Biostatistics

Applications in Medicine

Nuno Sepúlveda, 21.10.2024

Syllabus

1. General review
    a. What is Biostatistics?
    b. Population/Sample/Sample size
    c. Type of Data – quantitative and qualitative variables
    d. Common probability distributions
    e. Work example – Malaria in Tanzania
2. Applications in Medicine
    a. Construction and analysis of diagnostic tools – Binomial distribution, sensitivity, specificity, ROC curve,Rogal-Gladen estimator
    b. Estimation of treatment effects - generalized linear models
    c. Survival analysis - Kaplan-Meier curve, log-rank test, Cox's proportional hazards model
3. Applications in Genetics, Genomics, and other 'omics data
    a. Genetic association studies – Hardy-Weinberg test, homozygosity, minor allele frequencies, additive model, multiple testing correction
    b. Methylation association studies – M versus beta values, estimation of biological age
    c. Gene expression studies based on RNA-seq experiments – Tests based on Poisson and Negative-Binomial
4. Other Topics
    a. Estimation of Species diversity – Diversity indexes, Poisson mixture models
    b. Serological analysis – Gaussian (skew-normal) mixture models
    c. Advanced sample size and power calculations

What is a test perfect for diagnosis?

What is a test perfect for diagnosis?

Positive for all true cases

P(+|true case)=1

Sensitivity=1

Negative for all true non-cases

P(-|true non case)=1

Specificity=1

(Almost) Perfect diagnostic = Gold Standard

Hypothetical data

Disease status	Positive	Negative
True Cases	n_1	0
True Non-cases	0	n_2

Some mathematics (sampling model)

True cases

$$x_1 \mid n_1, \pi_{Se} \rightsquigarrow \text{Binomial}\left(n_1; \pi_{Se}\right)$$

What are the assumptions?

True non-cases

$$x_2 \mid n_2, \pi_{Sp} \rightsquigarrow \text{Binomial}\left(n_2; \pi_{Sp}\right)$$

Random versus deterministic

$$H_0 : \text{Sensitivity} < 1 \text{ vs } H_1 : \text{Sensitivity} = 1$$

$$H_0 : \pi_{Se} < 1 \text{ vs } H_1 : \pi_{Se} = 1$$

$$H_0 : \text{Specificity} < 1 \text{ vs } H_1 : \text{Specificity} = 1$$

$$H_0 : \pi_{Sp} < 1 \text{ vs } H_1 : \pi_{Sp} = 1$$

Technical problem I

$\pi_{Se} \in (0,1)$ and $\pi_{Sp} \in (0,1)$

Practical Solution

$$H_0 : \pi_{Se} = 0.999 \text{ vs } H_1 : \pi_{Se} > 0.999$$

$$H_0 : \pi_{Sp} = 0.999 \text{ vs } H_1 : \pi_{Sp} > 0.999$$

Simple test: calculate confidence intervals and check whether 0.999 is within it.

Should the confidence intervals be one-tail or two-tail?

Random versus deterministic

$$IC_{95\%}\left(\pi_{Se}\right) = \hat{\pi}_{Se} \pm 1.96 \times se\left(\hat{\pi}_{Se}\right)$$

$$IC_{95\%}\left(\pi_{Se}\right) = \left(\hat{\pi}_{Se} - 1.65 \times se\left(\hat{\pi}_{Se}\right); 1\right)$$

Technical problem II (standard error undefined)

$$se\left(\hat{\pi}_{Se}\right) = \sqrt{\frac{\pi_{Se}\left(1-\pi_{Se}\right)}{n}}$$

$$se\left(\hat{\pi}_{Se}\right) = \sqrt{\frac{\hat{\pi}_{Se}\left(1-\hat{\pi}_{Se}\right)}{n}}$$

Solution: use an alternative confidence interval

Clopper-Pearson exact interval — Exact version — Over-coverage

---

Wilson score interval

Agresti-Coull interval

Better approximations to the Normal distributions, especially when the true proportion is close to 0 or 1

Arcsin approximation interval — Under-coverage

time

Disease status	Positive	Negative
True Cases	30	0
True Non-cases	0	40

Calculate one-tailed 95% confidence intervals and test the null hypothesis

Use Wilson score test (prop.test)

Sample size determination

What is the sample size required for us to be confidence we are close to a deterministic situation?

time

1. Use prop.test function
2. Calculate (one tailed test) 95% confidence intervals for perfect sensitivity/specificity with different methods
3. Determine the minimum sample size required to obtain the one-tailed 95% confidence interval that does not contain 99.9%

Rogan-Gladen Estimator

You observe x positive tests in a sample of n individuals

The diagnostic test is imperfect

What is the estimate for the proportion of true positive individuals?

Rogan-Gladen Estimator

$x \mid n, \pi \rightsquigarrow \text{Binomial}\,(n; \pi)$

$\pi = \pi_{Se}\pi_{+} + (1 - \pi_{Sp})\pi_{-}$

$= \pi_{Se}\pi_{+} + (1 - \pi_{Sp})(1 - \pi_{+})$

π = probability of a positive test

π_{Se} = Sensitivity

π_{Sp} = Specificity

π_{+} = proportion of sampling a true positive

Can you estimate $\pi_{Se}, \pi_{Sp},$ and π_{+} from the data?

Rogan-Gladen Estimator

$x \mid n, \pi \rightsquigarrow \text{Binomial}\,(n; \pi)$

$\pi = \pi_{Se}\pi_{+} + (1 - \pi_{Sp})\pi_{-}$

$= \pi_{Se}\pi_{+} + (1 - \pi_{Sp})(1 - \pi_{+})$

π = probability of a positive test

π_{Se} = Sensitivity

π_{Sp} = Specificity

π_{+} = proportion of sampling a true positive

Can you estimate $\pi_{Se}, \pi_{Sp},$ and π_{+} from the data?

No, overparametrization and lack of identifiability!

Rogan-Gladen Estimator

$x \mid n, \pi \rightsquigarrow \text{Binomial}\,(n; \pi)$

$\pi = \pi_{Se}\pi_{+} + (1 - \pi_{Sp})\pi_{-}$

$= \pi_{Se}\pi_{+} + (1 - \pi_{Sp})(1 - \pi_{+})$

π = probability of a positive test

π_{Se} = Sensitivity

π_{Sp} = Specificity

π_{+} = proportion of sampling a true positive

Can you estimate $\pi_{Se}, \pi_{Sp},$ and π_{+} from the data?

No, overparametrization and lack of identifiability!

Do you have any solution?

Rogan-Gladen Estimator

$x \mid n, \pi \rightsquigarrow \text{Binomial}(n; \pi)$

$\pi = \pi_{Se}\pi_{+} + (1 - \pi_{Sp})\pi_{-}$

$= \pi_{Se}\pi_{+} + (1 - \pi_{Sp})(1 - \pi_{+})$

π = probability of a positive test

π_{Se} = Sensitivity

π_{Sp} = Specificity

π_{+} = proportion of sampling a true positive

Can you estimate $\pi_{Se}, \pi_{Sp},$ and π_{+} from the data?

No, overparametrization and lack of identifiability!

Do you have any solution? Use external estimates for π_{Se} and π_{Sp}

Rogan-Gladen Estimator

$$\hat{\pi}_+ = \frac{\hat{\pi} + \hat{\pi}_{Sp} - 1}{\hat{\pi}_{Se} + \hat{\pi}_{Sp} - 1}$$

$$\hat{\pi} = \frac{x}{n}$$

Method of moments

$$\hat{\pi} = \hat{\pi}_{Se}\pi_+ + (1 - \hat{\pi}_{Sp})(1 - \pi_+)$$

$$\hat{\pi} = (\hat{\pi}_{Se} + \hat{\pi}_{Sp} - 1)\hat{\pi}_+ + (1 - \hat{\pi}_{Sp})$$

$$\hat{\pi} + \pi_{Sp} - 1 = (\pi_{Se} + \pi_{Sp} - 1)\pi_+$$

$$\pi_+ = \frac{\hat{\pi} + \hat{\pi}_{Sp} - 1}{\hat{\pi}_{Se} + \hat{\pi}_{Sp} - 1}$$

Invariant property of MLE

Rogan and Gladen (1978). Estimating prevalence from the results of a screening test. Am J Epidemiol 107, 71-76.

