelli et al, 2001	Italy	TT	57.9 (15.3)	237 (72.7)	326	54.9	41.0	90.0	6.20
Naranjo et al, 2007	Spain	TT	47 (11)	56 (82.3)	105	76.2	73.6	40.0	1.87
Sadanandan & Rijesh, 2017	India	TT	44.99 (28-64)[b]	71 (87.3)	73	97.3	41.0	90.0	1.38
Kasundra et al, 2015	India	TT	43.9 (14)	50 (86.2)	116	80	78.5	91.3	38.33
Stewart & Eisen, 1978	Canada	TT	54 (30-84)[b]	82 (80)	103	49.51	45.0	71.0	2.03
Szabo et al, 1999	USA	TT	18-73	71 (71)	187	46.5	64.0	99.0	178.84
Tetro et al, 1998	USA	TT	48.1 (21.5-83.9)[b]	36 (31.6)	191	49.7	74.0	91.0	27.07
Wainner et al, 2005	USA	TT	45.7 (10.6)	41 (50)	118	32.2	48.0	67.0	1.87
Walters & Rice, 2002	USA	TT	42.5 (24-61)[b]	54 (70)	77	61.03	63.8	40.0	1.18
Wiesman et al, 2003	USA	TT	46.5 (10)	33 (55.9)	107	43.9	88.0	76.0	22.56
Zhang et al, 2020	USA	TT	59 (13)	39 (70.9)	85	89.4	47	56	1.13
Bueno-Gracia et al, 2016	Spain	ULNT1	54.3 (14.5)	58 (72.4)	95	60.0	58.0	84.0	7.33
Trillos et al, 2018	Colombia	ULNT1	50.51 (11.1)	120 (83.1)	230	86.9	93.0	6.0	0.95
Vanti et al, 2011	Italy	ULNT1	46.3 (10.8)	44 (75.0)	44	54.5	91.7	15.0	1.94
Vanti et al, 2012	Italy	ULNT1	45.91 (10.66)	35 (74.5)	84	41.66	40.0	79.0	2.60
Wainner et al, 2005	USA	ULNT1	45.7 (10.6)	41 (50)	118	32.2	75.0	13.0	0.46

Abbreviations: dOR, diagnostic odds ratio; DT, Durkan test; HET, hand elevation test; NR, not reported; PT, Phalen test; Sens, sensitivity; Spec, specificity; TT, Tinel test; ULNT1, upper-limb neurodynamic test specific to the median nerve.
[a]*Values are presented as mean ffl SD unless otherwise indicated.*
[b]*Values are mean (range).*

Test Accuracy.[32] It was prospectively registered in the PROSPERO database (registration number CRD42021237602).

Data Sources and Searches

PubMed (via MEDLINE), Scopus, Web of Science, and the Cochrane databases were systematically searched from inception to March 2021, using the following search strategy: ("carpal tunnel

syndrome" OR CTS) AND ("upper limb neurodynamic test" OR "upper-limb neurodynamic test" OR "median nerve test" OR ULNT OR ULNT1 OR "Tinel test" OR "Tinel sign" OR Tinel OR "Phalen test" OR "Phalen sign" OR Phalen OR "Hand elevation test" OR HET OR "Durkan test" OR "Durkan sign" OR "Hand compression test" OR Durkan) AND (threshold OR cut-off OR "cut off" OR "cut point" OR sensitivity OR specificity OR diagnostic OR diagnosis OR "differential diagnosis" OR prognostic).

Study Selection

Eligible articles were original studies measuring provocation test thresholds and their association with the diagnosis of CTS, ie, the performance of CTS diagnosis. The inclusion criteria were as follows: (1) study participants: individuals aged 18 years and older; (2) index tests used: DT, HET, PT, TT, or ULNT1; (3) reference standard: electrodiagnosis; (4) outcome: CTS diagnosis; and (5) study design: cross-sectional, case-control, or cohort studies, with either prospective or retrospective data collection. Studies were excluded if they reported insufficient data for a 2 × 2 table.

The literature search, screening, and trial selection were conducted independently by 2 reviewers (S.N.A.-A. and I.C.-R.). When there were disagreements, a third researcher made the final decision (V.M.-V.).

Data Extraction and Risk-of-Bias Assessment

Data were extracted independently by 2 reviewers (S.N.A.-A. and S.R.-G.). When there were disagreements, a third researcher made the final decision (V.M.-V.). The following data were extracted from each included study: (1) author identification and year of publication, (2) country of study, (3) provocation test used, (4) age of participants, (5) number of participants, (6) number of wrists, (7) prevalence of CTS, and (8) parameters summarizing the accuracy of the test (sensitivity, specificity, and dOR) per wrist. Trial authors were contacted up to 3 times to retrieve missing information.

