Endocrine (2017) 57:18–34
DOI 10.1007/s12020-017-1298-1

META-ANALYSIS

Systematic reviews of diagnostic tests in endocrinology: an audit of methods, reporting, and performance

Gabriela Spencer-Bonilla[1,2] · Naykky Singh Ospina[1,3] · Rene Rodriguez-Gutierrez[1,4] · Juan P. Brito[1,5] · Nicole Iñiguez-Ariza[5] · Shrikant Tamhane[1,5] · Patricia J. Erwin[6] · M. Hassan Murad[1,7] · Victor M. Montori[1,5]

Received: 24 January 2017 / Accepted: 1 April 2017 / Published online: 5 June 2017

Abstract

Background Systematic reviews provide clinicians and policymakers estimates of diagnostic test accuracy and their usefulness in clinical practice. We identified all available systematic reviews of diagnosis in endocrinology, summarized the diagnostic accuracy of the tests included, and assessed the credibility and clinical usefulness of the methods and reporting.

Methods We searched Ovid MEDLINE, EMBASE, and Cochrane CENTRAL from inception to December 2015 for systematic reviews and meta-analyses reporting accuracy measures of diagnostic tests in endocrinology. Experienced reviewers independently screened for eligible studies and collected data. We summarized the results, methods, and reporting of the reviews. We performed subgroup analyses to categorize diagnostic tests as most useful based on their accuracy.

Results We identified 84 systematic reviews; half of the tests included were classified as helpful when positive, one-fourth as helpful when negative. Most authors adequately reported how studies were identified and selected and how their trustworthiness (risk of bias) was judged. Only one in three reviews, however, reported an overall judgment about trustworthiness and one in five reported using adequate meta-analytic methods. One in four reported contacting authors for further information and about half included only patients with diagnostic uncertainty.

Conclusion Up to half of the diagnostic endocrine tests in which the likelihood ratio was calculated or provided are likely to be helpful in practice when positive as are one-quarter when negative. Most diagnostic systematic reviews in endocrine lack methodological rigor, protection against bias, and offer limited credibility. Substantial efforts, therefore, seem necessary to improve the quality of diagnostic systematic reviews in endocrinology.

Keywords Diagnostic accuracy · Diagnostic systematic review · Systematic review methodology

Electronic supplementary material The online version of this article (doi:10.1007/s12020-017-1298-1) contains supplementary material, which is available to authorized users.

Naykky Singh Ospina and Gabriela Spencer-Bonilla contributed equally to this work.

✉ Victor M. Montori
montori.victor@mayo.edu

1 Knowledge and Evaluation Research Unit, Mayo Clinic, Rochester, MN, USA

2 School of Medicine, University of Puerto Rico Medical Sciences Campus, San Juan, PR, USA

3 Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, University of Florida, Gainesville, FL, USA

4 Division of Endocrinology, Department of Internal Medicine, University Hospital "Dr. Jose E. Gonzalez", Autonomous University of Nuevo Leon, Monterrey, MX, USA

5 Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Mayo Clinic, Rochester, MN, USA

6 Mayo Medical Library, Mayo Clinic, Rochester, MN 55905, USA

7 Division of Preventive, Occupational, and Aerospace Medicine, Mayo Clinic, Rochester, MN, USA

Introduction

Accurate diagnosis informs adequate treatment in clinical practice. Understanding the accuracy of diagnostic tests

allows clinicians to discern between futile tests and those that can significantly advance the diagnostic process. Yet, clinicians often have inaccurate expectations regarding the benefits and harms of medical tests [1].

Despite the clinical importance of diagnostic tests, research regarding diagnosis usually lags behind its treatment counterpart, both in terms of quantity and quality [2, 3]. An analysis of research in the field of endocrinology suggests knowledge gaps (including those related to diagnosis) are not adequately addressed by ongoing research studies [4]. The Institute of Medicine's (IOM) report on diagnostic errors, recognizes this knowledge gap, and highlights the need for the research community to improve the understanding of the diagnostic process [2].

Systematic reviews (SRs), summarizing the body of evidence, provide clinicians, patients, guideline panelists, and policymakers with best estimates of the accuracy of diagnostic tests and their usefulness in clinical practice. They allow clinicians to evaluate the body of available evidence instead of using only the latest, largest, or most well-known study to inform their practice. In order for clinicians to apply the results in practice, SRs should: (1) produce results that are clinically useful and (2) have credible and reproducible methods for synthesizing and reporting the evidence. Our logic model (Fig. 1) evaluates the presence of criteria that can help clinicians evaluate SRs in general (e.g., performing an assessment of risk of bias) and diagnostic reviews in particular (e.g., including only patients with diagnostic uncertainty) [5, 6].

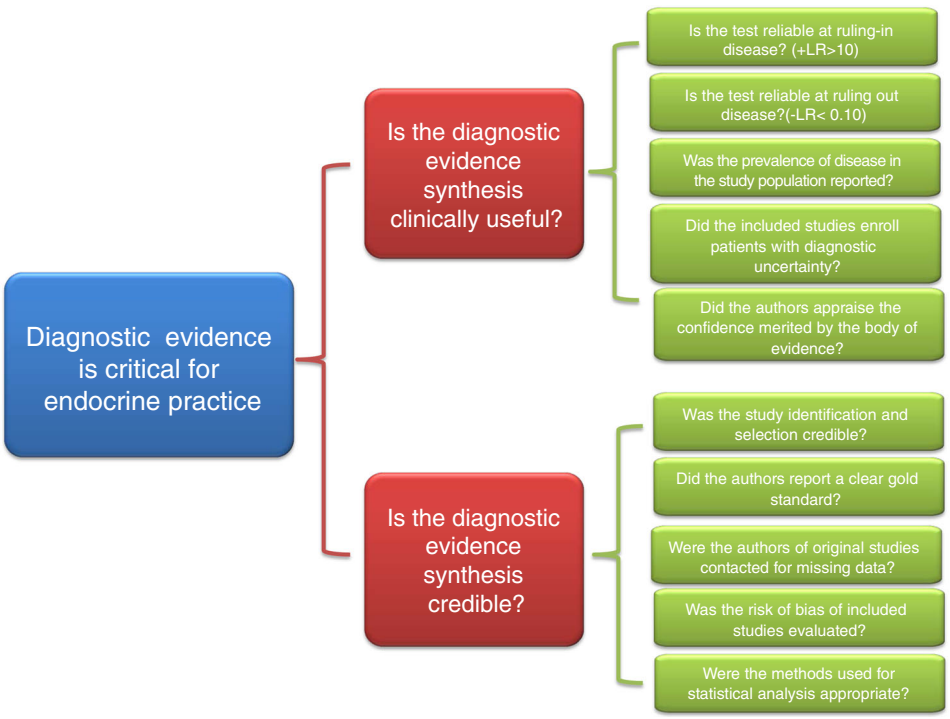
In accordance with IOM's recommendations, an assessment of the literature can help identify useful diagnostic tests in endocrinology and the conditions that warrant investment of scarce research resources to improve the diagnostic process. To this end, we identified all SRs and meta-analyses addressing the diagnostic accuracy of different tests in the field of endocrinology. Here, we summarize their findings along with the quality and reporting of the methods used in order to determine if the diagnostic evidence syntheses in endocrinology are: (a) clinically useful and (b) credible.

Methods

Eligibility criteria

We included SRs and meta-analyses reporting accuracy measures of diagnostic tests used in patients under evaluation for any endocrine condition. Endocrine conditions were defined as those that belonged to one of the following categories: (1) bone, (2) diabetes/glycemia, (3) neuroendocrine tumors, (4) pituitary–gonadal–adrenal, and (5) thyroid. Eligible studies reported clinically applicable diagnostic accuracy measures using: sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, diagnostic odds ratio, diagnostic accuracy, or the area under the curve.

Fig. 1 Logic model



Studies not reporting any of these diagnostic measures were excluded without author contact.

Search methods for identification of studies

A comprehensive search of MEDLINE, EMBASE, and Cochrane CENTRAL from inception to December 2015 without language restrictions was designed by an experienced medical librarian (P.J.E.) with input from the study's principal investigators (N.S.O. and R.R.G.). Controlled vocabulary supplemented with keywords was used to search for SRs of diagnostic tests in endocrinology (Online Resource 1). We consulted Mayo Clinic experts in the field to identify references that could have been missed by our search.

Selection of studies and data management

The search results were uploaded into SR software (DistillerSR, Ottawa, Canada, https://distillercer.com/products/distillersr-systematic-review-software/). Six reviewers, working independently and in duplicate, reviewed all abstracts and titles for inclusion (J.P.B., N.I.A., R.R.G., N.S.O., G.S.B., and S.T.) and assessed the eligibility of full-text publications retrieved if at least one reviewer considered the abstract eligible. Disagreements were resolved by arbitration (a third reviewer decided whether the study should be included). Chance-adjusted agreement was quantified using the kappa statistic [7].

Data collection

Reviewers performed data collection independently and in duplicate using a standardized form and data extraction instructions after calibrating with two included full texts. Extracted data included: (a) general information about the review (first author, year of publication, country, condition of interest, diagnostic test, gold standard used, prevalence of the disease in the studied population, number of articles and patients included); (b) clinical performance of the diagnostic test (pooled summary statistics); and (c) methods and data reporting (search strategy, language restrictions, review of references, independent duplicate process, clear clinical question, selection criteria, summary of included studies, predefined subgroup analyses, methods used for data representation, method used for analysis, evaluation of risk of bias and tool used, author contact description, publication bias assessment, heterogeneity assessment, and assessment of the confidence warranted by the evidence).

In order to determine the clinical usefulness of diagnostic tests, subgroup analyses were performed including only studies in which a likelihood ratio (a measure of diagnostic accuracy) was provided or could be calculated from the available data (e.g., when sensitivity and specificity were provided). We chose likelihood ratios for this analysis since this diagnostic variable allows comparison of results across studies, can be applied regardless of the prevalence of a condition, and is easily understood and applied by physicians when expressed in non-technical language [8]. In cases in which a single diagnostic test had more than one review available, we included the study with the best clinical performance: a best-case scenario analysis.

We used descriptive statistics to summarize extracted variables. Authors' descriptions of "risk of bias" and "quality" were considered to be reflecting the same construct [9]. Author ratings of studies as "high" or "moderate to high" quality were coded as low-risk of bias and "low" or "acceptable" quality ratings were coded as high-risk of bias. We considered assessments of "confidence in the evidence" to be only those that used the results of the review to assess the quality of the *body* of evidence. Bivariate and hierarchical models were considered to be adequate methods for statistical analyses [10]. When a description of methods for statistical analysis was provided, reviews that did not use these methods were coded as "other" if another method (e.g., Moses–Littenberg random effects) was used or "unclear" when the description was insufficient to ascertain the method used.

Results

The database search generated 2122 eligible reports. After abstract and title screening, 115 reports were identified for full-text review. After reproducible full-text screening (κ statistic = 0.89) [7], 84 studies were found to be eligible [11–95] (Fig. 2). The majority of the included studies evaluated diagnostic tests related to thyroid conditions (*n*-42, 50%), most commonly thyroid cancer, or diabetes/glycemia (*n*-19, 23%). Five SRs did not include a meta-analysis. General characteristics of the included studies are reported in Table 1.