Rogan-Gladen Estimator

Wald's Confidence interval

$$IC_{95\%}\left(\pi_{+}\right) = \hat{\pi}_{+} \pm 1.96 \times se\left(\hat{\pi}_{+}\right)$$

$$\hat{\pi} = \frac{x}{n}$$

$$\text{Var}\left[\hat{\pi}_{+} \mid \hat{\pi}_{Se}, \hat{\pi}_{Sp}\right] = \text{Var}\left[\frac{\hat{\pi} + \hat{\pi}_{Sp} - 1}{\hat{\pi}_{Se} + \hat{\pi}_{Sp} - 1} \mid \hat{\pi}_{Se}, \hat{\pi}_{Sp}\right] = \frac{\text{Var}\left[\hat{\pi}\right]}{\left(\hat{\pi}_{Se} + \hat{\pi}_{Sp} - 1\right)^{2}} = \frac{\pi\left(1 - \pi\right)}{n\left(\hat{\pi}_{Se} + \hat{\pi}_{Sp} - 1\right)^{2}}$$

$$\hat{\text{Var}}\left[\hat{\pi}_{+} \mid \hat{\pi}_{Se}, \hat{\pi}_{Sp}\right] = \frac{\hat{\pi}\left(1 - \hat{\pi}\right)}{n\left(\hat{\pi}_{Se} + \hat{\pi}_{Sp} - 1\right)^{2}}$$

$$se(\hat{\pi}_{+}) = \sqrt{\hat{\text{Var}}\left[\hat{\pi}_{+} \mid \hat{\pi}_{Se}, \hat{\pi}_{Sp}\right]}$$

Rogan and Gladen (1978). Estimating prevalence from the results of a screening test. Am J Epidemiol 107, 71-76.

Limitation and extension

$\hat{\pi}_{Se}$

are assumed to be known constant

$\hat{\pi}_{Sp}$

$\hat{\pi}_{Se}$

are assumed to be estimators

$\hat{\pi}_{Sp}$

$$\hat{\text{Var}}\left[\hat{\pi}_{+}\right] = E\left[Var\left[\hat{\pi}_{+} \mid \hat{\pi}_{Se}, \hat{\pi}_{Sp}\right]\right] + \text{Var}\left[E\left[\hat{\pi}_{+} \mid \hat{\pi}_{Se}, \hat{\pi}_{Sp}\right]\right]$$

Rogan and Gladen (1978). Estimating prevalence from the results of a screening test. Am J Epidemiol 107, 71-76.

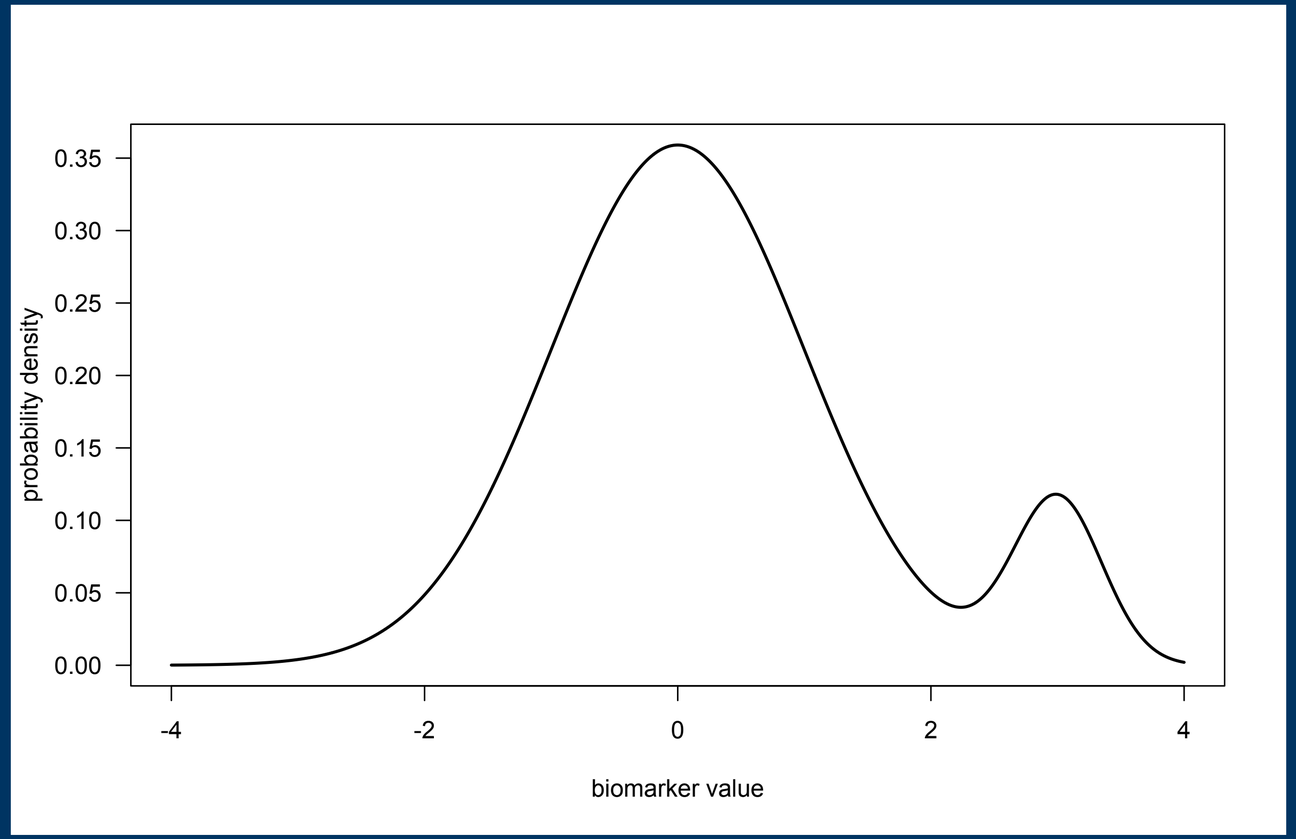
Exercise

	Positive	n	Sensitivity	Specificity
Anti-S antibodies	1271	2474	0.665 (India)	0.99 (India)
Anti-N antibodies	848	2474	0.995 (manufacturer)	0.997 (manufacturer)

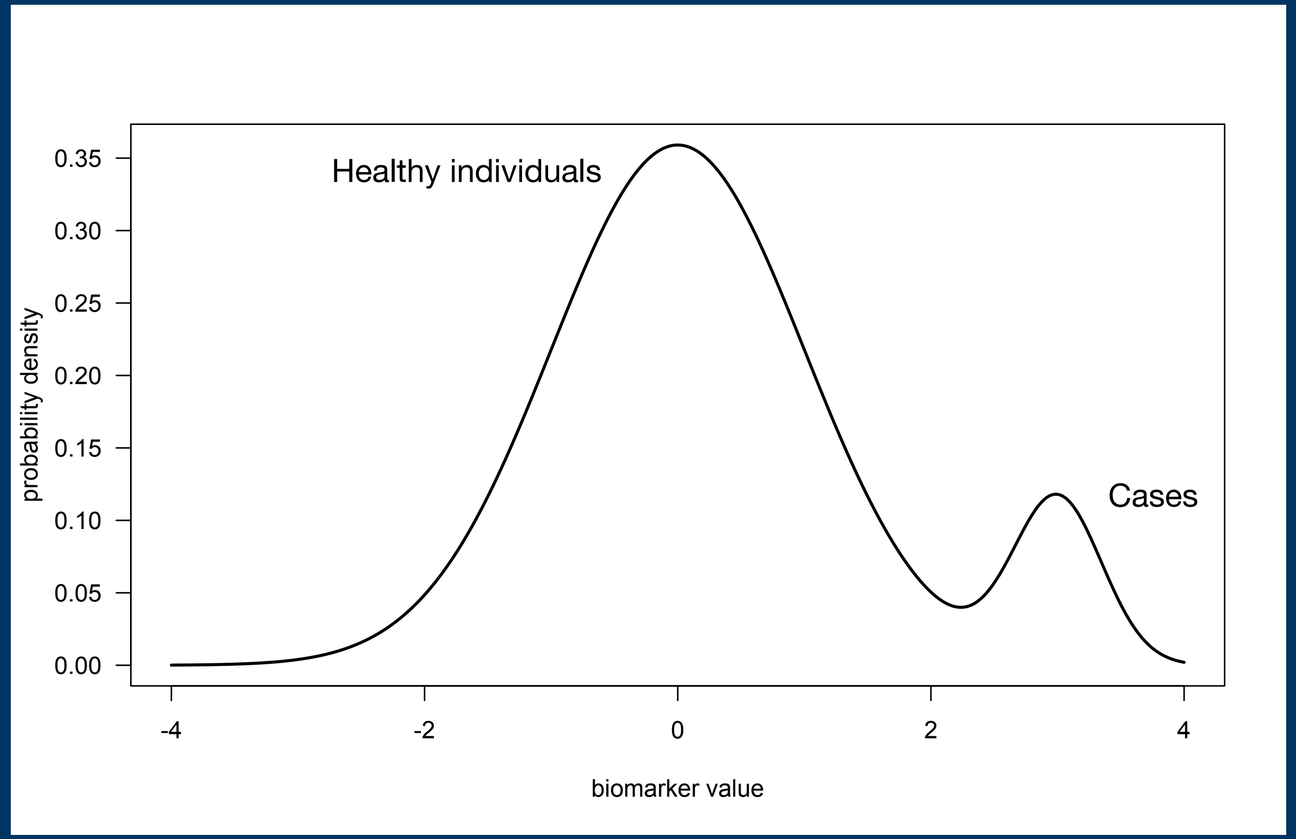
Estimate the proportion of true positive using the Rogan-Gladen estimator and its 95% confidence interval

