

PROTOCOL

Open Access

# Diagnostic accuracy of fine needle aspiration biopsy for detection of malignancy in pediatric thyroid nodules: protocol for a systematic review and meta-analysis



Sarah W. Lai<sup>1,2,6\*</sup>, Derek J. Roberts<sup>1,3,7</sup>, Doreen M. Rabi<sup>4</sup> and Karin Y. Winston<sup>3,5</sup>

## Abstract

**Background:** Fine needle aspiration biopsy (FNAB) is an accurate test commonly used to determine whether thyroid nodules are malignant in adults. However, less is known about its diagnostic accuracy for this purpose in children, where conduct of FNAB is less frequent, more technically challenging, and pre-test probabilities of malignancy are often higher. The purpose of this systematic review is to evaluate the diagnostic accuracy of FNAB for the detection of malignancy in pediatric thyroid nodules.

**Methods:** We will search electronic bibliographic databases (MEDLINE, EMBASE, the Cochrane Library, and Evidence-Based Medicine) from their date of inception, reference lists of included articles, proceedings from relevant conferences, and the table of contents of the *Journal of Pediatric Surgery* (January 2007–present). Two reviewers will independently screen titles and abstracts and identify diagnostic accuracy studies involving FNAB of the thyroid in children. We will include studies comparing FNAB to a reference standard of surgical histopathology or clinical follow-up for detection of malignancy in pediatric thyroid nodules. Two investigators will independently extract data and assess risk of bias using the Quality of Diagnostic Accuracy Studies-II tool. Pooled estimates of sensitivity, specificity, and positive and negative likelihood ratios will be calculated using bivariate random-effects and hierarchical summary receiver operating characteristic models. In the presence of between-study heterogeneity, we will conduct stratified meta-analyses and meta-regression to determine whether diagnostic accuracy estimates vary by country of origin, use of ultrasound guidance during FNAB, qualifications of the individuals performing/interpreting FNAB, adherence to the Bethesda criteria for cytology classification, length of clinical follow-up, timing of data collection, patient selection methods, and presence of verification bias.

**Discussion:** This meta-analysis will determine the diagnostic accuracy of FNAB for detection of malignancy in pediatric thyroid nodules and explore whether heterogeneity observed across studies may be explained by variations in patient population, FNAB technique or interpretation, and/or study-level risks of bias. This will be the first study to determine the accuracy of Bethesda cytological classification levels of FNAB (benign, atypical, follicular, suspicious, malignant). We expect that our results will help in guiding clinical decision-making in children with thyroid nodules.

**Systematic review registration:** PROSPERO No. CRD42014007140

**Keywords:** Fine needle biopsy, Pediatric, Thyroid nodule, Thyroid cancer, Meta-analysis, Systematic review, Diagnostic accuracy, Sensitivity, Specificity, Likelihood ratio

\* Correspondence: dr.sarahlai@gmail.com

<sup>1</sup>Department of Surgery, University of Calgary, Calgary, AB, Canada

<sup>2</sup>Department of Gastrointestinal Sciences, University of Calgary, Calgary, AB, Canada

Full list of author information is available at the end of the article

## Background

Thyroid nodules are uncommon in children, with a prevalence ranging from 0.05 to 2% [1–6]. Nodules are more likely to be found in girls than boys and in adolescents compared to their younger counterparts [7, 8]. Although nodules have a low risk of malignant transformation in adults (5 to 15%), the incidence in pediatric patients is estimated to be as high as 70% [2, 5, 7, 9]. Risk factors for thyroid malignancy in children include family history of thyroid cancer, certain genetic mutations, and exposure to therapeutic or environmental irradiation.

Some authors have advocated that the increased malignant potential of thyroid nodules in children justifies the liberal use of surgical exploration in several pediatric populations [5]. However, although thyroid surgery is typically well-tolerated, the potential for associated complications deters many clinicians from proceeding directly to operation [10–12]. Risks of thyroid surgery include hypothyroidism, hypoparathyroidism, recurrent laryngeal nerve injury, and postoperative bleeding and infection. These risks increase during completion thyroidectomy if a malignancy is found after hemithyroidectomy [13]. Thus, an accurate diagnostic test is essential to facilitate pre-operative decisions regarding management of pediatric thyroid nodules.

Fine needle aspiration biopsy (FNAB), also known as fine needle aspiration cytology, has been used since the early 1980s to classify the cytology of (and thereby diagnose) suspicious superficial soft tissue lesions. Improvements in ultrasound (US) technology have led to increased detection of incidental thyroid nodules and, consequently, more frequent use of FNAB [14]. A generalist (family practitioner, pediatrician, or internist) or a specialist (endocrinologist, surgeon, radiologist, or pathologist) may perform this procedure, with or without US guidance (which, in theory, may lead to heightened accuracy and increased safety). As comfort levels with FNAB have increased, greater confidence in the accuracy of cytology results has reduced the number of thyroid surgeries for benign nodules [15–17]. However, most diagnostic accuracy studies of FNAB for prediction of malignancy in thyroid nodules have focused on adult subjects, leading pediatric clinicians to question whether its reported accuracy is generalizable to children [18–20].

In 2007, the Thyroid FNAB State of the Science Conference addressed the varying terminology in FNAB reporting, concluding that inconsistencies prevented comparisons of diagnoses across different sites. Prior to the conference, most pathologists classified FNAB cytology as inadequate, benign, malignant, or indeterminate using variable definitions. Discussions at this conference resulted in the publication of the Bethesda System for Reporting Thyroid Cytopathology (also known as the Bethesda criteria) in 2009. The Bethesda criteria classify

FNAB samples as non-diagnostic, benign, atypia/follicular lesion of undetermined significance, follicular neoplasm or suspicious for follicular neoplasm, suspicious for malignancy, or malignant [21]. The largest benefit of these criteria is that they clearly describe and link each of these categories to a risk of malignancy, facilitating prognostication and clinical decision-making regarding surgery or non-operative/conservative management [22]. After introduction of this classification scheme, the American Thyroid Association endorsed FNAB as the standard of care in North America for evaluation of thyroid nodules in their clinical practice guidelines [23].

