METHODS OF DISEASE PREDICTION
&
PREDICTION OF TYPE 2 DIABETES:

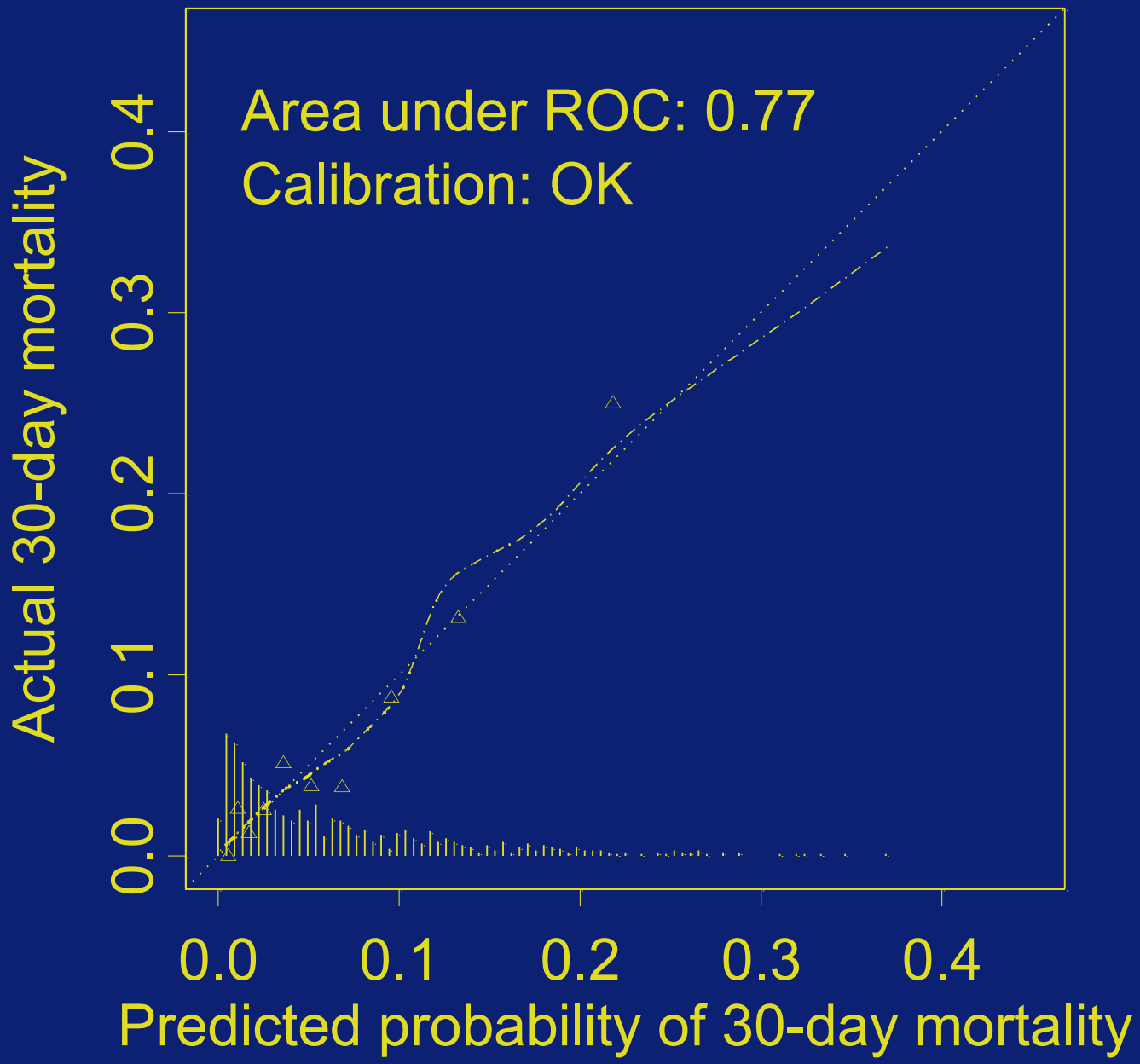
Jaakko Tuomilehto

Department of Public Health
University of Helsinki
Helsinki, Finland

Evaluation of predictions

- Calibration
  - is the average of predictions correct?
  - are predictions at low and high level both correct?
- Discrimination
  - can one distinguish low risk individuals from high risk individuals?

Example: predicted probabilities



3 types of validation

- Apparent: performance on sample used to develop model
- Internal: performance on population underlying the sample
- External: performance on related but slightly different population

Apparent validity

- Easy to calculate
- Gives optimistic performance estimates

Apparent estimates are optimistic since same data are used for:

- Definition of model structure:

  e.g. selection and coding of variables

- Estimation of model parameters:

  e.g. regression coefficients

- Evaluation of model performance:

  e.g. calibration and discrimination

Internal validity

- More difficult to calculate

- Test model in new data, **random/different** from the underlying population

Why internal validation?

- Honest estimate of performance should be obtained, at least for a population similar to the development sample

- Internal validated performance sets an upper limit to what may be expected in other settings (external validity)

External validity

- Moderately easy to calculate when new data are available

- Test model in new data, different from development population

Why external validation?

- Various factors may differ from the development population, including:
  - different selection of participants
  - different definitions of variables
  - different measurement or diagnostic procedures

Internal validation techniques

- Split-sample:
  - development / validation
- Cross-validation:
  - alternating development / validation
  - extreme: n-1 develop / 1 validate ('jack-knife')
- Bootstrap

Bootstrap is the preferred internal validation technique

- bootstrap sample for model development: *n* people drawn *with replacement*
- original sample for validation: *n* people
- difference: optimism
- efficiency: development and validation on *n* people

Example: bootstrap results for logistic regression model

- 30-day mortality ~ a + b_1*sex + b_2*age

Apparent area under the ROC curve:	0.77
Mean area of 200 bootstrap samples:	0.772
Mean area of 200 tests in original:	0.762
Optimism in apparent performance:	0.01
Optimism-corrected area:	0.76

External validation techniques

- Temporal validation: same investigators, validate in recent years
- Spatial validation (other place): same investigators, cross-validate in centers
- Fully external: other investigators, other centers

Example: external validity of logistic regression model

- 30-day mortality ~ a + b_1*sex + b_2*age

Apparent area in 785 patients:	0.77
Tested in 20,318 other patients:	0.74
Tested by other investigators:	?

Summary

- Apparent validity gives an optimistic estimate of model performance
- Internal validity may be estimated for instance by bootstrapping
- External validity should be determined in other populations

The difference between aetiological research and disease prediction in practice

- **Aetiological research** uses maximum efforts to detect true associations between predictor parameters and outcome

- **Disease prediction** in practice uses findings from aetiological research but uses only selected, most powerful predictors in a simplified format

Strategies for diabetes and hyperglycaemia risk identification

- Testing with OGTT
- FPG testing only
- A1C screening
- Random capillary BG screening
- Questionnaire comprising aetiological factors for diabetes

T2D SCREENING TEST PERFORMANCE AND VALIDITY

- **Main questions:**

1. **What is the chance that those who have positive results are affected with diabetes/hyperglycaemia (prediabetes)?**

2. **How well a positive test result will predict the development of future diabetes?**

SCREENING versus DIAGNOSIS

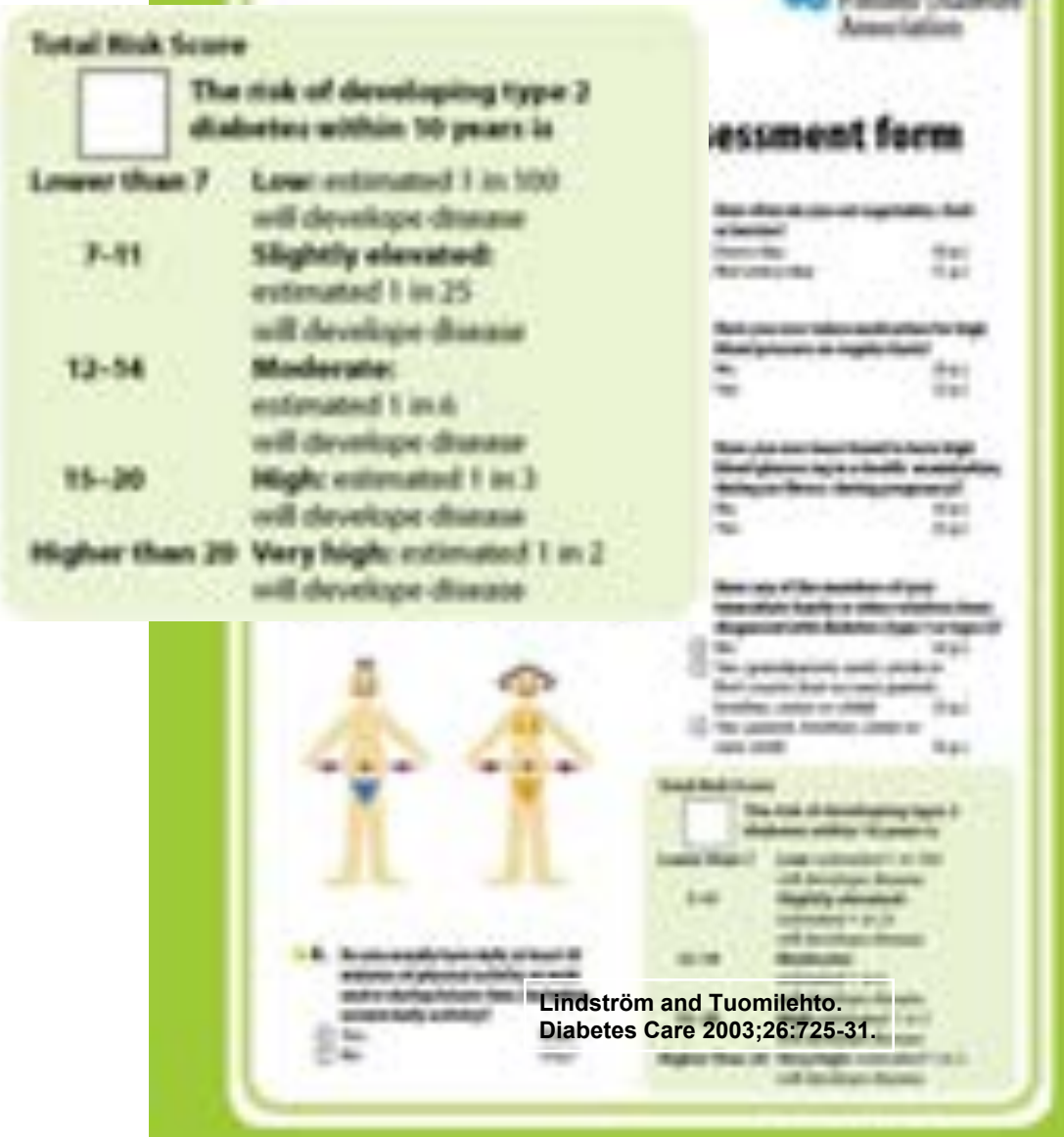
- A screening test is not intended to be diagnostic, but it should be reliable
- Screening procedures are easier to perform and cheaper than diagnostic tests
- A positive screening results requires confirmation through definitive diagnostic tests

Identification of high-risk individuals: FINnish Diabetes RIsk SCore

The FINDRISC:

- Age
- BMI
- Waist
- Physical activity
- Nutrition (f+v)
- Hypertension
- Hyperglycaemia
- Family history

www.diabetes.fi



Lindström and Tuomilehto. Diabetes Care 2003;26:725-31.

