Screening tests for detecting open-angle glaucoma: Systematic Review and Meta-analysis

Mowatt, G. et al. IOVS, 49 (12), 2008

Rüdiger Bock

Chair of Pattern Recognition (Computer Science 5)

Friedrich-Alexander-Universität Erlangen-Nürnberg

Goal

The purpose of the study is to assess the comparative accuracy of candidate screening tests by extensive literature review

Inclusion / exclusion criteria

- Included studies:
  - Screening like situation and participants older than 40 years or
  - group of patients with suspected glaucoma (e.g. increased IOP)
  - Reported sensitivity and specificity
- Excluded studies:
  - non-english language
  - studies investigating technical aspects
  - studies of participants
    - without an eye disease
    - specifically excluded patients with other ocular diseases
- Reference standard
  - follow up
  - ophthalmologist-diagnosed open angle glaucoma (OAG)

Candidate tests

- Structure
  - ophthalmoscopy
  - optic disc photography
  - RNFL photography
  - HRT II
- Function
  - oculokinetic perimetry (OKP)
  - white-on-white standard automated perimetry (SAP) (suprathres, thres)
  - Frequency doubling technique (FDT)
- Intro Ocular Pressure (IOP)
  - Goldmann opplanation tonometry (GAT)

- For missing modalities (e.g. OCT, Gdx VCC) no studies were found meeting inclusion criteria

Methods

- Summary ROC curve (SROC)
  - Mean square error estimate for each test (>1 studies) from logarithmized sensitivities and specificities at a common cutoff
  - Most frequently reported **cutoff** for each test is used
  - Used to estimate sensitivity and specificity for defined cutoff

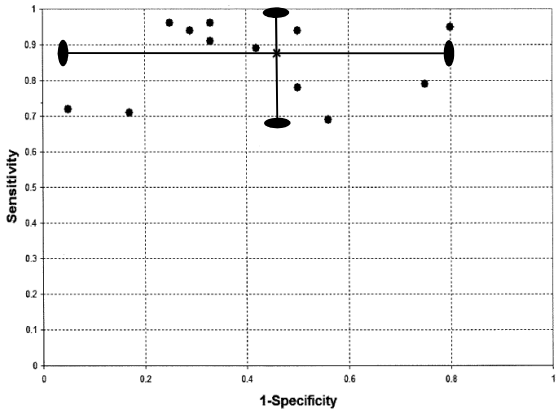


Fig 2. Sensitivity and (1-specificity) results from 12 studies (●) of accuracy of virtual bronchoscopy (contrived data). Pooled sensitivity and (1-specificity) are shown as X.

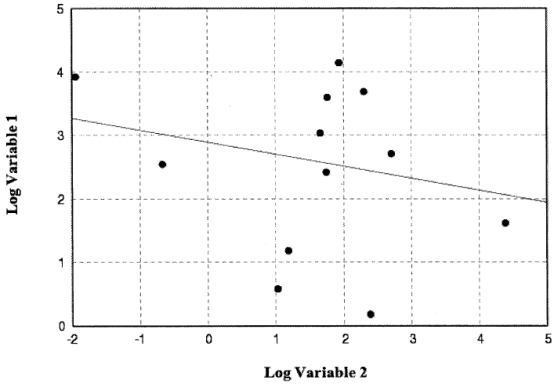


Fig 3. The regression model of the transformed data points (●) from Figure 2, shown on logarithmic axes, with the regression line shown.

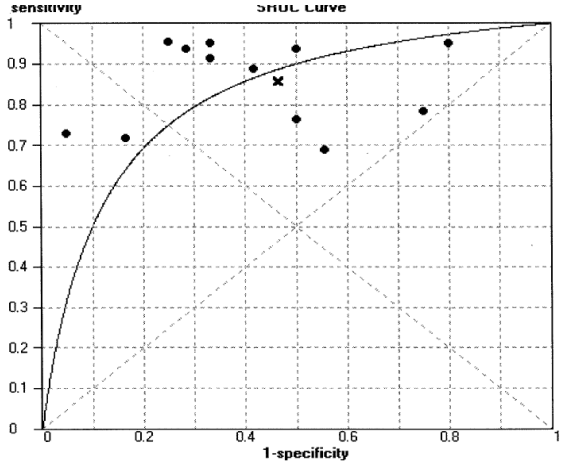


Fig 4. The points (●) from Figure 2, shown with the SROC curve superimposed. The regression line in Figure 3 has been transformed into the SROC curve, and the points in Figure 3 have been reverted back to the points from Figure 2. Pooled sensitivity and (1-specificity) are shown as X. (SROC = summary receiver operating characteristic.)