Although a meta-analysis was published in 2009 evaluating the accuracy of FNAB for detection of malignancy in pediatric thyroid nodules, another systematic review is urgently required for several reasons [24]. First, multiple relevant articles have been published since the last review by Stevens et al. [24], potentially altering conclusions of the study. Second, their meta-analysis reviewed literature published prior to January 2007 (that is, before introduction of the Bethesda criteria) and included minimal data on the use of US guidance during FNAB. Third, Stevens et al. [24] did not directly address the risk of design-related biases among the included articles—biases that have previously been shown to overestimate the reported accuracy of a diagnostic test—potentially limiting or even preventing clinical application of their findings [24–26]. In particular, as clinicians may elect to follow patients clinically rather than proceed to thyroid surgery after a non-malignant FNAB result, this will prevent comparison against the gold standard of surgical histopathology. Thus, partial verification bias is expected to be a major limiting factor in pediatric FNAB diagnostic accuracy studies. As the previous study did not assess these potential sources of bias and heterogeneity, an updated and more elaborate systematic review and meta-analysis could verify or potentially refute the applicability of their findings to current pediatric clinical practices. The objective of this study is to systematically review the diagnostic accuracy of FNAB for the detection of thyroid malignancy.

## Methods

### Protocol

This study adopts recommendations on the conduct and reporting of systematic reviews and meta-analyses outlined by the Preferred Reporting Items in Systematic Reviews and Meta-Analyses statement, the Meta-Analysis of Observational Studies in Epidemiology proposal, and the Cochrane Diagnostic Test Accuracy Working Group [27–30]. The protocol is registered in the PROSPERO

International Prospective Register of Systematic Reviews (Registration No. CRD42014007140).

### Focused clinical question

In pediatric patients with a thyroid nodule, is FNAB as accurate as surgical histopathology or clinical follow-up for the detection of thyroid malignancy?

PICOD components

- Population
  - Patients  $\leq 18$  years of age, or those defined as exclusively pediatric patients by the authors, with a thyroid nodule that is palpable or seen on diagnostic imaging
- Intervention
  - FNAB of the thyroid nodule, with or without US guidance
- Comparison
  - Surgical histopathology or clinical follow-up
- Outcome
  - Test accuracy for detection of thyroid malignancy as defined by the authors, including true and false positives and negatives, sensitivity and specificity, and positive and negative likelihood ratios
- Design
  - Diagnostic accuracy studies [30]

Primary outcome

- Test accuracy of FNAB for as defined by the authors

Secondary outcomes

- Test accuracy of FNAB for classification of lesions according to the Bethesda criteria (non-diagnostic, benign, atypia/follicular lesion of undetermined significance, follicular neoplasm or suspicious for follicular neoplasm, suspicious for malignancy, malignant). This outcome was chosen as secondary instead of primary to allow for a comprehensive evaluation of the accuracy of FNAB for classifying thyroid nodules in children (according to both Bethesda and non-Bethesda criteria)
- Test accuracy of FNAB with or without US guidance for detection of thyroid malignancy

### Search strategy

We will search Ovid MEDLINE and EMBASE, the Cochrane Database of Systematic Reviews, and Evidence-Based Medicine from their date of first inception, without language, publication date, or other restrictions. PubMed will also be searched to capture articles not yet indexed in MEDLINE. We will also use the PubMed “related articles” feature for articles included in the systematic review and

manually search the table of contents for the *Journal of Pediatric Surgery* from January 2007 onward. To identify unpublished and/or ongoing studies, we will contact experts in the field and search clinical trials registries (ClinicalTrials.gov and Current Controlled Trials), reference lists of included articles, and conference proceedings of major pediatric surgery (American Pediatric Surgical Association, Canadian Association of Pediatric Surgeons, and Pacific Association of Pediatric Surgeons) and pediatric endocrinology (European Society for Pediatric Endocrinology and Pediatric Endocrine Society/Lawson Wilkins Pediatric Endocrine Society) meetings from 2007 to 2015.

With the assistance of an information scientist/medical librarian, we developed search filters encompassing the themes thyroid, biopsy, and pediatrics, using a combination of keywords and Medical Subject Heading (MeSH)/Emtree terms (Table 1). These three themes will be combined in MEDLINE and EMBASE using the Boolean operator “AND.” A diagnostic accuracy theme will not be used as it has been shown to potentially lead to the exclusion of relevant articles in systematic reviews of diagnostic accuracy studies [30–32]. A similar search strategy using themes and Boolean operators will be performed in remaining databases.

### Inclusion and exclusion criteria

After removing duplicate citations, two investigators (SWL, KYW) will independently screen all remaining titles and abstracts in duplicate. This initial screen will be broad intentionally to avoid missing potentially relevant citations. We will subsequently review the full text of any citations that appear to satisfy the following criteria:

- Patients  $\leq 18$  years of age or described to be pediatric by the author(s)
- FNAB performed on the thyroid

Those articles identified for full text review will subsequently be read independently in full by the same two investigators (SWL, KYW) to determine their eligibility for inclusion in the systematic review. We will use the following inclusion/exclusion criteria based on PICOD:

Inclusion criteria

- Population
  - The study population consisted of patients  $\leq 18$  years of age (or patient populations where the study authors did not provide summary estimates describing age, but did report that the included patients were exclusively children), with a thyroid nodule that is palpable or seen on diagnostic imaging