FINnish Diabetes RIsk SCore (FINDRISC)

- Developed based on the real prospective data (baseline examination in 1987 and 10-year follow-up)
- Validated in cross-sectional and independent prospective data sets
- Scoring weights for the individual items derived from the empirical data: multivariate logistic model

FINDRISC: The aim

To develop a tool that:

- is simple, inexpensive and reliable way to identify people at high risk of T2D
- can be applied in the general population by lay people
- does not require blood drawing or other measurements that require trained personnel or special equipments

Risk model development:
FINRISK87 - SURVEY

Excluded if
- age < 35 yrs.
- DM medication
- missing variables

4435 subjects with baseline Risk Score

10 years follow-up (drug register)

182 DM cases identified

Risk model validation:
FINRISK92 - SURVEY

Excluded if
- age < 35 yrs.
- DM medication
- missing variables

4586 subjects with baseline Risk Score

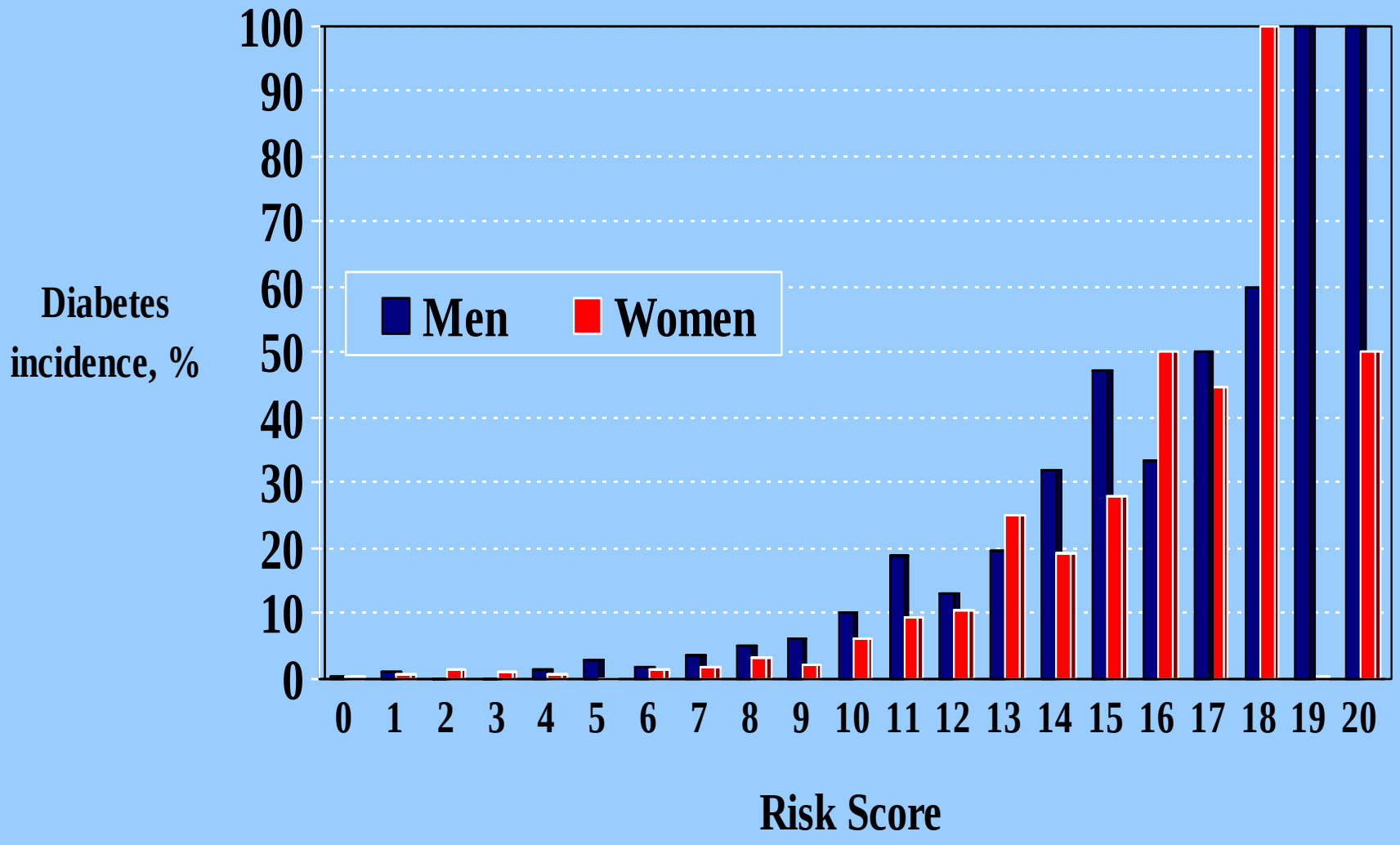
5 years follow-up (drug register)

67 DM cases identified

Analysis of Maximum Likelihood Estimates

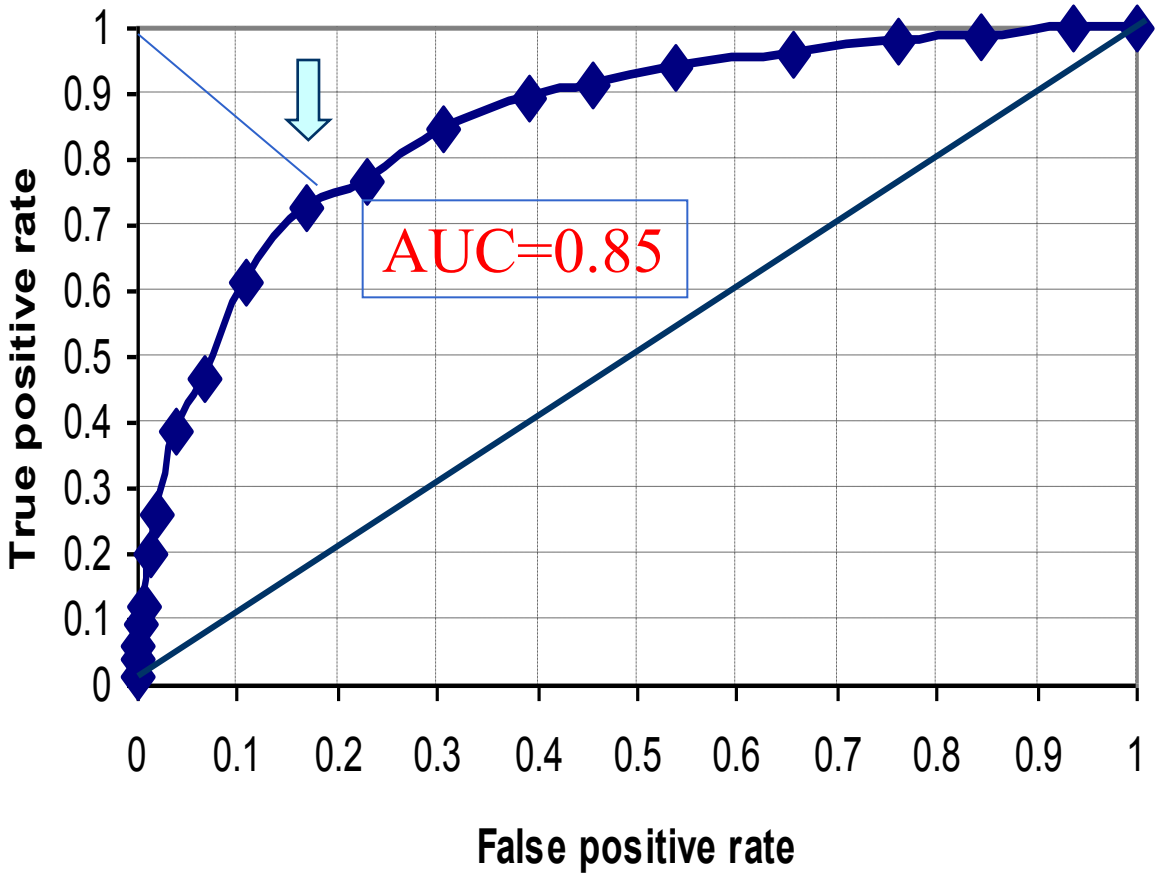
Variable	Parameter Estimate	p	Odds Ratio	RISK SCORE
INTERCEPT	**-5.671**	0.0001		
BMI_D1	**0.011**	0.9777	1.01	**1**
BMI_D2	**0.928**	0.0299	2.53	**3**
WAIST_D1	**1.037**	0.0022	2.82	**3**
WAIST_D2	**1.445**	0.0001	4.24	**4**
AGE_D1	**0.654**	0.0150	1.92	**2**
AGE_D2	**0.945**	0.0003	2.57	**3**
GLUCOSE	**2.261**	0.0001	9.59	**5**
BP_MED	**0.711**	0.0001	2.04	**2**
FRUIT+VEGET	**0.165**	0.3248	1.18	**1**
EXERCISE	**0.264**	0.1964	1.30	**2**

Diabetes incidence during 10-year follow-up by baseline FINDRISC value



ROC - curve for FINDRISC (Score 0-20)

Finrisk87 - Prospective 10-year follow-up



Cutpoint: score ≥10

sensitivity = 0.73

specificity = 0.83

positive predictive value = 0.16

negative predictive value = 0.99

AUC = Area Under the Curve

Sensitivity:

The probability that the people with disease will be test positive.