Two reviewers (S.N.A.-A. and A.T.-C.) independently assessed the risk of bias for each study included using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.[60] Any discrepancies were resolved by consensus; a third reviewer (V.M.-V.) resolved any discrepancies if consensus could not be reached. This tool assesses 4 domains: (1) patient selection, (2) index test, (3) reference standard, and (4) flow of patients and timing of the tests. Each domain was evaluated as having an unclear, a low, or a high risk of bias. The QUADAS-2 tool also assesses the applicability of the results with respect to patient selection, the index test, and the reference standard.

Statistical Analysis and Data Synthesis

We calculated sensitivity, specificity, PLR, NLR, and dOR as well as their corresponding 95% confidence intervals (CIs) for the provocation tests. HSROC curves summarized the overall test performance as multivariate methods that jointly analyze sensitivity and specificity. These curves have been proposed to be able to estimate the diagnostic performance of tests in meta-analyses, where the prediction region is useful for estimating the magnitude of heterogeneity such that wider prediction regions suggest greater heterogeneity.[29]

The dOR is a measure of the accuracy of a diagnostic test that combines sensitivity and specificity into a single value. The value ranges from zero to infinity, with a value of 1 corresponding to zero diagnostic ability and a higher value corresponding to better discriminatory test performance. The dOR was computed using Moses' constant from a linear model. This approach is based on the linear regression of the logarithm of the dOR of a study as a dependent variable and the expression of the positivity threshold of a study as an independent variable.[44]

The DerSimonian and Laird[14] method was used to compute the pooled estimates of dOR for each included study. The heterogeneity of the results across studies was assessed using the I^2 statistical parameter and the corresponding *P* values. I^2 values were considered as follows: might not be important (0%-30%), may represent moderate heterogeneity (30%-50%), substantial heterogeneity (50%-75%), and considerable heterogeneity (75%-100%).[23]

TABLE 2 Pooled Accuracy Parameters in the Diagnosis of Carpal Tunnel Syndrome by Test[a]

Test	n	Sensitivity (%)	Specificity (%)	PLR	NLR	dOR
Durkan test	14	62.00 (49.00, 77.00)	77.00 (63.00, 93.00)	6.86 (3.45, 13.62)	0.25 (0.13, 0.49)	15.84 (3.78, 66.38)
Hand elevation test	5	85.00 (77.00, 94.00)	95.00 (86.00, 105.00)	15.93 (6.02, 42.18)	0.10 (0.04, 0.27)	128.63 (40.64, 407.12)
Phalen test	32	68.00 (63.00, 74.00)	71.00 (63.00, 80.00)	4.39 (2.78, 6.93)	0.32 (0.21, 0.51)	7.23 (4.06, 12.86)
Tinel test	28	55.00 (48.00, 63.00)	75.00 (67.00, 83.00)	3.61 (2.13, 6.11)	0.48 (0.30, 0.79)	5.31 (3.49, 8.09)
ULNT1	5	71.00 (54.00, 94.00)	28.00 (12.00, 65.00)	1.46 (0.41, 5.21)	0.68 (0.14, 3.39)	1.78 (0.61, 5.19)

Abbreviations: dOR, diagnostic odds ratio; NLR, negative likelihood ratio; PLR, positive likelihood ratio; ULNT1, upper-limb neurodynamic test specific to the median nerve.
[a]*Values in parentheses are 95% confidence intervals.*

Sensitivity analyses were performed using 2 approaches: (1) repeating all the analyses using only studies that directly compared 2 or more tests and (2) estimating the individual influence of each study on the pooled dOR by removing studies one by one. Publication bias was evaluated by a visual examination of the funnel plots and through Deeks' method, with $P<.10$ considered statistically significant.[13]

All statistical analyses were performed using Stata SE software (Version 16; StataCorp LLC, College Station, TX).

Protocol Deviations: Unplanned Analyses

In the initial PROSPERO registry, we planned to assess risk of bias using QUADAS-2. In studies where 2 or more tests were directly compared, we decided to include the risk-of-bias assessment using the QUADAS-C,[64] an extension of QUADAS-2, that was developed to assess the risk of bias in comparative diagnostic accuracy studies. A sensitivity analysis was performed, excluding the studies with high risk of bias, due to

