



..insufficient evidence...

De meeste diagnostische tests (DCT) worden gebruikt om de aanwezigheid van een ziekte te bevestigen of te weerleggen. Het is belangrijk om te weten of de test betrouwbaar is. Dit wordt bepaald door de gevoeligheid en de specificiteit van de test. De gevoeligheid is de kans dat de test positief is bij aanwezigheid van de ziekte. De specificiteit is de kans dat de test negatief is bij afwezigheid van de ziekte.

De meeste tests zijn niet perfect. Ze kunnen fouten maken. Dit kan leiden tot een verkeerd resultaat. Het is daarom belangrijk om de test te interpreteren in combinatie met andere informatie, zoals de symptomen van de patiënt en de resultaten van andere tests.

## Evaluating Medical Tests

Colon polyps

© 2005 Teresa Winslow  
LSD. Alle rechten voorbehouden.

## Colonoscopy

ntv g | Nederlands Tijdschrift voor Geneeskunde

COMMENTAAR  
**Ontlastingstest bevolkingsonderzoek darmkanker**  
GOEDE KWALITEIT TEGEN EEN GOEDE PRIJS

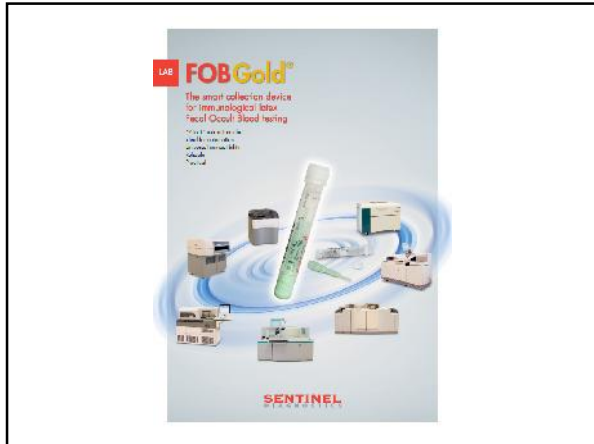
Wanneer u verwacht dat de eerste markt van het bevolkingsonderzoek naar darmkanker met een onderzoek naar de kwaliteit van de test zal worden uitgevoerd, wordt het belangrijk om te weten of de test betrouwbaar is. Dit wordt bepaald door de gevoeligheid en de specificiteit van de test. De gevoeligheid is de kans dat de test positief is bij aanwezigheid van de ziekte. De specificiteit is de kans dat de test negatief is bij afwezigheid van de ziekte.

De meeste tests zijn niet perfect. Ze kunnen fouten maken. Dit kan leiden tot een verkeerd resultaat. Het is daarom belangrijk om de test te interpreteren in combinatie met andere informatie, zoals de symptomen van de patiënt en de resultaten van andere tests.

## Fecal hemoglobijn testing

Hemocult

De test is de meest nauwkeurige manier om darmkanker te detecteren. Het is een eenvoudige test die thuis kan worden uitgevoerd. Het is belangrijk om de test te interpreteren in combinatie met andere informatie, zoals de symptomen van de patiënt en de resultaten van andere tests.



Test Evaluation: Three questions

Question	Feature

Test Evaluation: Three questions

Question	Feature
Is it true?	Technical Performance

Test Evaluation: Three questions

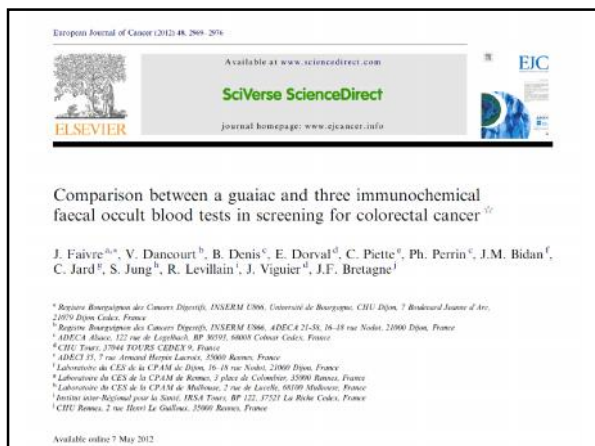
Question	Feature
Is it true?	Technical Performance
Is it meaningful?	Clinical Performance