Forty-four of the 84 included reviews, evaluating 65 diagnostic tests, provided a pooled estimate of the likelihood ratio or enough data to calculate it. The positive likelihood ratio (+LR) of these tests is shown in Fig. 3 and the negative likelihood ratio (−LR) in Fig. 4. Forty-seven percent of the 65 diagnostic tests had a positive likelihood ratio of ≥10. Only 35% of these studies (reporting helpful +LR), however, reported an overall statement of the risk of bias of the included studies. Online Resource 2 shows a summary of diagnostic test with +LR ≥ 10, considered to be most useful to "rule-in" a diagnosis in endocrinology.

Twenty-six percent of the tests had a negative likelihood ratio of ≤0.10, only five of which reported an overall statement of risk of bias. Online Resource 3 shows a

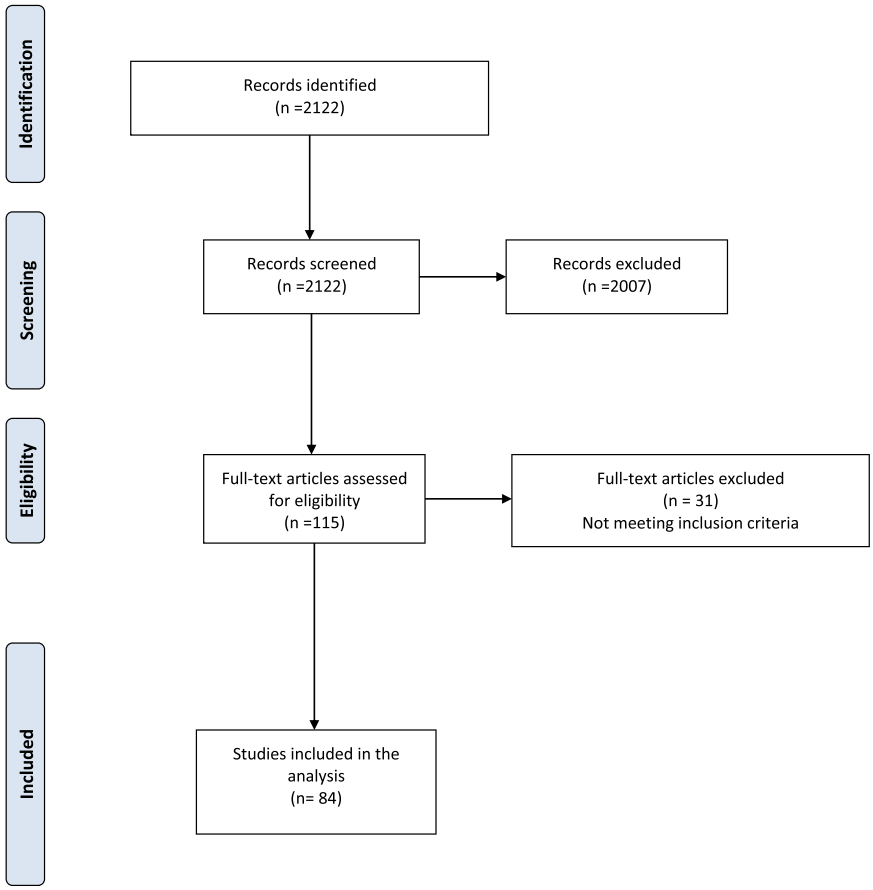


Fig. 2 PRISMA flowchart (study selection process)

summary of diagnostic tests with −LR ≤ 0.10, considered to be most useful to "rule-out" a diagnosis in endocrinology. The most useful test, when positive or negative was the composite of free cortisol, salivary midnight cortisol, and dexamethasone suppression test for ruling in or out Cushing's syndrome.

The most commonly reported summary statistics were sensitivity and specificity (94 and 87%, respectively) (Online Resource 4). The most common way of representing findings was the use of forest plots (68%) followed by receiver operator curves (65%). Only one meta-analysis presented the results using a Fagan nomogram, and three used other graphical representations (Online Resource 5).

About half of the reviews clearly reported including only patients in which there was diagnostic uncertainty and one in three reported the prevalence of the condition in the study population. Only five studies assessed the overall confidence merited by the body of evidence with unclear methods and limited conclusions provided (Fig. 5, Online Resource 6).

Most of the reviews reported a clear gold standard (93%), a comprehensive search strategy, a well-defined clinical question, and clear inclusion criteria; reviewed primary studies for references; and provided a summary table of the included studies. Only about 31% reported that their search was unrestricted by language. Approximately one in five of the reviews reported contacting authors, with a mean author response rate of 57% (range 16–91%) (Fig. 5, Online Resource 6).

The majority of the reviews reported assessing the risk of bias of the included studies (69%), but only one-third provided readers with an overall statement regarding the risk of bias of the included studies. When available, this information was reported in the abstract of the study in 7% of the cases and in the conclusion of the manuscript in 22%. Assessment of publication bias was reported by 40% of authors; the most common method was use of funnel plots. When a conclusion about publication bias was made, most (87%) were found to have insignificant publication bias.

The statistical method used for the diagnostic meta-analysis was not reported or was unclear in 8 (10%) of the reviews. Adequate methods of statistical analysis were reported in 16 (20%) of the meta-analyses. Other methods were reported in 70% of the reviews; the most common description was "random effects model". Pre-specified subgroup analyses were reported in 34% of the studies. The majority of the reviews assessed for heterogeneity (77%). Most commonly, the Cochrane's Q or the I^2 statistic method

Table 1 Summary of included systematic reviews

	Review characteristics						
	Author, year	Country	Condition of interest	Diagnostic test of interest	Gold standard	Number of studies	Number of patients included
Bone	Gotthardt, 2004	Germany	Parathyroid adenoma or hyperplasia in hyperparathyroidism	99mTc methoxyisobutyli-sonitrile scintigraphy	Histologic examination or follow-up examination	51	1987
	Nayak, 2006	USA	Osteoporosis	Calcaneal quantitative US	DXA	21	9061
	Caldarella, 2012	Italy	Secondary hyperparathyroidism	Planar scintigraphy using 99mTc-MIBI	Histological examination	24	471
	Cheung, 2012	Canada	Primary hyperparathyroidism	US, SPECT, 4D-CT	Surgical pathology	43	Not reported
	Cardarella, 2013	Italy	Parathyroid adenoma	C-methionine PET	Histological examination or clinical follow-up	9	258
	Wei, 2015	China	Primary hyperparathyroidism lesion localization	SPET/CT, SPET, and planar imaging using 99mTc-MIBI	Surgical and/or histopathology findings and/or close clinical and imaging follow-up	15	1641
	Wong, 2015	USA	Parathyroid adenoma	Parathyroid scintigraphy with (99mTc)-sestamibi (SPECT/CT)	Histopathology	24	1276
Diabetes	Windeler, 1990	Germany	Diabetes mellitus	Fructosamine assay	HbA1c	65	1718
	Jensen, 1996	Denmark	Microalbuminuria in diabetes	MICRAL dipstick	Standard laboratory measurements of albumin in urine	11	2904 urine samples
	Peters, 1996	USA, multinational research group	Diabetes mellitus	Glycosylated hemoglobin	OGTT	18	11,276
	Whitsel, 2000	USA	Diabetic autonomic failure	QTc prolongation	Cardiovascular reflex test	17	4584
	Ewald, 2004	UK	Microalbuminuria in diabetes	Random spot urine test	AER on a 24-h timed urine specimen	10	1470
	Virgili, 2007	Italy	Diabetic macular edema	Optical coherence tomography	Stereoscopic fundus photography and contact lens or non-contact lens biomicroscopy of the fundus	15	1297
	Dinh, 2008	USA	Osteomyelitis diabetic foot ulcers	PE (probe to bone test), plain radiograpy, MRI, bone scan	Histopathologic findings or bone culture	9	1054
	Feng, 2009	USA	Diabetic neuropathy	SW	No clear gold standard	30	8365
	Blomberg, 2012	USA	Congenital hyperinsulinism	PVS, ASVS, and 18F-DOPA PET	Immunohistochemistry or histology	13	415
	Ye, 2012	China	Diabetes mellitus	OGTT	No clear gold standard	15	8027

Table 1 continued

	Review characteristics						
	Author, year	Country	Condition of interest	Diagnostic test of interest	Gold standard	Number of studies	Number of patients included
	Tang, 2013	China	Diabetes mellitus	Hemoglobin A1C	OGTT	11	15,995
	Tian, 2013	China, Japan	Gestational diabetes mellitus	Hemoglobin A1C	No clear gold standard	41	2812
	Yan, 2013	China	Diabetes mellitus	Hemoglobin A1C	WHO criteria for DM	25	82,689
	Yang, 2013	China	Congenital hyperinsulinism	18F-DOPA PET	Pathology	10	181
	Hirschfeld, 2014	Germany	Peripheral neuropathy children and adolescents with DM1	Vibration perception, SW Rydel-Seiffer tuning fork	Neuroconductive studies	5	699
	Su, 2014	China	Postpartum abnormal glucose tolerance	Hemoglobin A1C	75 g OGTT	6	1086
	Tsapas, 2014	Greece, UK	Diabetic neuropathy	Plaster	Several neuropathy scores and indices	18	3470
	Wu, 2014	Taiwan	Microalbuminuria	Albumin concentration, ratio of albumin/creatinine urine collection	30–300 mg/d by 24 h times urine collection of urinary AER	14	2078
	Xu, 2014	China	Diabetes mellitus	Hemoglobin A1C	75 g OGTT	9	25,932
	Shi, 2015	China	Diabetic retinopathy, diabetic macular edema	Telemedicine with digital photographs of the fundus	Telemedicine of standard field stereoscopic color slides of fundus	20	1960
NeT	Treglia, 2012	Italy	Gastroenteropancreatic neuroendocrine tumors	Gallium-68 somatostatin receptor PET and PET/CT	SRS, usually performed using Indium-111 DTPA-octreotide	15	567
	Puli, 2013	USA	Pancreatic neuroendocrine tumors	Endoscopic US	Histology or clinical follow-up	13	456
	Rufini, 2013	Italy	Neuroendocrine tumors	F-DOPA PER and PET/CT	Histology and/or clinical/imaging follow-up	28	851
	Yang, 2015	China	Neuroendocrine tumors	Chromogranin A	Histology	13	2227
Pituitary–gonadal–adrenal	Boland, 1998	USA	Malignant adrenal masses	Unenhanced CT	No clear gold standard	10	495 lesions
	Dorin, 2003	USA	Primary adrenal insufficiency	250-μg cosyntropin stimulation test	ITT or metyrapone test	Unclear	Unclear
	Elamin, 2008	USA	Cushing's syndrome	UFC, salivary midnight cortisol, 1 mg overnight DST, 2 day 2 mg DST and combined strategies	Pathological diagnosis, response to therapy targeting CS, or clinical follow-up	27	8631
	Kazlauskaite, 2008	USA	Hypothalamic–pituitary–adrenal insufficiency	Low dose and high dose corticotropin tests	Insulin hypoglycemia, metyrapone test	13	679