Specificity:

The probability that the test will be negative if the disease is truly absent.

ROC curve

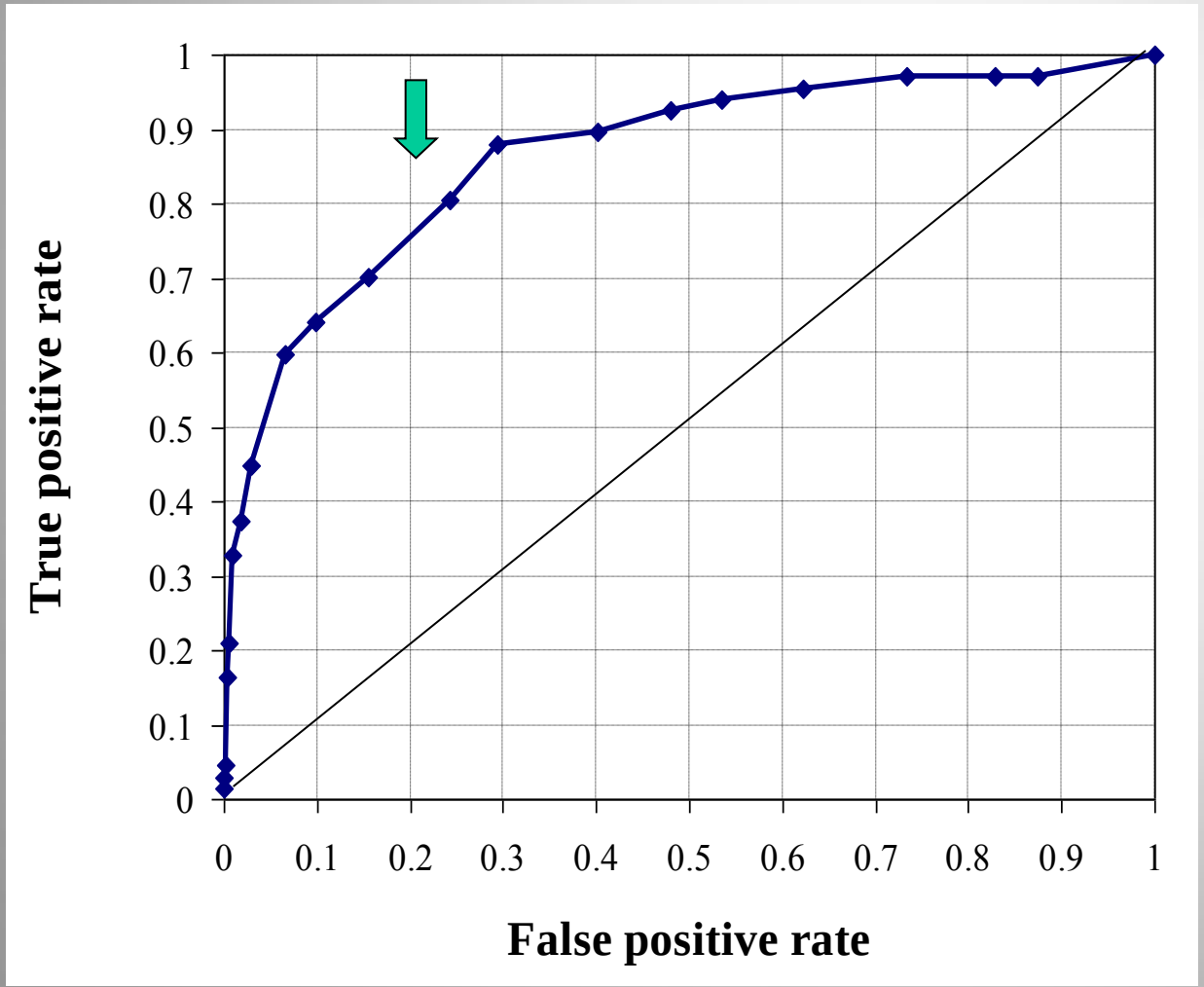
A **receiver operating characteristic curve**, or **ROC curve**, is a graphical plot that illustrates the diagnostic ability of a binary classifier system.

The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate(FPR) at various threshold settings. The true-positive rate is also known as sensitivity, recall or *probability of detection* in machine learning.

The false-positive rate is also known as the fall-out or *probability of false alarm* and can be calculated as (1−specificity).

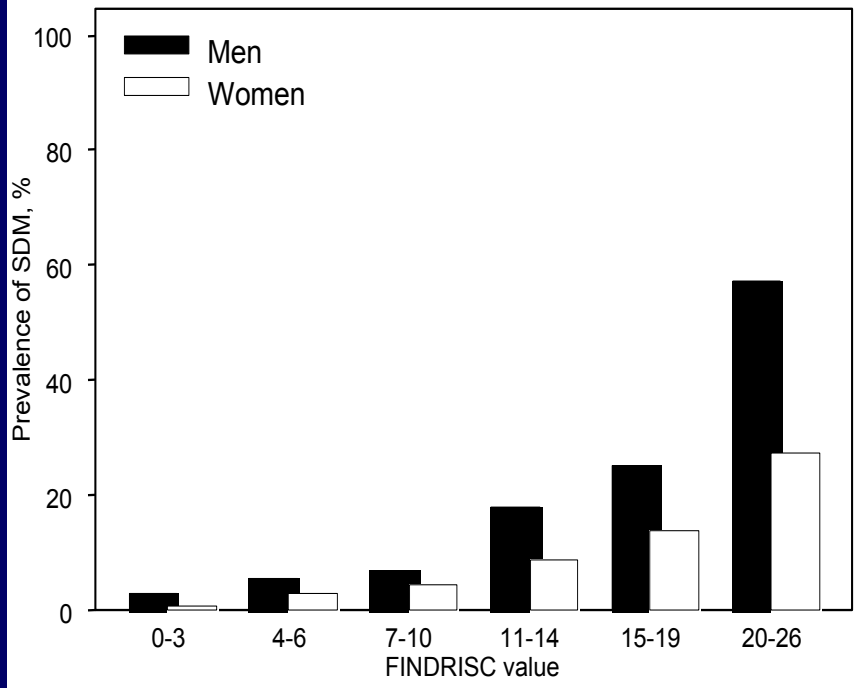
ROC - curve for DM Risk Score validation (Score 0-20):

Finrisk92 - Prospective data

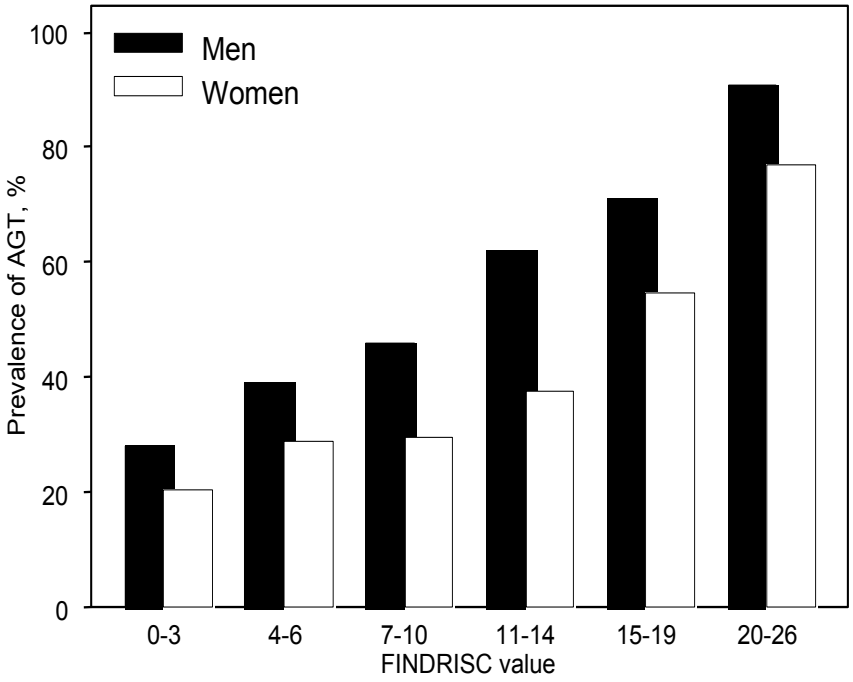


Score	≥9 →
sensitivity=	0.81
specificity=	0.76
Pos. predictive value=	0.05
AUC =	0.87