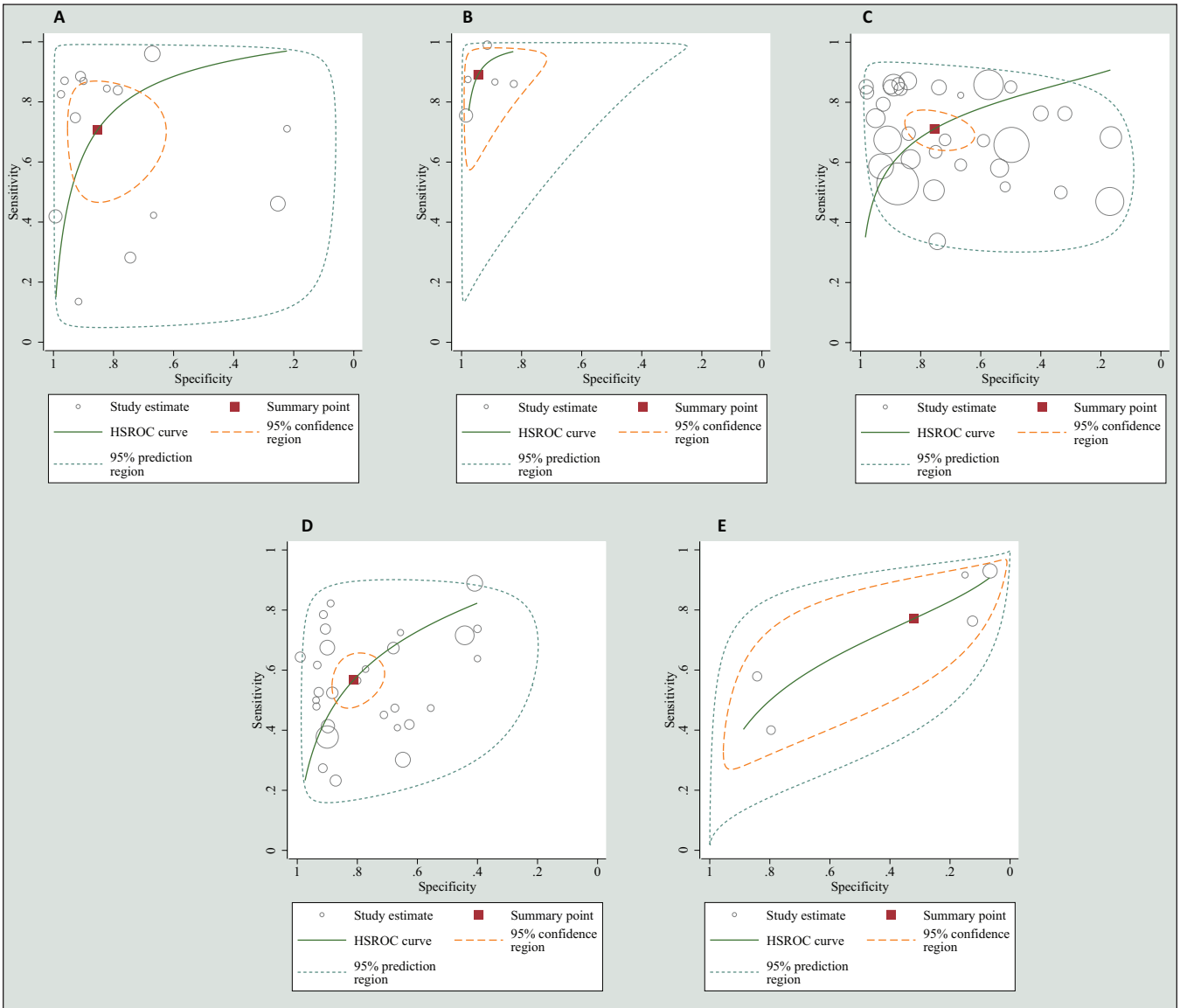

FIGURE 2. HSROC curve for the Durkan test (A), hand elevation test (B), Phalen test (C), Tinel test (D), and ULNT1 test (E). HSROC curves were used to summarize overall test performance as multivariate methods that jointly analyze sensitivity and specificity. These curves have been proposed to be able to estimate the diagnostic performance of tests in meta-analyses, where the prediction region is useful for estimating the magnitude of heterogeneity such that wider prediction regions suggest greater heterogeneity. Abbreviations: HSROC, hierarchical summary receiver operating characteristic; ULNT1, upper-limb neurodynamic test specific to the median nerve.

some concerns about bias in the included studies, particularly with the index test and with participant selection.

RESULTS

Baseline Characteristics

The literature search yielded 1164 articles. After removing duplicates, the titles and abstracts of 420 articles were screened. Following the full-text reviews, 37 studies were included[1,3,4,6-11,15-19,21,22,24,27,28,31,33,37,38,45-47,51-53,55-59,61,62,65] (**FIGURE 1**) (the reasons for excluding studies are shown in **SUPPLEMENTAL TABLE 1**). Regarding the different tests examined in the studies included, 14 studies provided information about DT,[3,6,15-17,27,28,31,37,45,46,52,53,65] 5 studies provided information about HET,[1,3,4,24,31] 31 studies provided information about PT,[1,3,4,6-8,10,11,15-19,21,22,24,27,28,31,33,37,38,45-47,52,53,58,59,61,62,65] 28 studies provided information about TT,[1,3, 4,7,8,10,11,15,16,18,19,21,22,24,27,28,31,33,37,38,46,51-53,58,59,62,65] and 5 studies provided information about ULNT1.[9,55-58]

Most of the included studies were longitudinal designs, but 1 was a cross-sectional analysis from a longitudinal study.[21] Two studies had a descriptive design.[18,46] The studies were published between 1974 and 2020 and were performed in 18 countries: Austria, Australia, Canada, Colombia, Cuba, Egypt, France, Germany, India, Israel, Italy, Korea, Mexico, Poland, Spain, Turkey, the United Kingdom, and the United States.