Table 1 continued

	Review characteristics						
	Author, year	Country	Condition of interest	Diagnostic test of interest	Gold standard	Number of studies	Number of patients included
	Carroll, 2009	USA	Cushing's syndrome	Late-night salivary cortisol	Other clinical and biochemical lab tests concordant with Cushing's or pathology	7	947
	Jacobson, 2010	USA	Pheochromocytoma and neuroblastoma	123I-meta-iodobenzylguanidine scintigraphy	Histological and follow-up	22	Unclear
	Boland, 2011	USA	Adrenal masses	FDG PET	Histological and follow-up	21	1217
	Hazem, 2011	USA	Growth hormone deficiency	Any test to diagnose growth hormone deficiency	Several: mainly ITT	23	1100
	Iliodromiti, 2013	UK	Polycystic ovary syndrome	Anti-mullerian hormone	Rotterdam criteria	10	683
	Rufini, 2013	Italy	Pheochromocytoma and paraganglioma	Metaiodobenzylguanidine scintigraphy with PET	Histology/clinical or imaging follow-up/study description	28	852
	Shen, 2014	China	Growth hormone deficiency	IGF-1, IGFBP-3	Hormone provocative test	12	1633 measurements
Thyroid	Eustatia-Rutten, 2004	The Netherlands	Thyroid cancer recurrence	Serum thyroglobulin (RIA and IMA)	Radioiodine scintigraphy, radiological examinations (CT, MRI, PET), histology and follow-up	46	9094
	Peng, 2007	China	Thyroid nodules	Conventional FNAB or US guided FNAB	Conventional biopsy	39	11,194
	Peng, 2008	USA	Thyroid nodules	FNA and frozen section evaluation	No clear gold standard	52	Unclear
	Raijmakers, 2008	The Netherlands	Lymph node metastases in thyroid cancer	Sentinel node biopsy	Histopathological findings	14	457
	Dong, 2009	China	Recurrent or metastatic DTC and a radioiodine-negative whole-body scan	18F-FDG-PET/PET-CT	Histopathologic results	17	571
	Stevens, 2009	Canada	Thyroid nodules	Fine-needle aspiration	Histological and follow-up	12	530
	Bojunga, 2010	Germany	Thyroid nodules	Real-time elastography	Histopathologic results	8	639
	Iared, 2010	Brazil	Follicular thyroid neoplasm	Color doppler US	Histopathologic results	4	457
	Miller, 2011	USA	Papillary thyroid cancer	FDG PET	Pathology, clinical	11	498
	Vriens, 2011	The Netherlands	Thyroid nodules with indeterminate fine-needle aspiration biopsy	FDG PET	Histopathologic results	6	225

Table 1 continued

Review characteristics						
Author, year	Country	Condition of interest	Diagnostic test of interest	Gold standard	Number of studies	Number of patients included
Bongiovanni, 2012	Switzerland	Thyroid nodules	The Bethesda system for reporting thyroid cytopathology	Histopathology	8	25,445
Cheng, 2012	China	Medullary thyroid carcinoma	18F-FDG-PET and 18F-FDG-PET/CT	Histopathologic/cytological analysis and/or close clinical and/or imaging follow-up	15	815
De Matos, 2012	Brazil	Thyroid nodules	Cytokeratin-19, galectin-3 and hbme-1	Histopathology	66	5168
Tozzoli, 2012	Italy	Graves' disease	TSH receptor antibodies	No clear gold standard	21	3081
Wu, 2012	China	Cervical lymph node metastasis in patients with papillary thyroid carcinoma	US	Histopathology	13	1020
Razavi, 2013	USA	Malignant thyroid nodules	Elastography & B mode US	FNA histology, histopathology	24	2624
Treglia, 2013	Switzerland, Italy	Thyroid cancer	99mTC-MIBI	Histology or FNA	21	2060
Wang, 2013	China	Thyroid nodules	18F PET	Histopathologic analysis and/or follow-up	7	267
Zhang, 2013	China	Thyroid cancer	Shear wave elastography	Histology or FNA	5	469
Brito, 2014	USA	Thyroid cancer	US features	Histology, core biopsy, two FNABs, follow-up at 6 months	31	13,736
Ghajarzadeh, 2014	Iran	Thyroid cancer	Sonoelastography	FNA cytology or histopathology	12	1180 nodules
Giovanella, 2014	Switzerland, Italy, Germany	Thyroid cancer	Unstimulated highly sensitive Tg	Stimulated Tg measurement	9	3178
Grani, 2014	Italy	Lymph node metastases in thyroid cancer	Thyroglobulin washout in needle after FNA	Surgical histology	24	2865 lymph nodes
Jia, 2014	China	Thyroid cancer	BRAF V600E	Histology	16	1131
Li, 2014	China	Malignant thyroid nodules	FNAB/core needle biopsy	Histology or follow-up	5	1264
Lin, 2014	China	Malignant thyroid nodules	Shear wave elastography	USFNA and/or histology	15	1525
Ma, 2014	China	Thyroid nodules	Elastography	Histology or FNA pathology	35	4127
Qu, 2014	China	Malignant thyroid nodules	SUV_{max} on F-FDG PET or PET/CT	Histopathology	29	5715
Sheffield, 2014	Canada	Thyroid nodules	Bethesda system for reporting thyroid cytopathology	Histological diagnoses after thyroid surgery	13	28,904

Table 1 continued

Review characteristics						
Author, year	Country	Condition of interest	Diagnostic test of interest	Gold standard	Number of studies	Number of patients included
Sun, 2014	China	Thyroid cancer	Real-time US elastography	Fine-needle aspiration cytology or histopathology	31	5451
Wale, 2014	UK, Singapore	Thyroid cancer	99 m TC methoxyisobutylisonitrile scintigraphy (MIBI)	Histology	10	712
Wei, 2014	China	Malignant thyroid nodules	Thyroid imaging reporting and data system	Histology or FNAB	5	7753
Wolinski, 2014	Poland	Malignant thyroid nodules	US features	Histology	14	5439
Wu, 2014	China	Thyroid cancer	Diffusion-weighted MR	FNA cytology or histopathology	7	358
Yu, 2014	China	Malignant thyroid nodules	Contrast-enhanced US	Histology or cytology	7	597
Zhang, 2014	China	Malignant thyroid nodules	MicroRNA	Histology	7	543
Dong, 2015	China	Thyroid nodules	SWV using VTQ ARFI technology	Histopathology or FNA	13	1451
Nell, 2015	The Netherlands	Thyroid cancer	Qualitative elastography	Thyroid nodule cytology or histology	27	3908
Pyo, 2015	South Korea	Papillary thyroid carcinoma	BRAF immunohisto-chemistry using clone VE1	Histopathology of surgical specimen	11	1141
Remonti, 2015	Brazil	Thyroid cancer	US features	Histopathology of surgical specimen	52	12,786
Veer, 2015	UK	Thyroid nodules	Elastography	Histology and/or FNA cytology	38	5942 lesions
Zhou, 2015	China	Thyroid cancer	MicroRNAs	FNAB	7	543

18F/FDG fludeoxyglucose, *4d-CT* four-dimensional computed tomography, *99mTc* technetium-99m methoxyisobutylisonitrile, *AER* albumin excretion rate, *ARFI* acoustic radiation force impulse imaging, *ASVS* selective pancreatic arterial calcium stimulation with hepatic venous sampling, *DM* Diabetes mellitus, *DM1* type 1 diabetes mellitus, *DOPA* dihydroxyphenylalanine, *DST* Dexamethasone suppression test, *DXA* dual-energy X-ray absorptiometry, *FNA* fine-needle aspiration, *FNAB* fine-needle aspiration biopsy, *ITT* insulin tolerance test, *SPECT* sestamibi single-photon emission computed tomography, *IGFBP-3* insulin-like growth factor-binding protein, *IGF-1* insulin-like growth factor 1, *IMA* immunometric assay, *NeT* neuroendocrine tumors, *OGTT* oral glucose tolerance test, *PE* physical exam, *NC* not clear, *PET* positron emission tomography, *PVS* pancreatic venous sampling, *QTc* corrected QT interval, *RIA* radioimmunoassay, *SRS* somatostatin receptor scintigraphy, SUV_{max} standardized uptake value, *SW* Semmes Weinstein monofilament exam, *SWV* shear wave velocity, *Tg* thyroglobulin, *UFC* urinary free cortisol, *US* ultrasound, *VTQ* virtual touch tissue quantification, *WHO* World Health Organization

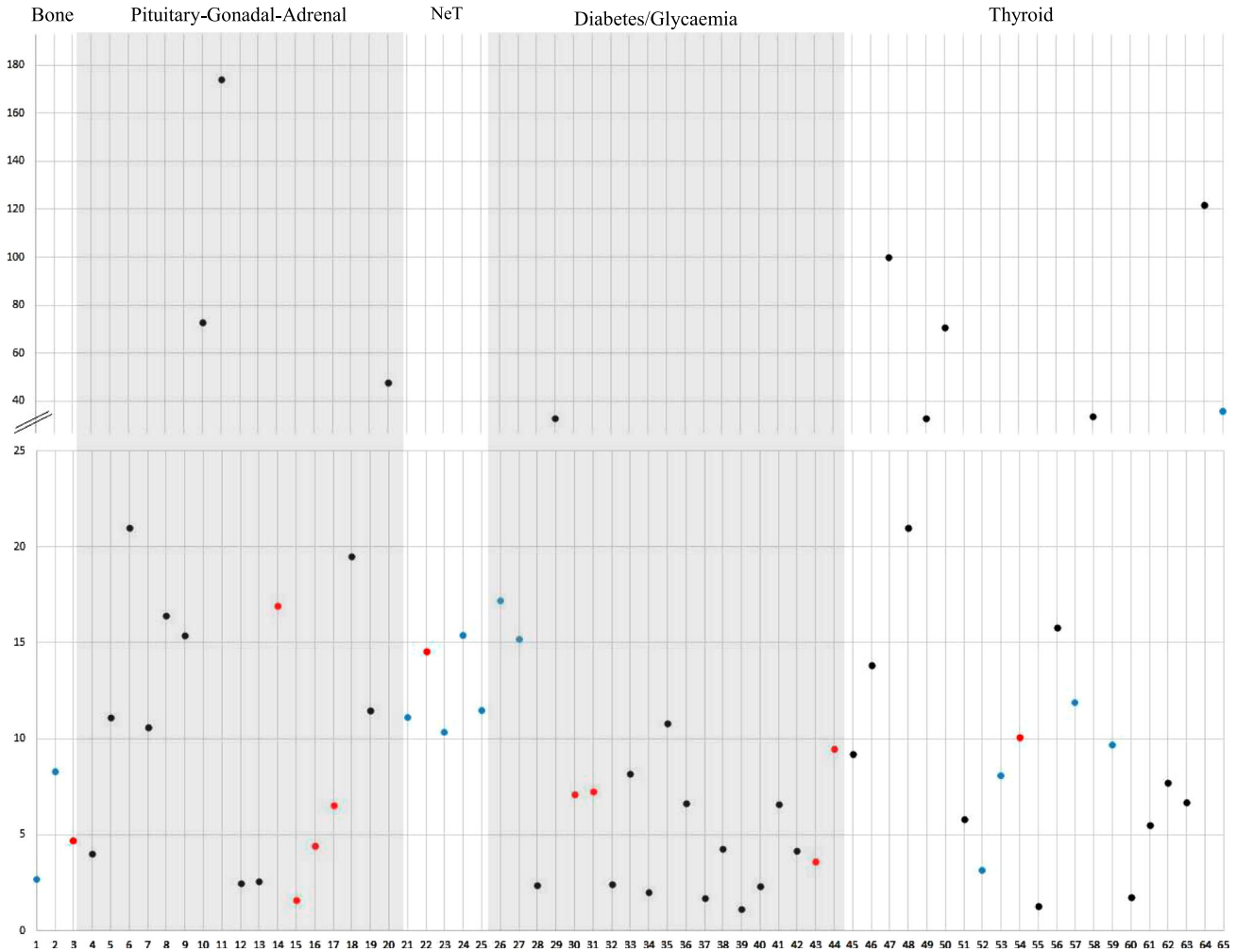
was used; and 86% of those with a conclusion reported significant or high heterogeneity. Exploration of heterogeneity was performed in 38% of meta-analyses mostly through subgroup analyses or meta-regression. Results changed in about half of the reviews that reported the results of exploring heterogeneity (14/30).