Test Evaluation: Three questions

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

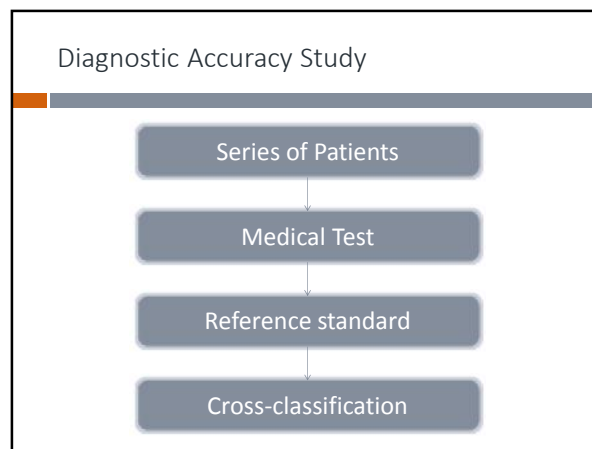
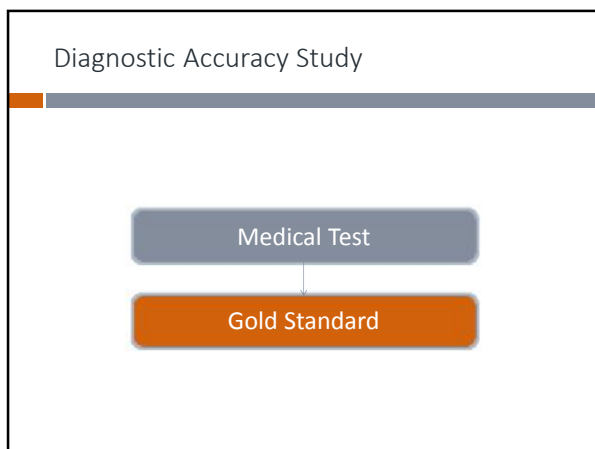
1. Technical Performance

Is it true?



## 2. Clinical Performance

Is it meaningful?



### The results

		Reference Standard	
		Target Condition	Other Condition
BioMarker	Positive		
	Negative		

### The results

		Reference Standard	
		Target Condition	Other Condition
BioMarker	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 Witte, M.D., E.M. Stess, M.D., T.M. Bossuyt, Ph.D., G.A. Meier, M.D., Ph.D., M. von Esge, Ph.D., S.H.C. van Roon, M.D., Ph.D., J. Oosterhoff, S.B. Kruiswijk, M.D., Ph.D., P. Fockens, M.D., Ph.D., M.E. van Leeuwen, M.D., Ph.D., E. Dekker, M.D., Ph.D. and E.J. Kuipers, M.D., Ph.D.

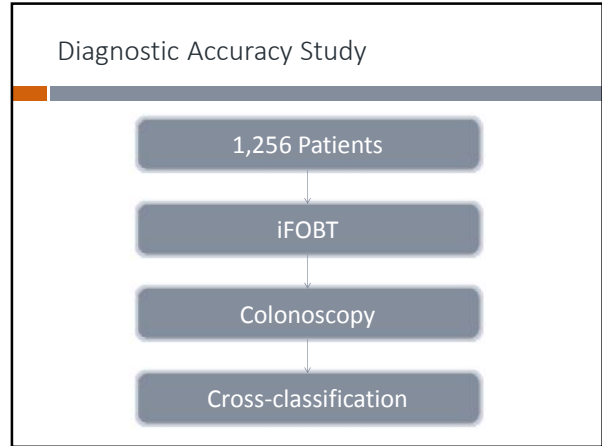
**OBJECTIVE:** Fecal immunochemical testing (FIT) is increasingly used for colorectal cancer (CRC) screening. We aimed to estimate its diagnostic accuracy in institutional population screening measured against colonoscopy.

**METHODS:** Participants (50–75 years) in an institutional primary colorectal cancer 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 (adenomas and advanced adenomas) for cutoff levels of 50 (FIT50), 75 (FIT75), and 100 (FIT100) ng hemoglobin (Hb) per gram, 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%), 10 (0.8%) of those had CRC. All 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 FIT75 neoplasia (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: 67–92) 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: 98–100) for CRC, and 37% (95% CI: 29–44) 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 94% (95% CI: 94–97). All FIT100, 71 screens (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: 90–98), and for CRC 25% (95% CI: 16–36) 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 (1.0% vs. 3.7%, P=0.93), 75 (1.3% vs. 3.1%, P=0.80) and 100 (0.8% vs. 2.9%, P=0.68) ng Hb/g feces.

**CONCLUSION:** None had of the 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.

Am J Gastroenterol 2012; 107: 1070–1076. doi:10.1039/1076.1070.g published online 11 July 2012



### The results

Colonoscopy

		Adv Neoplasia	Other
iFOBT	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 useful?

- ### 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<sup>13</sup>

- Medical Tests
  - should be treated like other interventions in EBM
  - should be evaluated like other interventions (up to a point...)

