Biomarker & Test Evaluation Program

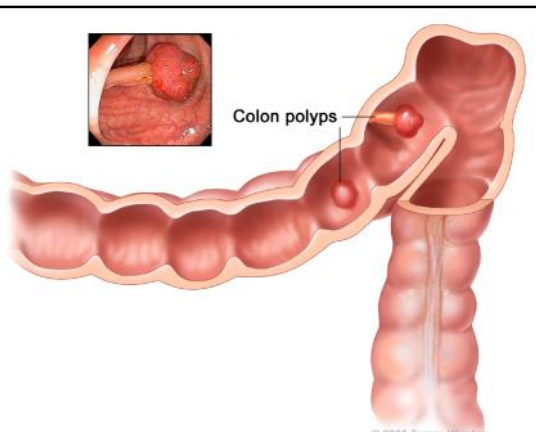
EVALUATING MEDICAL TESTS:

FROM PERFORMANCE
TO EFFECTIVENESS

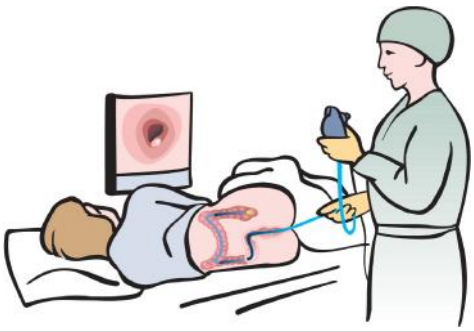
Patrick BOSSUYT

..insufficient evidence…

Evaluating Medical Tests



Colonoscopy



ntvg Nederlands Tijdschrift voor Geneeskunde

COMMENTAAR

Ontlastingstest bevolkingsonderzoek darmkanker

GOEDE KWALITEIT TEGEN EEN GOEDE PRIJS

COMMENTAAR

Bevolkingsonderzoek naar dikkedarmkanker: welke ontlastingstest?

Fecal hemoglobin testing

Test Evaluation: Three questions

Question	Feature

Test Evaluation: Three questions

Question	Feature
Is it true?	Technical Performance

Test Evaluation: Three questions

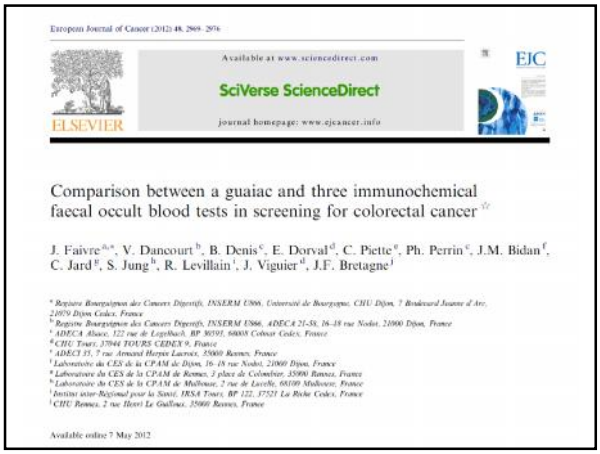
Question	Feature
Is it true?	Technical Performance
Is it meaningful?	Clinical Performance

Test Evaluation: Three questions

Question	Feature
Is it true?	Technical Performance
Is it meaningful?	Clinical Performance
Is it useful?	Clinical Effectiveness

1. Technical Performance

Is it true?



European Journal of Cancer (2012) 48, 2969–2976

Available at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.ejcancer.info

EJC

ELSEVIER

Comparison between a guaiac and three immunochemical faecal occult blood tests in screening for colorectal cancer ☆

J. Faivre [a,*], V. Dancourt [b], B. Denis [c], E. Dorval [d], C. Piette [e], Ph. Perrin [c], J.M. Bidan [f], C. Jard [g], S. Jung [h], R. Levillain [i], J. Viguier [d], J.F. Bretagne [j]

Available online 7 May 2012

2. Clinical Performance

Is it meaningfull?