There was a total sample of 2662 wrists for DT, 864 wrists for HET, 6361 wrists for PT, 6094 wrists for TT, and 571 wrists for ULNT1; mean participant age ranged from 42 to 59 years; CTS prevalence ranged from 11.5% to 97.3% (**TABLE 1**). The analysis and the 2 × 2 data (TP, FP, FN, TN) of each study are shown in **SUPPLEMENTAL TABLES 2 TO 6**.

Risk of Bias

Most studies had shortcomings in the index test and patient selection domains (43% and 41%, respectively). Over a third of the studies had shortcomings in the index test domain (**SUPPLEMENTAL FIGURES 1 AND 2**). Most studies had shortcomings in the index test and flow and timing domains (82.75% and 62.06%, respectively).

Meta-analysis

The pooled dORs for CTS were 15.84 (95% CI: 3.78, 66.38) for DT, 128.63 (95% CI: 40.64, 407.12) for HET, 7.23 (95% CI: 4.06, 12.86) for PT, 5.31 (95% CI: 3.49, 8.09) for TT, and 1.78 (95% CI: 0.61, 5.19) for ULNT1. There was substantial heterogeneity in the dOR of CTS for HET (I^2 = 65.9%; P = .019) and ULNT1 (I^2 = 74.5%; P = .003). There was considerable heterogeneity in the dOR of CTS for DT (I^2 = 64.4; P = .000), PT (I^2 = 88.1; P = .000), and TT (I^2 = 88.1; P = .000). The pooled sensitivity, specificity, PLR, NLR, and dOR are presented in **TABLE 2**. The HSROC curves are displayed in **FIGURE 2**. The forest plots of each test are presented in **SUPPLEMENTAL FIGURES 3 TO 7**.

Sensitivity Analyses for the Effect of Individual Studies

The pooled dOR was not affected after removing any individual study for DT, HET, PT, TT, and ULNT1. The sensitivity analyses performed using only studies that directly compared 2 or more tests are presented in **TABLE 3**.

When the analyses were performed excluding studies with a high risk of bias, the pooled dORs for the diagnosis of CTS were 18.98 (95% CI: 1.59, 225.85) for DT, 81.78 (95% CI: 7.98, 838.23) for HET, 7.43 (95% CI: 3.36, 16.36) for PT, 6.59 (95% CI: 3.77, 11.50) for TT, and 2.82 (95% CI: 1.20, 6.64) for ULNT1.

Publication Bias

The asymmetry test using Deek's method suggested an absence of publication bias for DT (P = .830), HET (P = .570), PT (P = .370), TT (P = .500), and ULNT1 (P = .660).

DISCUSSION

Physical assessment has been routinely used as the initial step in diagnosing CTS; electrodiagnosis is the gold standard for confirming CTS diagnosis in clinical practice. Because of the high prevalence of CTS, the high cost of electrodiagnosis, and the inconvenience

TABLE 3 Sensitivity Analyses Using Studies That Compared 2 or More Tests[a]

Test	n	Sensitivity (%)	Specificity (%)	PLR	NLR	dOR
Durkan test	14	62.00 (49.00, 77.00)	77.00 (63.00, 93.00)	6.86 (3.45, 13.62)	0.25 (0.13, 0.49)	15.84 (3.78, 66.38)
Hand elevation test	5	85.00 (77.00, 94.00)	95.00 (86.00, 105.00)	15.93 (6.02, 42.18)	0.10 (0.04, 0.27)	128.63 (40.64, 407.12)
Phalen test	30	69.00 (63.00, 74.00)	71.00 (64.00, 80.00)	4.53 (2.86, 7.19)	0.32 (0.20, 0.52)	7.65 (4.22, 13.86)
Tinel test	27	55.00 (49.00, 63.00)	75.00 (67.00, 83.00)	3.70 (2.17, 6.31)	0.48 (0.29, 0.78)	5.53 (3.58, 8.53)
ULNT1	1	75.00 (52.00, 108)	0.13 (7.00, 24.00)	0.86 (0.05, 15.81)	1.92 (0.01, 481.79)	0.46 (0.17, 1.25)

Abbreviations: dOR, diagnostic odds ratio; NLR, negative likelihood ratio; PLR, positive likelihood ratio; ULNT1, upper-limb neurodynamic test specific to the median nerve.
[a]*Values in parentheses are 95% confidence intervals.*

of this technique, it is important to know which is the most useful provocation test for patients with CTS. We synthesize the evidence regarding the utility of the main provocation tests as diagnostic methods of CTS in clinical settings and provide pooled dOR and HSROC to assess the accuracy of DT, HET, PT, TT, and ULNT1 for diagnosing CTS. HET has the best accuracy for CTS diagnosis: dOR: 128.63 (95% CI: 40.64, 407.12), sensitivity: 85.00% (95% CI: 77.00, 94.00), and specificity: 95.00% (95% CI: 86.00, 105.00); DT, PT, and TT are also accurate.