Prevalence of abnormal glucose tolerance



Unrecognized type 2 diabetes



IGT, IFG or unrecognized T2DM

FINDRISC predicts the risk of myocardial infarction

One point increase in score =

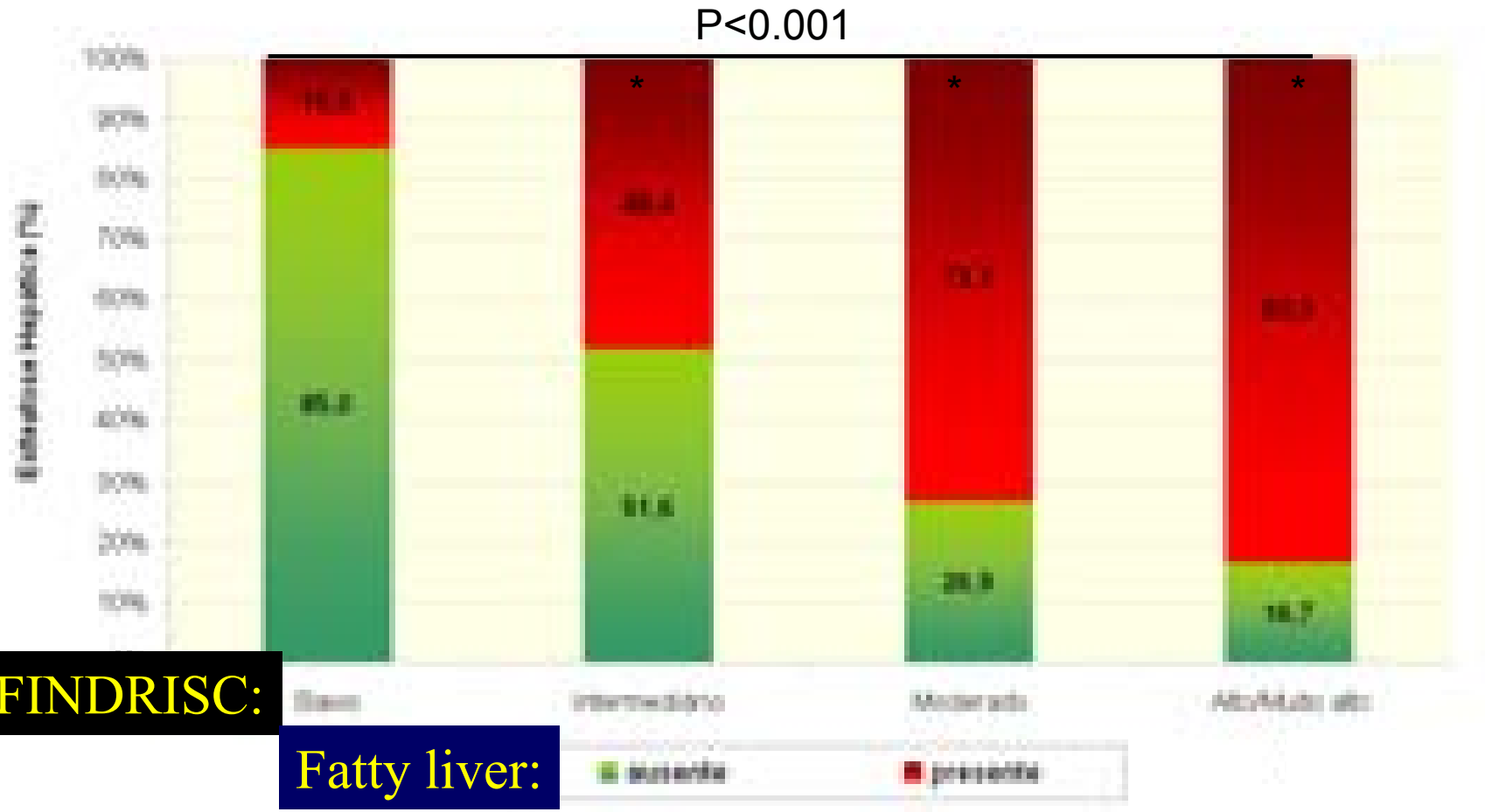
- 15% risk increase in men
- 19% risk increase in in women

	Men		Women	
	Hazard ratio	95% CI	Hazard ratio	95 % CI
Model 1	**1.15**	**1.12–1.19**	**1.19**	**1.14–1.24**
Model 2	**1.15**	**1.11–1.19**	**1.19**	**1.14–1.24**
Model 3	**1.04**	**1.00–1.08**	**1.06**	**1.00–1.12**

Model 1: Diabetes Risk Score alone
Model 2: Model 1 + adjusted for smoking
Model 3: Model 2 + SBP + total and HDL-cholesterol

Silventoinen K, et al. Eur J Cardiovasc Prev Rehabil 2005;12:451–8.