Diagnostic Accuracy Study

Medical Test

Gold Standard

Diagnostic Accuracy Study

Series of Patients

Medical Test

Reference standard

Cross-classification

The results

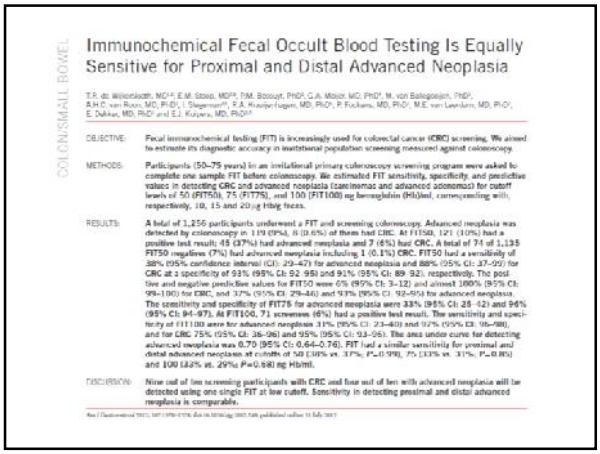
Reference Standard

BioMarker		Target Condition	Other Condition
	Positive		
	Negative		

The results

Reference Standard

BioMarker		Target Condition	Other Condition
	Positive	TP	FP
	Negative	FN	TN



COLON/SMALL BOWEL

Immunochemical Fecal Occult Blood Testing Is Equally Sensitive for Proximal and Distal Advanced Neoplasia

T.R. de Wijkerslooth, MD[a], E.M. Stoop, MD[a], P.M. Bossuyt, PhD[b], C.A. Meijer, MD, PhD[a], M. van Ballegooijen, PhD[c], A.M.E. van Roon, MD, PhD[a], I. Stegeman[a], F.A. Kraaijenhagen, MD, PhD[a], P. Fockens, MD, PhD[a], M.E. van Leerdam, MD, PhD[a], E. Dekker, MD, PhD[a] and E.J. Kuipers, MD, PhD[a]

OBJECTIVE: Fecal immunochemical testing (FIT) is increasingly used for colorectal cancer (CRC) screening. We aimed to estimate its diagnostic accuracy in invitational population screening measured against colonoscopy.

METHODS: Participants (50–75 years) in an invitational primary colonoscopy screening program were asked to complete one sample FIT before colonoscopy. We estimated FIT sensitivity, specificity, and predictive values in detecting CRC and advanced neoplasia (carcinomas and advanced adenomas) for cutoff levels of 50 (FIT50), 75 (FIT75), and 100 (FIT100) ng hemoglobin (Hb)/ml, corresponding with, respectively, 10, 15 and 20 μg Hb/g feces.

RESULTS: A total of 1,256 participants underwent a FIT and screening colonoscopy. Advanced neoplasia was detected by colonoscopy in 119 (9%), 8 (0.6%) of them had CRC. At FIT50, 121 (10%) had a positive test result; 45 (37%) had advanced neoplasia and 7 (6%) had CRC. A total of 74 of 1,135 FIT50 negatives (7%) had advanced neoplasia including 1 (0.1%) CRC. FIT50 had a sensitivity of 38% (95% confidence interval (CI): 29–47) for advanced neoplasia and 88% (95% CI: 37–99) for CRC at a specificity of 93% (95% CI: 92–95) and 91% (95% CI: 89–92), respectively. The positive and negative predictive values for FIT50 were 6% (95% CI: 3–12) and almost 100% (95% CI: 99–100) for CRC, and 37% (95% CI: 29–46) and 93% (95% CI: 92–95) for advanced neoplasia. The sensitivity and specificity of FIT75 for advanced neoplasia were 33% (95% CI: 25–42) and 96% (95% CI: 94–97). At FIT100, 71 screenees (6%) had a positive test result. The sensitivity and specificity of FIT100 were for advanced neoplasia 31% (95% CI: 23–40) and 92% (95% CI: 90–93), and for CRC 75% (95% CI: 36–96) and 95% (95% CI: 93–96). The area under curve for detecting advanced neoplasia was 0.70 (95% CI: 0.64–0.76). FIT had a similar sensitivity for proximal and distal advanced neoplasia at cutoffs of 50 (38% vs. 37%, P=0.99), 75 (33% vs. 31%, P=0.85) and 100 (33% vs. 29% P=0.68) ng Hb/ml.