**Table 1** Electronic database search strategies

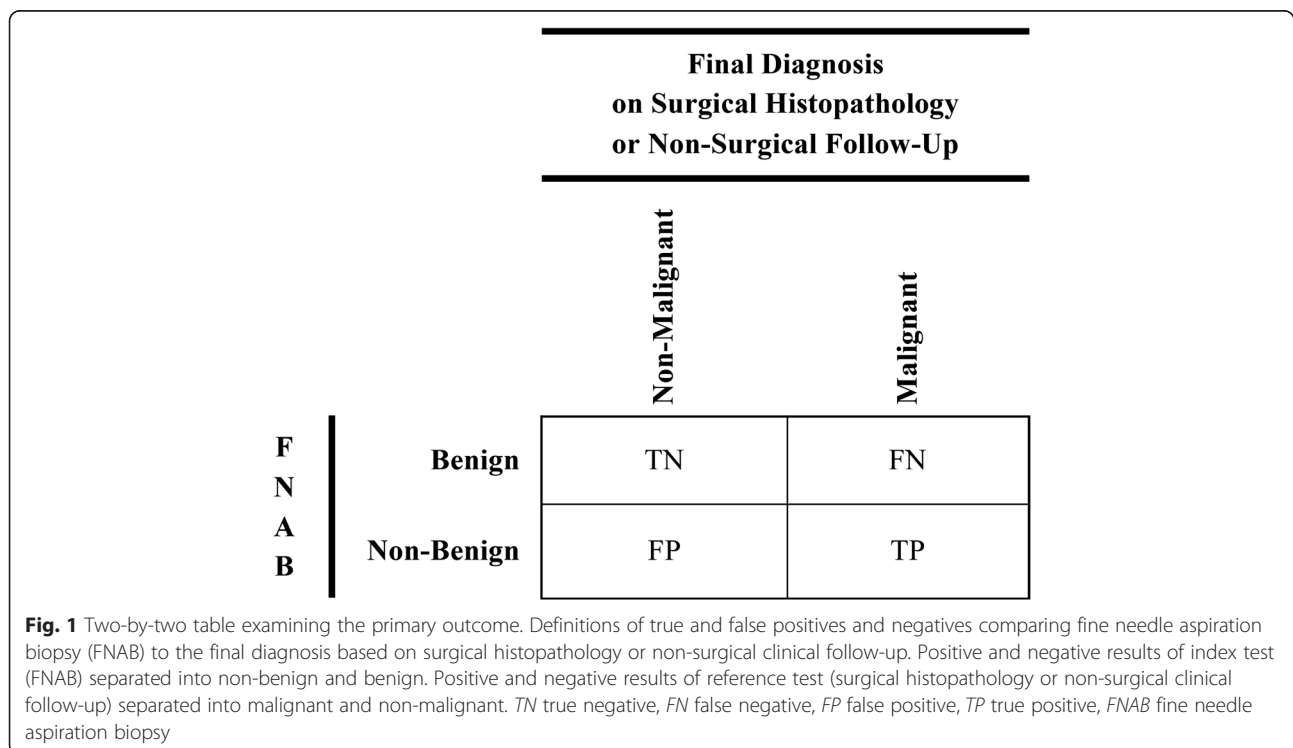
Search filter	Ovid MEDLINE, Cochrane Database of Systematic Reviews, PubMed, Evidence-Based Medicine reviews		Ovid EMBASE	
	MeSH terms	Text words	Emtree terms	Text words
Thyroid	Thyroid gland, thyroid nodule, thyroid neoplasm,	Thyroid*	Thyroid gland, thyroid nodule, thyroid tumor, thyroid carcinoma, thyroid cancer	Thyroid*
Biopsy	Biopsy; biopsy, fine-needle; biopsy, needle; image-guided biopsy; biopsy, large-core needle;	FNA, core needle biopsy, biopsy*, aspirat*, core, needle, cytology	Needle biopsy, biopsy, image guided biopsy, biopsy needle, large core needle biopsy, tumor biopsy, biopsy technique, fine biopsy needle, core biopsy needle, fine needle aspiration biopsy	Fna, core needle biopsy, biops* aspirat*, core, needle, cytology
Pediatric	Pediatrics, adolescent, child	Adol*, child*, ped*, paed*	Pediatrics, adolescent, child	Adol*, child* ped* paed*

MeSH medical subject headings

- Data for at least ten pediatric patients were reported (to exclude case reports and small case series)
- Intervention
  - The index test was FNAB of a thyroid nodule, with or without US guidance
- Comparison
  - The reference standard was surgical histopathology or clinical follow-up
- Outcome
  - The studies examined test accuracy for detection of thyroid malignancy as defined by the authors, including true and false positives and negatives, sensitivity and specificity, and positive and negative likelihood ratios
- Sufficient data were presented to tabulate the results comparing FNAB to surgical pathology or clinical follow-up into two-by-two contingency tables (Fig. 1)
- Design
  - Diagnostic accuracy studies (single gated) that compare the results of an index test to the results of a reference standard on the same subjects [33–35]

Exclusion criteria

- Non original data
  - Duplicate data sets
  - Overlapping data sets
  - Articles with smaller cohorts will be excluded



- Authors will be contacted to clarify their patient population if the degree of overlap is unclear
- Non-human studies
- Studies involving patients with exclusively malignant or benign thyroid surgical histopathology

Two investigators (SWL, KYW) will pilot test inclusion and exclusion criteria using 20 randomly selected articles to ensure complete investigator agreement of the criteria. Agreement regarding inclusion and exclusion of full-text articles between the two investigators (SWL, KYW) will be quantified using the kappa statistic. A kappa statistic greater than 0.6 will be considered moderate agreement [36]. Disagreements will be resolved by consensus or arbitration by a third party (DJR or DMR) after the article of interest has been re-read in full by all investigators [37].