Willeit et al (2022). Seroprevalence of SARS-CoV-2 infection in the Tyrolean district of Schwaz at the time of the rapid mass vaccination in March 2021 following B.1.351-variant outbreak. Front Public Health 10, 989337

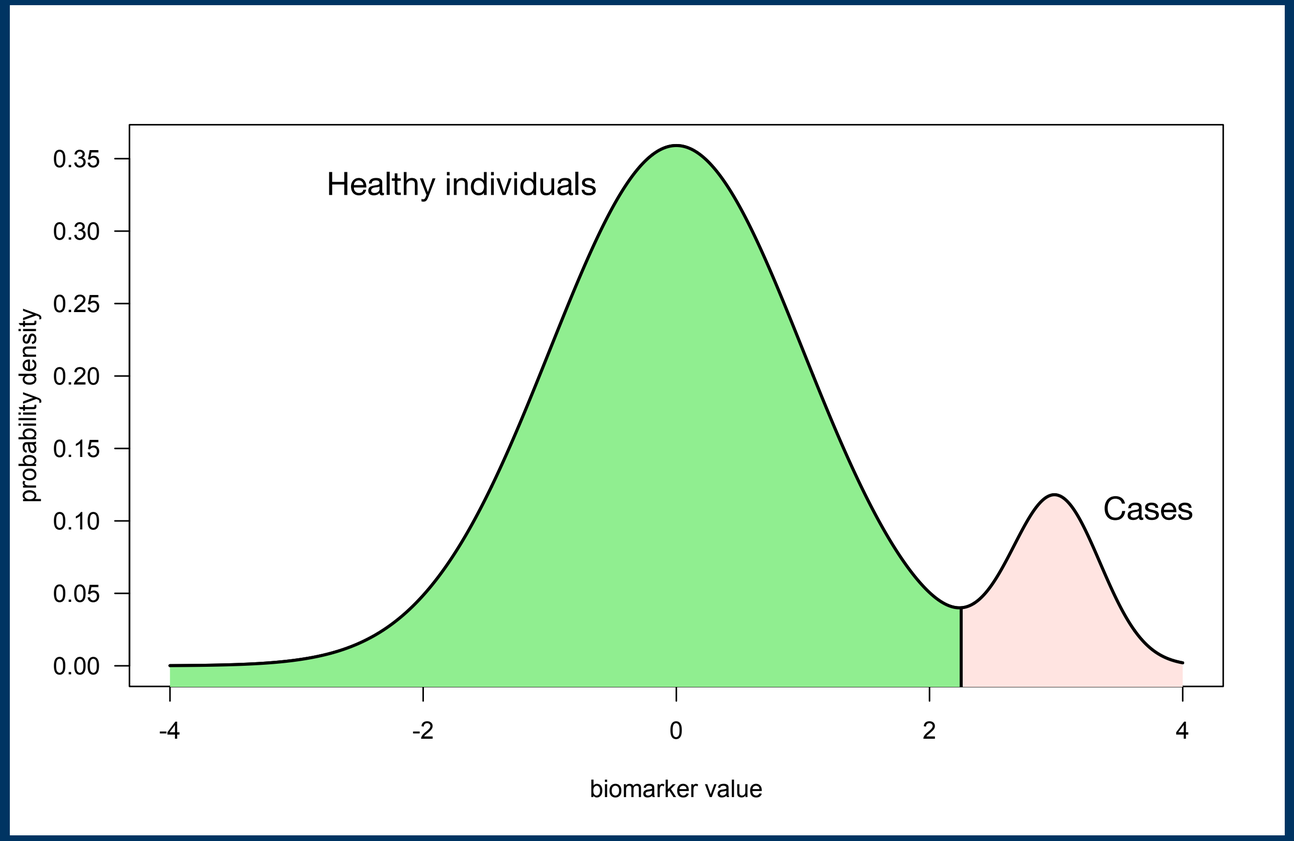
Constructing a diagnostic tool based on quantitative biomarker



Constructing a diagnostic tool based on quantitative biomarker



Constructing a diagnostic tool based on quantitative biomarker



How to estimate the optimal cutoff to distinguish cases from healthy controls

Evaluating the diagnostic potential of a biomarker

Two samples

True cases

True healthy

Estimate sensitivity and specificity using all possible cutoffs for the biomarker

Evaluating the diagnostic potential of a biomarker

Two samples

True cases

True healthy

Estimate sensitivity and specificity using all possible cutoffs for the biomarker

$$x_1 \,|\, n_1, \pi_{Se} \rightsquigarrow \text{Binomial}\left(n_1; \pi_{Se}\right)$$

$$\pi_{Se}(\kappa) = P\left[X \geq \kappa \,|\, n_1, \text{True case}\right]$$

$$\hat{\pi}_{Se}(\kappa) = \frac{\text{\# individuals with X} \geq \kappa}{n_1}$$

Evaluating the diagnostic potential of a biomarker

Two samples

True cases

True healthy

Estimate sensitivity and specificity using all possible cutoffs for the biomarker

$$x_1 \,|\, n_1, \pi_{Se} \rightsquigarrow \text{Binomial}\left(n_1; \pi_{Se}\right)$$

$$x_2 \,|\, n_2, \pi_{Sp} \rightsquigarrow \text{Binomial}\left(n_2; \pi_{Sp}\right)$$

$$\pi_{Se}(\kappa) = P\left[X \geq \kappa \,|\, n_1, \text{True case}\right]$$

$$\pi_{Sp}(\kappa) = P\left[X < \kappa \,|\, n_2, \text{True healthy}\right]$$

$$\hat{\pi}_{Se}(\kappa) = \frac{\text{\# individuals with X} \geq \kappa}{n_1}$$

$$\hat{\pi}_{Sp}(\kappa) = \frac{\text{\# individuals with X} < \kappa}{n_2}$$

Evaluating the diagnostic potential of a biomarker

Two samples

True cases

True healthy

Estimate sensitivity and specificity using all possible cutoffs for the biomarker

Construct a plot with the respective estimated sensitivities and 1-specificities

Receiver Operating Characteristic (ROC) curve

Real world data

ELSEVIER

Journal of Psychosomatic Research 69 (2010) 17–22

Journal of Psychosomatic Research

Measuring fatigue in clinical and community settings

Matteo Cella*, Trudie Chalder

Institute of Psychiatry, King's College London, London, United Kingdom

Received 11 June 2009; received in revised form 8 October 2009; accepted 13 October 2009

Abstract

Objective: The Chalder Fatigue Scale (CFQ) is a widely used instrument to assess fatigue in both clinical and nonclinical settings. Psychometric properties of the scale and discriminative abilities were examined. **Methods:** A total of 361 patients with CFS and 1615 individuals in the community were assessed with the CFQ. Principal component analysis (PCA) was used to explore the structure of the scale. Receiver-operating characteristic curve (ROC) was used to investigate the discriminative properties. **Results:** Two components, physical and mental fatigue, were identified in the CFS patient group and in the general population samples. Area under the curve for ROC was .91. The fatigue scale effectively discriminates, at high scores, between CFS patients and the general population. **Conclusion:** Physical and mental fatigue are clearly separable components of fatigue. The CFQ can discriminate reliably between clinical and nonclinical conditions.