DISCUSSION: Nine out of ten screening participants with CRC and four out of ten with advanced neoplasia will be detected using one single FIT at low cutoff. Sensitivity in detecting proximal and distal advanced neoplasia is comparable.

Diagnostic Accuracy Study

1,256 Patients

↓

iFOBT

↓

Colonoscopy

↓

Cross-classification

The results

Colonoscopy

iFOBT	Adv Neoplasia	Other
Positive	45	76
Negative	74	1,061

Measures of Diagnostic Test Accuracy

- Sensitivity & Specificity 38 % 93 %
- Predictive Values 37 % 93 %
- Likelihood Ratios 5.66 .67
- Diagnostic Odds Ratio 8.5

	PE	Non PE
Positive	45	76
Negative	74	1,061

3. Clinical Effectiveness

Is it usefull?

New interventions should

- ..extend life
- ..reduce morbidity
- ..maintain health

- ..reduce costs
- ..reduce risks
- ..make things simpler

Testing should

- ..extend life
- ..reduce morbidity
- ..maintain health

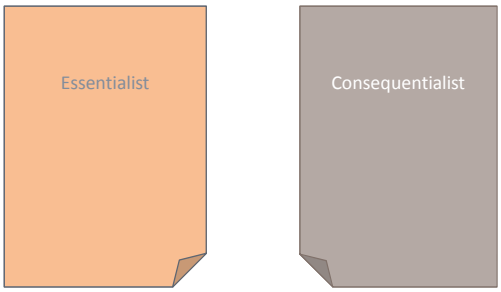
- ..reduce costs
- ..reduce risks
- ..make things simpler

EvidenceLive 13

- Medical Tests
  - should be treated like other interventions in EBM
  - should be evaluated like other interventions

    (up to a point…)

Value of Medical Tests



Essentialism

the theory that
the value of a marker or a medical test
should be judged
by the 'trueness'
of its results

Consequentialism

the theory that
the value of a marker or a medical test
should be judged
by the value
of its consequences

Two views on tests

	Essentialism	Consequentialism
Focus		
Value		
Emphasis		
Statistics		

Two views on tests

	Essentialism	Consequentialism
Focus	Results	Consequences
Value	Truth	Usefulness
Emphasis	Validity	Utility
Statistics	Accuracy	Effectiveness