Previous systematic reviews[26,34,39] provided inconclusive results due to variability in included studies. Our meta-analysis quantifies the accuracy of these tests and confirms that DT, HET, PT, and TT are useful tests. Clinical practice guidelines[35] strongly recommend against using DT, PT, TT, or ULNT1 as stand-alone provocation tests to diagnose CTS. Our findings support using DT, PT, and TT for CTS screening and provide a quantitative synthesis of the accuracy of HET.

Although DT, PT, and TT were accurate for CTS diagnosis, the large dOR of HET could be due to the reduction of blood supply to the nerve by reducing local blood pressure after the elevation of the hand above the heart and the recurrence of symptoms on the radial side of the hand.[4] Although our results showed moderate ULNT1 sensitivity (71.00; 95% CI: 54.00, 94.00), the sensitivity might decrease in severe CTS.[5,49] Due to the low dOR, the ULNT1 is unsuitable for diagnosing CTS when used in isolation. This could be because the reference standard assumes that conduction loss is always present in peripheral neuropathic pain in patients with CTS. ULNT1 detects the increased median nerve mechanosensitivity, which is associated with the increased excitability of small-diameter afferents,[39] nervous system pathways,[39] and central nervous system pathways,[39] but does not always present with conduction loss.[63] This could explain the low specificity and dOR value of this test. In addition, the unclear definition of a positive ULNT1 may also play a role.

Clinical Implications

The positive dORs for DT, HET, PT, and TT and the fact that provocative tests are low-cost, time-efficient, and noninvasive alternatives allow us to recommend these tests for diagnosing CTS in clinical practice. A correct performance of these tests by clinicians could have great advantages for the health care system, decreasing health expenditure by reducing the excessive number of nerve conduction studies in people with suspected CTS and reducing the waiting list for specific CTS rehabilitation programs.

Limitations

The wide CIs of HET and DT could be due to the large number of case-control studies used in the analyses[30]; thus, the results of these tests should be considered with caution. Not all the included studies performed the provocation tests using the same protocol in HET and ULNT1, which may bias the estimates.

Several studies were excluded because their measurements were not well defined. The number of studies reporting ULNT1 measurements is still scarce. Future studies must determine the usefulness of the test in CTS diagnosis. Electrodiagnosis has a substantial number of false-negative and false-positive results. Ultrasonography may offer a superior approach to diagnosing CTS.[54] Thus, the findings of this meta-analysis should be considered with caution.

CONCLUSION

Clinicians should consider using the HET for diagnosing CTS. The most common diagnostic tests for CTS (DT, PT, and TT) have good accuracy. Our findings update the results of previous systematic reviews and should be considered when developing future clinical practice guidelines. ◉

KEY POINTS

FINDINGS: The hand elevation test (HET) has good accuracy for diagnosing carpal tunnel syndrome (CTS) and should be considered for use in the physical examination for CTS diagnosis by clinicians as HET is highly accurate for diagnosing CTS. The most popular tests—Durkan test (DT), Phalen test, and Tinel test—have good accuracy in screening for CTS.

IMPLICATIONS: We encourage clinicians to consider using the HET in practice for diagnosing CTS given its accuracy, and we encourage guideline developers to consider including HET when developing future clinical practice guidelines.

CAUTION: The confidence intervals of HET and DT could be overestimated because a large number of case-control studies were included in the analyses. More research is needed to assess the impact of the limitations of the current systematic review and to investigate the diagnostic abilities of different provocation tests (eg, upper-limb neurodynamic test specific to the median nerve) for CTS.

STUDY DETAILS

AUTHOR CONTRIBUTIONS: All authors were involved in the conception and design of the review. Sergio Núñez de Arenas-Arroyo and Dr Cavero-Redondo developed the search strategy and performed study selection. Dr Torres-Costoso and Sara Reina-Gutiérrez extracted data from included studies. Drs Cavero-Redondo and Martínez-Vizcaíno and María José Guzmán-Pavón were involved in the data analysis. All authors were involved in the interpretation and discussion of the results. Sergio Núñez de Arenas-Arroyo drafted the manuscript, and all authors revised the draft. All authors approved the final version of the article. All authors had access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

DATA SHARING: Data are available on request.

PATIENT AND PUBLIC INVOLVEMENT: Patients were not involved in this review.