Keywords: CFS; Chalder fatigue scale; Chronic fatigue syndrome; Fatigue; ROC

Table 3
Sensitivity and 1−specificity for the fatigue scales scores

CFQ Score	Sensitivity	1−Specificity
0	1.000	1.000
1	1.000	.999
2	.997	.998
3	.997	.997
4	.997	.993
5	.994	.985
6	.989	.979
7	.989	.964
8	.989	.940
9	.989	.913
10	.978	.866
11	.966	.637
12	.966	.565
13	.958	.484
14	.936	.408
15	.922	.330
16	.899	.260
17	.888	.207
18	.854	.167
19	.824	.136
20	.773	.107
21	.728	.084
22	.658	.062
23	.605	.043
24	.546	.029
25	.473	.018
26	.415	.014
27	.339	.010
28	.286	.007
29	.202	.004
30	.151	.000
31	.095	.000
32	.062	.000
33	.000	.000

ROC curve

ELSEVIER

Journal of Psychosomatic Research 69 (2010) 17–22

Journal of Psychosomatic Research

Measuring fatigue in clinical and community settings

Matteo Cella*, Trudie Chalder

Institute of Psychiatry, King's College London, London, United Kingdom

Received 11 June 2009; received in revised form 8 October 2009; accepted 13 October 2009

Abstract

Objective: The Chalder Fatigue Scale (CFQ) is a widely used instrument to assess fatigue in both clinical and nonclinical settings. Psychometric properties of the scale and discriminative abilities were examined. **Methods:** A total of 361 patients with CFS and 1615 individuals in the community were assessed with the CFQ. Principal component analysis (PCA) was used to explore the structure of the scale. Receiver-operating characteristic curve (ROC) was used to investigate the discriminative properties. **Results:** Two components, physical and mental fatigue, were identified in the CFS patient group and in the general population samples. Area under the curve for ROC was .91. The fatigue scale effectively discriminates, at high scores, between CFS patients and the general population. **Conclusion:** Physical and mental fatigue are clearly separable components of fatigue. The CFQ can discriminate reliably between clinical and nonclinical conditions.

Keywords: CFS; Chalder fatigue scale; Chronic fatigue syndrome; Fatigue; ROC

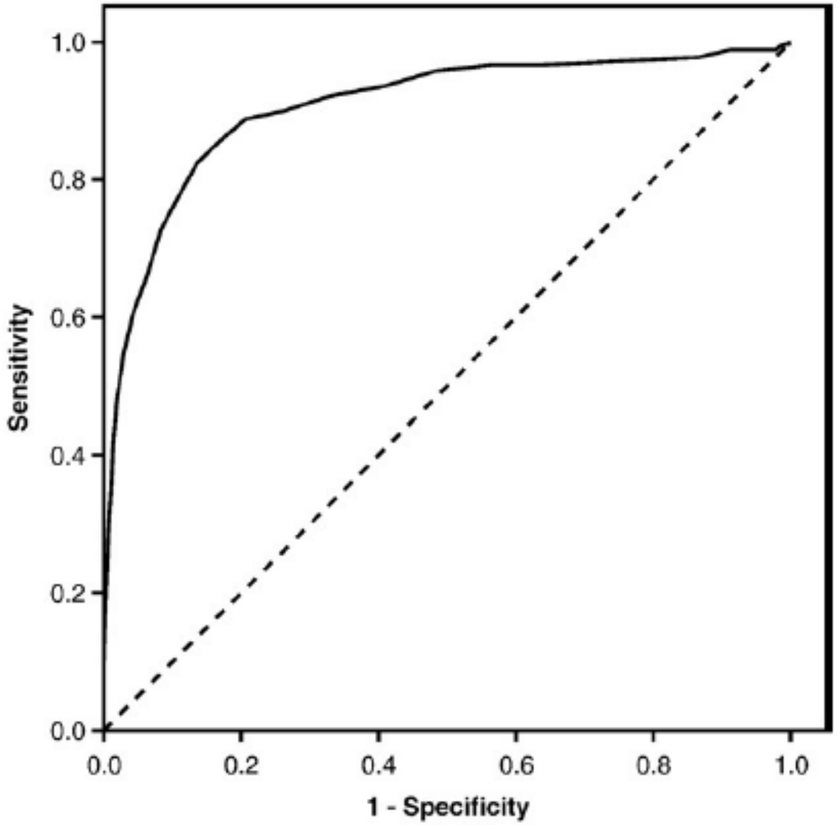


Fig. 1. Fatigue scale ROC plot.

Real practical questions

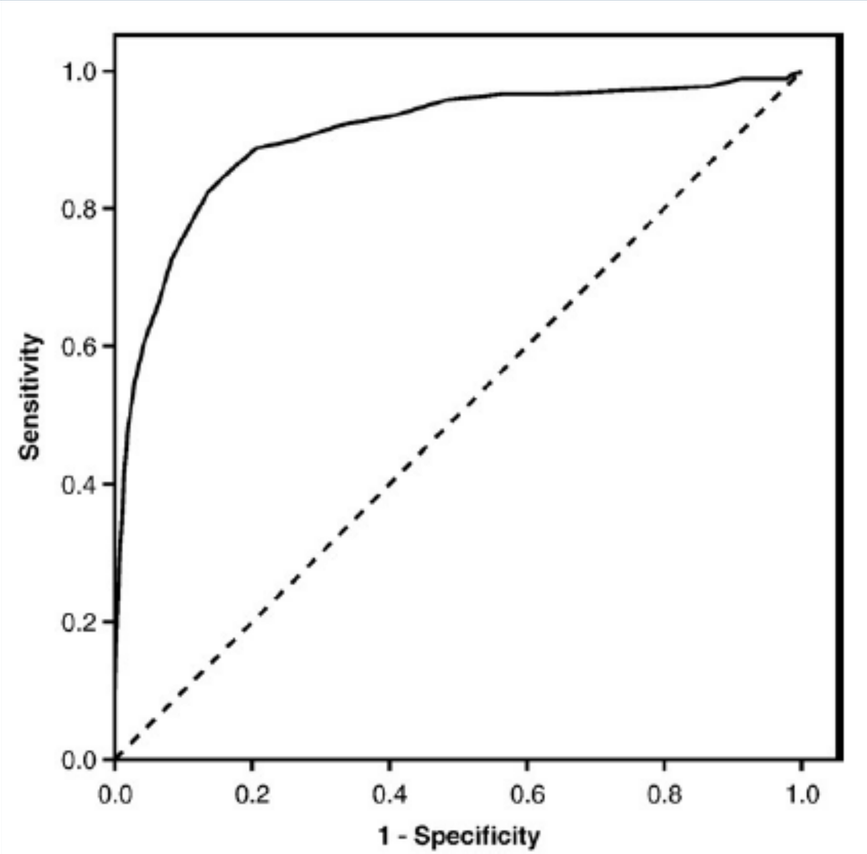


Fig. 1. Fatigue scale ROC plot.

How good is a diagnostic tool?

What is the best cutoff for discriminating cases from controls?

How good is a diagnostic tool?