Discussion

We performed a systematic evaluation of reviews and meta-analyses evaluating diagnostic tests in endocrinology to determine the clinical usefulness of the included tests and the credibility of the syntheses. Most of the included reviews focused on glycemia/diabetes and thyroid conditions. Clinically, ~50% of the diagnostic tests had a positive likelihood ratio equal to or greater than 10, significantly changing the probability of disease and helping clinicians rule in a disease if the test is positive. About 25% of the tests had a negative likelihood ratio ≤0.10, and can significantly decrease the probability of disease and help clinicians rule out a disease process. These diagnostic tests (Online Resources 2 and 3) can therefore be considered among the most useful tests in day-to-day endocrine clinical practice. This conclusion, however, is limited by the quality of information regarding the risk of bias of the included studies.

Likelihood ratios were used as summary statistics in 43% of the reviews and applied in a Fagan Nomogram in only one. As in previous evaluations, the most common measures used to report diagnostic accuracy were sensitivity and specificity [96]. Studies using hypothetical scenarios have shown that physicians integrate diagnostic performance information into their clinical assessment more accurately when it is provided as a likelihood ratio described in non-technical wording [8] and that probability modifying plots and natural frequency trees might help clinicians more accurately interpret diagnostic test results [97].

Empirical evidence suggests that characteristics of the sampled population can bias the results of evaluations of diagnostic test performance [98, 99]. About half of systematic reviewers reported only including primary studies



◀ **Fig. 3** Positive likelihood ratios of the included studies. *Y*-axis, LR+ (positive likelihood ratio). Risk of bias: *blue*: low to moderate; *red*: high; *black*: not reported. NeT: Neuroendocrine tumors. Diagnostic tests included: (1) MET-PET for localization of parathyroid adenoma. (2) MIBG for localization in secondary hyperparathyroidism. (3) Calcaneal quantitative ultrasound (*T*-score threshold − 2.5) for osteoporosis. (4) Anti-mullerian hormone for polycystic ovarian syndrome. (5) FDG PET for malignant adrenal mass. (6) Late night salivary cortisol for Cushing's syndrome. (7) Urinary free cortisol for Cushing's syndrome. (8) 2 mg dexamethasone suppression test for Cushing's syndrome. (9) Urinary free cortisol plus dexamethasone suppression test for Cushing's syndrome. (10) Urinary free cortisol plus salivary midnight cortisol for Cushing's syndrome. (11) Urinary free cortisol plus salivary midnight cortisol plus dexamethasone suppression test (all tests) for Cushing's syndrome. (12) IGF-1 for growth hormone deficiency. (13) IGFBP-3 for growth hormone deficiency. (14) Growth hormone releasing peptide 6 for growth hormone deficiency. (15) Serum growth hormone levels for growth hormone deficiency. (16) Glucagon stimulation test for growth hormone deficiency. (17) ITT for growth hormone deficiency. (18) 250 µg cosyntropin stimulation test for primary adrenal insufficiency. (19) 250 µg cosyntropin stimulation test for secondary adrenal insufficiency. (20) 123I-MIBG for localization of pheochromocytoma. (21) Endoscopic ultrasound for pancreatic neuroendocrine tumors. (22) Chromogranin A for neuroendocrine tumors. (23) Gallium-68 somatostatin receptor PET and PET/CT for thoracic and gastroenteropancreatic neuroendocrine tumors. (24) 18F DOPA PET or PET/CT for thoracic and gastroenteropancreatic neuroendocrine tumor. (25) 18F DOPA PET or PET/CT for pheochromocytoma/paraganglioma. (26) Telemedicine for diabetic retinopathy (absence). (27) Telemedicine for diabetic macular edema. (28) Hemoglobin A1c for postpartum abnormal glucose tolerance. (29) HbA1c+ 2 SD for diabetes. (30) Albumin urine concentration for microalbuminuria. (31) Ratio of albumin to creatinine for microalbuminuria. (32) Plaster for diabetic neuropathy. (33) Hemoglobin A1C for gestational diabetes. (34) QTc prolongation for diabetic autonomic failure. (35) MICRAL dipstick for diabetic microalbuminuria. (36) Probe to bone test for osteomyelitis in patients with diabetes. (37) Plain radiography for osteomyelitis in patients with diabetes. (38) MRI for osteomyelitis in patients with diabetes. (39) Bone scan Tc for osteomyelitis in patients with diabetes. (40) Bone scan for osteomyelitis in patients with diabetes. (41) Optical coherence tomography for diabetic macular edema. (42) 18F-dihydroxyphenylalanine positron emission tomography for congenital hyperinsulinism. (43) Pancreatic venous sampling for focal congenital hyperinsulinism. (44) 18F-DOPA PET for focal congenital hyperinsulinism. (45) Real-time elastography for thyroid cancer. (46) Core needle biopsy for thyroid malignancy. (47) Bethesda system for reporting thyroid cytopathology (only malignant and benign) for thyroid cancer. (48) Frozen section for follicular lesions, thyroid cancer. (49) Frozen section for non-follicular lesions (thyroid cancer). (50) Frozen section for thyroid lesions, not otherwise specified (thyroid cancer). (51) Contrast-enhanced ultrasound for thyroid cancer. (52) MicroRNA for thyroid cancer. (53) Thyroid ultrasound features (taller than wide) for thyroid cancer. (54) Thyroid US features spongiform for thyroid cancer. (55) Standardized uptake value (SUV_{max}) for thyroid cancer. (56) Thyroglobulin washout for thyroid cancer. (57) Diffusion weighted MR imaging for thyroid cancer. (58) Serum thyroglobulin with immunometric assay (with ablation, during thyroxine) for thyroid cancer recurrence. (59) B-RAF V600E for thyroid cancer. (60) 99mTC methoxyisobutylisonitrile scintigraphy for thyroid cancer. (61) FDG PET for thyroid cancer. (62) GAL-3 for thyroid cancer. (63) HBME-1 for thyroid cancer. (64) 3G TRAB assay for Graves' disease. (65) Ultrasonography for cervical lymph node metastases in papillary thyroid cancer

with patients with diagnostic uncertainty and one-third reported the prevalence of the condition in the study population. These findings limit clinician's abilities to assess the applicability of findings of SRs in their clinical settings.

Although most reviews reported credible methods for study identification and selection, included a clear gold standard, and evaluated the risk of bias of the studies, a limited number used appropriate statistical methods, contacted authors, or provided an overall summary of the risk of bias of the included studies. In 2006, a SR that evaluated 89 reviews of diagnostic tests in oncology found that reviews commonly defined inclusion criteria, provided a summary of the included studies, and reported performing crucial steps in duplicate. Quality assessment of the included studies was performed in 61% of the studies and a formal assessment provided in only 30% [6]. Our results, in a different field of medicine and 10 years later, are quite similar.

The number of reviews of diagnostic tests in endocrinology that reported author contact (~20%) to obtain primary information was lower than in SRs in general (50–85%) [100], though this procedure can lead to changes in the estimates of clinical utility of tests [101]. In a study of 114 SRs, half of the authors conducted an explicit assessment of the quality of the individual studies that were included and in 64 of the cases the results of these evaluation were presented in a table [9]. We found that although most reviews evaluated the quality of the included studies, most failed to provide an overall statement for the reader and a minority (7%) reported this information in the abstract. In addition, the majority of the studies included some description of statistical methods used, but only 20% used bivariate or hierarchical models, recommended by experts for diagnostic meta-analyses. These methods are preferred because they simultaneously summarize sensitivity and specificity [10, 102–104].

SRs and meta-analyses should include assessments of the body of evidence based on the study design and risk of bias of the included studies, indirectness, inconsistency across results, imprecision, and risk of publication bias [105]. We found that most diagnostic SRs in endocrinology fail to provide an overall statement about study quality, do not commonly evaluate the risk of publication bias, explore inconsistencies, or provide an overall assessment of the confidence in the evidence. This limits the results of these reviews to a summary estimate without assessing the confidence merited by the results or providing actionable insight to improve the quality of the body of evidence (identification of knowledge gaps and areas for research).

Diagnostic SRs are critical in clinical practice to: (1) identify the most useful clinical tests and (2) help identify

