Use of meta-regression methods in the evaluation of heterogeneity in meta-analyses of diagnostic studies

Madhukar Pai, Laura L. Flores, Alan Hubbard, Lee W. Riley, John M Colford

Divisions of Epidemiology, Infectious Diseases & Biostatistics
University of California, Berkeley
Cochrane HIV/AIDS Review Group, San Francisco

Heterogeneity in diagnostic meta-analyses

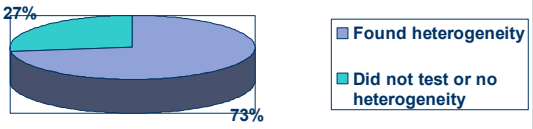
- Heterogeneity in measures of test accuracy is an important concern with all meta-analyses
  - A bigger concern with diagnostic meta-analyses?
- Sources of heterogeneity include variability in:
  - Disease
  - Index tests
  - Reference standards
  - Thresholds used
  - Populations and disease spectrum
  - Study quality
  - Random error

Lijmer et al. Stat Med. 2002 Jun 15;21(11):1525-37.

Heterogeneity in diagnostic meta-analyses

- How often is significant heterogeneity found in diagnostic meta-analyses?
- Rapid survey of recent, published meta-analyses:
  - Searched PubMed [Jan – Oct 2003]
  - meta-analysis AND (accuracy OR (sensitivity AND specificity))
  - Total hits = 92
    - Diagnostic meta-analyses = 26
      - Median number of studies included = 27 (range 6 – 64)
      - SROC analysis = 14/26 (54%)

Heterogeneity in diagnostic meta-analyses



N = 26 meta-analyses

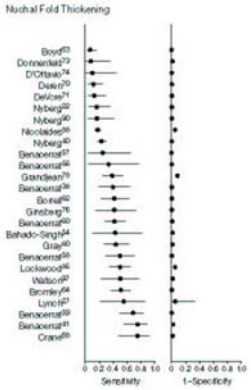
Heterogeneity in diagnostic meta-analyses

- Which measures of test accuracy were heterogeneous?
  - Sensitivity only in 2 meta-analyses
  - Specificity only in 2 meta-analyses
  - DOR in 2 meta-analyses
  - Both sensitivity and specificity in 13 meta-analyses
    - Sensitivity tended to be more heterogeneous than specificity

N = 19 meta-analyses

Ultrasound for Down's syndrome

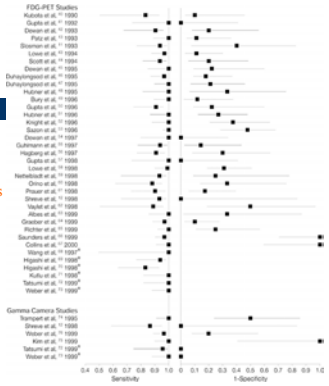
Sensitivity heterogeneous



Smith-Bindman et al. JAMA. 2001 Feb 28;285(8):1044-55.

PET for lung nodules

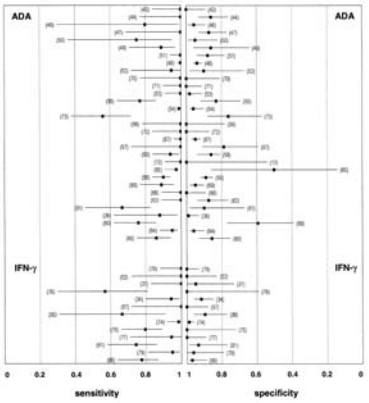
Specificity heterogeneous



Gould et al. JAMA. 2001;285:914-924.

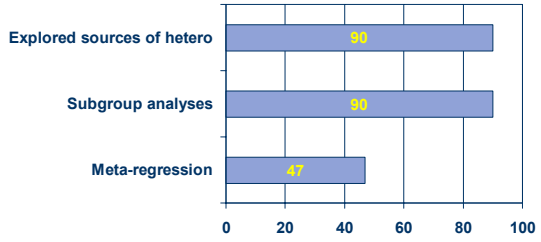
ADA and IFN for pleural TB

Both sensitivity & specificity heterogeneous



Greco et al. Int J Tuberc Lung Dis. 2003 Aug;7(8):777-86.

Heterogeneity in diagnostic meta-analyses



N = 19 meta-analyses

Methods

- We conducted two meta-analyses on the accuracy of nucleic acid amplification tests (e.g. PCR) for the diagnosis of:
  - Tuberculous meningitis* [49 studies]
  - Tuberculous pleuritis (pleurisy)** [40 studies]
- Significant heterogeneity was found in both reviews
- We describe the use of meta-regression methods for exploration of heterogeneity with these reviews as case studies
- We also discuss some methodological issues that we encountered while performing the analyses

*Pai M, et al. Diagnostic accuracy of nucleic acid amplification tests for tuberculous meningitis: a systematic review and meta-analysis. *Lancet Infect Dis* 2003;3:633-43.