### Value of Medical Tests

Essentialist

Consequentialist

### Essentialism

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

### Consequentialism

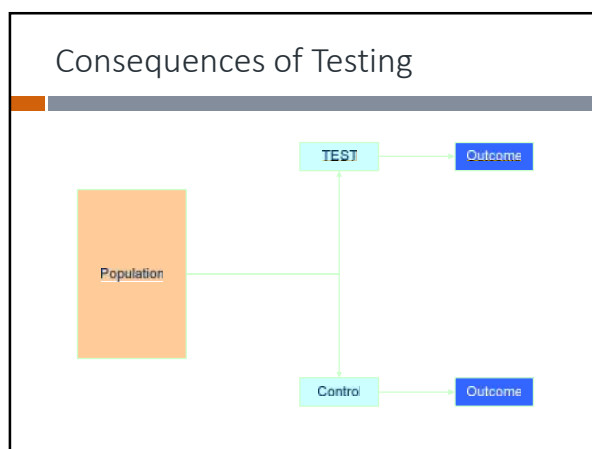
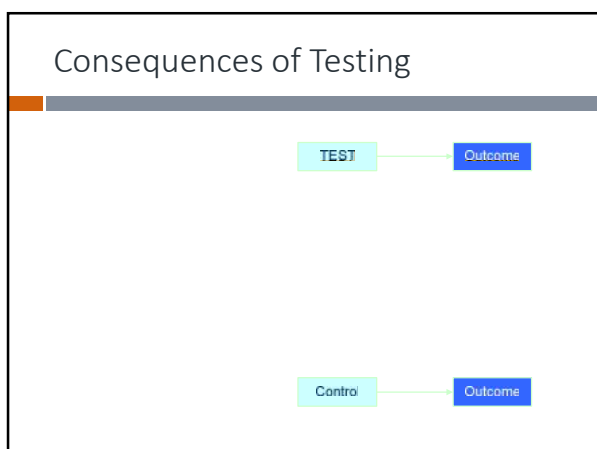
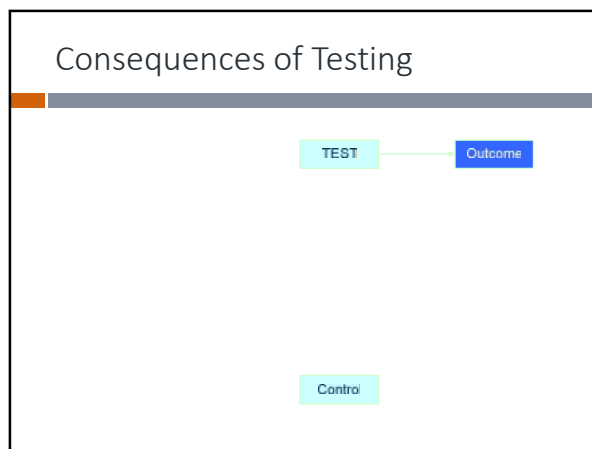
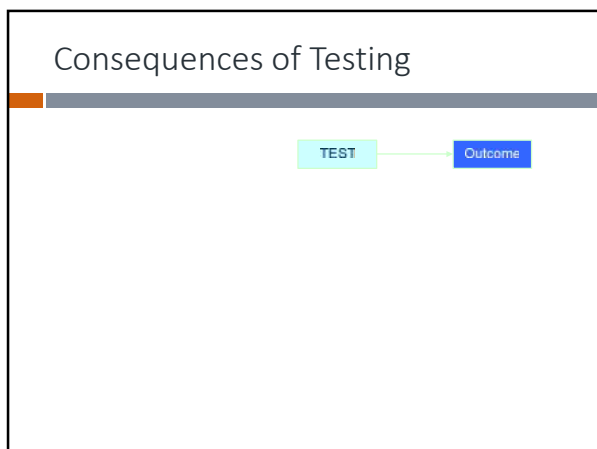
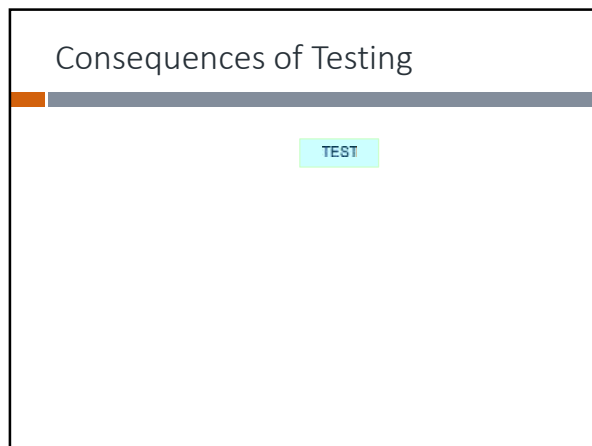
the theory that the value of a marker or a medical test should be judged by the value of its consequences