REFERENCES

1. Ahn DS. Hand elevation: a new test for carpal tunnel syndrome. *Ann Plast Surg*. 2001;46:120-124. https://doi.org/10.1097/00000637-200102000-00005
2. Alfonso MI, Dzwierzynski W. Hoffman-Tinel sign: the realities. *Phys Med Rehabil Clin N Am*. 1998;9:721-736. https://doi.org/10.1016/S1047-9651(18)30229-8
3. Amirfeyz R, Clark D, Parsons B, et al. Clinical tests for carpal tunnel syndrome in contemporary practice. *Arch Orthop Trauma Surg*. 2011;131:471-474. https://doi.org/10.1007/s00402-010-1150-z
4. Amirfeyz R, Gozzard C, Leslie IJ. Hand elevation test for assessment of carpal tunnel syndrome. *J Hand Surg*. 2005;30:361-364. https://doi.org/10.1016/J.JHSB.2005.04.007
5. Baselgia LT, Bennett DL, Silbiger RM, Schmid AB. Negative neurodynamic tests do not exclude neural dysfunction in patients with entrapment neuropathies. *Arch Phys Med Rehabil*. 2017;98:480-486. https://doi.org/10.1016/j.apmr.2016.06.019
6. Boland RA, Kiernan MC. Assessing the accuracy of a combination of clinical tests for identifying carpal tunnel syndrome. *J Clin Neurosci*. 2009;16:929-933. https://doi.org/10.1016/j.jocn.2008.09.004
7. Brüske J, Bednarski M, Grzelec H, Zyluk A. The usefulness of the Phalen test and the Hoffmann-Tinel sign in the diagnosis of carpal tunnel syndrome. *Acta Orthop Belg*. 2002;68:141-145.
8. Buch-Jaeger N, Foucher G. Correlation of clinical signs with nerve conduction tests in the diagnosis of carpal tunnel syndrome. *J Hand Surg*. 1994;19:720-724. https://doi.org/10.1016/0266-7681(94)90244-5
9. Bueno-Gracia E, Tricás-Moreno JM, Fanlo-Mazas P, et al. Validity of the Upper Limb Neurodynamic Test 1 for the diagnosis of carpal tunnel syndrome. The role of structural differentiation. *Man Ther*. 2016;22:190-195. https://doi.org/10.1016/j.math.2015.12.007
10. Campos-Serna I, Garcia Espinoza JA, Cahuana-Quispe AI, et al. Diagnosis and treatment of carpal tunnel syndrome in a tertiary care center in Mexico City. *World J Plast Surg*. 2020;9:213-218. https://doi.org/10.29252/wjps.9.2.213
11. Chiquete E, Miranda-Muñoz A, Castañeda-Moreno V, et al. Clinical prediction rule for diagnosis of the carpal tunnel syndrome. *Rev Mex de Neurocienc*. 2011;12.
12. Dale AM, Harris-Adamson C, Rempel D, et al. Prevalence and incidence of carpal tunnel syndrome in US working populations: pooled analysis of six prospective studies. *Scand J Work Environ Health*. 2013;39:495-505. https://doi.org/10.5271/sjweh.3351
13. 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;58:882-893. https://doi.org/10.1016/j.jclinepi.2005.01.016
14. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177-188. https://doi.org/10.1016/0197-2456(86)90046-2
15. Durkan J. A new diagnostic test for carpal tunnel syndrome. *J Bone Joint Surg Am*. 1991;74:1111. https://doi.org/10.2106/00004623-199173040-00009
16. El Miedany Y, Ashour S, Youssef S, Mehanna A, Meky FA. Clinical diagnosis of carpal tunnel syndrome: old tests–new concepts. *Jt Bone Spine*. 2008;75:451-457. https://doi.org/10.1016/j.jbspin.2007.09.014
17. Fertl E, Wöber C, Zeitlhofer J. The serial use of two provocative tests in the clinical diagnosis of carpal tunnel syndrome. *Acta Neurol Scand*. 1998;98:328-332. https://doi.org/10.1111/j.1600-0404.1998.tb01743.x
18. González-Roig JL, Cubero-Rego L, Santos-Anzorandia C. Relationship between clinical probability of carpal tunnel syndrome and neurophysiological studies. *Rev Esp Cir Ortop Traumatol*. 2008;52:353-358. https://doi.org/10.1016/S1988-8856(08)70121-4
19. Hansen PA, Micklesen P, Robinson LR. Clinical utility of the flick maneuver in diagnosing carpal tunnel syndrome. *Am J Phys Med Rehabil*. 2004;83:363-367. https://doi.org/10.1097/01.PHM.0000124439.14757.99