Consequences of Testing

TEST

Consequences of Testing

TEST → Outcome

Consequences of Testing

TEST → Outcome

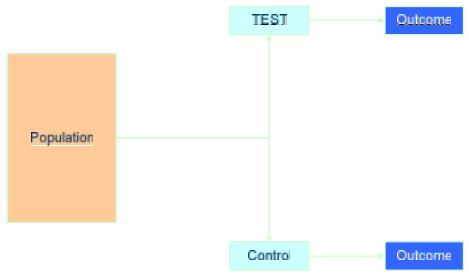
Control

Consequences of Testing

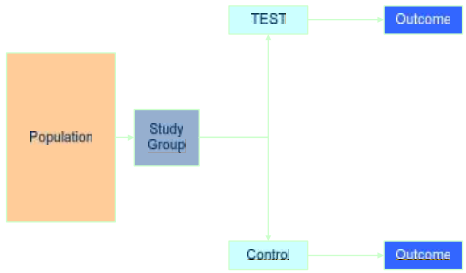
TEST → Outcome

Control → Outcome

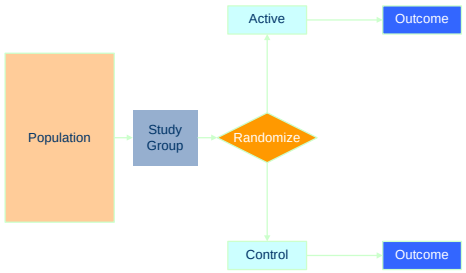
Consequences of Testing



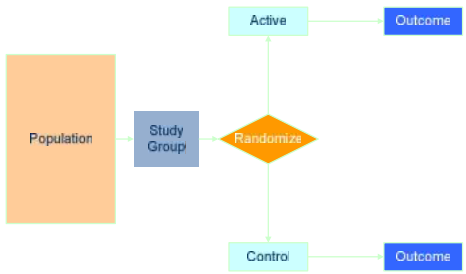
Consequences of Testing



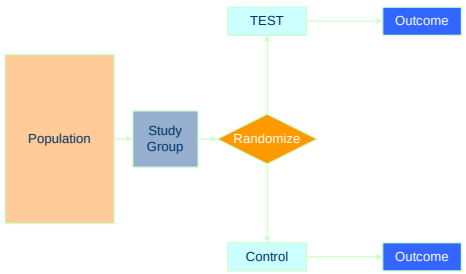
Consequences of Testing



Randomized Clinical Trial



Medical Test RCT



Clinical Effectiveness

	Explanation
Health Outcome	Health outcomes that matter to patients and society: to prevent premature death, to restore or maintain functional health.
Probabilistic	Not all outcomes will be observed in everyone tested; evaluations will be made at the group level, and expressed in terms of a distribution of outcomes.
Comparative	Clinical utility from a testing or marker based strategy is defined relative to a comparator strategy: current best standard practice.

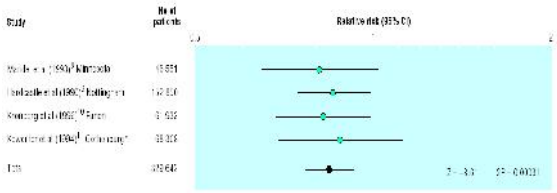
A national colorectal cancer screening programme

Screening for colorectal cancer using the faecal occult blood test, Hemoccult (Review)

Hewitson P, Glasziou PP, Irwig L, Towler B, Watson E

THE COCHRANE COLLABORATION®

Meta-analysis CRC Screening RCT



Relative Risk CRC Mortality: 0.84

(95% CI: 0.77 to 0.93)

BMJ 1998;317:559–65

ntvg Nederlands Tijdschrift voor Geneeskunde

COMMENTAAR

Ontlastingstest bevolkingsonderzoek darmkanker

GOEDE KWALITEIT TEGEN EEN GOEDE PRIJS

COMMENTAAR

Bevolkingsonderzoek naar dikkedarmkanker: welke ontlastingstest?

i-FOBT: What Evidence?

Three questions

Question	Feature
Is it true?	Technical Performance
Is it meaningful?	Clinical Performance
Is it useful?	Clinical Effectiveness

Decisions about Tests

- Guided by consequences
- Technical and Clinical Performance not sufficient
- But…
- Often can be redefined as necessary (and sufficient) conditions for effectiveness

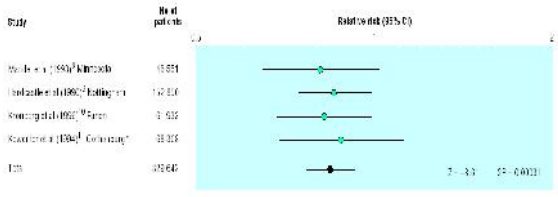
AHRC EFFECTIVE HEALTH CARE PROGRAM WHITE PAPER SERIES

Using the Principles of Randomized Controlled Trial Design to Guide Test Evaluation

Sarah J. Lord, MBBS, MS, Les Irwig, MBBCh, PhD, Patrick M. M. Bossuyt, PhD

A national colorectal cancer screening programme

Meta-analysis CRC Screening RCT

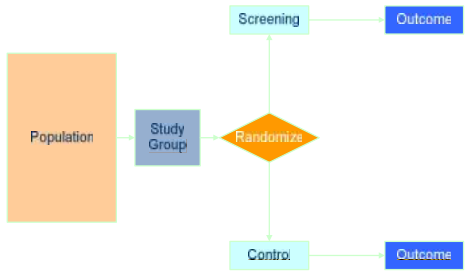


Relative Risk CRC Mortality: 0.84

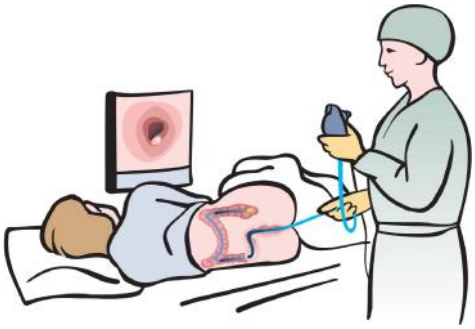
(95% CI: 0.77 to 0.93)