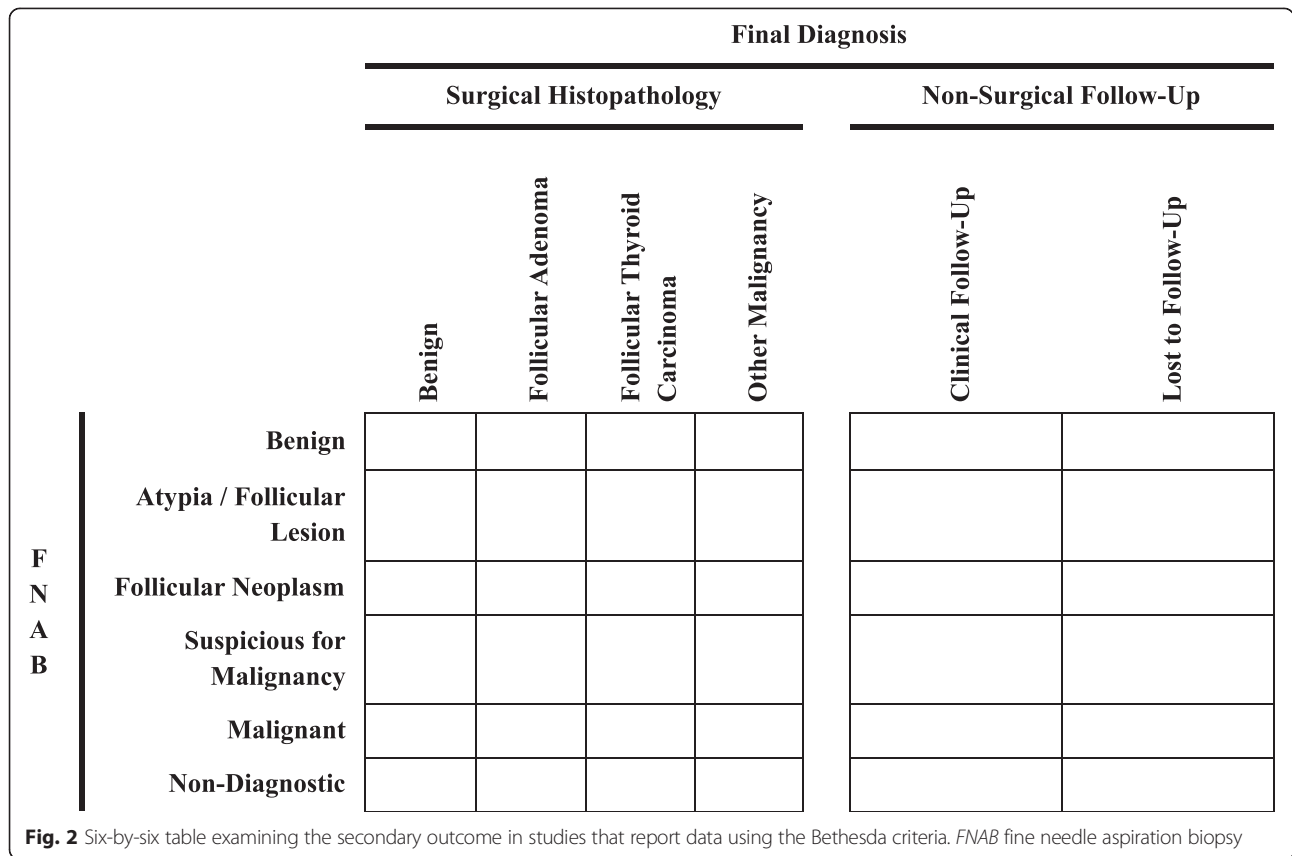
#### Data extraction

Two investigators (SWL, KYW) will extract data from all eligible diagnostic studies independently and in duplicate using a predesigned Microsoft Access 2010 (Microsoft, Redmond, WA) database form. This database form will be pilot tested on a random sample of five included studies until reliable data extraction is confirmed (kappa statistic > 0.6) [36]. We will extract the following data from included studies:

1. Study information
  - First author
  - Title
  - Year of publication
2. Study design and methodology
  - Directionality of data collection
    - Retrospective, prospective
  - Participant selection method
    - Consecutive, random
  - Inclusion and exclusion criteria
    - Including whether thyroid surgery was listed as a prerequisite for enrolment
  - Study setting
    - Country of origin, single versus multi-site
3. Patient sample information
  - Sample size
  - Participant characteristics
    - Age, gender
4. Experimental (index) test (FNAB)
  - FNAB description
    - Number of biopsies, complications, use of US guidance, qualifications of the individual performing FNAB (general practitioner,

- pediatrician, endocrinologist, surgeon, radiologist, pathologist)
  - FNAB reporting
    - Adherence to Bethesda or other criteria, qualifications of pathologist reporting results (pathologist, cytopathologist, pediatric pathologist, pediatric cytopathologist)
5. Reference standard test
    - Type of surgery performed (total thyroidectomy, hemithyroidectomy, surgical biopsy)
    - Length of time between FNAB and surgery
    - Results of surgical histopathology (benign versus malignant, type of malignancy)
    - Qualifications of pathologist reporting results (pathologist, pediatric pathologist)
    - Number of patients who did not proceed from FNAB to surgery
    - Length and type of follow-up (clinical, radiological)
    - Number of patients lost to follow-up
  6. Blinding of the pathologists to the results of FNAB and surgical histopathology
  7. Study results and analysis
    - Data to populate a two-by-two table (Fig. 1) to assess the primary outcome for FNAB
      - Figure 1 defines true and false positives and negatives based on the two-by-two table. Positive and negative results of index test (FNAB) will be separated into benign and non-benign. Positive and negative results of gold standard reference test (surgical histopathology) and surrogate reference test (clinical follow-up) will be separated into malignant and non-malignant.
      - This table will be used to generate pooled estimates of diagnostic accuracy (sensitivity, specificity, positive and negative likelihood ratios) as our primary outcome
    - For our secondary outcome analysis, where possible, we will extract data to populate six-by-six tables (Fig. 2) to assess the accuracy of FNAB using the six Bethesda classifications (non-diagnostic, benign, atypia/follicular lesion of undetermined significance, follicular neoplasm or suspicious for follicular neoplasm, suspicious for malignancy, malignant), compared with six potential outcomes: four surgical (benign, follicular adenoma, follicular thyroid carcinoma, other malignancy), and two non-operative (clinical follow-up, loss to follow-up).
    - To evaluate the test accuracy of each FNAB diagnostic category to predict malignancy, the six-by-six data will be condensed into multiple two-by-two contingency tables by altering the threshold of interpretation of FNAB results as test negative or positive. Figure 3 shows the





sliding thresholds used for FNAB interpretation, stratified into four separate comparisons. All non-diagnostic biopsies will be removed from the diagnostic accuracy meta-analysis as initial and final diagnosis of malignant or non-malignant disease is unclear in patients clinically followed or lost to follow-up

- Figure 4 defines true and false positives and negatives based on the sliding thresholds for all four comparisons. Positive and negative results of the gold standard reference test (surgical histopathology) will be separated into malignant and non-malignant. Positive and negative results of the surrogate reference test (clinical follow-up) will be separated into final diagnoses based on FNAB results. We will assume that non-malignant FNAB would be followed clinically and converted to surgical management if malignancy developed. Positive and negative results of patients lost to follow-up will be separated into final diagnoses based on the assumption that non-malignant FNAB would be followed clinically and that malignant FNAB lost to follow-up would subsequently be managed at a different facility

- These tables will be used to generate multiple pooled estimates of diagnostic accuracy (sensitivity, specificity, positive and negative likelihood ratios) for each comparison

Non-English language literature will be translated by interpreters. Agreement between the two investigators (SWL, KYW) will be ensured by consensus or arbitration by a third party (DJR or DMR) as needed.