Methods

- Diagnostic odds ratios (DOR)
  - single indicator of test performance

$$DOR = \left(\frac{sensitivity}{1-sensitivity}\right)\left(\frac{speci\dot{f}icity}{1-specificity}\right)$$

All studies vs. high quality studies

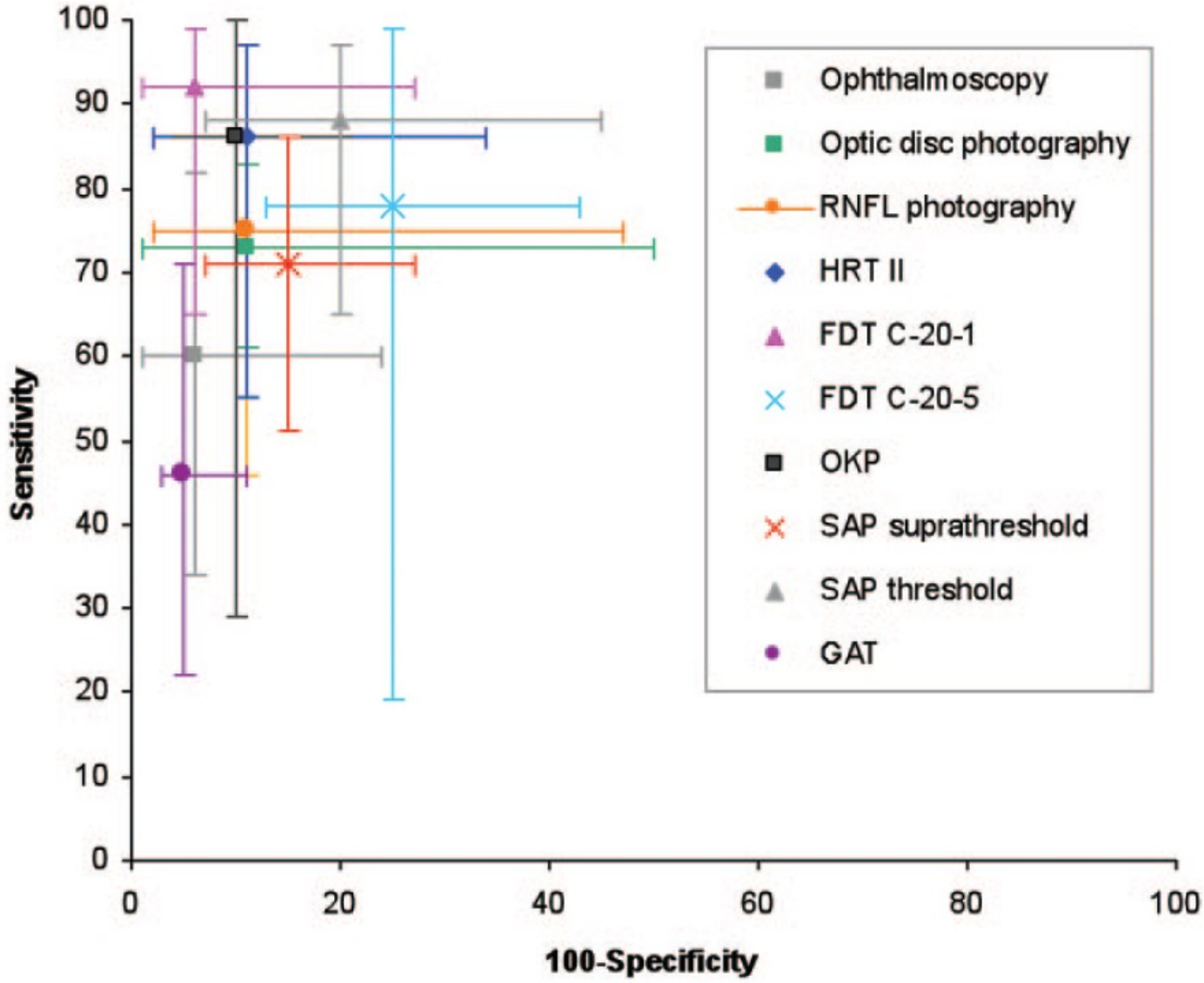
Table 2. HSROC Analysis: All Studies Compared with Higher Quality Studies

	Optic Disc Photography		HRT II		FDT C-20-5		SAP Threshold	
	Sensitivity % (95% CrI)	Specificity % (95% CrI)	Sensitivity % (95% CrI)	Specificity % (95% CrI)	Sensitivity % (95% CrI)	Specificity % (95% CrI)	Sensitivity % (95% CrI)	Specificity % (95% CrI)
All studies	73 (61–83)	89 (50–99)	86 (55–97)	89 (66–98)	78 (19–99)	75 (57–87)	88 (65–97)	80 (55–93)
Higher quality	74 (30–95)	82 (45–97)	93 (58–99)	85 (47–97)	72 (26–96)	60 (17–92)	73 (28–95)	64 (22–92)