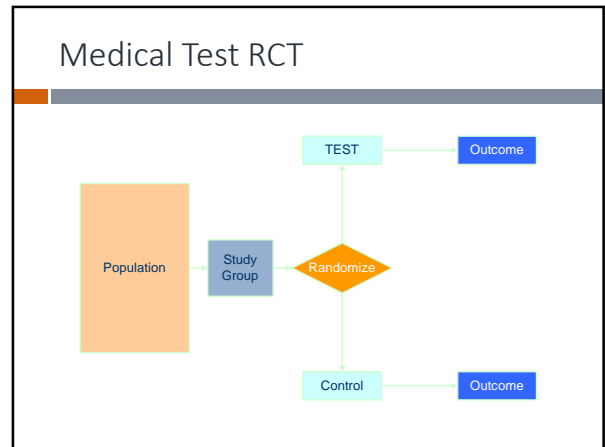
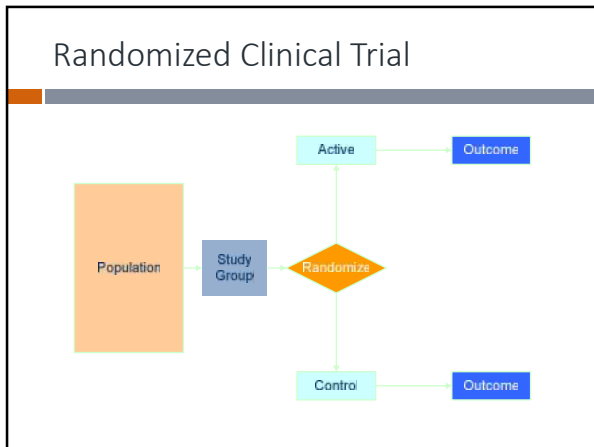
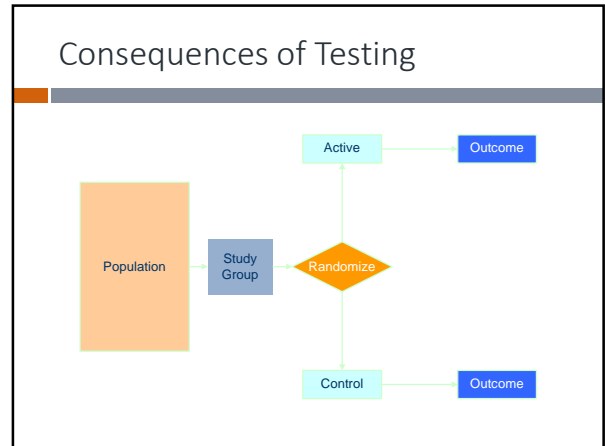
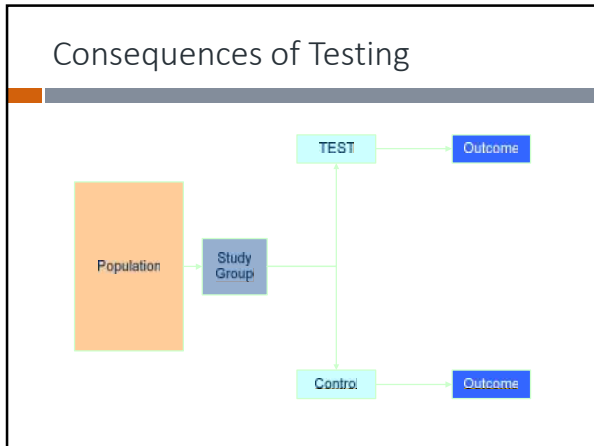
### Two views on tests

	Essentialism	Consequentialism
Focus		
Value		
Emphasis		
Statistics		

### Two views on tests

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





### 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.



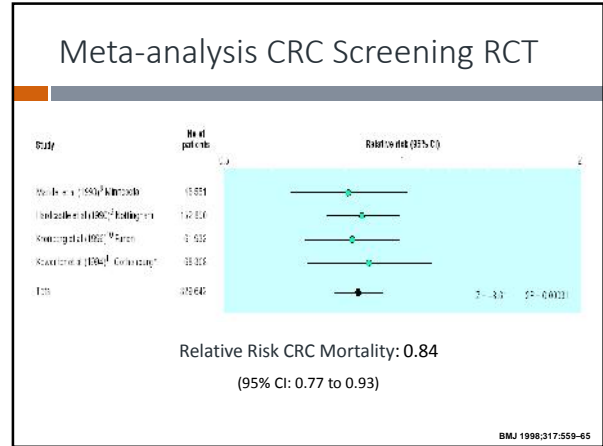


**Screening for colorectal cancer using the faecal occult blood test, Hemoccult (Review)**

Hewison P, Glasziou PR, Irwig L, Tondor B, Watson F.