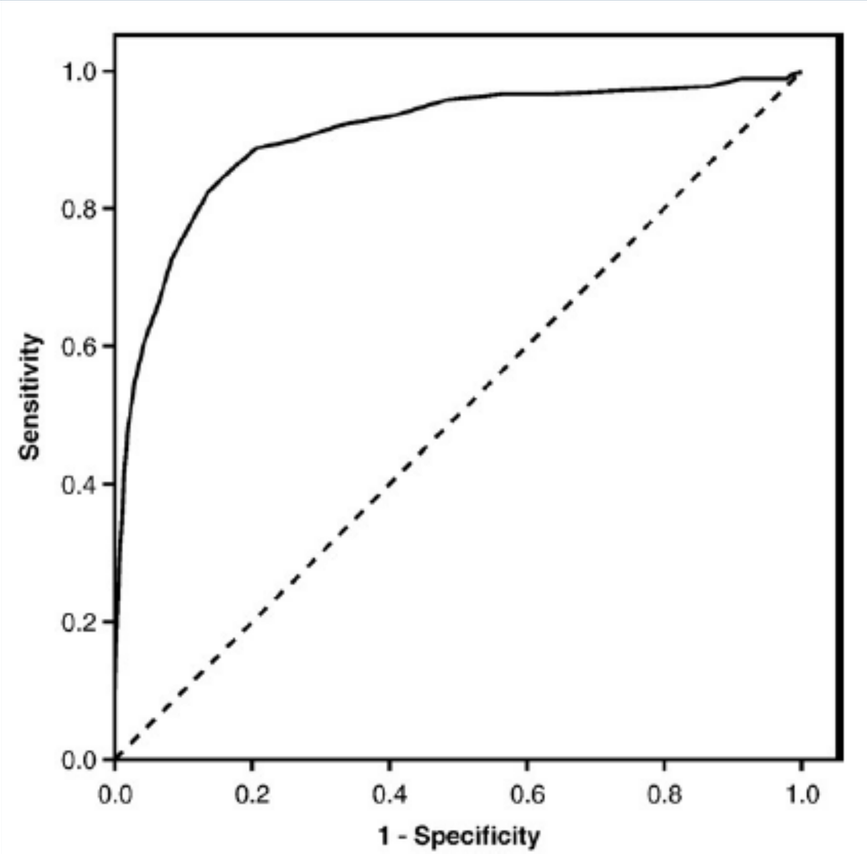


Fig. 1. Fatigue scale ROC plot.

Area under the curve (AUC)