Fatty Liver by FINDRISC in Brazil

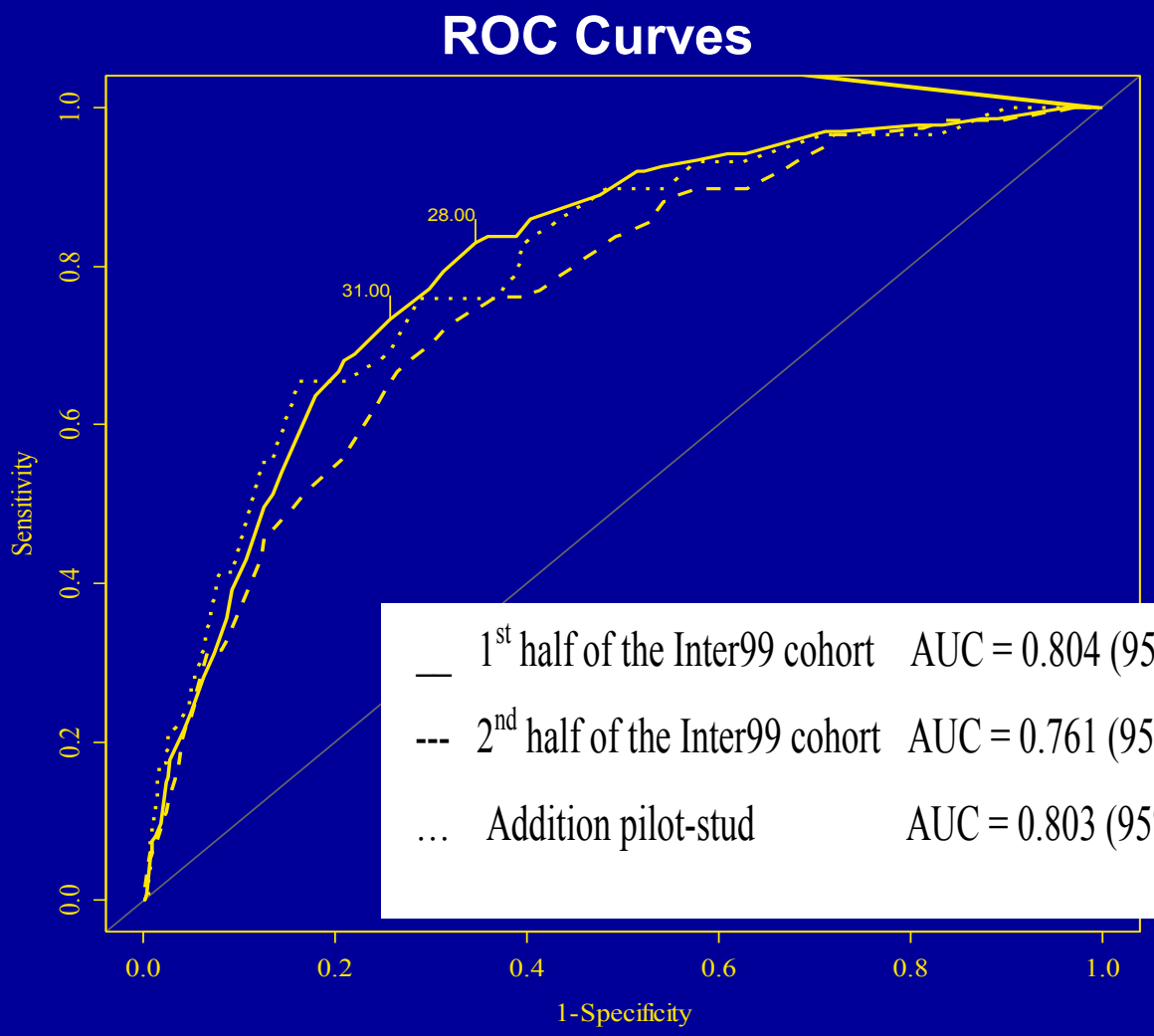


de Carvalho JM, et al Ann Med 2011; 43: 487-94

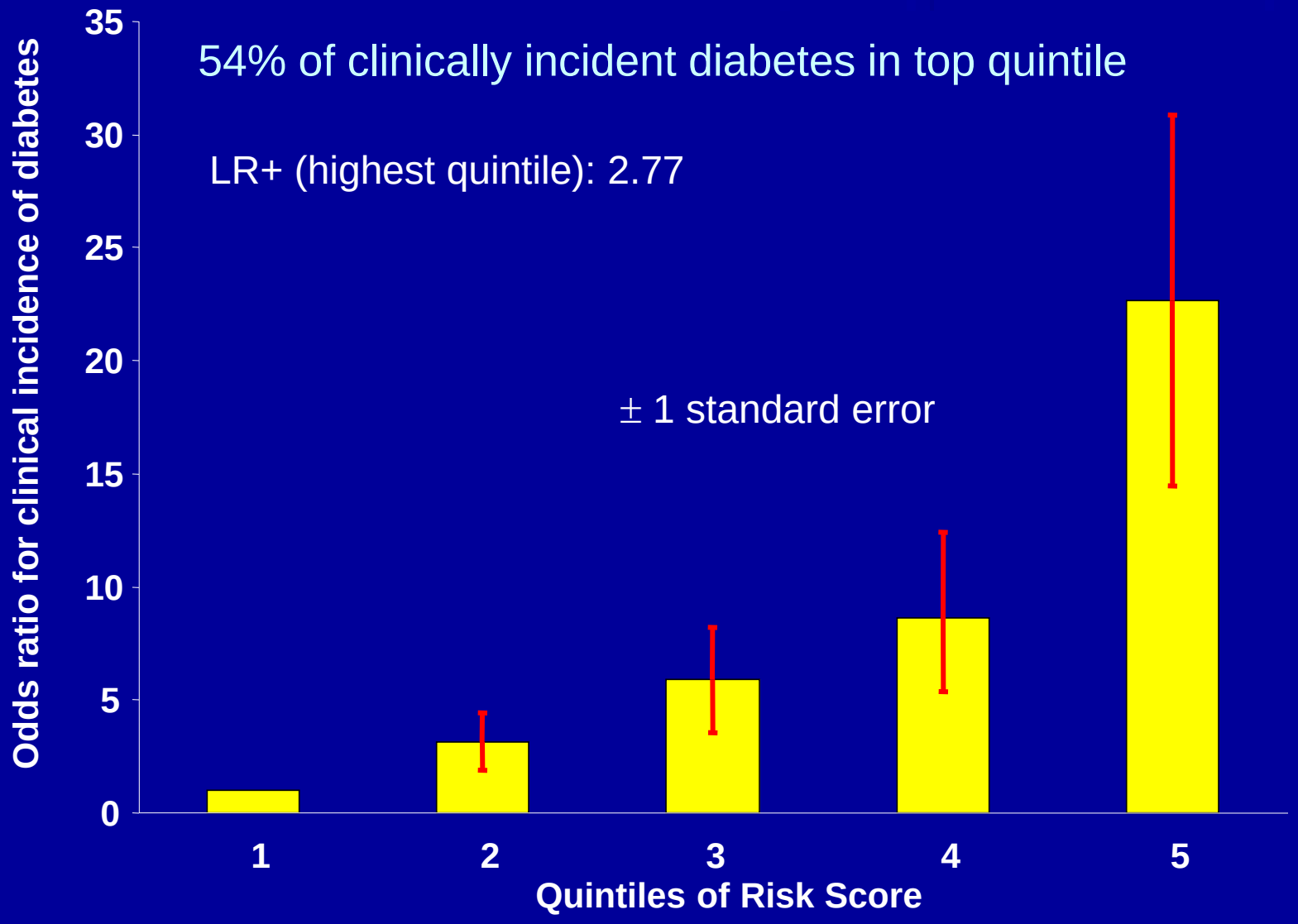
The Danish Risk Score

Variable	β-coeff	OR	95% CI	Risk score
Age (45 vs. 30-40)	0.6926	2.0	(1.0-4.1)	7
Age (50 vs. 30-40)	1.3111	3.7	(2.0-7.0)	13
Age (55-60 vs. 30-40)	1.8475	6.3	(3.5-11.5)	18
Gender (m vs. f)	0.3970	1.5	(1.0-2.2)	4
BMI 25-29 vs. < 25	0.7401	2.1	(1.3-3.5)	7
BMI ≥ 30 vs. <25	1.4672	4.4	(2.6-7.3)	15
Known hypertension (y vs. n)	0.9832	2.7	(1.8-4.0)	10
PAL (inactive vs. active)	0.6488	1.9	(1.0-3.5)	6
Parent diabetic: (y vs. n)	0.6835	2.0	(1.3-3.0)	7

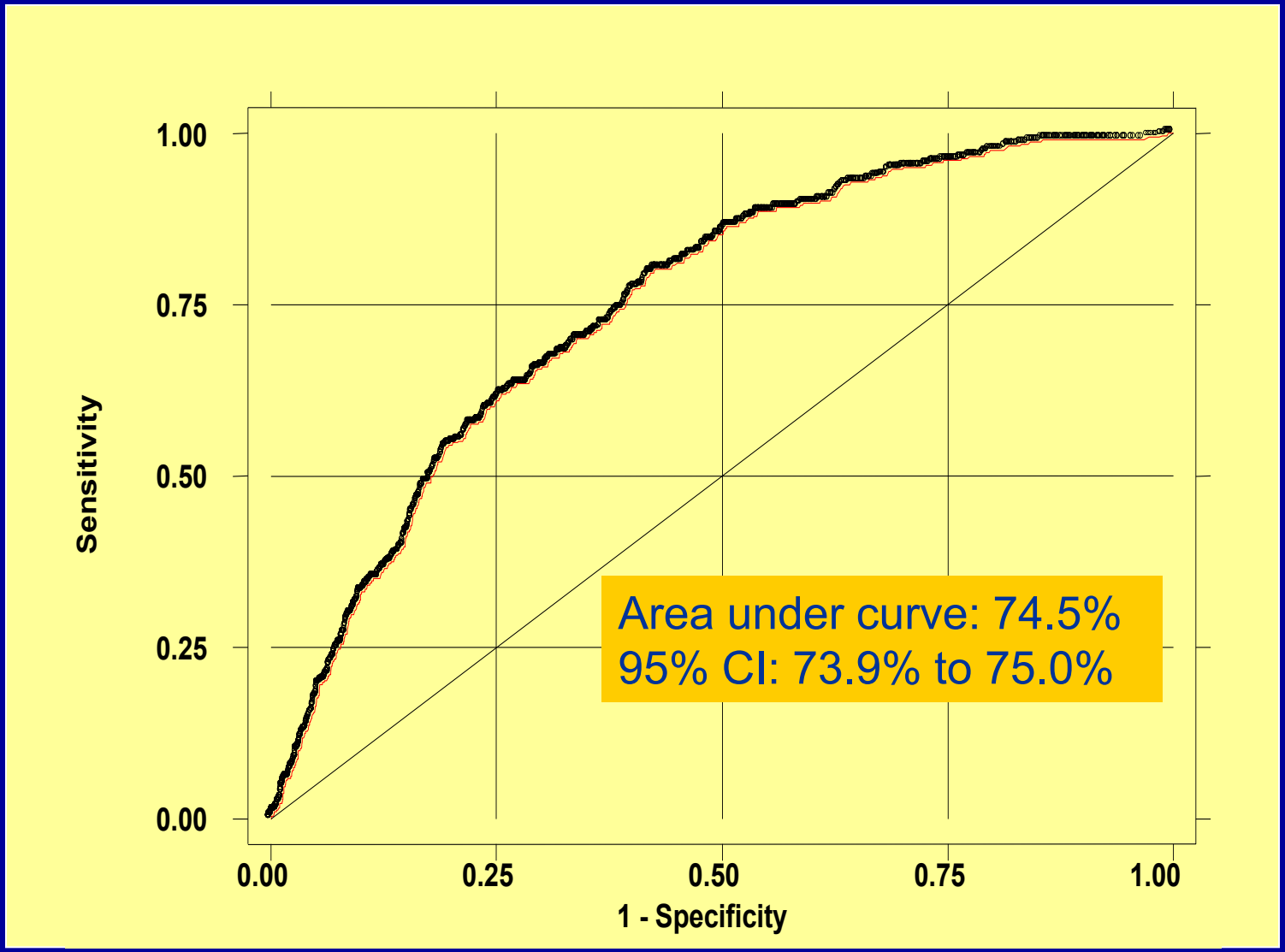
Performance of the Danish Risk Score



Association of quintiles of risk score with clinical incidence of diabetes – Epic-Norfolk study



ROC curve for the detection of clinically incident diabetes using the risk score – Epic-Norfolk study



- The Leicester South Asian Score can be used to identify those at high risk of IGR and T2DM in UK multi-ethnic populations
- The score is simple (7 questions), non invasive and inexpensive
- This score may be used to increase the uptake to screening

A risk score for predicting incident diabetes in a Thai population in a 10-year follow-up

Risk factor	Coefficient	Diabetes risk score
Age (years)		
30–39		0
40–44	−0.07	0
45–49	0.27	1
≥50	0.60	2
Sex		
Women		0
Men	0.44	2
BMI (kg/m^2)		
<23		0
≥23 but <27.5	0.69	3
≥27.5	1.28	5
Waist circumference (cm)		
<90 in men, <80 women		0
≥90 in men, ≥80 in women	0.56	2
Hypertension		
No		0
Yes	0.68	2
History of diabetes in parent or sibling		
No		0
Yes	1.09	4

-The ability to predict diabetes risk correctly (AUC_{roc}: 78%)

- Adding fasting glucose into the model did not improve the prediction

Indian Diabetes risk score

Variables	Risk score
Age (30 – 44) yrs	10
Age (45 – 59) yrs	18
Age (>59) yrs	19
Family history of diabetes	7
Body mass index ($\geq$ 25) kg/m^2	7
Waist (M = >85 ,W = >80 cm)	5
Sedentary physical activity	4
Maximum Score	42

A person with a score $\geq$ 21 has high probability of having Diabetes undetected