**THE COCHRANE COLLABORATION®**



**Ontlastingsstest bevolkingsonderzoek darmkanker**

**GOEDE KWALITEIT TEGEN EEN GOEDE PRIJS**

Hans van Hattum, LUC, Hans Bouter en Dr. J. Buijs

**COMMENTAAR**

**Ontlastingsstest bevolkingsonderzoek darmkanker**

**GOEDE KWALITEIT TEGEN EEN GOEDE PRIJS**

Hans van Hattum, LUC, Hans Bouter en Dr. J. Buijs

**ntvg** | Nederlands Tijdschrift Geneeskunde

**COMMENTAAR**

**Ontlastingsstest bevolkingsonderzoek darmkanker**

**GOEDE KWALITEIT TEGEN EEN GOEDE PRIJS**

Hans van Hattum, LUC, Hans Bouter en Dr. J. Buijs

**COMMENTAAR**

**Bevolkingsonderzoek naar dikedarmkanker: welke ontlastingsstest?**

Hans van Hattum, LUC, Hans Bouter en Dr. J. Buijs

## i-FOBT: What Evidence?

### Three questions

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

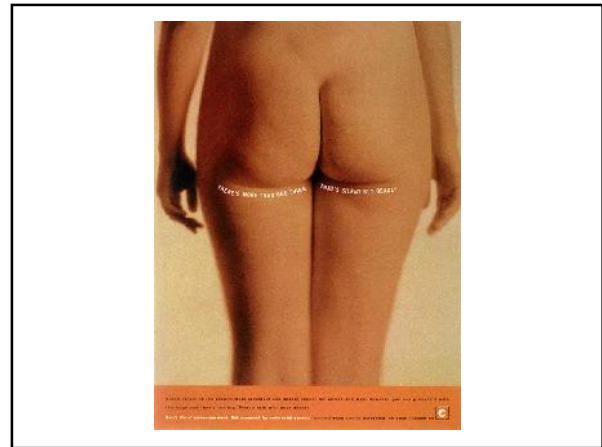
- ### Decisions about Tests
- Guided by consequences
  - Technical and Clinical Performance not sufficient
  - But...
  - Often can be redefined as necessary (and sufficient) conditions for effectiveness

AHRQ EFFECTIVE HEALTH CARE PROGRAM WHITE PAPER SERIES

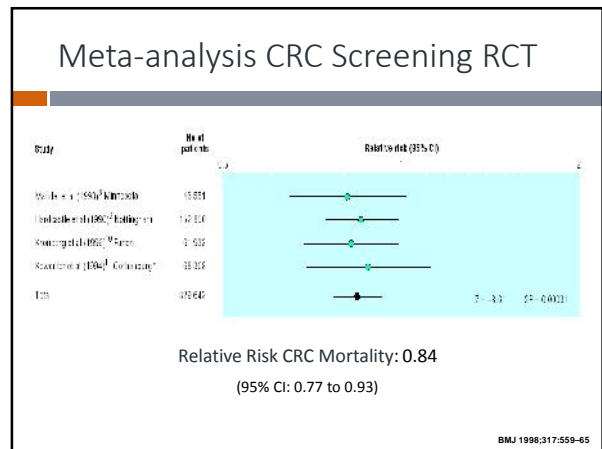
### Using the Principles of Randomized Controlled Trial Design to Guide Test Evaluation