$$\text{AUC} \in [0,1]$$

$$\text{AUC} = \pi_{Se} \times \pi_{Sp} \text{ for a binary predictor}$$

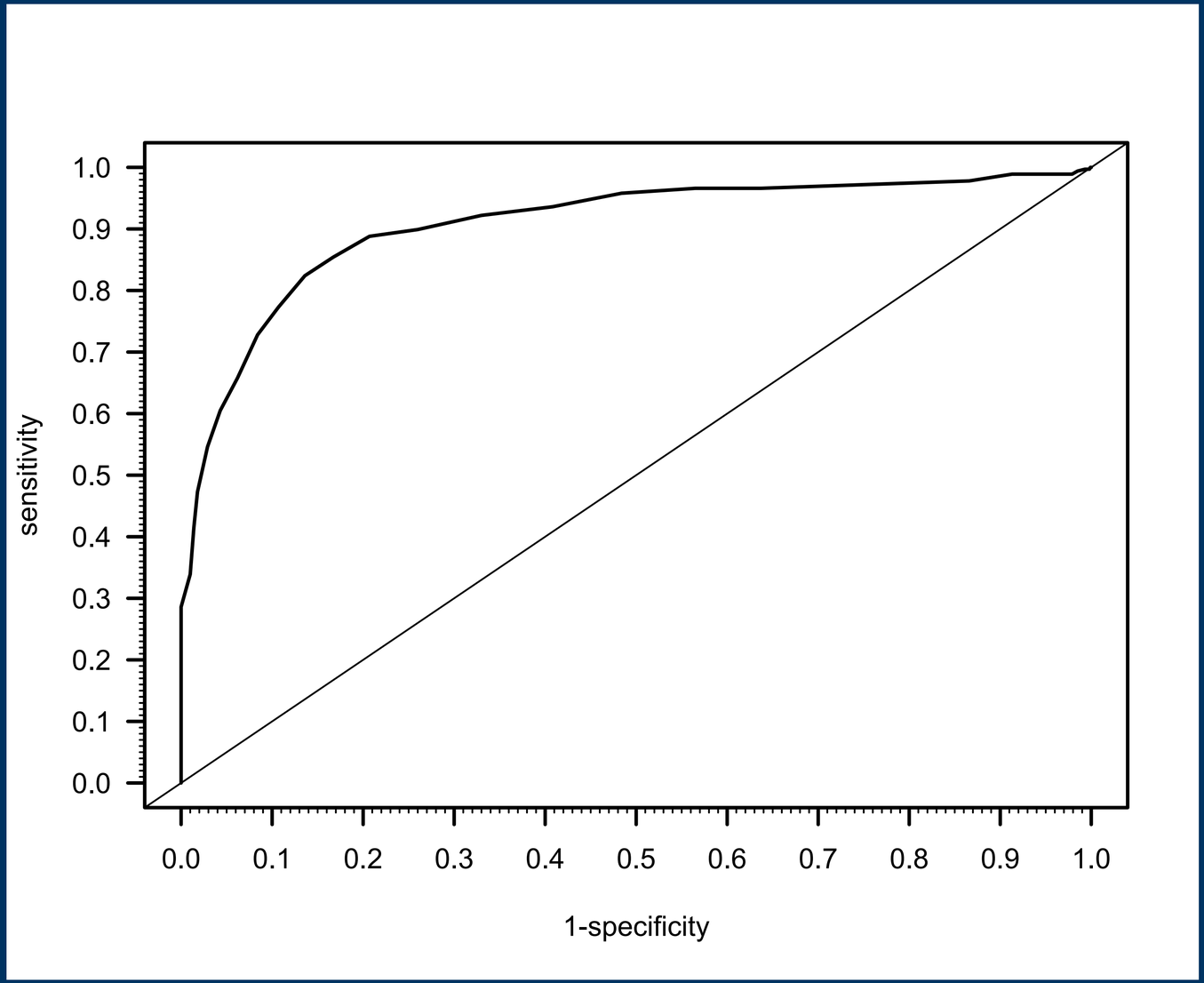
Warm-up Exercise

$\text{AUC} = 0$ $\text{AUC} = 0.5$ $\text{AUC} = 1$

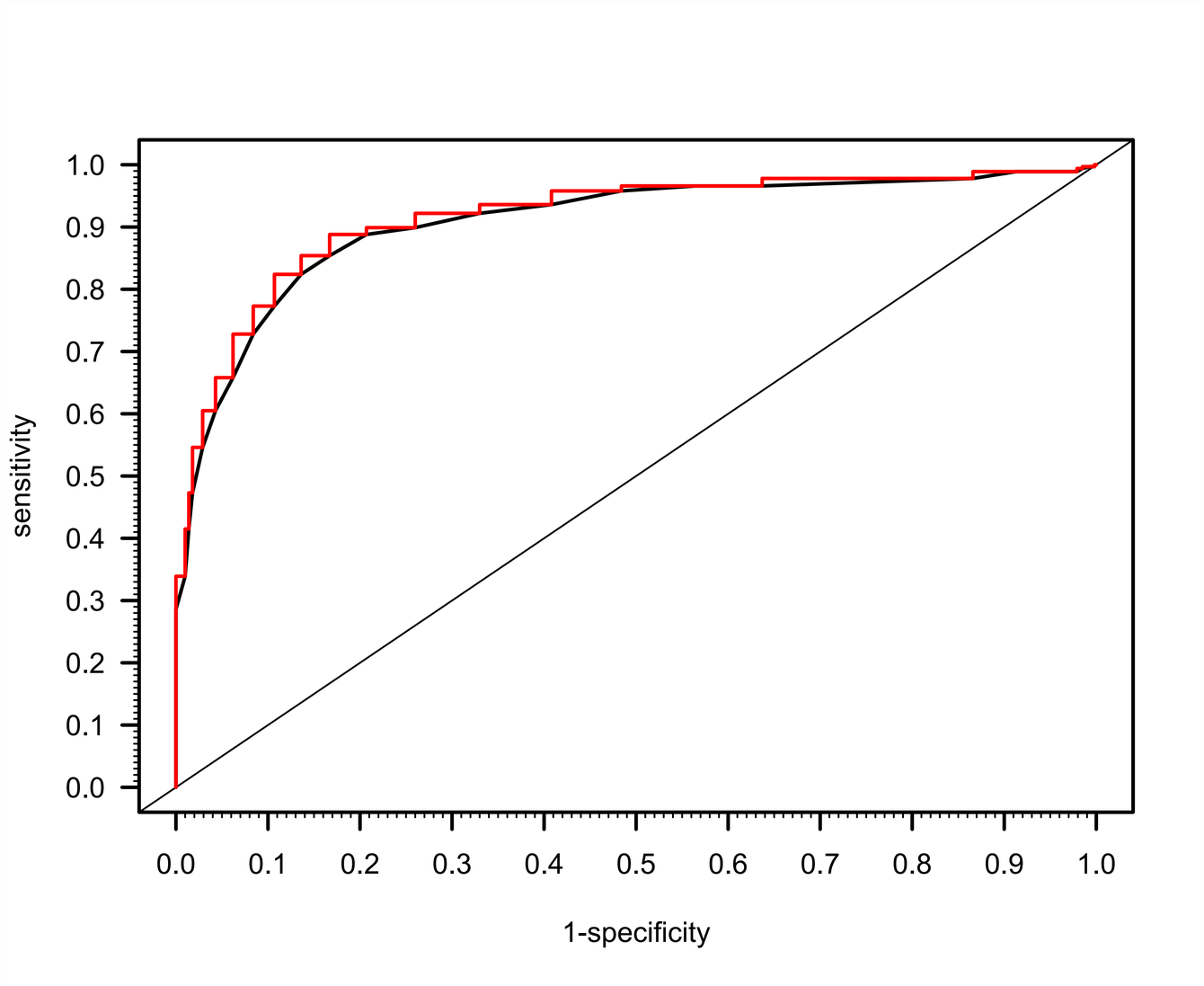
Draw the respective ROC curves

What is the respective interpretation?

A nuance conversation about ROC curve visualisation and AUC



A nuance conversation about ROC curve visualisation and AUC



How good is a diagnostic tool?

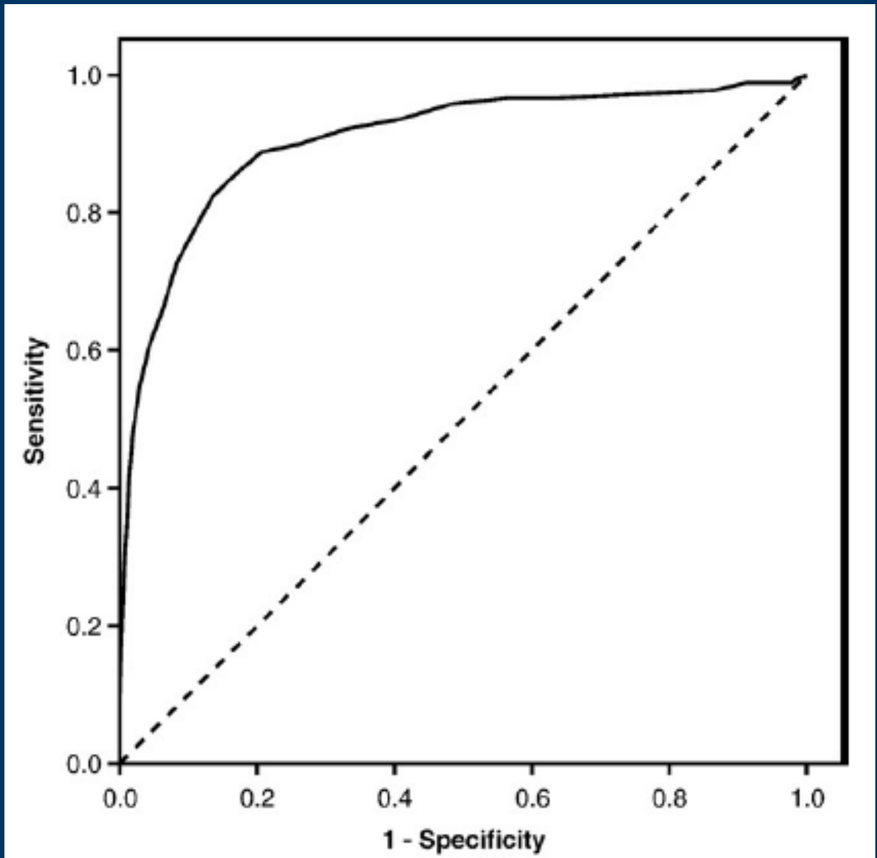


Fig. 1. Fatigue scale ROC plot.

In practice, what is the most important hypothesis to test against the data?

How good is a diagnostic tool?

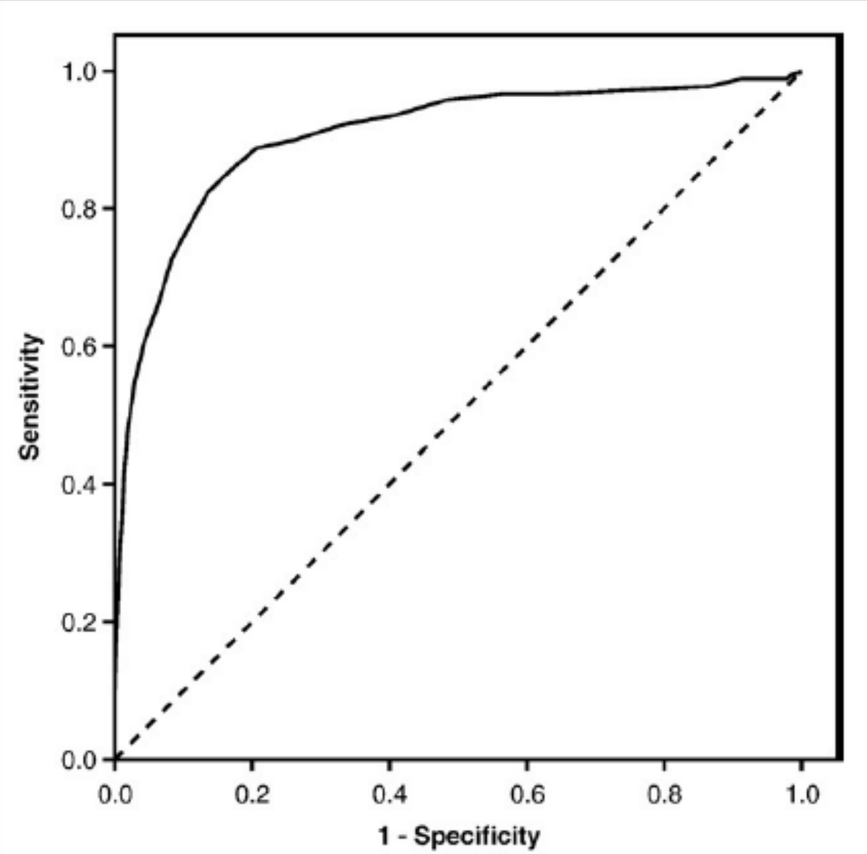


Fig. 1. Fatigue scale ROC plot.

In practice, what is the most important hypothesis to test against the data?

$$H_0 : \text{AUC} = 0.5 \text{ versus AUC} > 0.5$$

Wald's Confidence interval

$$IC_{95\%}\left(\text{AUC}\right) = \hat{\text{AUC}} \pm 1.96 \times se\left(\hat{\text{AUC}}\right)$$

Accept H_0 if $0.5 \in IC_{95\%}\left(\text{AUC}\right)$

Reject H_0, otherwise

How good is a diagnostic tool?