**Study quality assessment and risk of bias**

The risk of bias of each article will be evaluated independently by two investigators (SWL, KYW) and reported according to the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [38]. The presence of spectrum, threshold, disease progression, and verification bias (partial or differential) will be specifically assessed, as defined below.

Spectrum bias occurs when study participants do not represent the population of interest due to inappropriate patient selection. This is an anticipated source of bias in articles where thyroid surgery forms part of the inclusion criteria. These exclusively surgical cohorts likely represent a distinct subset of the population with more worrisome findings and a higher pre-test probability for malignancy, leading to potential differences in diagnostic accuracy

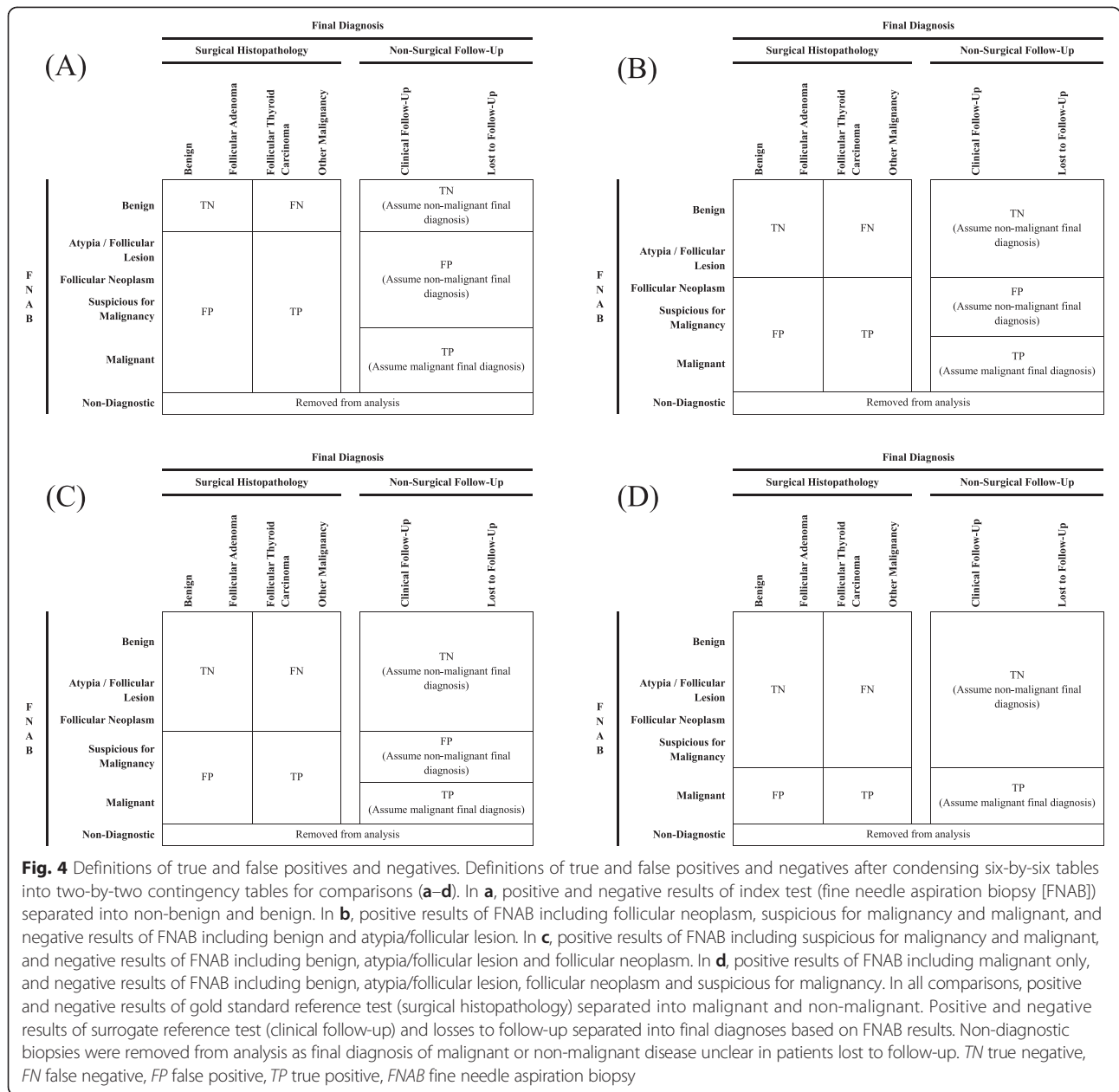
		Comparisons Derived by Sliding Thresholds of FNAB Interpretation			
		A	B	C	D
FNAB	Negative				Benign
					Atypia / Follicular Lesion
				Benign	Follicular Neoplasm
				Atypia / Follicular Lesion	Suspicious for Malignancy
	Positive	Benign	Atypia / Follicular Lesion	Follicular Neoplasm	Suspicious for Malignancy
		Atypia / Follicular Lesion	Follicular Neoplasm	Suspicious for Malignancy	Malignant
		Follicular Neoplasm	Suspicious for Malignancy	Malignant	
		Suspicious for Malignancy	Malignant		

**Fig. 3** Sliding thresholds. Sliding thresholds used for fine needle aspiration biopsy (FNAB) interpretation as test negative or positive. Four separate comparisons (A, B, C, D) evaluate the test accuracy of each FNAB diagnostic category to predict thyroid nodule malignancy. FNAB fine needle aspiration biopsy

results. Another example of spectrum bias includes studies targeting hypothyroid or hyperthyroid patients specifically, where extrapolation to the general pediatric population with thyroid nodules may be inappropriate.