**Pai M, et al. Nucleic acid amplification tests in the diagnosis of tuberculous pleuritis: a systematic review and meta-analysis. Under review: *Int J Tubercl & Lung Dis*.

Methods

- To evaluate heterogeneity, we used the SROC regression analysis (Littenberg & Moses 1993) that models:
  - the log of the diagnostic odds ratio (DOR) as the dependent variable
  - independent variables (covariates) were components of study quality, and test characteristics defined *a priori* as potential sources of heterogeneity
    - Covariates significant at p=0.2 on bivariate analyses were included in the initial model
    - Covariates with the largest p-values were dropped sequentially to decide the final model
  - regression coefficients were exponentiated and interpreted as relative DORs (RDOR)
  - R^2 were used to determine proportion of variability in DOR explained by covariates in the model

Littenberg & Moses. Med Des Making 1993;13:313-321

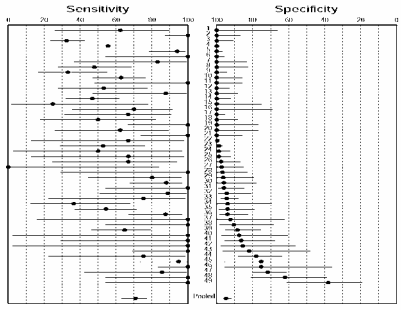
Methods

- Quality criteria (Lijmer 1999) evaluated as sources of heterogeneity:
  - Study design (spectrum)
    - Case-control vs. cross-sectional
  - Verification bias
    - Complete, partial, differential verification
  - Blinding
    - Single or double blinded interpretation of test and reference standard
  - Patient sampling
    - Consecutive or random vs non-consecutive/non-random
  - Data collection strategy
    - Prospective vs retrospective
  - Study size
- The following test characteristics were also evaluated:
  - PCR target sequence (IS6110 vs others)
  - DNA extraction method (phenol-chloroform vs others)
  - Smear status (smear positive vs negative)
  - Nested PCR vs simple PCR
  - Detection using hybridization (yes or no)

Lijmer JG, et al. JAMA 1999;282:1061-6

Nucleic acid amplification tests for TB meningitis

All 49 studies

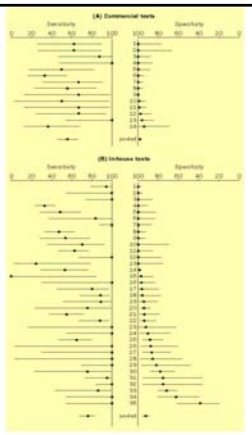


Pooled Se = 0.71
Heterogeneity p<0.001

Pooled Sp = 0.95
Heterogeneity p<0.001

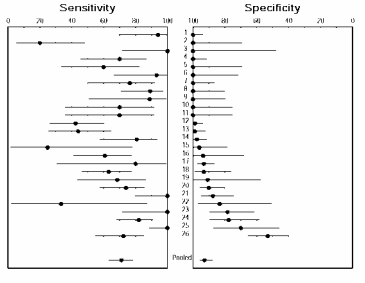
Nucleic acid amplification tests for TB meningitis

Subgroups:
commercial tests
Vs.
in-house PCR



Pai M, et al. Diagnostic accuracy of nucleic acid amplification tests for tuberculous meningitis: a systematic review and meta-analysis. *Lancet Infect Dis* 2003;3:633-43.

Accuracy of nucleic acid amplification tests (e.g. PCR) for the diagnosis of TB pleuritis



Pooled Se = 0.71
Heterogeneity p <0.001

Pooled Sp = 0.93
Heterogeneity p <0.001

Pai M, et al. Nucleic acid amplification tests in the diagnosis of tuberculous pleuritis: a systematic review and meta-analysis. Under review: *Int J Tubercl & Lung Dis.*

Case study 1: TB meningitis meta-analysis: preliminary meta-regression (unadjusted)