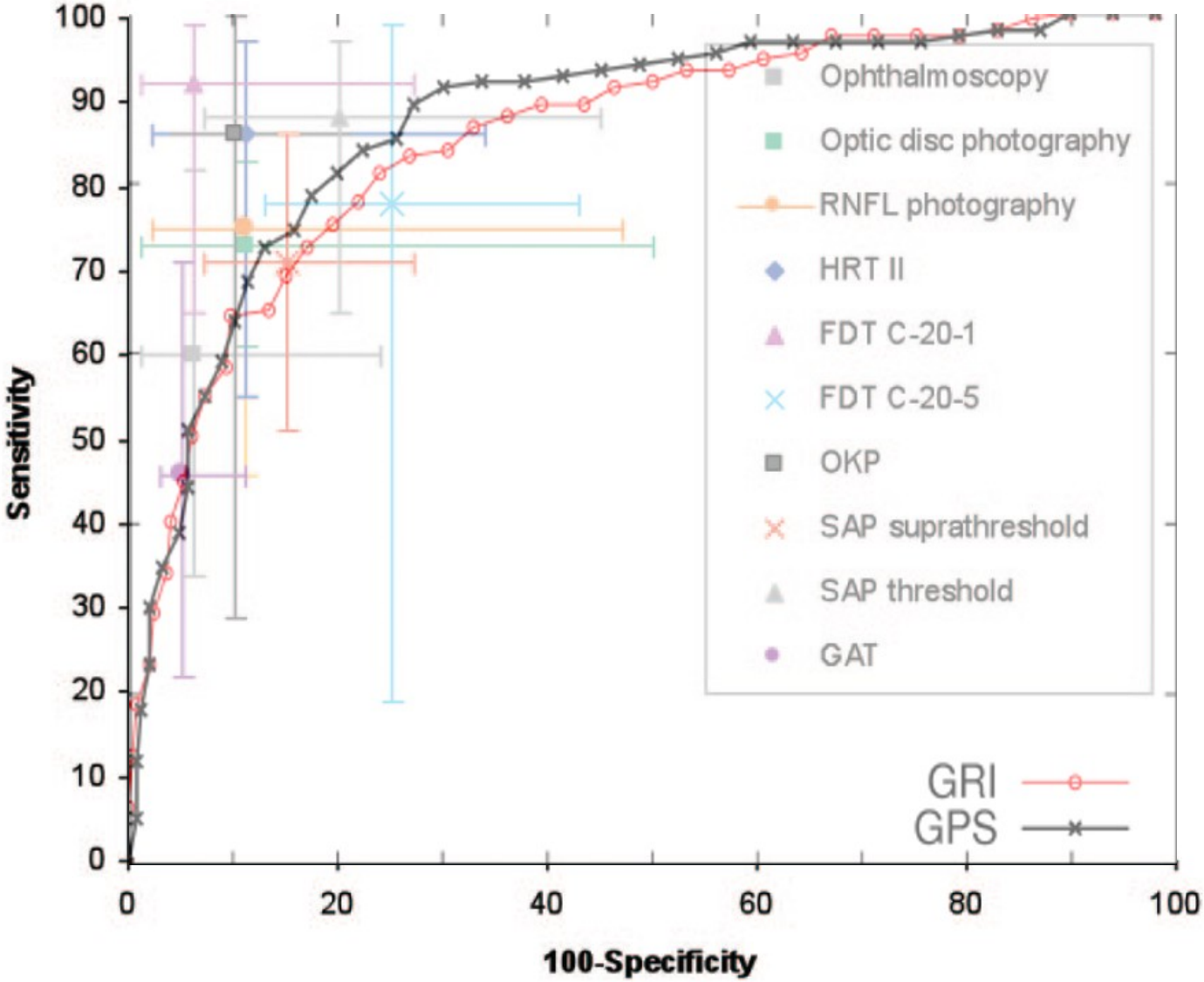
Optic disc photography (all studies $n = 6$, higher quality studies $n = 3$); HRT II (all studies $n = 3$, higher quality studies $n = 2$); FDT C-20-5 (all studies $n = 5$, higher quality studies $n = 2$); SAP threshold (all studies $n = 5$, higher quality studies $n = 2$).

- Study quality determined by QUADAS quality assessment tool / questionaire

HSROC


Sensitivity
100
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
100-Specificity
Ophthalmoscopy
Optic disc photography
RNFL photography
HRT II
FDT C-20-1
FDT C-20-5
OKP
SAP suprathreshold
SAP threshold
GAT

HSROC: Established methods vs. GRI / GPS



Conclusions

- Many candidate tests
  - No group of tests was clearly more accurate
- Poor performance tests (based on limited data)
  - ophthalmoscopy, SAP, retinal photography, GA tonometer
- Better diagnostic performance
  - FDT, HRT II, OKP
  - -> Directly comparative study in relevant population
- Limitations
  - Only 6 of 40 studies directly compared two or more tests
  - Out of date: Gdx, OCT missing
    - But, methods are transparent and reproducible
- Possible common statistical framework to evaluate and compare GRI (over several studies)