BMJ 1998;317:559–65

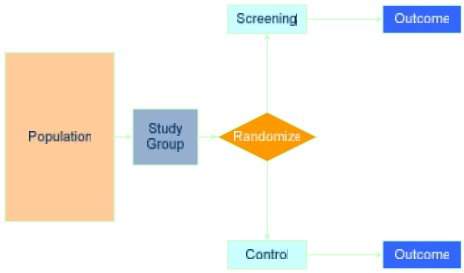
CRC Screening RCT



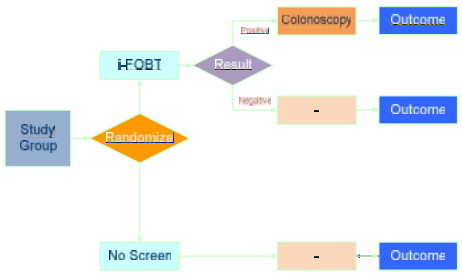
Colonoscopy

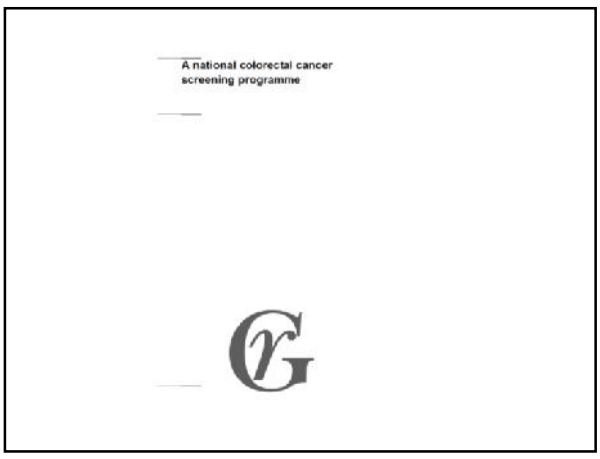


CRC Screening RCT



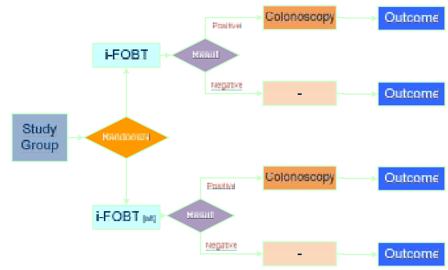
RCT Screening



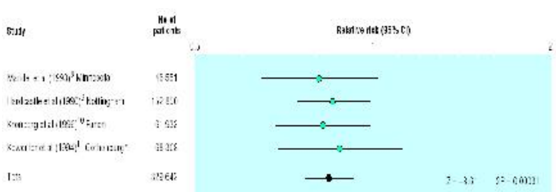


Fecal hemoglobin testing

RCT Screening



Meta-analysis CRC Screening RCT

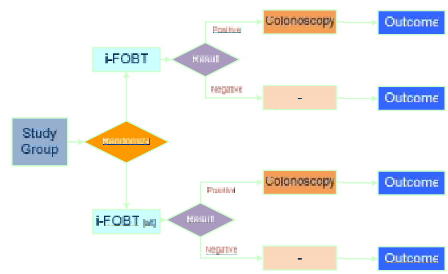


Relative Risk CRC Mortality: 0.84

(95% CI: 0.77 to 0.93)

BMJ 1998;317:559-65

RCT Screening

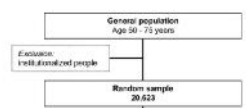


Random Comparison of Guaiac and Immunochemical Fecal Occult Blood Tests for Colorectal Cancer in a Screening Population

Background & Aims: Despite poor performance, guaiac-based fecal occult blood tests (G-FOBT) are most frequently implemented for colorectal cancer screening. Immunochemical fecal occult blood tests (I-FOBT) are claimed to perform better, without randomized comparison in screening populations. Our aim was to randomly compare G-FOBT with I-FOBT in a screening population. ***Methods:*** We conducted a population-based study on a random sample of 20,623 individuals 50–75 years of age, randomized to either G-FOBT (Hemoccult-II) or I-FOBT (OC-Sensor). Tests and invitations were sent together. For I-FOBT, the standard cutoff of 100 ng/ml was used. Positive FOBTs were verified with colonoscopy. Advanced adenomas …

blood and quality control on the evaluation of the tests is hardly possible. Despite these disadvantages, the G-FOBT is still the most implemented test for CRC screening.