OMANI DIABETES RISK SCORE

1. Age

0 p. 20 - 39 years
7 p. 40 - 59 years
9 p. 60 + years

2. Body mass index

0 p. < 25 kg/m^2
2 p. 25 - 29 kg/m^2
3 p. ≥ 30 kg/m^2

3. Waist circumference

	MEN	WOMEN
0 p.	< 94 cm	< 80 cm
2 p.	94 + cm	80 + cm

4. Family history of diabetes

0 p. No
8 p. Yes

5. Current hypertension status

0 p. No
3 p. Yes

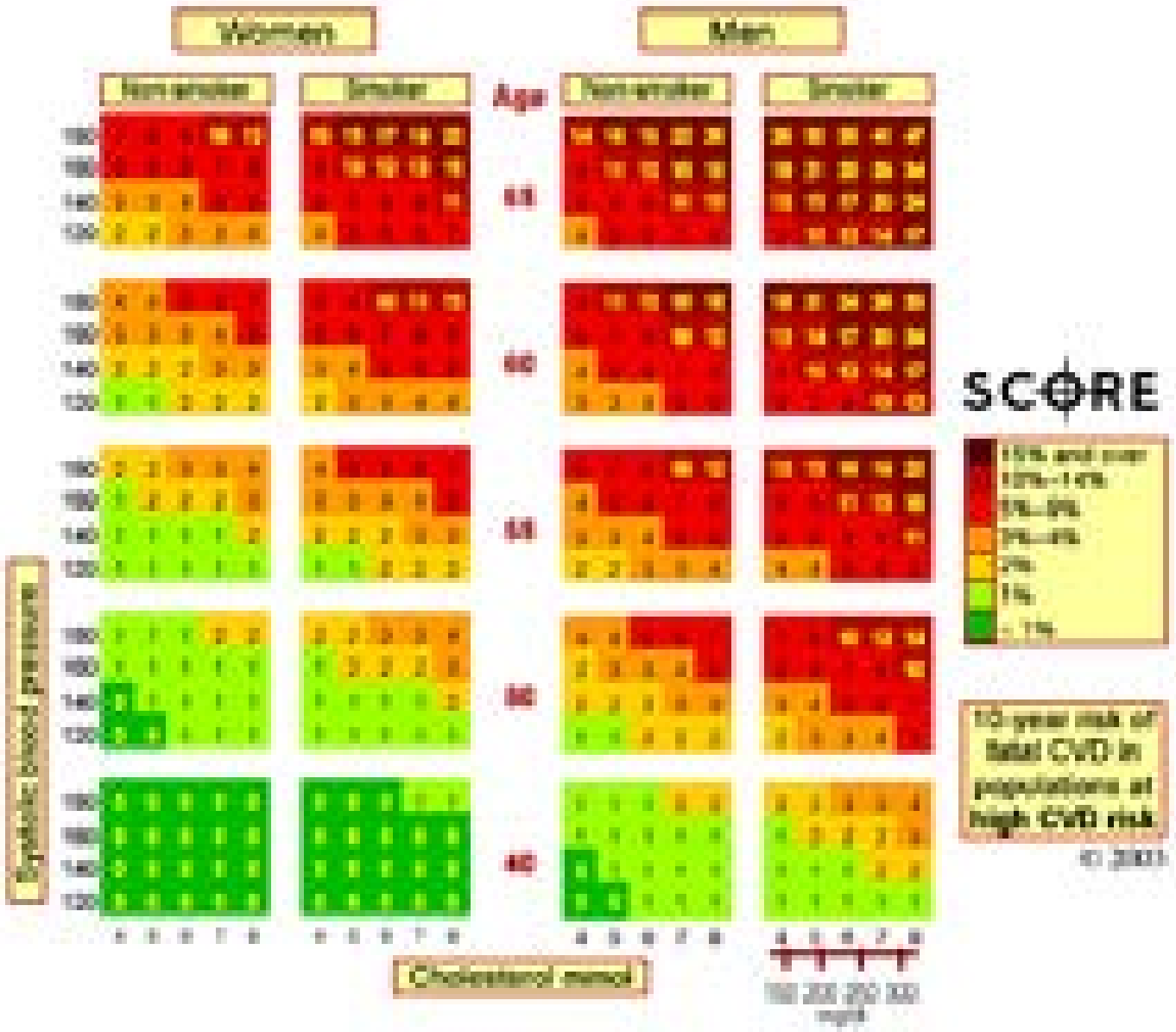
- The area under the curve:

1991: 0.83 (95%CI 0.82 to 0.84);
2001: 0.76 (95%CI 0.74 to 0.79).

- The cut-point of Diabetes Risk Score >10 in the 1991 cohort :
  - sensitivity 78.6% (74.6% - 82.1%)
  - specificity 73.4% (72.0% - 74.7%)

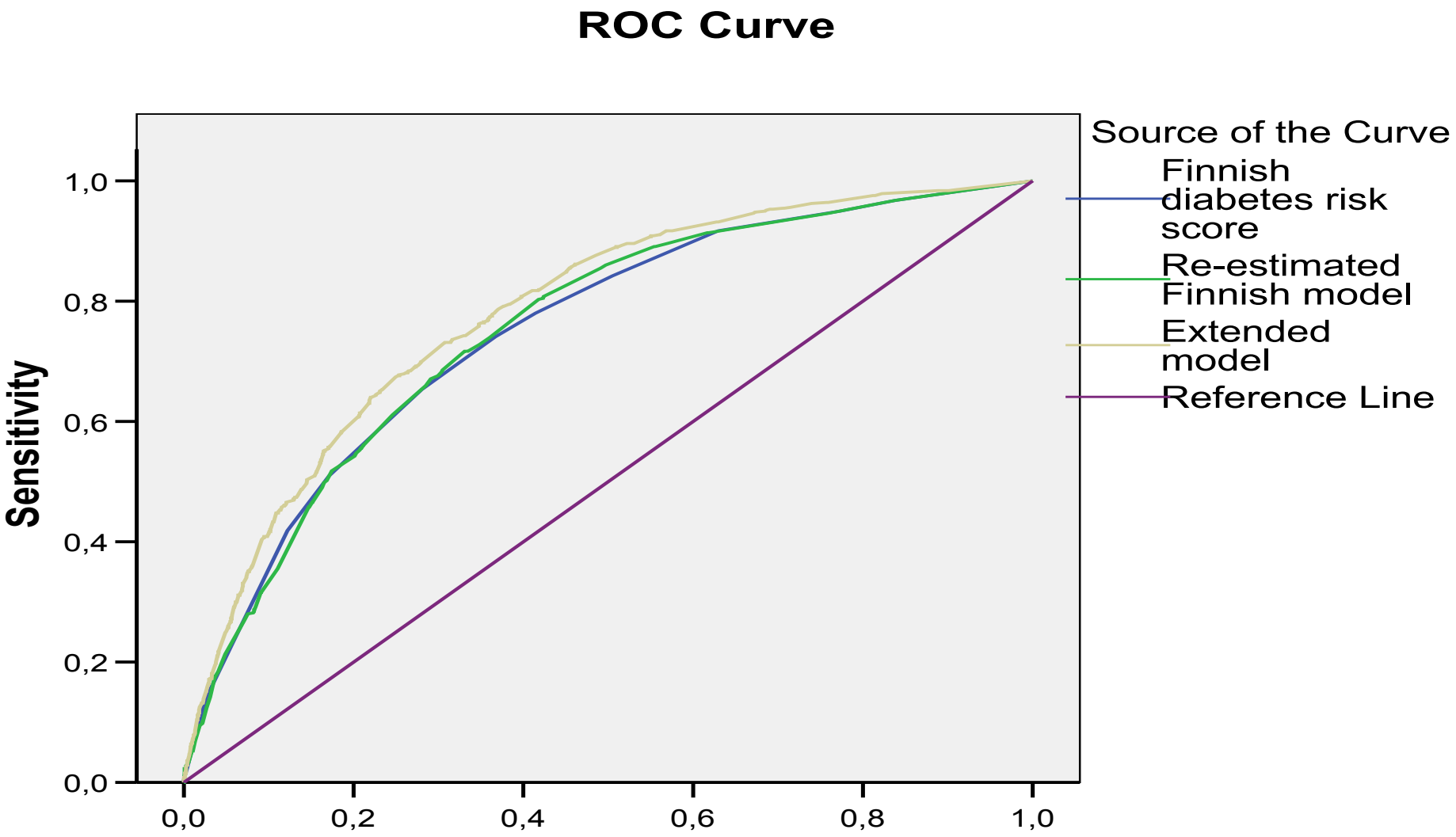
Total score points: 0-25

European risk score for prediction of CVD events: SCORE



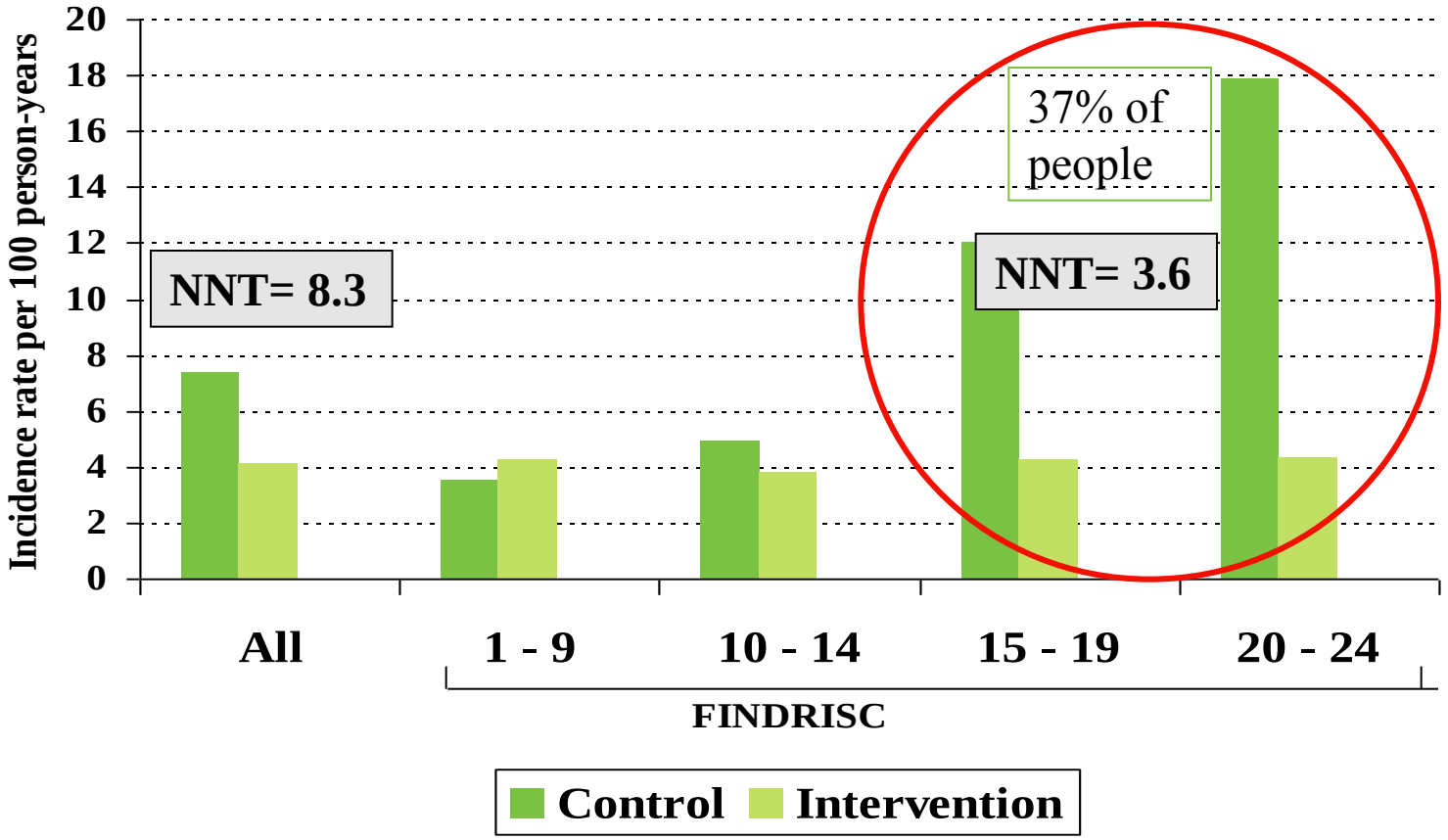
Diabetes Risk Scoring system in Indian men in Mauritius

DETECT -2: Prospective analysis to identify diabetes



Extended model (area under the ROC-curve 0.781 [95% CI 0.763-0.800])
Re-estimated Finnish model (area under the ROC-curve 0.753 [95% CI 0.734-0.773])
Finnish diabetes risk score (area under the ROC-curve 0.750 [95% CI 0.730-0.770])

DPS: Diabetes incidence in the intervention vs. control group by baseline FINDRISC



Lindström et al. Diabetes Care 2008

Will determining the diabetes risk help to prevent the development of T2D in high-risk individuals?