Threshold bias develops when pathologists use varying definitions to report FNAB results. This leads to a greater likelihood to diagnose benign or malignant disease based on an individual pathologist’s threshold of concern. The Bethesda criteria were introduced to standardize FNAB classification and minimize threshold bias. Studies reporting results by Bethesda versus other criteria will be compared to evaluate the potential contribution of threshold bias to diagnostic accuracy.

Disease progression bias is a concern when the interval between the index test and reference standard is long enough to potentially allow progression of disease from benign to malignant or from one type of malignant disease to another. An index test may be negative and the reference test positive due to rapid development of malignancy, rather than signify an inaccurate index test. To assess the risk of disease progression bias, an appropriate time frame between FNAB and surgery and length of clinical follow-up would need to be defined. However, this interval is not well described in the literature as the latency period for development of thyroid malignancy after discovery of a nodule may extend for years, despite exposure to known



risk factors [39, 40]. As such, we will collect data regarding these parameters without imposing predefined intervals such that studies may later be categorized into those with shorter versus longer intervals for stratified meta-regression.

Partial verification bias occurs when results of the index test influence whether or not the patient receives the reference standard. There is significant potential for this type of bias among the studies that will be included in this systematic review as benign cytology may decrease the likelihood of any type of follow-up, whether surgical or clinical, unless there are other significant risk factors for malignancy. Partial verification bias frequently

leads to inflated diagnostic accuracy as benign FNAB results may be assumed inappropriately to represent true negative disease [41]. Differential verification bias arises when results of the index test determine which reference standard is used to confirm the diagnosis. Using clinical follow-up as a surrogate reference standard, many studies will be prone to differential verification bias with benign cytology followed clinically instead of with surgery. Verification bias, whether partial or differential, is expected to be the primary limiting factor affecting the validity of pooled estimates across the diagnostic accuracy studies that will be included in this systematic review. In order to eliminate verification bias, in an ideal diagnostic accuracy



study, all patients presenting with a nodule must undergo both FNAB and surgical excision to definitively diagnosis benign or malignant disease. However, this practice does not occur as most low-risk patients are observed in follow-up to avoid the risks of surgery. Ethically, inclusion of patients with FNAB who undergo surgery and lifelong clinical follow-up provides the best case scenario for confirming diagnostic accuracy. Verification bias may be reduced, but not eliminated, with serial clinical and radiological examinations for several years to capture any false negative FNAB, though the required duration of follow-up is unclear. It is anticipated that this systematic review will find a mixture of studies with different biases. The interconnectedness of spectrum and verification bias in this setting will also be assessed, since studies with surgical cohorts prone to spectrum bias are also at low risk of verification bias (i.e., all patients will have definitive surgical histopathology).

As a supplement to the QUADAS-2 tool, we will also examine the timing of data collection (prospective, retrospective), the qualifications of the individual performing the FNAB (general practitioner, pediatrician, endocrinologist, surgeon, radiologist, pathologist) or the interpreting pathologist (general pathologist, cytopathologist, pediatric pathologist, pediatric cytopathologist), and adherence of cytology reporting to the Bethesda versus other criteria.

Disagreements between the two investigators (SWL, KYW) will be resolved by consensus or arbitration by a third party (DJR or DMR).

#### Data synthesis and analysis

True and false positives and negatives will be defined by two-by-two contingency tables (Fig. 1) for the primary outcome. True and false positives and negatives will be defined after condensing six-by-six tables (Fig. 2) into two-by-two contingency tables (Fig. 4) for the secondary outcome that will examine each Bethesda classification level. These tables will be used to calculate study-level estimates of sensitivity, specificity, and positive and negative likelihood ratios for detection of thyroid malignancy. Hierarchical summary receiver operating characteristic (HSROC) curves will be generated to depict the bivariate relationship between individual study estimates of sensitivity and specificity [30, 42, 43]. We will also use this model to calculate the proportion of between-study heterogeneity that may be due to diagnostic threshold variability using the between-study covariance parameter [30, 43–45].

Bivariate random-effects models will be used to derive pooled estimates of sensitivity, specificity, and positive and negative likelihood ratios for detection of malignancy with FNAB [43, 45, 46]. These models incorporate the degree of negative correlation that may exist between sensitivity

and specificity across studies [43, 45, 46]. This joint synthesis of diagnostic accuracy estimates is unbiased despite diagnostic threshold variability and facilitates the development of Bayesian probability modifying and Fagan plots [42, 43, 45–47]. These two plots will allow for an assessment of the likely post-test probability obtained after applying FNAB to samples of patients with varying ranges of pre-test probabilities of thyroid malignancy. These models will also allow us to determine the extent of heterogeneity (due to diagnostic threshold variability or study-level covariates) in our pooled estimates through the production of forest plots and the computation of  $I^2$ - and  $Q$ -statistics [45–49].

In the presence of inter-study heterogeneity, we will use the bivariate model to conduct subgroup analyses and meta-regression to determine whether a number of pre-defined covariates may explain variation in reported diagnostic performance results across studies [43–46, 48–50]. Covariates of interest will include those describing the study setting (country of origin, single versus multi-site), risk of bias (prospective versus retrospective data collection, random versus consecutive method of selection, thyroidectomy as part of the inclusion criteria, presence of verification bias, length of follow-up, loss to follow-up greater than 15 %), FNAB implementation and interpretation (use of US guidance, qualifications of individual performing and interpreting FNAB, use of Bethesda or other criteria), and length of clinical follow-up. We will also examine whether any studies exert undue influence on our pooled diagnostic accuracy estimates by performing a sensitivity analysis, removing those that appear to be influential outliers or those which may include potentially overlapping patients. Influential studies will be identified using spike plots of Cook's distance and scatter plots of standardized residuals [42, 43, 51–53]. Finally, to assess for the presence of small study effects potentially due to publication bias, we will create funnel plots using the diagnostic odds ratio and conduct Deek's asymmetry tests [54].

All statistical analyses will be performed using Stata version 13.1 (Stata Corp, College Station, TX), including the "midas" and "metandi" command packages [42, 43, 55].