A promising alternative is the immunochemical fecal occult blood test (I-FOBT). I-FOBTs are also inexpensive and noninvasive; in addition, these tests are often easier to carry out than G-FOBTs. Another advantage of I-FOBTs is that they are specific for human blood. The most prominent advantage is that many I-FOBTs make quality control possible. At least in theory, they also promise better diagnostic performance than G-FOBTs. In several studies I-FOBTs, seem to have higher specificity compared with G-FOBTs.



ntvg Nederlands Tijdschrift voor Geneeskunde

COMMENTAAR

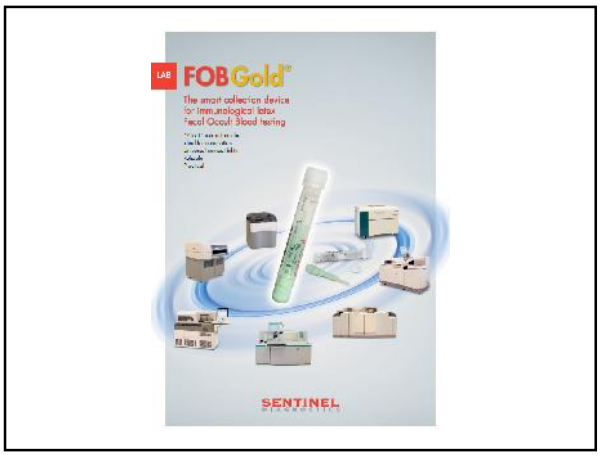
Ontlastingstest bevolkingsonderzoek darmkanker

GOEDE KWALITEIT TEGEN EEN GOEDE PRIJS

COMMENTAAR

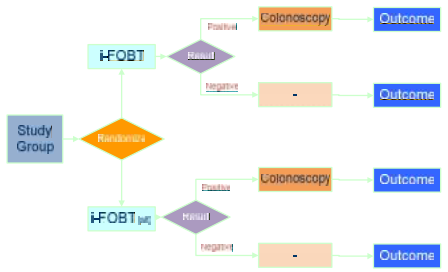
Bevolkingsonderzoek naar dikkedarmkanker: welke ontlastingstest?

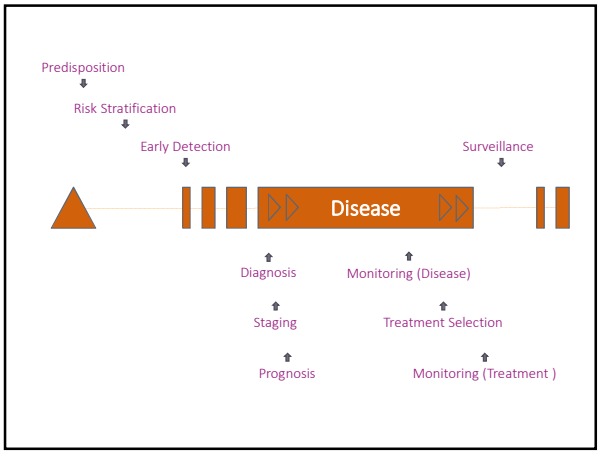
i-FOBT: What Evidence?



Fecal hemoglobin testing

RCT Screening





Medical Tests in EBM

- Should be treated like other interventions
- Should be evaluated like other interventions

  (up to a point…)

Take Home Messages

- Technical Performance,
  Clinical Performance,
  Clinical Effectiveness
- Decisions about medical tests are based on consequences, not on performance (only).
- Performance requirements can - and should - be defined as necessary conditions for effectiveness.