Study characteristic	b	*P* value	Relative Diagnostic Odds Ratio (RDOR) (95% CI)
Study design (case-control vs. cross-sectional)	1. 19	0.03	3.28 (1.14, 9.5)
Blinding (unblinded/unknown vs. blinded)	0.514	0.37	1.67 (0.53, 5.31)
Sampling (non-consecutive/non-random/unknown vs. consecutive or random)	-0.216	0. 70	0.80 (0.26, 2.54)
Data collection (prospective vs. retrospective/both/unknown)	-0.04	0.94	0.96 (0.30, 3.0)
Study size (<10 TBM specimens vs. 10 or more)	0.964	0.09	2.62 (0.85, 8.0)
Smear status (smear positive and negative/unknown vs. neg)	-0.143	0.80	0.86 (0.26, 2.83)
Target sequence (IS6110 vs. others)	0.06	0.91	1.06 (0.32, 3.42)
DNA extraction (phenol-chloroform vs. others)	-0.76	0.17	0.46 (0.15, 1.42)
Nested PCR (nested vs. simple)	0.80	0.25	2.23 (0.54, 9.20)
Detection using hybridization (yes or no)	-0.42	0.45	0.65 (0.21, 2.0)

Case study 1: TB meningitis meta-analysis: final meta-regression model

Covariate	b	P-value	RDOR
Design: Case-control vs. Cross-sectional	1.08	0.04	2.9 (1.0, 8.5)
Study size: <10 vs. 10+ TBM specimens	0.64	0.26	1.9 (0.6, 5.9)
DNA extraction method: Phenol-chloroform vs. other	-0.53	0.33	0.6 (0.2, 1.8)

$R^2 = 0.23$

Case study 2: TB pleuritis meta-analysis: meta-regression final model

Covariate	b	P-value	RDOR (95% CI)
Design: Case-control vs. Cross-sectional	1.00	0.09	2.7 (0.83, 8.93)
Target sequence: IS6110 vs. others	1.03	0.07	2.8 (0.88, 8.76)

$R^2 = 0.25$

Discussion

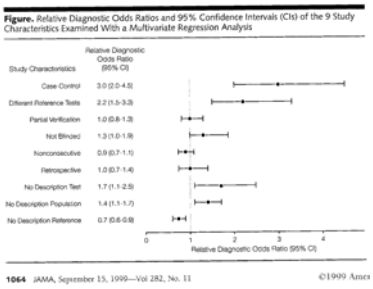
- In both meta-analyses, we identified potential sources of heterogeneity using the meta-regression approach
  - Study design (case-control) appeared to be strongly associated with test accuracy

Discussion: design-related bias

- Empiric research* suggests that some design features might be associated with diagnostic accuracy:
  - Case-control study design
  - Verification bias
  - Lack of blinding
  - Inadequate description of index test and study population

*Lijmer JG, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282:1061-6

Discussion: design-related bias



*Lijmer JG, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282:1061-6

21

Discussion: methodological issues

- In both meta-analyses, meta-regression could account for only 25% of the variability in DOR
  - Considerable heterogeneity persisted after regression
  - What methods can be used in such situations?
    - Multilevel, hierarchical models?
    - IPD meta-analyses?
- Poor quality of reporting can affect analyses
  - Quality of reporting was poor in our reviews
  - How should one handle covariates where a large proportion are coded as "not reported" in the meta-regression?
  - We contacted authors for addition information on covariates
    - Contacting authors had a large impact on some (but not all) covariates
  - What impact will author contact have on meta-regression results?
    - Work in progress
- Results may not be understood by all peer reviewers and editors
  - We had to revert to simpler, subgroup analyses in publications

22

Discussion: impact of contacting authors in the TB meningitis review

Characteristic	Before contact % [N = 49]	After contact % [N = 49]
Blinding		
Double or single blind	26	59
Not blinded	0	10
Not reported	74	31
Sampling		
Consecutive/random	18	49
Not consecutive/random	6	20
Not reported	76	31
Data collection		
Prospective	51	61
Retrospective	0	4
Both	2	10
Not reported	47	25

23

Can meta-regression be used more often?

- Heterogeneity may be a very common finding in diagnostic meta-analyses
  - Subgroup (stratified) analysis appears to be the most popular method for identifying sources of heterogeneity
- Meta-regression could be used more often for exploring heterogeneity (currently used in <50% of meta-analyses?)
  - Due to lack of power?
    - Median # of studies = 27 in the rapid survey; power may not be a concern
  - Due to poor quality of reporting in primary studies?
    - Need to contact authors to overcome this problem
  - Due to difficulties in communicating results to target audience?
    - Need to identify simpler methods of communicating meta-regression findings to consumers, clinicians, editors/reviewers.
  - Lack of awareness and training?
    - Not many simple, clearly written publications on how to perform meta-regression in diagnostic meta-analyses
    - Few workshops/courses on diagnostic reviews

24


Thank
you!