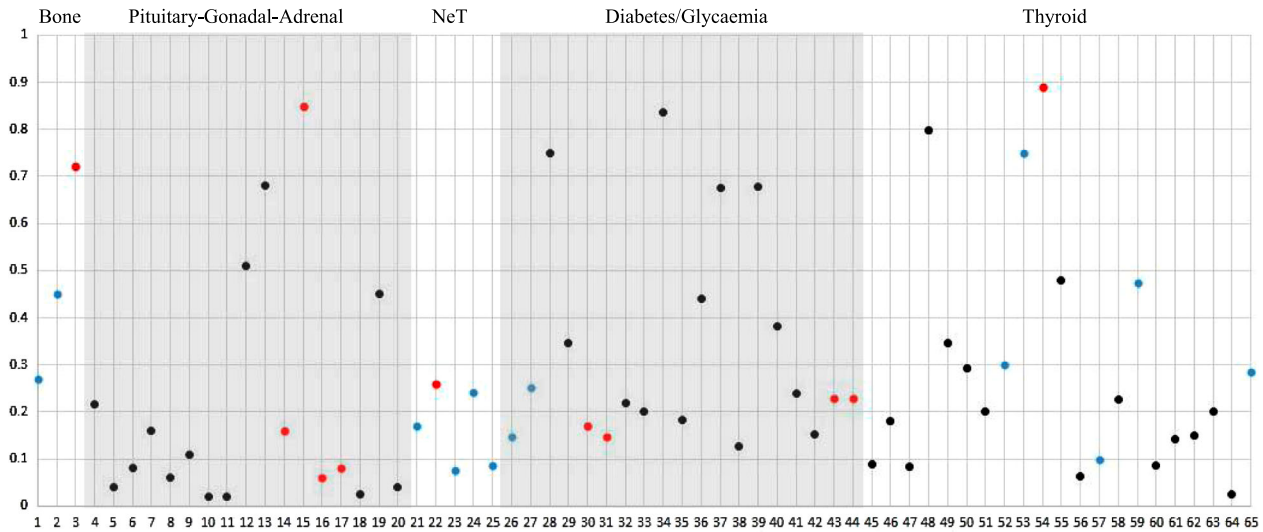


Fig. 4 Negative likelihood ratios of the included studies. *Y*-axis: LR− (negative likelihood ratio). Risk of bias: *blue*: low to moderate; *red*: high; *black*: not reported. NeT: Neuroendocrine tumors. Diagnostic tests included: (1) MET-PET for localization of parathyroid adenoma. (2) MIBG for localization in secondary hyperparathyroidism. (3) Calcaneal quantitative ultrasound (*T*-score threshold −2.5) for osteoporosis. (4) Anti-mullerian hormone for polycystic ovarian syndrome. (5) FDG PET for malignant adrenal mass. (6) Late night salivary cortisol for Cushing's syndrome. (7) Urinary free cortisol for Cushing's syndrome. (8) 2 mg dexamethasone suppression test for Cushing's syndrome. (9) Urinary free cortisol plus dexamethasone suppression test for Cushing's syndrome. (10) Urinary free cortisol plus salivary midnight cortisol for Cushing's syndrome. (11) Urinary free cortisol plus salivary midnight cortisol plus dexamethasone suppression test (all tests) for Cushing's syndrome. (12) IGF-1 for growth hormone deficiency. (13) IGFBP-3 for growth hormone deficiency. (14) Growth hormone releasing peptide 6 for growth hormone deficiency. (15) Serum growth hormone levels for growth hormone deficiency. (16) Glucagon stimulation test for growth hormone deficiency. (17) ITT for growth hormone deficiency. (18) 250 µg cosyntropin stimulation test for primary adrenal insufficiency. (19) 250 µg cosyntropin stimulation test for secondary adrenal insufficiency. (20) 123I-MIBG for localization of pheochromocytoma. (21) Endoscopic ultrasound for pancreatic neuroendocrine tumors. (22) Chromogranin A for neuroendocrine tumors. (23) Gallium-68 somatostatin receptor PET and PET/CT for thoracic and gastroenteropancreatic neuroendocrine tumors. (24) 18F DOPA PET or PET/CT for thoracic and gastroenteropancreatic neuroendocrine tumor. (25) 18F DOPA PET or PET/CT for pheochromocytoma/paraganglioma. (26) Telemedicine for diabetic retinopathy (absence). (27) Telemedicine for diabetic macular edema. (28) Hemoglobin A1C for postpartum abnormal glucose tolerance. (29) HbA1c+ 2 SD for diabetes. (30) Albumin urine concentration for microalbuminuria. (31) Ratio of albumin to creatinine for microalbuminuria. (32) Plaster for diabetic neuropathy. (33) Hemoglobin A1C for gestational diabetes. (34) QTc prolongation for diabetic autonomic failure. (35) MICRAL dipstick for diabetic microalbuminuria. (36) Probe to bone test for osteomyelitis in patients with diabetes. (37) Plain radiography for osteomyelitis in patients with diabetes. (38) MRI for osteomyelitis in patients with diabetes. (39) Bone scan Tc for osteomyelitis in patients with diabetes. (40) Bone scan for osteomyelitis in patients with diabetes. (41) Optical coherence tomography for diabetic macular edema. (42) 18F-dihydroxyphenylalanine positron emission tomography for congenital hyperinsulinism. (43) Pancreatic venous sampling for focal congenital hyperinsulinism. (44) 18F-DOPA PET for focal congenital hyperinsulinism. (45) Real-time elastography for thyroid cancer. (46) Core needle Biopsy for thyroid malignancy. (47) Bethesda system for reporting thyroid cytopathology (only malignant and benign) for thyroid cancer. (48) Frozen section for follicular lesions, thyroid cancer. (49) Frozen section for non-follicular lesions (thyroid cancer). (50) Frozen section for thyroid lesions, not otherwise specified (thyroid cancer). (51) Contrast-enhanced ultrasound for thyroid cancer. (52) MicroRNA for thyroid cancer. (53) Thyroid ultrasound features (taller than wide) for thyroid cancer. (54) Thyroid US features spongiform for thyroid cancer. (55) Standardized uptake value (SUV_{max}) for thyroid cancer. (56) Thyroglobulin washout for thyroid cancer. (57) Diffusion weighted MR imaging for thyroid cancer. (58) Serum thyroglobulin with immunometric assay (with ablation, during thyroxine) for thyroid cancer recurrence. (59) B-RAF V600E for thyroid cancer. (60) 99mTC methoxyisobutylisonitrile scintigraphy for thyroid cancer. (61) FDG PET for thyroid cancer. (62) GAL-3 for thyroid cancer. (63) HBME-1 for thyroid cancer. (64) 3G TRAB assay for Graves' disease. (65) Ultrasonography for cervical lymph node metastases in papillary thyroid cancer

areas where further research is needed. These benefits can be hampered if the methodology used to perform and report these reviews is not adequate. Our review suggests multiple areas where authors of SRs of diagnostic tests could improve such as: author contact, exploration of heterogeneity (e.g., sensitivity analysis), use of advanced statistical methods, reporting of the overall risk of bias of included studies, and assessing the confidence in the body of the evidence.

We performed subgroup analyses to identify those studies that would be more useful in clinical practice; however, this was limited to the included tests (useful tests for which a SR has not been done are missing) and those that provided information that allowed us to calculate a likelihood ratio. Diagnostic accuracy estimates and risk of bias assessment are just a few of the components that can be used to determine if a test would actually benefit a patient. Diagnostic test studies are not usually linked to a patient

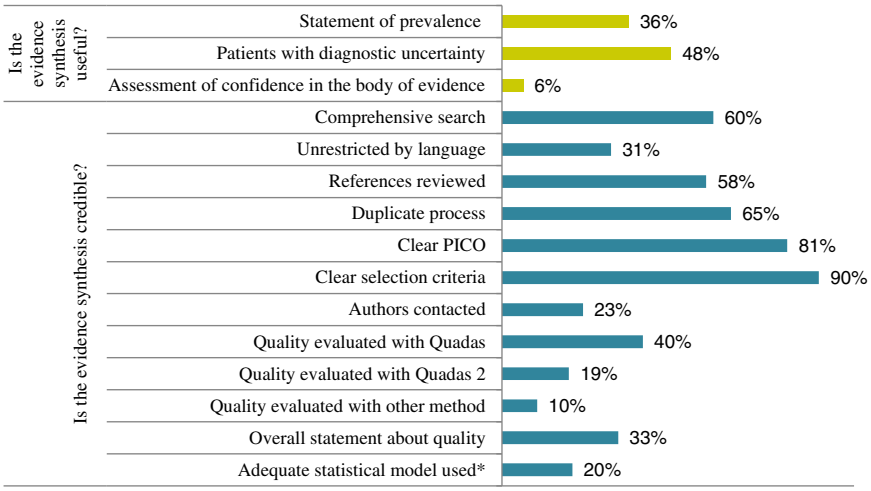


Fig. 5 Is the evidence synthesis credible and useful? *Asterisk* only applies to meta-analyses

important outcome; they provide only indirect evidence of potential benefits that depend on other factors such as: availability of treatment for a potential disease, patient context, and so on [105].

This SR audit summarizes the accuracy of diagnostic tests in endocrinology, the reporting of these reviews, and helps to identify knowledge gaps in diagnosis. Several limitations exist to the application of our results in clinical practice. First, our search may have missed SRs not indexed under endocrinology. Second, we did not assess the quality of studies included in SRs and relied (when available) on authors' assessments of quality for our analysis. Importantly, while best methods to perform and report a SR and meta-analysis of healthcare interventions are delineated by PRISMA [106], best standards for SRs of diagnostic tests may differ [107].

Conclusions

Almost half of the tests in which a LR could be calculated or was provided produced significant variations in the pretest probability of disease and are very likely to be helpful when positive; ~25% of the tests are very likely to be helpful when negative. In general, most reviews of diagnostic tests in endocrinology followed acceptable general methods for study identification, screening, and extraction. Most of the reviews reported evaluating the risk of bias (rarely providing an overall statement), but only 23% contacted authors and 20% used adequate statistical methods. As a result, the overall confidence in the diagnostic estimates provided by these studies is limited.

Progress in the field of diagnostic tests in endocrinology should be supported by standardized methods and reporting for SRs and meta-analyses. These standards should include use of adequate statistical methods and overall statements about the confidence in the body of evidence evaluating these tests. These evidence summaries should provide evidence to help clinicians and patients discuss the usefulness of the tests and the trustworthiness in the evidence producing these estimates.

Acknowledgements G.S.B. was supported by CTSA Grant Number TL1TR000137 from the National Center for Advancing Translational Science (NCATS) and Grant Number 3R01HL131535-01S1 from the National Heart Lung and Blood Institute (NHLBI). V.M.M. was partially supported by Grant Number UL1TR000135 from the NCATS, a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the author and do not necessarily represent the official view of the NIH.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. T.C. Hoffmann, C. Del Mar: Clinicians' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA Intern. Med. (2017). doi:10.1001/jamainternmed.2016.8254
2. J.R. Ball, E. Balogh, Improving diagnosis in health care: highlights of a report from the national academies of sciences, engineering, and medicine. Ann. Intern. Med. **164**(1), 59–61 (2016). doi:10.7326/m15-2256
3. P. Tricoci, J.M. Allen, J.M. Kramer, R.M. Califf, S.C. Smith Jr., Scientific evidence underlying the ACC/AHA clinical practice guidelines. J. Am. Med. Assoc. **301**(8), 831–841 (2009). doi:10.1001/jama.2009.205
4. N. Singh Ospina, R. Rodriguez-Gutierrez, J.P. Brito, W.F. Young Jr., V.M. Montori, Is the endocrine research pipeline broken? A systematic evaluation of the endocrine society clinical practice guidelines and trial registration. BMC Med. **13**, 187 (2015). doi:10.1186/s12916-015-0435-z