#### Discussion

Thyroid nodules can provoke anxiety in children, families, and physicians alike due to diagnostic uncertainty in the setting of greater potential for malignancy. The ability of a diagnostic test to distinguish malignant from benign disease is paramount for clinicians to provide appropriate counselling regarding treatment and prognosis. In addition to providing a systematic review and meta-analysis of the diagnostic accuracy of FNAB in pediatric thyroid nodules for the detection of malignancy, this will be the first study to determine the

accuracy of FNAB according to the Bethesda criteria. In doing this, our results may serve as a better guide for clinical decision-making in children with thyroid nodules.

Although the American Thyroid Association endorses FNAB as the standard of care in North America for the evaluation of thyroid nodules in adults and children, the evidence supporting this recommendation is likely based on the results of studies conducted among adults. Pediatric studies may be limited by several study-level biases. Thus, this systematic review and meta-analysis will rigorously examine the potential magnitude of influence that individual study-level biases may have on the diagnostic accuracy of FNAB. Other specific aims to be addressed by this study include determining the value of adherence to the Bethesda criteria, US guidance, and the qualifications of the individual performing and interpreting the FNAB on the diagnostic accuracy of FNAB. If these factors are found to enhance diagnostic accuracy, this may support the need for routine referral of children with thyroid nodules to specialty centres where US and FNAB-trained personnel are available to improve patient care and outcomes.

#### Abbreviations

CI: confidence interval; FN: false negative; FP: false positive; FNAB: fine needle aspiration biopsy; HSROC: hierarchical summary receiver operating characteristic; MeSH: Medical Subject Heading; QUADAS: Quality Assessment of Diagnostic Accuracy Studies; TN: true negative; TP: true positive; US: ultrasound.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

SWL and KYW conceived and designed the study and search strategy, which was refined by DJR and DMR. SWL, DJR, DMR, and KYW designed the statistical analysis plan. SWL and KYW wrote the first draft of the study protocol, which was critically revised by DJR and DMR. SWL registered the protocol with the PROSPERO database. All authors read and approved the final protocol.

#### Authors' information

SWL is a general surgeon, pediatric surgery fellow, and Clinician Investigator Program resident who is pursuing a Master of Science degree in the Gastrointestinal Sciences Program at the University of Calgary. DJR is a general surgery and Clinician Investigator Program resident who is pursuing a Doctor of Philosophy degree in epidemiology and knowledge translation at the University of Calgary. DMR is an endocrinologist in Calgary with an interest in systematic reviews and meta-analyses related to endocrine disease. KYW is a pediatric endocrinologist who is pursuing a Master of Science degree in the Community Health Sciences Program at the University of Calgary.

#### Acknowledgements

The authors would like to thank Diane Lorenzetti, MLS for her assistance in refining the search strategy. SWL was funded by the University of Calgary Clinician Investigator Program. DJR was funded by an Alberta Innovates-Health Solutions Clinician Fellowship Award, a Knowledge Translation Canada Strategic Funding in Health Research Fellowship, and the Canadian Institutes of Health Research. DMR was funded by an Alberta Innovates-Health Solutions Population Investigator Award. KYW was funded by a research fellowship from the Canadian Pediatric Endocrine Group. These sources of funding have had no input in this study's conception and design, nor will they have any role in its implementation, analysis, interpretation, or incorporation into a final manuscript.

#### Author details

<sup>1</sup>Department of Surgery, University of Calgary, Calgary, AB, Canada. <sup>2</sup>Department of Gastrointestinal Sciences, University of Calgary, Calgary, AB, Canada. <sup>3</sup>Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada. <sup>4</sup>Department of Medicine, University of Calgary, Calgary, AB, Canada. <sup>5</sup>Department of Pediatrics, University of Calgary, Calgary, AB, Canada. <sup>6</sup>Department of Pediatric Surgery, Alberta Children's Hospital, 2888 Shaganappi Trail NW, Calgary, AB, Canada T3B 6A8. <sup>7</sup>Intensive Care Unit Administration, Ground Floor McCaig Tower, Foothills Medical Centre, 3134 Hospital Drive NW, Calgary, AB, Canada T2N 5A1.