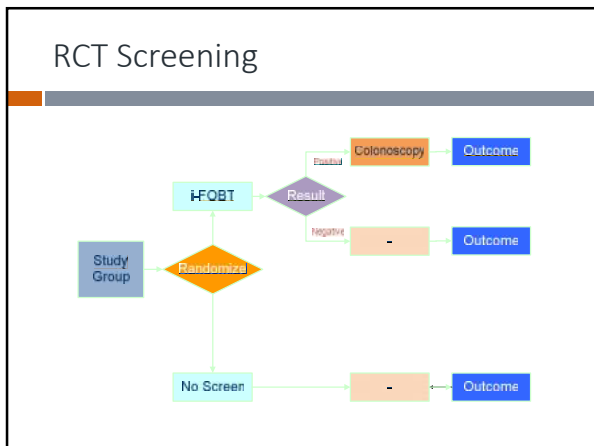
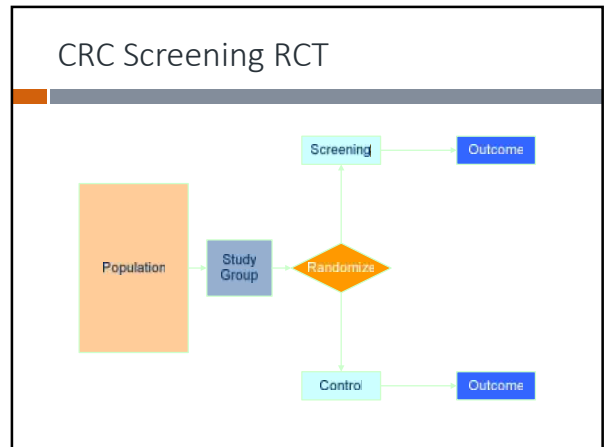
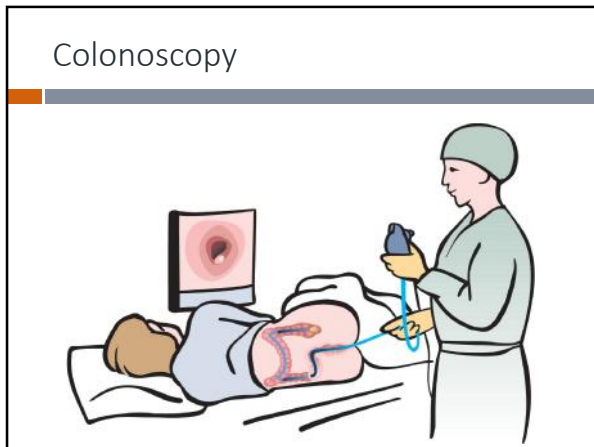
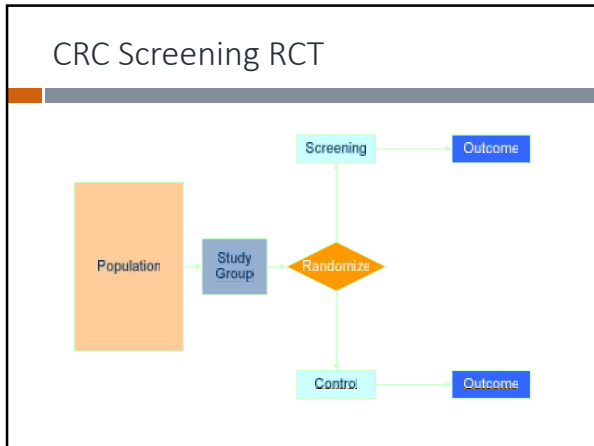
Sarah J. Vard, MD, MS, Lex Irving, MD, PhD, Patrick M. M. Bossuyt, PhD

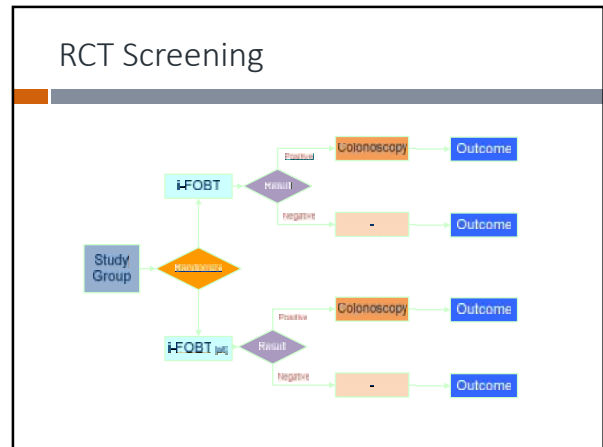
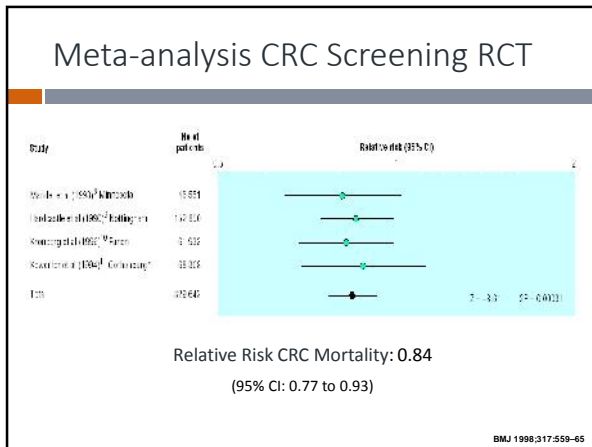
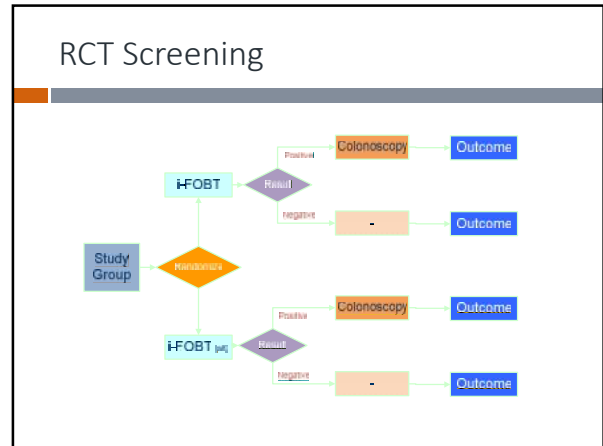
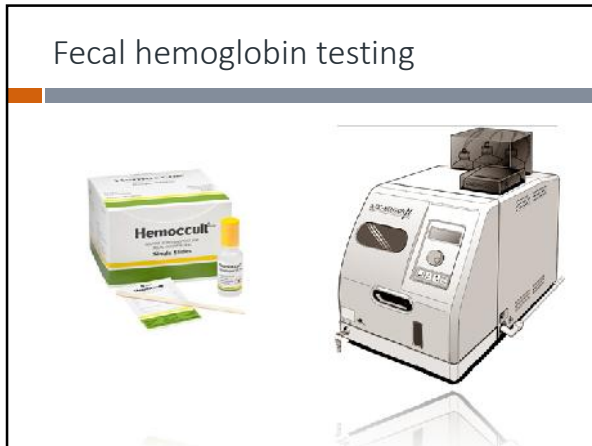
The decision to use a new test should be based on evidence that it will improve patient outcomes or produce other benefits without adversely affecting patients. In particular, decisions made about a new test should be based on the benefits of introducing a new test relative to current best practice. However, evidence that a new test is better than current best practice may not always be necessary. The authors recommend the hypothetical RCT as a conceptual framework to identify what types of comparative evidence are needed for test evaluation. Evaluation begins by asking the major claims for the new test and determining whether it will be used in a replacement, addition, or adjunct to address these claims. The objectives of the hypothetical RCT is constructed to show the essential design elements, including population, prior tests, new test and existing test strategies, and primary and secondary outcomes. Critical steps in the process between study and patient outcomes such as differences in test accuracy changes in treatment, or avoidance of other tests, are identified for each test strategy. An evidence base, the basis of the critical design, is identified and presented to determine the most important outcome for evaluation. Hypothetical RCTs will not be necessary if it is not to use other sources of evidence to address these questions. It will still depend on consensus as to the spectrum of patients included by the old and new test strategies. Key words: Diagnostic techniques and procedures; health care quality and quantity; randomized controlled trials; reproductive assessment; health care; risk; death; Making 2006;29:1-12