5. L. Ge, J.C. Wang, J.L. Li, L. Liang, N. An, X.T. Shi, Y.C. Liu, J. H. Tian, The assessment of the quality of reporting of systematic reviews/meta-analyses in diagnostic tests published by authors in China. PLoS ONE **9**(1), e85908 (2014). doi:10.1371/journal.pone.0085908
6. S. Mallett, J.J. Deeks, S. Halligan, S. Hopewell, V. Cornelius, D. G. Altman, Systematic reviews of diagnostic tests in cancer: review of methods and reporting. Br. Med. J. **333**(7565), 413 (2006). doi:10.1136/bmj.38895.467130.55
7. T. McGinn, P.C. Wyer, T.B. Newman, S. Keitz, R. Leipzig, G.G. For, Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). Can. Med. Assoc. J. **171** (11), 1369–1373 (2004). doi:10.1503/cmaj.1031981
8. L.M. Bachmann, D.B. Bischof, S.A. Bischofberger, M.G. Bonani, F.M. Osann, J. Steurer, Systematic quantitative overviews of the literature to determine the value of diagnostic tests for predicting acute appendicitis: study protocol. BMC Surg. **2**, 2 (2002)
9. P. Whiting, A.W. Rutjes, J. Dinnes, J.B. Reitsma, P.M. Bossuyt, J. Kleijnen, A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. J. Clin. Epidemiol. **58**(1), 1–12 (2005). doi:10.1016/j.jclinepi.2004.04.008
10. R.M. Harbord, P. Whiting, J.A. Sterne, M. Egger, J.J. Deeks, A. Shang, L.M. Bachmann, An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. J. Clin. Epidemiol. **61**(11), 1095–1103 (2008). doi:10.1016/j.jclinepi.2007.09.013
11. M. Gotthardt, B. Lohmann, T.M. Behr, A. Bauhofer, C. Franzius, M.L. Schipper, M. Wagner, H. Hoffken, H. Sitter, M. Rothmund, K. Joseph, C. Nies, Clinical value of parathyroid scintigraphy with technetium-99m methoxyisobutylisonitrile: discrepancies in clinical data and a systematic metaanalysis of the literature. World J. Surg. **28**(1), 100–107 (2004)
12. S. Nayak, I. Olkin, H. Liu, M. Grabe, M.K. Gould, I.E. Allen, D. K. Owens, D.M. Bravata, Meta-analysis: accuracy of quantitative ultrasound for identifying patients with osteoporosis. Ann. Intern. Med. **144**(11), 832–841 (2006)
13. C. Caldarella, G. Treglia, M.A. Isgro, A. Giordano, Diagnostic performance of positron emission tomography using 11C-methionine in patients with suspected parathyroid adenoma: a meta-analysis. Endocrine **43**(1), 78–83 (2013). doi:10.1007/s12020-012-9746-4
14. C. Caldarella, G. Treglia, A. Pontecorvi, A. Giordano, Diagnostic performance of planar scintigraphy using 99mTc-MIBI in patients with secondary hyperparathyroidism: a meta-analysis. Ann. Nucl. Med. **26**(10), 794–803 (2012). doi:10.1007/s12149-012-0643-y
15. K. Cheung, T.S. Wang, F. Farrokhyar, S.A. Roman, J.A. Sosa, A meta-analysis of preoperative localization techniques for patients with primary hyperparathyroidism. Ann. Surg. Oncol. **19**(2), 577–583 (2012). doi:10.1245/s10434-011-1870-5
16. W.J. Wei, C.T. Shen, H.J. Song, Z.L. Qiu, Q.Y. Luo, Comparison of SPET/CT, SPET and planar imaging using 99mTc-MIBI as independent techniques to support minimally invasive parathyroidectomy in primary hyperparathyroidism: a meta-analysis. Hell. J. Nucl. Med. **18**(2), 127–135 (2015). doi:10.1967/s002449910207
17. K.K. Wong, L.M. Fig, M.D. Gross, B.A. Dwamena, Parathyroid adenoma localization with 99mTc-sestamibi SPECT/CT: a meta-analysis. Nucl. Med. Commun. **36**(4), 363–375 (2015). doi:10.1097/MNM.0000000000000262
18. J. Windeler, J. Kobberling, The fructosamine assay in diagnosis and control of diabetes mellitus scientific evidence for its clinical usefulness? J. Clin. Chem. Clin. Biochem. **28**(3), 129–138 (1990)
19. J.E. Jensen, S.H. Nielsen, L. Foged, S.N. Holmegaard, E. Magid, The MICRAL test for diabetic microalbuminuria: predictive values as a function of prevalence. Scand. J. Clin. Lab. Invest. **56** (2), 117–122 (1996)
20. A.L. Peters, M.B. Davidson, D.L. Schriger, V. Hasselblad, A clinical approach for the diagnosis of diabetes mellitus: an analysis using glycosylated hemoglobin levels. Metaanalysis Research Group on the Diagnosis of Diabetes Using Glycated Hemoglobin Levels. [Erratum appears in JAMA 1997 Apr 9;277 (14):1125]. J. Am. Med. Assoc. **276**(15), 1246–1252 (1996)
21. E.A. Whitsel, E.J. Boyko, D.S. Siscovick, Reassessing the role of QTc in the diagnosis of autonomic failure among patients with diabetes: a meta-analysis. Diabetes Care **23**(2), 241–247 (2000)
22. B. Ewald, J. Attia, Which test to detect microalbuminuria in diabetic patients? A systematic review. Aust. Fam. Physician **33** (7), 565–567, 571 (2004).
23. G. Virgili, F. Menchini, A.F. Dimastrogiovanni, E. Rapizzi, U. Menchini, F. Bandello, R.G. Chiodini, Optical coherence tomography versus stereoscopic fundus photography or biomicroscopy for diagnosing diabetic macular edema: a systematic review. Invest. Ophthalmol. Vis. Sci. **48**(11), 4963–4973 (2007)
24. M.T. Dinh, C.L. Abad, N. Safdar, Diagnostic accuracy of the physical examination and imaging tests for osteomyelitis underlying diabetic foot ulcers: meta-analysis. Clin. Infect. Dis. **47**(4), 519–527 (2008). doi:10.1086/590011
25. Y. Feng, F.J. Schlosser, B.E. Sumpio, The Semmes Weinstein monofilament examination as a screening tool for diabetic peripheral neuropathy. J. Vasc. Surg. **50**(3), 675–682, 682.e671 (2009). doi:10.1016/j.jvs.2009.05.017.
26. B.A. Blomberg, M.C. Moghbel, B. Saboury, C.A. Stanley, A. Alavi, The value of radiologic interventions and 18F-DOPA PET in diagnosing and localizing focal congenital hyperinsulinism: systematic review and meta-analysis. Mol. Imaging Biol. **15**(1), 97–105 (2013). doi:10.1007/s11307-012-0572-0
27. Y. Ye, H. Xie, X. Zhao, S. Zhang, The oral glucose tolerance test for the diagnosis of diabetes mellitus in patients during acute coronary syndrome hospitalization: a meta-analysis of diagnostic test accuracy. Cardiovasc. Diabetol. **11**, 155 (2012). doi:10.1186/1475-2840-11-155
28. S.T. Tang, Q. Zhang, C.J. Wang, H.Q. Tang, T.X. Wu, [Glycosylated hemoglobin A1c for the diagnosis of diabetes mellitus: a meta-analysis]. Chung-Hua Nei Ko Tsa Chih Chinese. J. Intern. Med. **52**(1), 21–25 (2013)
29. Q.W. Tian, C. Xuan, H.W. Wang, J.X. Zhao, W.L. Yu, G. Gao, B.B. Zhang, L.M. Lun, Diagnostic accuracy of glycosylated hemoglobin in Chinese patients with gestational diabetes mellitus: a meta-analysis based on 2,812 patients and 5,918 controls. Genet. Test. Mol. Biomarkers **17**(9), 687–695 (2013). doi:10.1089/gtmb.2013.0099
30. S. Yan, S. Liu, Y. Zhao, W. Zhang, X. Sun, J. Li, F. Jiang, J. Ju, N. Lang, Y. Zhang, W. Zhou, Q. Li, Diagnostic accuracy of HbA1c in diabetes between Eastern and Western. Eur. J. Clin. Invest. **43**(7), 716–726 (2013). doi:10.1111/eci.12098
31. J. Yang, R. Hao, X. Zhu, Diagnostic role of 18F-dihydroxyphenylalanine positron emission tomography in patients with congenital hyperinsulinism: a meta-analysis. Nucl. Med. Commun. **34**(4), 347–353 (2013). doi:10.1097/MNM.0b013e32835e6ac6
32. G. Hirschfeld, Mv Glischinski, M. Blankenburg, B. Zernikow, Screening for peripheral neuropathies in children with diabetes: a systematic review. Pediatrics **133**(5), e1324–e1330 (2014). doi:10.1542/peds.2013-3645
33. X. Su, Z. Zhang, X. Qu, Y. Tian, G. Zhang, Hemoglobin A1c for diagnosis of postpartum abnormal glucose tolerance among women with gestational diabetes mellitus: diagnostic meta-