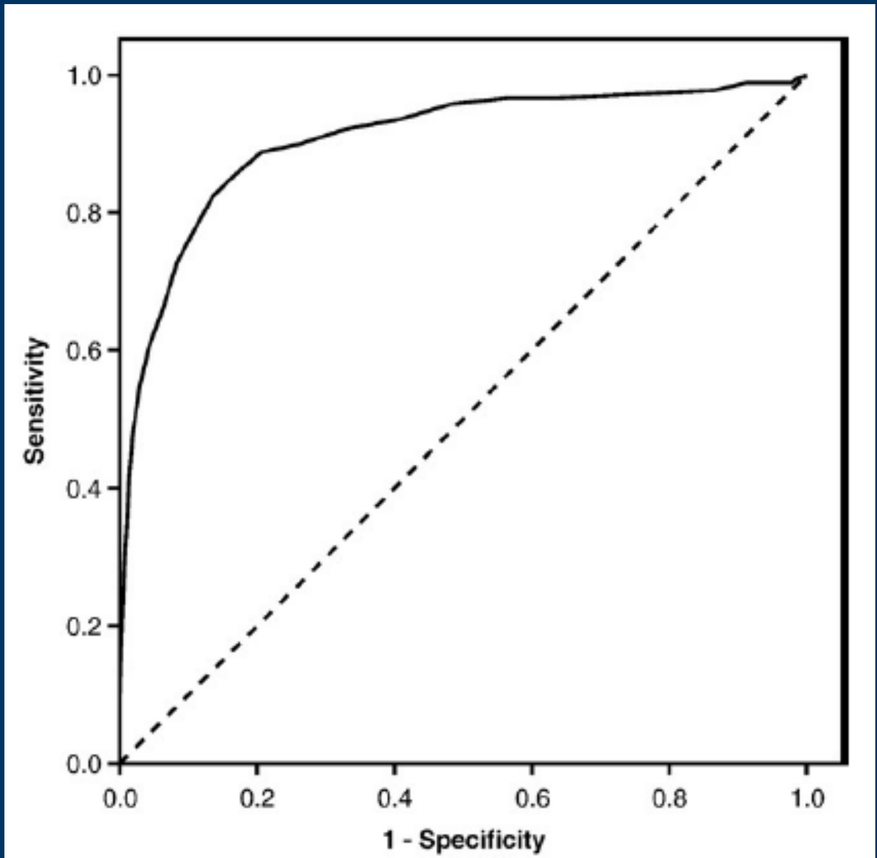


Fig. 1. Fatigue scale ROC plot.

Wald's Confidence interval

$$IC_{95\%}\left(\text{AUC}\right) = \hat{\text{AUC}} \pm 1.96 \times se\left(\hat{\text{AUC}}\right)$$

Problem: the upper limit might be larger than 1

Solution: use bootstrap (beyond the scope of the course).

Best cutoff for discriminating cases from controls

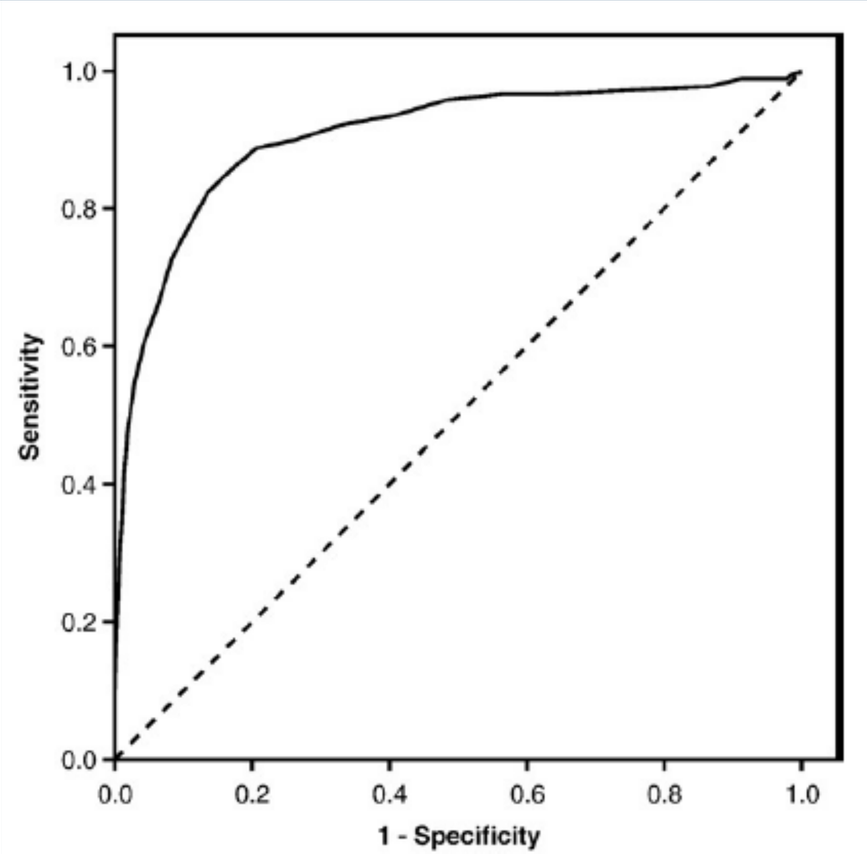


Fig. 1. Fatigue scale ROC plot.

Which point choose we choose from the ROC curve?

Which possible criteria can we use?

Best cutoff for discriminating cases from controls

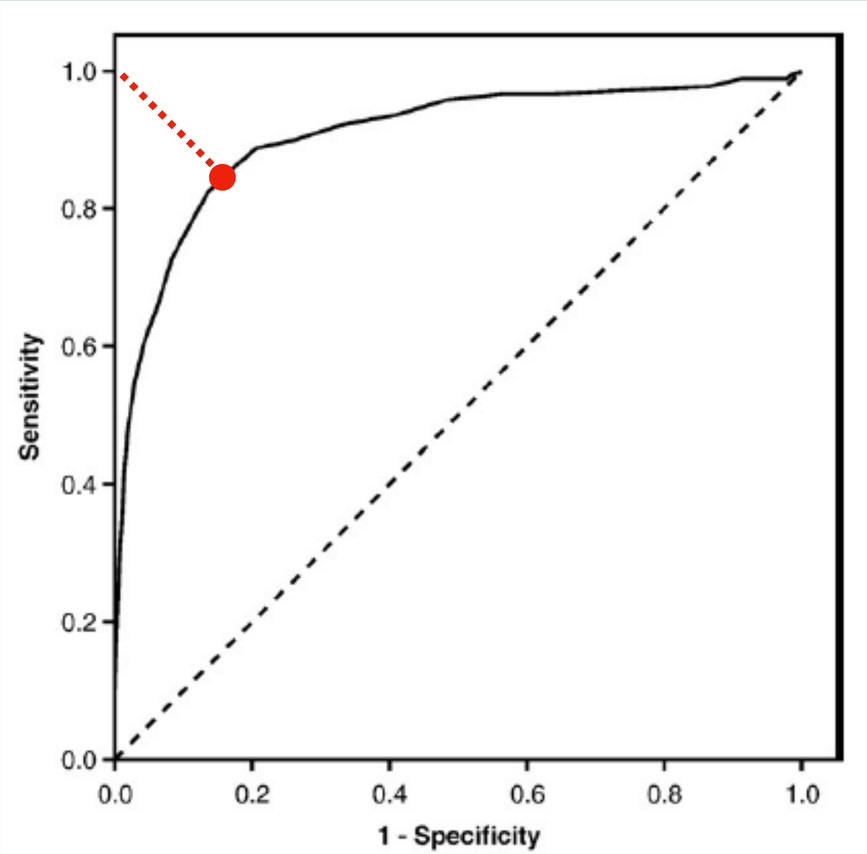


Fig. 1. Fatigue scale ROC plot.

Which point choose we choose from the ROC curve?

Which possible criteria can we use?

Criterion I: ROC01

Point that is the closest to perfect classification (X=0,Y=1)