A national colorectal cancer screening programme



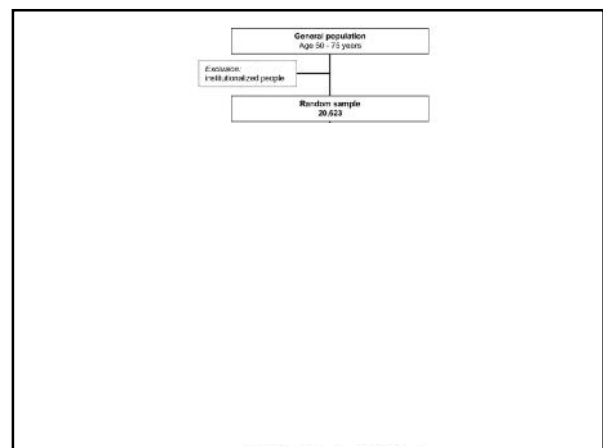


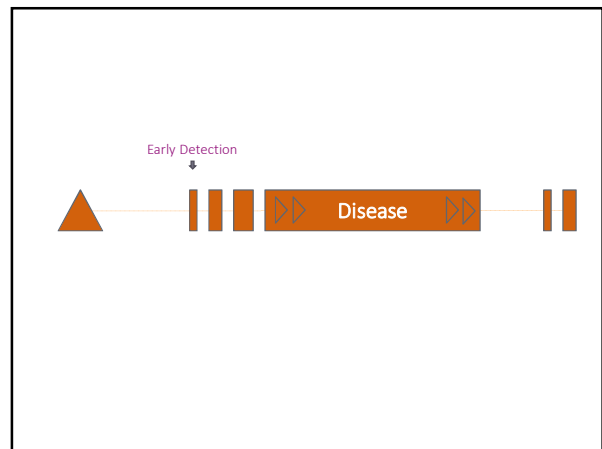
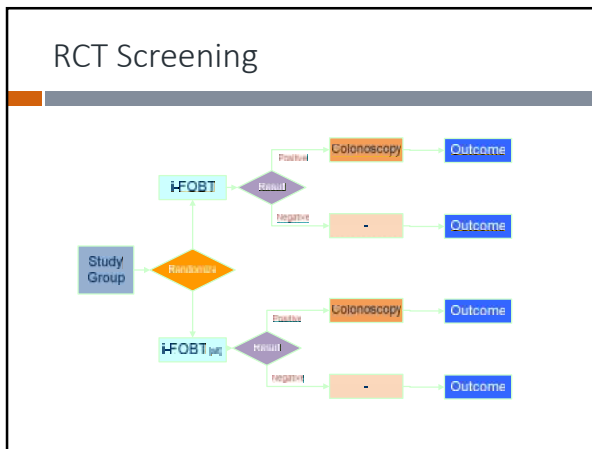
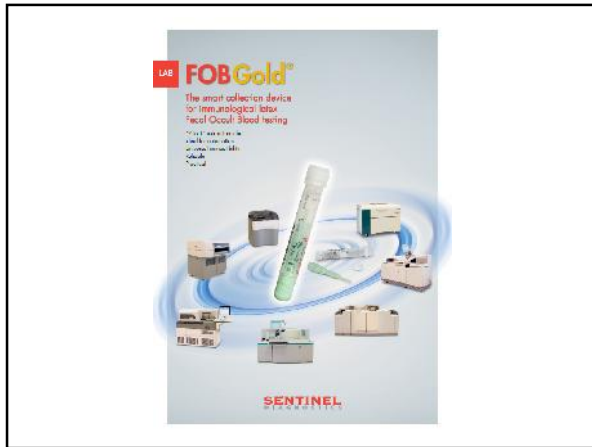
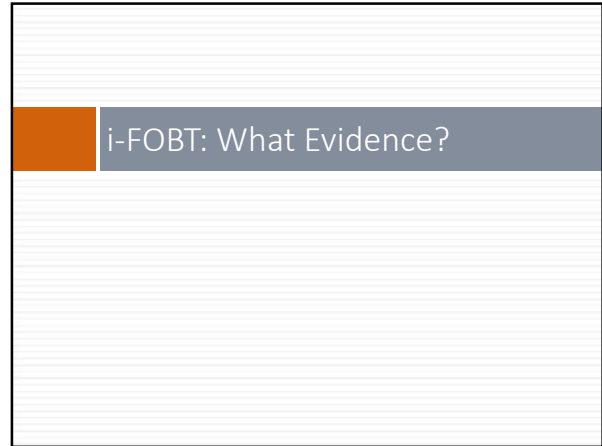


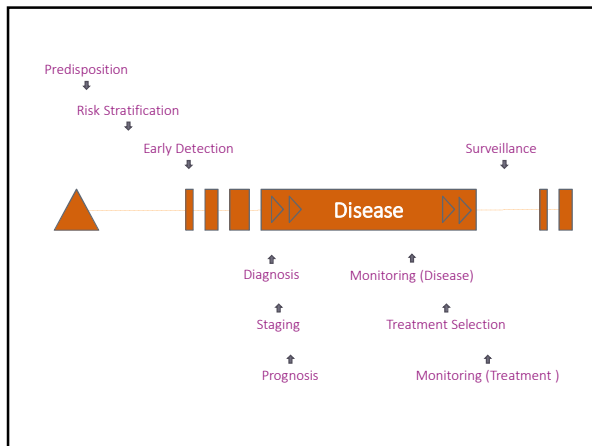
#### Random Comparison of Guaiac and Immunochemical Fecal Occult Blood Tests for Colorectal Cancer in a Screening Population

LEO G. VAN ROSSUM,\* ANNE F. VAN RIEL†, ROBERT J. LAHILL,† MARTIN G. VAN OLEN,\* PAUL. ROOKENS,† HWY H. VAN RIEL,N\* ANIEL L. VULLIUS,† JAN H. JANGEN,† and L. VILLEN DUSCH†

**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 comparisons in screening populations. Our aim was to more closely compare G-FOBT with i-FOBT in a screening population. **Methods:** We conducted a population-based study on a random sample of 30,222 individuals 50-75 years of age, randomized to either G-FOBT (Hemoccult ID or i-FOBT (OC-Sensor). Tests and investigations were sent together. For i-FOBT, the standard cutoff of 100 ng/ml was used. Positive FOBTs were verified with colonoscopy. Advanced adenomas (≥10 mm, ≥3, or ≥1-1-1) were also verified with colonoscopy. **Results:** The overall specificity of G-FOBT was 92.1% (95% CI 91.8-92.4) and 92.1% (95% CI 91.8-92.4) for i-FOBT. The overall sensitivity of G-FOBT was 12.1% (95% CI 11.8-12.4) and 12.1% (95% CI 11.8-12.4) for i-FOBT. The overall specificity of G-FOBT was 92.1% (95% CI 91.8-92.4) and 92.1% (95% CI 91.8-92.4) for i-FOBT. The overall sensitivity of G-FOBT was 12.1% (95% CI 11.8-12.4) and 12.1% (95% CI 11.8-12.4) for i-FOBT. **Conclusion:** The overall specificity of G-FOBT was 92.1% (95% CI 91.8-92.4) and 92.1% (95% CI 91.8-92.4) for i-FOBT. The overall sensitivity of G-FOBT was 12.1% (95% CI 11.8-12.4) and 12.1% (95% CI 11.8-12.4) for i-FOBT.







### 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.