analysis. PLoS ONE **9**(7), e102144 (2014). doi:10.1371/journal.pone.0102144
34. A. Tsapas, A. Liakos, P. Paschos, T. Karagiannis, E. Bekiari, N. Tentolouris, P. Boura, A simple plaster for screening for diabetic neuropathy: a diagnostic test accuracy systematic review and meta-analysis. Metabolism **63**(4), 584–592 (2014). doi:10.1016/j.metabol.2013.11.019
35. H.Y. Wu, Y.S. Peng, C.K. Chiang, J.W. Huang, K.Y. Hung, K. D. Wu, Y.K. Tu, K.L. Chien, Diagnostic performance of random urine samples using albumin concentration vs ratio of albumin to creatinine for microalbuminuria screening in patients with diabetes mellitus: a systematic review and meta-analysis. JAMA Intern. Med. **174**(7), 1108–1115 (2014). doi:10.1001/jamainternmed.2014.1363
36. N. Xu, H. Wu, D. Li, J. Wang, Diagnostic accuracy of glycated hemoglobin compared with oral glucose tolerance test for diagnosing diabetes mellitus in Chinese adults: a meta-analysis. Diabetes Res. Clin. Pract. **106**(1), 11–18 (2014). doi:10.1016/j.diabres.2014.04.010
37. L. Shi, H. Wu, J. Dong, K. Jiang, X. Lu, J. Shi, Telemedicine for detecting diabetic retinopathy: a systematic review and meta-analysis. Br. J. Ophthalmol. **99**(6), 823–831 (2015). doi:10.1136/bjophthalmol-2014-305631
38. G. Treglia, F. Bertagna, R. Sadeghi, F.A. Verburg, L. Ceriani, L. Giovanella, Focal thyroid incidental uptake detected by 18F-fluorodeoxyglucose positron emission tomography. Meta-analysis on prevalence and malignancy risk. Nucl. Med. **52**(4), 130–136 (2013). doi:10.3413/Nukmed-0568-13-03
39. S.R. Puli, N. Kalva, M.L. Bechtold, S.R. Pamulaparthy, M.D. Cashman, N.C. Estes, R.H. Pearl, F.H. Volmar, S. Dillon, M.F. Shekleton, D. Forcione, Diagnostic accuracy of endoscopic ultrasound in pancreatic neuroendocrine tumors: a systematic review and meta analysis. World J. Gastroenterol. **19**(23), 3678–3684 (2013). doi:10.3748/wjg.v19.i23.3678
40. X. Yang, Y. Yang, Z. Li, C. Cheng, T. Yang, C. Wang, L. Liu, S. Liu, Diagnostic value of circulating chromogranin a for neuroendocrine tumors: a systematic review and meta-analysis. PLoS ONE **10**(4), e124884 (2015). doi:10.1371/journal.pone.0124884
41. G.W. Boland, M.J. Lee, G.S. Gazelle, E.F. Halpern, M.M. McNicholas, P.R. Mueller, Characterization of adrenal masses using unenhanced CT: an analysis of the CT literature. Am. J. Roentgenol. **171**(1), 201–204 (1998)
42. R.I. Dorin, C.R. Qualls, L.M. Crapo, Diagnosis of adrenal insufficiency. Ann. Intern. Med. **139**(3), 194–204 (2003)
43. M.B. Elamin, M.H. Murad, R. Mullan, D. Erickson, K. Harris, S. Nadeem, R. Ennis, P.J. Erwin, V.M. Montori, Accuracy of diagnostic tests for Cushing's syndrome: a systematic review and metaanalyses. J. Clin. Endocrinol. Metab. **93**(5), 1553–1562 (2008). doi:10.1210/jc.2008-0139
44. R. Kazlauskaite, A.T. Evans, C.V. Villabona, T.A. Abdu, B. Ambrosi, A.B. Atkinson, C.H. Choi, R.N. Clayton, C.H. Courtney, E.N. Gonc, M. Maghnie, S.R. Rose, S.G. Soule, K. Tordjman; Adrenal, I.C.f.E.o.C.T.i.H.-P., Corticotropin tests for hypothalamic-pituitary-adrenal insufficiency: a metaanalysis. J. Clin. Endocrinol. Metab. **93**(11), 4245–4253 (2008). doi:10.1210/jc.2008-0710
45. T. Carroll, H. Raff, J.W. Findling, Late-night salivary cortisol for the diagnosis of Cushing syndrome: a meta-analysis. Endocr. Pract. **15**(4), 335–342 (2009). doi:10.4158/EP09023OR
46. A.F. Jacobson, H. Deng, J. Lombard, H.J. Lessig, R.R. Black, 123I-meta-iodobenzylguanidine scintigraphy for the detection of neuroblastoma and pheochromocytoma: results of a meta-analysis. J. Clin. Endocrinol. Metab. **95**(6), 2596–2606 (2010). doi:10.1210/jc.2009-2604
47. G.W. Boland, B.A. Dwamena, M.J. Sangwaiya, A.G. Goehler, M.A. Blake, P.F. Hahn, J.A. Scott, M.K. Kalra, Characterization of adrenal masses by using FDG PET: a systematic review and meta-analysis of diagnostic test performance. Radiology **259**(1), 117–126 (2011). doi:10.1148/radiol.11100569
48. A. Hazem, M.B. Elamin, G. Malaga, I. Bancos, Y. Prevost, C. Zeballos-Palacios, E.R. Velasquez, P.J. Erwin, N. Natt, V.M. Montori, M.H. Murad, The accuracy of diagnostic tests for GH deficiency in adults: a systematic review and meta-analysis. Eur. J. Endocrinol. **165**(6), 841–849 (2011). doi:10.1530/EJE-11-0476
49. S. Iliodromiti, T.W. Kelsey, R.A. Anderson, S.M. Nelson, Can anti-Mullerian hormone predict the diagnosis of polycystic ovary syndrome? A systematic review and meta-analysis of extracted data. J. Clin. Endocrinol. Metab. **98**(8), 3332–3340 (2013). doi:10.1210/jc.2013-1393
50. V. Rufini, G. Treglia, P. Castaldi, G. Perotti, A. Giordano, Comparison of metaiodobenzylguanidine scintigraphy with positron emission tomography in the diagnostic work-up of pheochromocytoma and paraganglioma: a systematic review. Q. J. Nucl. Med. Mol. Imaging **57**(2), 122–133 (2013)
51. Y. Shen, J. Zhang, Y. Zhao, Y. Yan, Y. Liu, J. Cai, Diagnostic value of serum IGF-1 and IGFBP-3 in growth hormone deficiency: a systematic review with meta-analysis. Eur. J. Pediatr. **174**(4), 419–427 (2015). doi:10.1007/s00431-014-2406-3
52. C.F. Eustatia-Rutten, J.W. Smit, J.A. Romijn, E.Pvd Kleij-Corssmit, A.M. Pereira, M.P. Stokkel, J. Kievit, Diagnostic value of serum thyroglobulin measurements in the follow-up of differentiated thyroid carcinoma, a structured meta-analysis. Clin. Endocrinol. **61**(1), 61–74 (2004)
53. V. Rufini, G. Treglia, F. Montravers, A. Giordano, Diagnostic accuracy of [18F]DOPA PET and PET/CT in patients with neuroendocrine tumors: a meta-analysis. Clin. Transl. Imaging **1**(2), 111–122 (2013). doi:10.1007/s40336-013-0005-3
54. L. Peng, M.J. Gu, Diagnostic value of conventional and ultrasound-guided fine-needle aspiration biopsy for thyroid nodules: a meta-analysis. [Chinese]. Acad. J. Second Mil. Med. Univ. **28**(9), 968–972 (2007)
55. Y. Peng, H.H. Wang, A meta-analysis of comparing fine-needle aspiration and frozen section for evaluating thyroid nodules. Diagn. Cytopathol. **36**(12), 916–920 (2008). doi:10.1002/dc.20943
56. P.G. Raijmakers, M.A. Paul, P. Lips, Sentinel node detection in patients with thyroid carcinoma: a meta-analysis. World J. Surg. **32**(9), 1961–1967 (2008). doi:10.1007/s00268-008-9657-y
57. M.J. Dong, Z.F. Liu, K. Zhao, L.X. Ruan, G.L. Wang, S.Y. Yang, F. Sun, X.G. Luo, Value of 18F-FDG-PET/PET-CT in differentiated thyroid carcinoma with radioiodine-negative whole-body scan: a meta-analysis. Nucl. Med. Commun. **30**(8), 639–650 (2009). doi:10.1097/MNM.0b013e32832dcfa7
58. C. Stevens, J.K. Lee, M. Sadatsafavi, G.K. Blair, Pediatric thyroid fine-needle aspiration cytology: a meta-analysis. J. Pediatr. Surg. **44**(11), 2184–2191 (2009). doi:10.1016/j.jpedsurg.2009.07.022
59. J. Bojunga, E. Herrmann, G. Meyer, S. Weber, S. Zeuzem, M. Friedrich-Rust, Real-time elastography for the differentiation of benign and malignant thyroid nodules: a meta-analysis. Thyroid **20**(10), 1145–1150 (2010). doi:10.1089/thy.2010.0079
60. W. Iared, D.C. Shigueoka, J.C. Cristofoli, R. Andriolo, A.N. Atallah, S.A. Ajzen, O. Valente, Use of color Doppler ultrasonography for the prediction of malignancy in follicular thyroid neoplasms: systematic review and meta-analysis. J. Ultrasound Med. **29**(3), 419–425 (2010)
61. M.E. Miller, Q. Chen, D. Elashoff, E. Abemayor, M.S. John, Positron emission tomography and positron emission tomography-CT evaluation for recurrent papillary thyroid