FINDRISC in the Finnish Diabetes Prevention Study (DPS) population

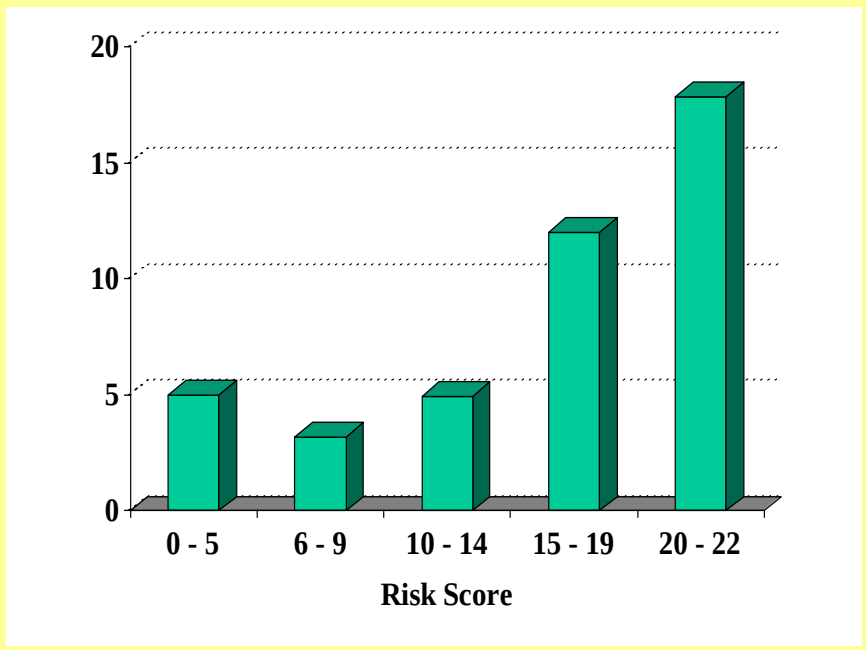
The DPS control group

- age 40-64, BMI >25, IGT
- annual laboratory visit
- n=236 with baseline FINDRISC
- median follow-up 3 years

The DPS intervention group

- age 40-64, BMI >25, IGT
- lifestyle intervention
- n=233 with baseline FINDRISC
- median follow-up 3 years

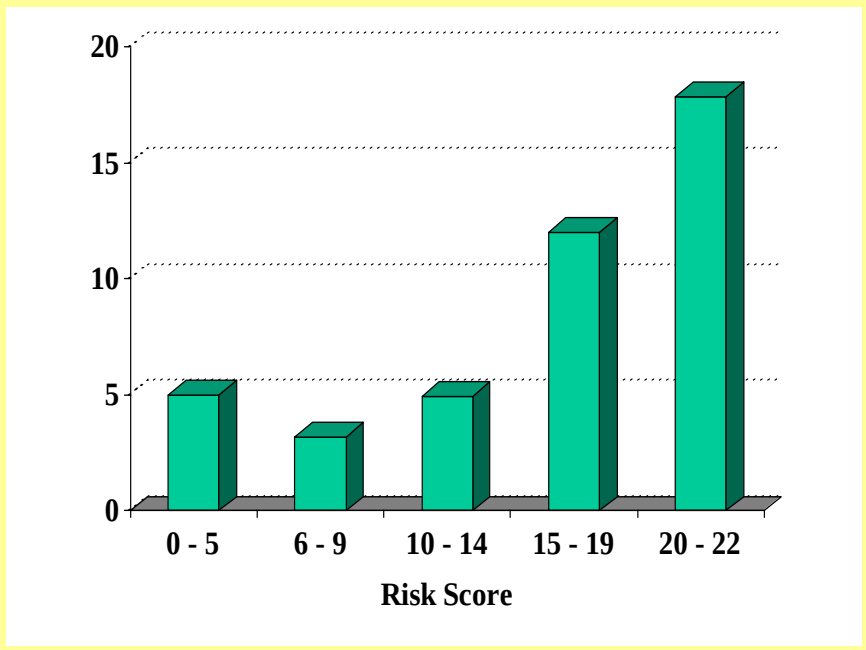
Diabetes rate (cases/100 person-years) by baseline FINDRISC value: the DPS control group



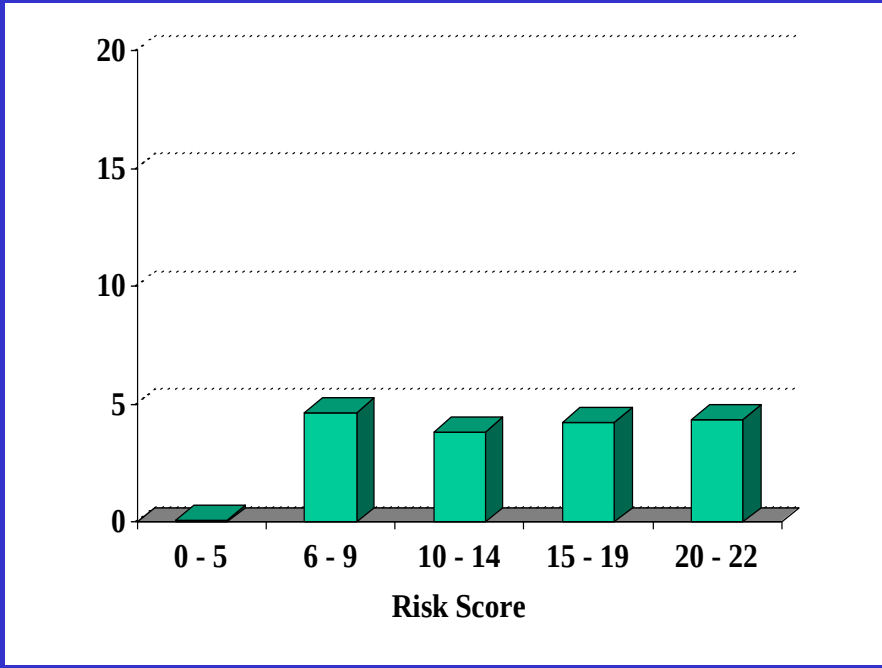
Control Group

Lindström et al. Diabetes Care 2007

Diabetes rate (cases/100 person-years) by baseline FINDRISC value: the DPS intervention group



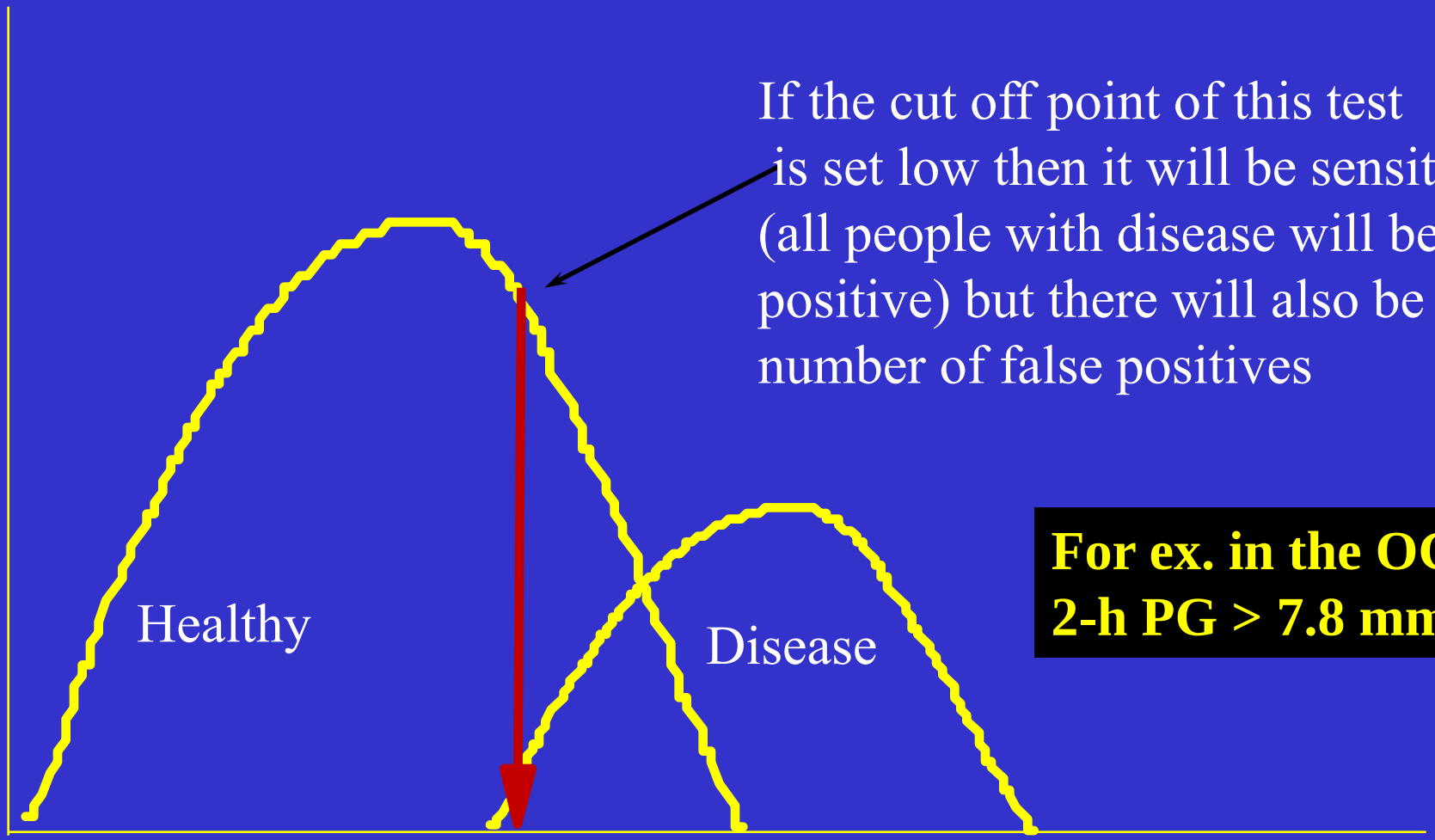
Control Group



Intervention Group

Lindström et al. Diabetes Care 2007

Sensitivity



If the cut off point of this test is set low then it will be sensitive (all people with disease will be test positive) but there will also be a number of false positives