20. Harbord RM, Whiting P, Sterne JAC, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J Clin Epidemiol*. 2008;61:1095-1103. https://doi.org/10.1016/j.jclinepi.2007.09.013
21. Hegmann KT, Merryweather A, Thiese MS, et al. Median nerve symptoms, signs, and electrodiagnostic abnormalities among working adults. *J Am Acad Orthop Surg*. 2018;26:576-584. https://doi.org/10.5435/JAAOS-D-17-00034
22. Heller L, Ring H, Costeff H, Solzi P. Evaluation of Tinel's and Phalen's signs in diagnosis of the carpal tunnel syndrome. *Eur Neurol*. 2008;25:40-42. https://doi.org/10.1159/000115985
23. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21:1539-1558. https://doi.org/10.1002/sim.1186
24. Kasundra GM, Sood I, Bhargava AN, et al. Carpal tunnel syndrome: analyzing efficacy and utility of clinical tests and various diagnostic modalities. *J Neurosci Rural Pract*. 2015;6:504-510. https://doi.org/10.4103/0976-3147.169867
25. Keith MW, Masear V, Chung K, et al. Diagnosis of carpal tunnel syndrome. *J Am Acad Orthop Surg*. 2009;17:389-396. https://doi.org/10.5435/00124635-200906000-00007
26. Koulidis K, Veremis Y, Anderson C, Heneghan NR. Diagnostic accuracy of upper limb neurodynamic tests for the assessment of peripheral neuropathic pain: a systematic review. *Musculoskelet Sci Pract*. 2019;40:21-33. https://doi.org/10.1016/j.msksp.2019.01.001
27. Küçükakkaş O, Yurdakul OV. The diagnostic value of clinical examinations when diagnosing carpal tunnel syndrome assisted by nerve conduction studies. *J Clin Neurosci*. 2019;61:136-141. https://doi.org/10.1016/j.jocn.2018.10.106
28. Kuhlman KA, Hennessey WJ. Sensitivity and specificity of carpal tunnel syndrome signs. *Am J Phys Med Rehabil*. 1997;76:451-457. https://doi.org/10.1097/00002060-199711000-00004
29. Lijmer JG, Bossuyt PMM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med*. 2002;21:1525-1537. https://doi.org/10.1002/sim.1185
30. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *J Am Med Assoc*. 1999;282:1061-1066. https://doi.org/10.1001/jama.282.11.1061
31. Ma HJ, Kim I. The diagnostic assessment of hand elevation test in carpal tunnel syndrome. *J Korean Neurosurg Soc*. 2012;52:472-475. https://doi.org/10.3340/jkns.2012.52.5.472
32. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Chapter 10 Analysing and Presenting Results*. Available at: http://srdta.cochrane.org/. Accessed February 22, 2021.
33. MacDermid JC, Kramer JF, McFarlane RM, Roth JH. Inter-rater agreement and accuracy of clinical tests used in diagnosis of carpal tunnel syndrome. *Work*. 1997;8:37-44. https://doi.org/10.3233/WOR-1997-8105
34. MacDermid JC, Wessel J. Clinical diagnosis of carpal tunnel syndrome: a systematic review. *J Hand Ther*. 2004;17:309-319. https://doi.org/10.1197/j.jht.2004.02.015
35. American Academy of Orthopaedic Surgeons. *Management of Carpal Tunnel Syndrome Evidence-Based Clinical Practice Guideline*; 2016. Available at: www.orthoguidelines.org. Accessed April 8, 2021.
36. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. 2018;319:388. https://doi.org/10.1001/jama.2017.19163
37. Mondelli M, Passero S, Giannini F. Provocative tests in different stages of carpal tunnel syndrome. *Clin Neurol Neurosurg*. 2001;103:178-183. https://doi.org/10.1016/S0303-8467(01)00140-8
38. Naranjo A, Ojeda S, Mendoza D, Francisco F, Quevedo JC, Erausquin C. What is the diagnostic value of ultrasonography compared to physical evaluation in patients with idiopathic carpal tunnel syndrome? *Clin Exp Rheumatol*. 2007;25:853-859.
39. Nee RJ, Jull GA, Vicenzino B, Coppieters MW. The validity of upper-limb neurodynamic tests for detecting peripheral neuropathic pain. *J Orthop*