carcinoma: meta-analysis and literature review. Head Neck **33** (4), 562–565 (2011). doi:10.1002/hed.21492
62. D. Vriens, J.Hd Wilt, G.Jvd Wilt, R.T. Netea-Maier, W.J. Oyen, L.Fd Geus-Oei, The role of [18F]-2-fluoro-2-deoxy-d-glucose-positron emission tomography in thyroid nodules with indeterminate fine-needle aspiration biopsy: systematic review and meta-analysis of the literature. Cancer **117**(20), 4582–4594 (2011). doi:10.1002/cncr.26085
63. M. Bongiovanni, A. Spitale, W.C. Faquin, L. Mazzucchelli, Z. W. Baloch, The Bethesda system for reporting thyroid cytopathology: a meta-analysis. Acta Cytol. **56**(4), 333–339 (2012). doi:10.1159/000339959
64. X. Cheng, L. Bao, Z. Xu, D. Li, J. Wang, Y. Li, 18F-FDG-PET and 18F-FDG-PET/CT in the detection of recurrent or metastatic medullary thyroid carcinoma: a systematic review and meta-analysis. J. Med. Imaging Radiat. Oncol. **56**(2), 136–142 (2012). doi:10.1111/j.1754-9485.2012.02344.x
65. L.Ld Matos, A.B.D. Giglio, C.O. Matsubayashi, Md.L. Farah, A. D. Giglio, M.Ad.S. Pinhal, Expression of CK-19, galectin-3 and HBME-1 in the differentiation of thyroid lesions: systematic review and diagnostic meta-analysis. Diagn. Pathol. **7**, 97 (2012). doi:10.1186/1746-1596-7-97
66. R. Tozzoli, M. Bagnasco, D. Giavarina, N. Bizzaro, TSH receptor autoantibody immunoassay in patients with Graves' disease: improvement of diagnostic accuracy over different generations of methods. Systematic review and meta-analysis. Autoimmun. Rev. **12**(2), 107–113 (2012). doi:10.1016/j.autrev.2012.07.003
67. L.M. Wu, H.Y. Gu, X.H. Qu, J. Zheng, W. Zhang, Y. Yin, J.R. Xu, The accuracy of ultrasonography in the preoperative diagnosis of cervical lymph node metastasis in patients with papillary thyroid carcinoma: a meta-analysis. Eur. J. Radiol. **81**(8), 1798–1805 (2012). doi:10.1016/j.ejrad.2011.04.028
68. S.A. Razavi, T.A. Hadduck, G. Sadigh, B.A. Dwamena, Comparative effectiveness of elastographic and B-mode ultrasound criteria for diagnostic discrimination of thyroid nodules: a meta-analysis. Am. J. Roentgenol. **200**(6), 1317–1326 (2013). doi:10.2214/AJR.12.9215
69. G. Treglia, C. Caldarella, E. Saggiorato, L. Ceriani, F. Orlandi, M. Salvatori, L. Giovanella, Diagnostic performance of (99m) Tc-MIBI scan in predicting the malignancy of thyroid nodules: a meta-analysis. Endocrine **44**(1), 70–78 (2013). doi:10.1007/s12020-013-9932-z
70. N. Wang, H. Zhai, Y. Lu, Is fluorine-18 fluorodeoxyglucose positron emission tomography useful for the thyroid nodules with indeterminate fine needle aspiration biopsy? A meta-analysis of the literature. [Erratum appears in J Otolaryngol Head Neck Surg.2014;43():43]. J. Otolaryngol. Head Neck Surg. **42**, 38 (2013). doi:10.1186/1916-0216-42-38
71. B. Zhang, X. Ma, N. Wu, L. Liu, X. Liu, J. Zhang, J. Yang, T. Niu, Shear wave elastography for differentiation of benign and malignant thyroid nodules: a meta-analysis. J. Ultrasound Med. **32**(12), 2163–2169 (2013). doi:10.7863/ultra.32.12.2163
72. J.P. Brito, M.R. Gionfriddo, A.A. Nofal, K.R. Boehmer, A.L. Leppin, C. Reading, M. Callstrom, T.A. Elraiyah, L.J. Prokop, M.N. Stan, M.H. Murad, J.C. Morris, V.M. Montori, The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. J. Clin. Endocrinol. Metab. **99**(4), 1253–1263 (2014). doi:10.1210/jc.2013-2928
73. M. Ghajarzadeh, F. Sodagari, M. Shakiba, Diagnostic accuracy of sonoelastography in detecting malignant thyroid nodules: a systematic review and meta-analysis. Am. J. Roentgenol. **202**(4), W379–W389 (2014). doi:10.2214/AJR.12.9785
74. L. Giovanella, G. Treglia, R. Sadeghi, P. Trimboli, L. Ceriani, F. A. Verburg, Unstimulated highly sensitive thyroglobulin in follow-up of differentiated thyroid cancer patients: a meta-analysis. J. Clin. Endocrinol. Metab. **99**(2), 440–447 (2014). doi:10.1210/jc.2013-3156
75. G. Grani, A. Fumarola, Thyroglobulin in lymph node fine-needle aspiration washout: a systematic review and meta-analysis of diagnostic accuracy. J. Clin. Endocrinol. Metab. **99**(6), 1970–1982 (2014). doi:10.1210/jc.2014-1098
76. Y. Jia, Y. Yu, X. Li, S. Wei, X. Zheng, X. Yang, J. Zhao, T. Xia, M. Gao, Diagnostic value of B-RAF(V600E) in difficult-to-diagnose thyroid nodules using fine-needle aspiration: systematic review and meta-analysis. Diagn. Cytopathol. **42**(1), 94–101 (2014). doi:10.1002/dc.23044
77. L. Li, B.D. Chen, H.F. Zhu, S. Wu, D. Wei, J.Q. Zhang, L. Yu, Comparison of pre-operation diagnosis of thyroid cancer with fine needle aspiration and core-needle biopsy: a meta-analysis. Asian Pac. J. Cancer Prev. **15**(17), 7187–7193 (2014)
78. Y.Y. Ma, X.G. Zhang, P. Paerhati, Y. Mu, N. Tieliewuhan, Value of ultrasonographic elastography in differential diagnosis of benign/malignant thyroid nodules in China: a meta-analysis. [Chinese]. Chin. J. Evid. Based Med. **14**(5), 584–591 (2014). doi:10.7507/1672-2531.20140097
79. N. Qu, L. Zhang, Z.W. Lu, W.J. Wei, Y. Zhang, Q.H. Ji, Risk of malignancy in focal thyroid lesions identified by (18)F-fluorodeoxyglucose positron emission tomography or positron emission tomography/computed tomography: evidence from a large series of studies. Tumour Biol. **35**(6), 6139–6147 (2014). doi:10.1007/s13277-014-1813-4
80. B.S. Sheffield, H. Masoudi, B. Walker, S.M. Wiseman, Preoperative diagnosis of thyroid nodules using the Bethesda system for reporting thyroid cytopathology: a comprehensive review and meta-analysis. Expert Rev. Endocrinol. Metab. **9**(2), 97–110 (2014). doi:10.1586/17446651.2014.887435
81. J. Sun, J. Cai, X. Wang, Real-time ultrasound elastography for differentiation of benign and malignant thyroid nodules: a meta-analysis. J. Ultrasound Med. **33**(3), 495–502 (2014). doi:10.7863/ultra.33.3.495
82. A. Wale, K.A. Miles, B. Young, C. Zammit, A. Williams, J. Quin, S. Dizdarevic, Combined (99m)Tc-methoxyisobutylisonitrile scintigraphy and fine-needle aspiration cytology offers an accurate and potentially cost-effective investigative strategy for the assessment of solitary or dominant thyroid nodules. Eur. J. Nuclear Med. Mol. Imaging **41**(1), 105–115 (2014). doi:10.1007/s00259-013-2546-0
83. X. Wei, Y. Li, S. Zhang, M. Gao, Thyroid imaging reporting and data system (TI-RADS) in the diagnostic value of thyroid nodules: a systematic review. Tumour Biol. **35**(7), 6769–6776 (2014). doi:10.1007/s13277-014-1837-9
84. K. Wolinski, M. Szkudlarek, E. Szczepanek-Parulska, M. Ruchala, Usefulness of different ultrasound features of malignancy in predicting the type of thyroid lesions: a meta-analysis of prospective studies. Pol. Arch. Med. Wewn. **124**(3), 97–104 (2014)
85. D. Yu, Y. Han, T. Chen, Contrast-enhanced ultrasound for differentiation of benign and malignant thyroid lesions: meta-analysis. Otolaryngol. Head Neck Surg. **151**(6), 909–915 (2014). doi:10.1177/0194599814555838
86. Y. Zhang, Q. Zhong, X. Chen, J. Fang, Z. Huang, Diagnostic value of microRNAs in discriminating malignant thyroid nodules from benign ones on fine-needle aspiration samples. Tumour Biol. **35**(9), 9343–9353 (2014). doi:10.1007/s13277-014-2209-1
87. F.J. Dong, M. Li, Y. Jiao, J.F. Xu, Y. Xiong, L. Zhang, H. Luo, Z.M. Ding, Acoustic radiation force impulse imaging for detecting thyroid nodules: a systematic review and pooled meta-analysis. Med. Ultrason. **17**(2), 192–199 (2015). doi:10.11152/mu.2013.2066.172.hyr
88. S. Nell, J.W. Kist, T.P. Debray, Bd Keizer, T.Jv Oostenbrugge, I. H.B. Rinkes, G.D. Valk, M.R. Vriens, Qualitative elastography

can replace thyroid nodule fine-needle aspiration in patients with soft thyroid nodules. A systematic review and meta-analysis. Eur. J. Radiol. **84**(4), 652–661 (2015). doi:10.1016/j.ejrad.2015.01.003
89. J.S. Pyo, J.H. Sohn, G. Kang, BRAF immunohistochemistry using clone VE1 is strongly concordant with BRAF(V600E) mutation test in papillary thyroid carcinoma. Endocr. Pathol. **26** (3), 211–217 (2015). doi:10.1007/s12022-015-9374-7
90. L.R. Remonti, C.K. Kramer, C.B. Leitao, L.C. Pinto, J.L. Gross, Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. Thyroid **25** (5), 538–550 (2015). doi:10.1089/thy.2014.0353
91. V. Veer, S. Puttagunta, The role of elastography in evaluating thyroid nodules: a literature review and meta-analysis. Eur. Arch. Otorhinolaryngol. **272**(8), 1845–1855 (2015). doi:10.1007/s00405-014-3155-7
92. G.J. Zhou, M. Xiao, L.N. Zhao, J.G. Tang, L. Zhang, MicroRNAs as novel biomarkers for the differentiation of malignant versus benign thyroid lesions: a meta-analysis. Genet. Mol. Res. **14**(3), 7279–7289 (2015). doi:10.4238/2015.July.3.3
93. G. Treglia, P. Castaldi, G. Rindi, A. Giordano, V. Rufini, Diagnostic performance of Gallium-68 somatostatin receptor PET and PET/CT in patients with thoracic and gastro-enteropancreatic neuroendocrine tumours: a meta-analysis. Endocrine **42**(1), 80–87 (2012). doi:10.1007/s12020-012-9631-1
94. P. Lin, M. Chen, B. Liu, S. Wang, X. Li, Diagnostic performance of shear wave elastography in the identification of malignant thyroid nodules: a meta-analysis. Eur. Radiol. **24**(11), 2729–2738 (2014). doi:10.1007/s00330-014-3320-9
95. L.M. Wu, X.X. Chen, Y.L. Li, J. Hua, J. Chen, J. Hu, J.R. Xu, On the utility of quantitative diffusion-weighted MR imaging as a tool in differentiation between malignant and benign thyroid nodules. Acad. Radiol. **21**(3), 355–363 (2014). doi:10.1016/j.acra.2013.10.008
96. H. Honest, K.S. Khan, Reporting of measures of accuracy in systematic reviews of diagnostic literature. BMC Health Serv. Res. **2**, 4 (2002)
97. Y. Ben-Shlomo, S.M. Collin, J. Quekett, J.A. Sterne, P. Whiting, Presentation of diagnostic information to doctors may change their interpretation and clinical management: a web-based randomised controlled trial. PLoS ONE **10**(7), e0128637 (2015). doi:10.1371/journal.pone.0128637
98. J.G. Lijmer, B.W. Mol, S. Heisterkamp, G.J. Bonsel, M.H. Prins, J.H. van der Meulen, P.M. Bossuyt, Empirical evidence of design-related bias in studies of diagnostic tests. J. Am. Med. Assoc. **282**(11), 1061–1066 (1999)
99. A.W. Rutjes, J.B. Reitsma, M. Di Nisio, N. Smidt, J.C. van Rijn, P.M. Bossuyt, Evidence of bias and variation in diagnostic accuracy studies. Can. Med. Assoc. J. **174**(4), 469–476 (2006). doi:10.1503/cmaj.050090
100. R.J. Mullan, D.N. Flynn, B. Carlberg, I.M. Tleyjeh, C.C. Kamath, M.L. LaBella, P.J. Erwin, G.H. Guyatt, V.M. Montori, Systematic reviewers commonly contact study authors but do so with limited rigor. J. Clin. Epidemiol. **62**(2), 138–142 (2009). doi:10.1016/j.jclinepi.2008.08.002
101. S.S. Selph, A.D. Ginsburg, R. Chou, Impact of contacting study authors to obtain additional data for systematic reviews: diagnostic accuracy studies for hepatic fibrosis. Syst. Rev. **3**, 107 (2014). doi:10.1186/2046-4053-3-107
102. M.M.G. Leeflang, A.W.S. Rutjes, J.B. Reitsma, L. Hooft, P.M. M. Bossuyt, Variation of a test's sensitivity and specificity with disease prevalence. Can. Med. Assoc. J. **185**(11), E537–E544 (2013). doi:10.1503/cmaj.121286
103. Y. Takwoingi, B. Guo, R.D. Riley, J.J. Deeks Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. Stat. Methods Med. Res. (2015). doi:10.1177/0962280215592269
104. E.A. Ochodo, J.B. Reitsma, P.M. Bossuyt, M.M.G. Leeflang, Survey revealed a lack of clarity about recommended methods for meta-analysis of diagnostic accuracy data. J. Clin. Epidemiol. **66**(11), 1281–1288 (2013). doi:10.1016/j.jclinepi.2013.05.015
105. H.J. Schunemann, A.D. Oxman, J. Brozek, P. Glasziou, R. Jaeschke, G.E. Vist, J.W. Williams Jr., R. Kunz, J. Craig, V.M. Montori, P. Bossuyt, G.H. Guyatt, G.W. Group, Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. Br. Med. J. **336**(7653), 1106–1110 (2008). doi:10.1136/bmj.39500.677199.AE
106. A. Liberati, D.G. Altman, J. Tetzlaff, C. Mulrow, P.C. Gotzsche, J.P. Ioannidis, M. Clarke, P.J. Devereaux, J. Kleijnen, D. Moher, The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J. Clin. Epidemiol. **62**(10), e1–e34 (2009). doi:10.1016/j.jclinepi.2009.06.006
107. W.L. Deville, F. Buntinx, L.M. Bouter, V.M. Montori, H.C. de Vet, D.A. van der Windt, P.D. Bezemer, Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med. Res. Methodol. **2**, 9 (2002)