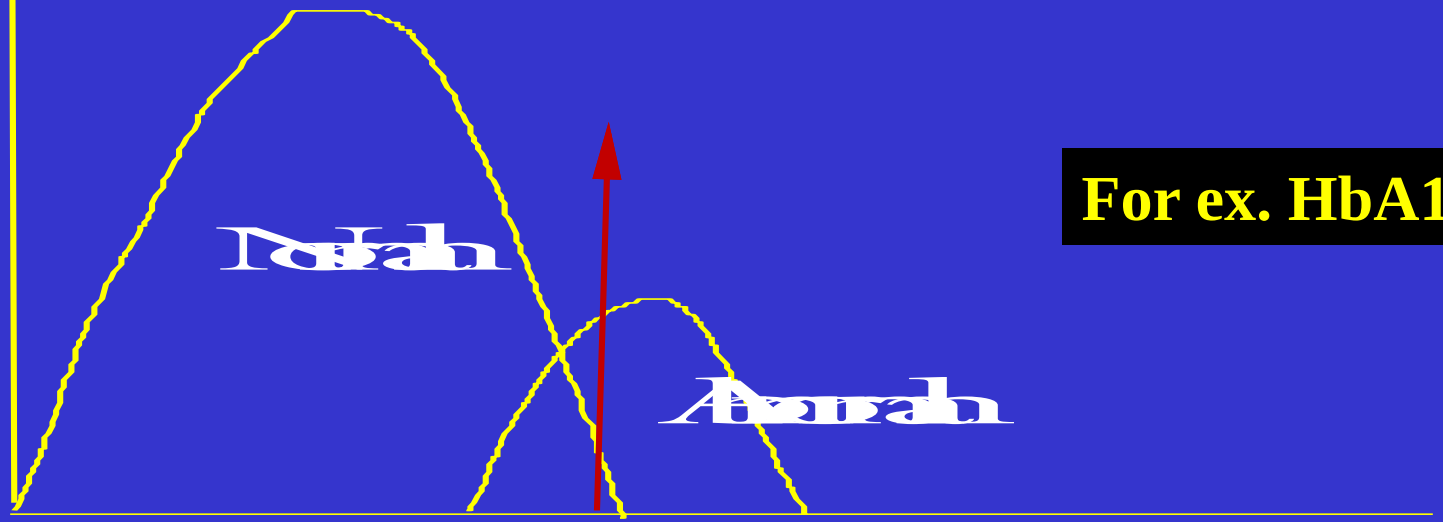
For ex. in the OGTT 2-h PG > 7.8 mmol/l

Specificity

- **The probability that the test will be negative if the disease is truly absent.**

- **A specific test has a high likelihood of <u>false negatives</u>.**

Glycosylation



For ex. HbA1c >6.5%

Performance of FINDRISC in identifying unrecognized T2DM

among 45-74-year old men and women (Finrisk-02, n=2966)

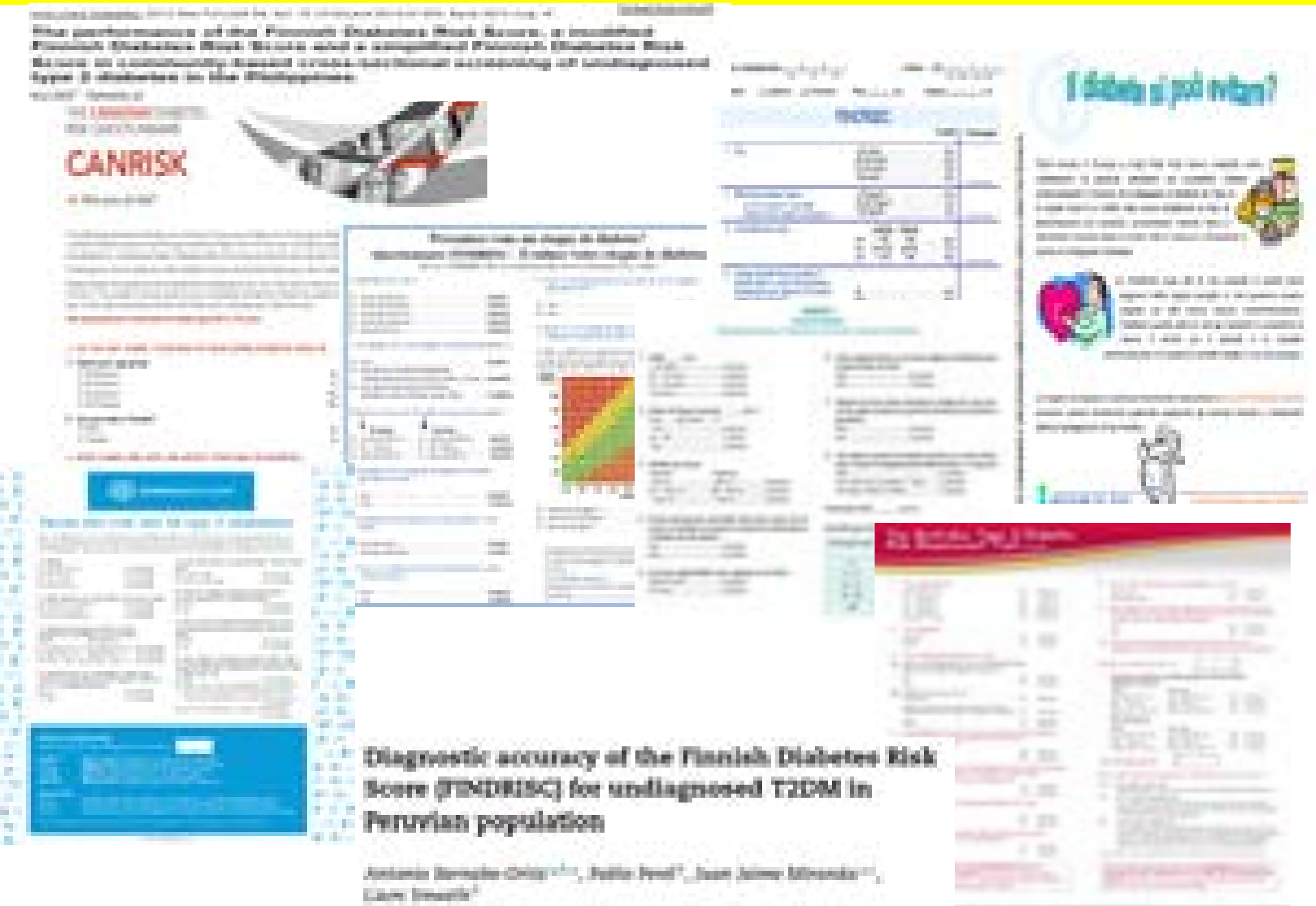
	Sensitivity	PPV	NPV	% of study sample
Cutoff value = 11				
Men	66%	22%	94%	**35%**
Women	70%	11%	96%	**41%**
Cutoff value = 13				
Men	45%	25%	92%	**21%**
Women	55%	14%	96%	**27%**
Cutoff value = 15				
Men	30%	30%	91%	**12%**
Women	38%	16%	95%	**16%**

Saaristo et al. Diabetes Vasc Dis Res 2005; 2:67-72

Performance of the FINDRISC to identify abnormal glucose tolerance in other populations

	Cutoff point	Sensitivity	Specificity	Ref.
The IGLOO Study, Italy				
n=1377, age 55-77	≥9	77% (DM/IGT)	45%	Franciosi et al. Diabetes Care 2005; 28:1187-1194
Krakow, Poland (DM/IGT)				
n=12496	≥9	82% (DM/IGT)	70%	Szurkowska et al. Przeglad Lekarski 2006; 63 (Suppl. 4):P42
The KORA Survey 2000, Germany (DM)				
n=1353, age 55-74	≥9	82% (DM)	43%	Rathmann et al. 2005; 165:436-441

FINDRISC has been translated to >30 languages and is used around the world either in original or adapted version, and used in over 200 publications



The most common diabetes risk score components

- **Age**
- **BMI**
- **Waist circumference**
- **Family history of diabetes**
- **Sex**
- **Fasting glucose**
- **Ethnicity**
- **Blood pressure**
- **Triglycerides**
- **Physical activity**
- + HDL-cholesterol, diet (fruit and vegetables, red meat, whole grain bread, coffee, alcohol), 2h glucose, smoking, height, social deprivation, LDL-cholesterol, steroid medication, delivery of macrosomic infant, education

CONCLUSION - ADVICE

Never use blood glucose testing in a non-diabetic person without determining her/his diabetes risk

CONCLUSIONS

- Diabetes risk scores have been developed/validated in several populations
- They work well in predicting future development of T2D, but may be to some extent population-specific
- Risk scores can also be used as primary screening tool to detect undignosed T2D
- The parameters included in various models and scores are more or less the same, but the cut-points and score weights (beta-coefficients) are different
- A universal diabetes risk score may not possible, but it is possible to implement diabetes risk scores in all populations
- There is good evidence that people at high risk identified by risk score benefit from healthy lifestyle advice