Received: 28 March 2015 Accepted: 4 September 2015

Published online: 24 September 2015

#### References

- Altincik A, Demir K, Abaci A, Bober E, Buyugebiz A. Fine-needle aspiration biopsy in the diagnosis and follow-up of thyroid nodules in childhood. *JCRPE*. 2010;2:78–80.
- Buryk MA, Monaco SE, Witchel SF, Mehta DK, Gurtunca N, Nikiforov YE, et al. Preoperative cytology with molecular analysis to help guide surgery for pediatric thyroid nodules. *Int J Pediatr Otorhinolaryngol*. 2013;77:1697–700.
- Kaur J, Srinivasan R, Arora SK, Rajwanshi A, Saikia UN, Dutta P, et al. Fine-needle aspiration in the evaluation of thyroid lesions in children. *Diagn Cytopathol*. 2012;40(S1):E33–7.
- Hoperia V, Larin A, Jensen K, Bauer A, Vasko V. Thyroid fine needle aspiration biopsies in children: study of cytological-histological correlation and immunostaining with thyroid peroxidase monoclonal antibodies. *Int J Pediatr Endocrinol*. 2010; doi:10.1155/2010/690108.
- Mirshemirani A, Roshanzamir F, Tabari AK, Ghorobi J, Salehpoor S, Gorji FA. Thyroid nodules in childhood: a single institute experience. *Iran J Pediatr*. 2010;20:91–6.
- Monaco SE, Pantanowitz L, Khalbuss WE, Benkovich VA, Ozolek J, Nikiforova MN, et al. Cytomorphological and molecular genetic findings in pediatric thyroid fine-needle aspiration. *Cancer Cytopathol*. 2012;120:342–50.
- Niedziela M. Pathogenesis, diagnosis and management of thyroid nodules in children. *Endocr Relat Cancer*. 2006;13:427–53.
- Khozeimeh N, Gingalewski C. Thyroid nodules in children: a single institution's experience. *J Oncol*. 2011. doi:10.1155/2011/974125.
- Gupta A, Ly S, Castroneves LA, Frates MC, Benson CB, Feldman HA, et al. A standardized assessment of thyroid nodules in children confirms higher cancer prevalence than in adults. *J Clin Endocrinol Metab*. 2013;98:3238–45.
- Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. *Acta Cytol*. 2012;56:333–9.
- Ogilvie JB, Piatigorsky EJ, Clark OH. Current status of fine needle aspiration for thyroid nodules. *Adv Surg*. 2006;40:223–38.
- Poller DN, Stelow EB, Yiangou C. Thyroid FNAC cytology: can we do it better? *Cytopathology*. 2008;19:4–10.
- Calo PG, Pisano G, Medas F, Tatti A, Tuveri M, Nicolosi A. Risk factors in reoperative thyroid surgery for recurrent goitre: our experience. *G Chir*. 2012;33:335–8.
- Roy R, Kouniavsky G, Schneider E, Allendorf JD, Chabot JA, Logerfo P, et al. Predictive factors of malignancy in pediatric thyroid nodules. *Surgery*. 2011;150:1228–33.
- Caplan RH, Strutt PJ, Kicken WA, Wester SM. Fine needle aspiration biopsy of thyroid nodules. *Wis Med J*. 1991;90:285–8.
- Hamburger JI. Consistency of sequential needle biopsy findings for thyroid nodules. Management implications. *Arch Intern Med*. 1987;147:97–9.
- Amrikachi M, Ponder TB, Wheeler TM, Smith D, Ramzy I. Thyroid fine-needle aspiration biopsy in children and adolescents: experience with 218 aspirates. *Diagn Cytopathol*. 2005;32:189–92.
- Bongiovanni M, Crippa S, Baloch Z, Piana S, Spitale A, Pagni F, et al. Comparison of 5-tiered and 6-tiered diagnostic systems for the reporting of thyroid cytopathology. *Cancer Cytopathol*. 2012;120:117–25.
- Sugino K, Ito K, Nagahama M, Kitagawa W, Shibuya H, Ohkuwa K, et al. Diagnostic accuracy of fine needle aspiration biopsy cytology and ultrasonography in patients with thyroid nodules diagnosed as benign or indeterminate before thyroidectomy. *Endocr J*. 2013;60:375–82.
- Lobo C, McQueen A, Beale T, Kocjan G. The UK Royal College of pathologists thyroid fine-needle aspiration diagnostic classification is a

- robust tool for the clinical management of abnormal thyroid nodules. *Acta Cytol.* 2011;55:499–506.
21. Cibas ES, Ali SZ. NCI Thyroid FNA state of the science, conference. The Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol.* 2009;132:658–65.
  22. Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, et al. Management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid.* 2006;16:109–42.
  23. Thyroid Nodules ATA(ATA)GTo, Differentiated Thyroid C, Cooper DS, Doherty GM, Haugen BR, Kloos RT, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid.* 2009;19:1167–214.
  24. Stevens C, Lee JKP, Sadatsafavi M, Blair GK. Pediatric thyroid fine-needle aspiration cytology: a meta-analysis. *J Pediatr Surg.* 2009;44:2184–91.
  25. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA.* 1999;282:1061–6.
  26. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med.* 2002;21:1525–37.
  27. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Open Med.* 2009;3:e123–30.
  28. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol.* 2009;62:e1–34.
  29. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA.* 2000;283:2008–12.
  30. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working, Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med.* 2008;149:889–97.
  31. Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *J Clin Epidemiol.* 2011;64:602–7.
  32. Leeflang MM, Deeks JJ, Takwoingi Y, Macaskill P. Cochrane diagnostic test accuracy reviews. *Syst Rev.* 2013;2:82.
  33. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med.* 2003;138(1):W1–12.
  34. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Cochrane Diagnostic Test Accuracy Working, Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med.* 2008;149(12):889–97.
  35. Handbook for DTA Reviews [Internet]; 2015 [cited Aug 16, 2015]. Available from: <http://dta.cochrane.org/dta-review-author-training>.
  36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–74.
  37. Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. 2nd ed. London: BMJ Publishing Group; 2001.
  38. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(5):529–36.
  39. Chiesa F, Tradati N, Calabrese L, Gibelli B, Giugliano G, Paganelli G, et al. Thyroid disease in northern Italian children born around the time of the Chernobyl nuclear accident. *Ann Oncol.* 2004;15:1842–6.
  40. Nikiforov Y, Gnepp DR. Pediatric thyroid cancer after the chernobyl disaster. Pathomorphologic study of 84 cases (1991-1992) from the Republic of Belarus. *Cancer.* 1994;74:748–66.
  41. Whiting PF, Rutjes AWS, Westwood ME, Mallett S. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol.* 2013;10(66):1093–104.
  42. Harbord RM, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. In: Sterne JAC, editor. Meta-analysis in stata: an updated collection from the stata journal. College Station, TX: Stata Press; 2000. p. 181–99.
  43. midas: A program for Meta-analytical Integration of Diagnostic Accuracy Studies in Stata [Internet]. College Station, TX: Stata Press; 2007. Available from: <http://fmwww.bc.edu/repec/bocode/m/midas>.
  44. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med.* 2001;20:2865–84.
  45. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol.* 2005;58:982–90.
  46. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol.* 2007;7:3.
  47. Gatsonis C, Paliwal P. Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. *AJR Am J Roentgenol.* 2006;187:271–81.
  48. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21:1539–58.
  49. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327:557–60.
  50. de Groot JA, Dendukuri N, Janssen KJ, Reitsma JB, Brophy J, Joseph L, et al. Adjusting for partial verification or workup bias in meta-analyses of diagnostic accuracy studies. *Am J Epidemiol.* 2012;175:847–53.
  51. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman GD, editors. Systematic reviews in health care: meta-analysis in context. London, UK: BMJ Publishing Group; 2001. p. 285–312.
  52. Cook DR. Influential observations in linear regression. *J Am Stat Assoc.* 1979;74:169–74.
  53. Cook DR. Detection of influential observation in linear regression. *Technometrics.* 1977;19:15–8.
  54. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol.* 2005;58:882–93.
  55. Sterne JAC, Bradburn MJ, Egger M. Meta-analysis in Stata(TM). In: Egger M, Smith GD, Altman GD, editors. Systematic reviews in health care: meta-analysis in context. London, UK: BMJ Publishing Group; 2001. p. 347–69.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