Sports Phys Ther. 2012;42:413-424. https://doi.org/10.2519/jospt.2012.3988
40. Nora DB, Becker J, Ehlers JA, Gomes I. Clinical features of 1039 patients with neurophysiological diagnosis of carpal tunnel syndrome. *Clin Neurol Neurosurg*. 2004;107:64-69. https://doi.org/10.1016/j.clineuro.2004.08.003
41. Padua L, Coraci D, Erra C, et al. Carpal tunnel syndrome: clinical features, diagnosis, and management. *Lancet Neurol*. 2016;15:1273-1284. https://doi.org/10.1016/S1474-4422(16)30231-9
42. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:2020-2021.
43. Phalen GS, Kendrick JI. Compression neuropathy of the median nerve in the carpal tunnel. *JAMA*. 1957;164:524-530. https://doi.org/10.1001/jama.1957.02980050014005
44. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58:982-990. https://doi.org/10.1016/j.jclinepi.2005.02.022
45. Richter M, Brüser J. Die wertigkeit der klinischen diagnostik beim karpaltunnelsyndrom [Value of clinical diagnosis in carpal tunnel syndrome]. *Handchir Mikrochir Plast Chir*. 1999;6:373-376. https://doi.org/10.1055/s-1999-13555
46. Sadanandan A, Rijesh P. Tinel's with Tourniquet test for the diagnosis of carpal tunnel syndrome—a new provocative clinical test. *J Evol Med Dent Sci*. 2017;6:4572-4575. https://doi.org/10.14260/Jemds/2017/989
47. Sawaya A, Sakr C. When is the Phalen's test of diagnostic value: an electrophysiologic analysis? *J Clin Neurophysiol*. 2009;26:132-133. https://doi.org/10.1097/WNP.0b013e31819d8046
48. Schmid AB, Fundaun J, Tampin B. Entrapment neuropathies: a contemporary approach to pathophysiology, clinical assessment, and management. *Pain Rep*. 2020;5:e829. https://doi.org/10.1097/PR9.0000000000000829
49. Schmid AB, Hailey L, Tampin B. Entrapment neuropathies: challenging common beliefs with novel evidence. *J Orthop Sports Phys Ther*. 2018;48:58-62. https://doi.org/10.2519/jospt.2018.0603
50. St John Smith E. Advances in understanding nociception and neuropathic pain. *J Neurol*. 2018;265:231-238. https://doi.org/10.1007/s00415-017-8641-6
51. Stewart J, Eisen A. Tinel's sign and the carpal tunnel syndrome. *Br Med J*. 1978;2:1436. https://doi.org/10.1136/bmj.2.6145.1125-a
52. Szabo RM, Slater RR, Farver TB, Breger Stanton D, Sharman WK. The value of diagnostic testing in carpal tunnel syndrome. *J Hand Surg*. 1999;24:704-714. https://doi.org/10.1053/jhsu.1999.0704
53. Tetro AM, Evanoff BA, Hollstien SB, Gelberman RH. A new provocative test for carpal tunnel syndrome. *J Bone Joint Surg Br*. 1998;80:493-498. https://doi.org/10.1302/0301-620X.80B3.0800493
54. Torres-Costoso A, Martínez-Vizcaíno V, Álvarez-Bueno C, Ferri-Morales A, Cavero-Redondo I. Accuracy of ultrasonography for the diagnosis of carpal tunnel syndrome: a systematic review and meta-analysis. *Arch Phys Med Rehabil*. 2018;99:758-765.e10. https://doi.org/10.1016/j.apmr.2017.08.489
55. Trillos MC, Soto F, Briceno-Ayala L. Upper limb neurodynamic test 1 in patients with clinical diagnosis of carpal tunnel syndrome: a diagnostic accuracy study. *J Hand Ther*. 2018;31:333-338. https://doi.org/10.1016/j.jht.2017.05.004
56. Vanti C, Bonfiglioli R, Calabrese M, Marinelli F, Violante FS, Pillastrini P. Relationship between interpretation and accuracy of the upper limb neurodynamic test 1 in carpal tunnel syndrome. *J Manipulative Physiol Ther*. 2012;35:54-63. https://doi.org/10.1016/j.jmpt.2011.09.008
57. Vanti C, Bonfiglioli R, Calabrese M, et al. Upper limb neurodynamic test 1 and symptoms reproduction in carpal tunnel syndrome. A validity study. *Man Ther*. 2011;16:258-263. https://doi.org/10.1016/j.math.2010.11.003
58. Wainner RS, Fritz JM, Irrgang JJ, Delitto A, Allison S, Boninger ML. Development of a clinical prediction rule for the diagnosis of carpal tunnel syndrome. *Arch Phys Med Rehabil*. 2005;86:609-618. https://doi.org/10.1016/j.apmr.2004.11.008
59. Walters C, Rice V. An evaluation of provocative testing in the diagnosis of carpal tunnel syndrome. *Mil Med*. 2002;167:647-652. https://doi.org/10.1093/milmed/167.8.647
60. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529-536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009
61. Widodo SA, Rochman F, Haryadi RD, Patricia MK. The diagnostic accuracy of Phalen test and Prayer test in carpal tunnel syndrome. *Indian J Forensic Med Toxicol*. 2020;14:1956-1961.
62. Wiesman IM, Novak CB, Mackinnon SE, Winograd JM. Sensitivity and specificity of clinical testing for carpal tunnel syndrome. *Plast Surg*. 2003;11:70-72. https://doi.org/10.4172/plastic-surgery.1000355
63. Witt JC, Hentz JG, Stevens JC. Carpal tunnel syndrome with normal nerve conduction studies. *Muscle Nerve*. 2004;29:515-522. https://doi.org/10.1002/mus.20019
64. Yang B, Whiting PF, Davenport CF, et al. Guidance on how to use QUADAS-C. 2020;1-21.
65. Zhang D, Chruscielski CM, Blazar P, Earp BE. Accuracy of provocative tests for carpal tunnel syndrome. *J Hand Surg Glob Online*. 2020;2:121-125. https://doi.org/10.1016/j.jhsg.2020.03.002