Best cutoff for discriminating cases from controls

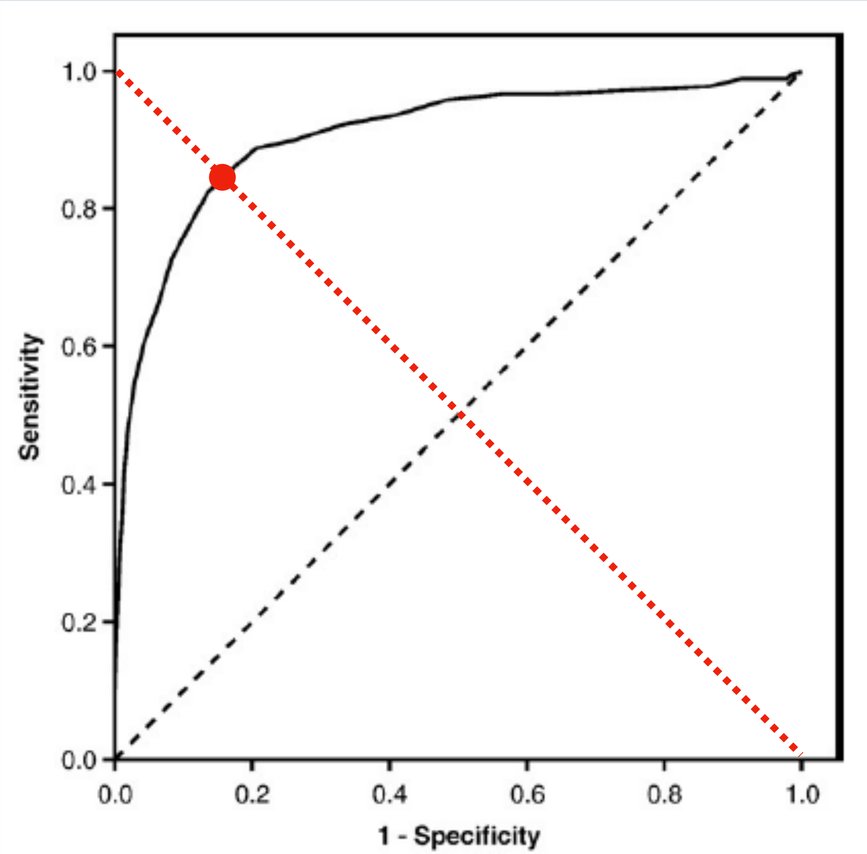


Fig. 1. Fatigue scale ROC plot.

Which point choose we choose from the ROC curve?

Which possible criteria can we use?

Criterion II: SeEqualSp

Point where the sensitivity and specificity are the same

Point where the distance between sensitivity and specificity is minimized

Best cutoff for discriminating cases from controls

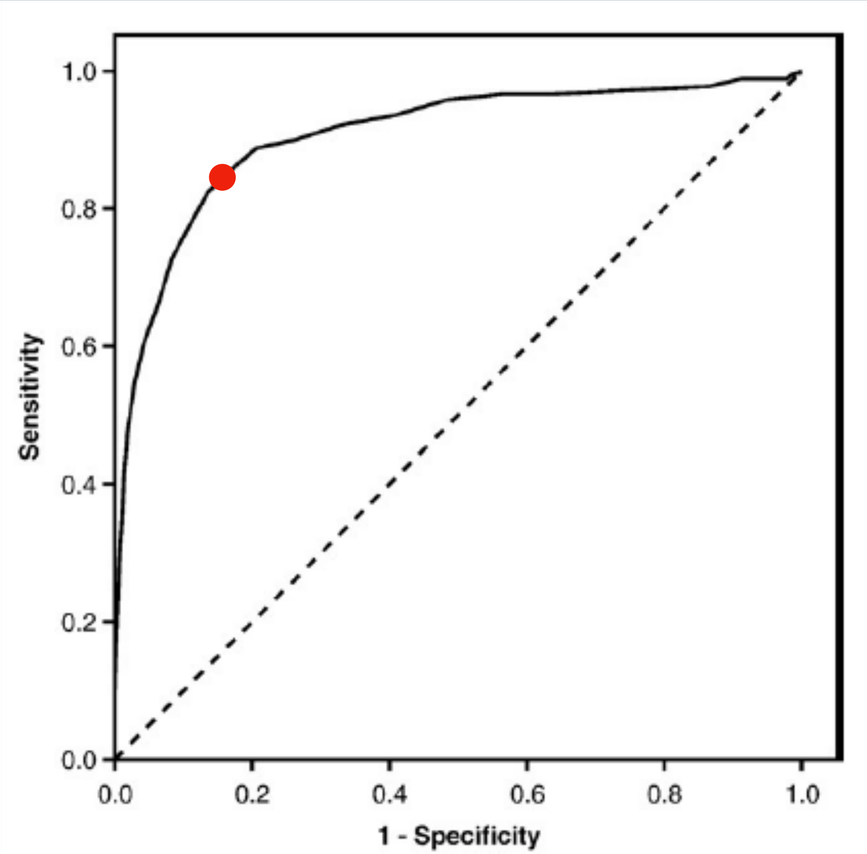


Fig. 1. Fatigue scale ROC plot.

Which point choose we choose from the ROC curve?

Which possible criteria can we use?

Criterion III: maxEfficiency

Point where the sensitivity and specificity provide the best accuracy

Exercise

ELSEVIER

Journal of Psychosomatic Research 69 (2010) 17–22

Journal of Psychosomatic Research

Measuring fatigue in clinical and community settings

Matteo Cella*, Trudie Chalder

Institute of Psychiatry, King's College London, London, United Kingdom

Received 11 June 2009; received in revised form 8 October 2009; accepted 13 October 2009

Abstract

Objective: The Chalder Fatigue Scale (CFQ) is a widely used instrument to assess fatigue in both clinical and nonclinical settings. Psychometric properties of the scale and discriminative abilities were examined. **Methods:** A total of 361 patients with CFS and 1615 individuals in the community were assessed with the CFQ. Principal component analysis (PCA) was used to explore the structure of the scale. Receiver-operating characteristic curve (ROC) was used to investigate the discriminative properties. **Results:** Two components, physical and mental fatigue, were identified in the CFS patient group and in the general population samples. Area under the curve for ROC was .91. The fatigue scale effectively discriminates, at high scores, between CFS patients and the general population. **Conclusion:** Physical and mental fatigue are clearly separable components of fatigue. The CFQ can discriminate reliably between clinical and nonclinical conditions.

Keywords: CFS; Chalder fatigue scale; Chronic fatigue syndrome; Fatigue; ROC

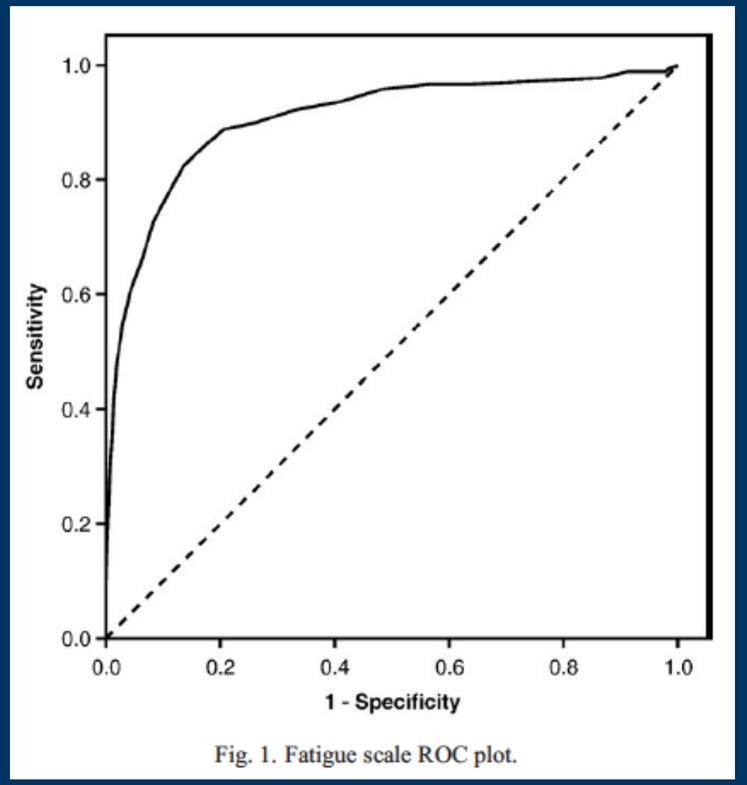


Fig. 1. Fatigue scale ROC plot.

Table 3
Sensitivity and 1−specificity for the fatigue scales scores

CFQ Score	Sensitivity	1−Specificity
0	1.000	1.000
1	1.000	.999
2	.997	.998
3	.997	.997
4	.997	.993
5	.994	.985
6	.989	.979
7	.989	.964
8	.989	.940
9	.989	.913
10	.978	.866
11	.966	.637
12	.966	.565
13	.958	.484
14	.936	.408
15	.922	.330
16	.899	.260
17	.888	.207
18	.854	.167
19	.824	.136
20	.773	.107
21	.728	.084
22	.658	.062
23	.605	.043
24	.546	.029
25	.473	.018
26	.415	.014
27	.339	.010
28	.286	.007
29	.202	.004
30	.151	.000
31	.095	.000
32	.062	.000
33	.000	.000

Calculate AUC (pROC package) and its 95% confidence interval.
Estimate the optimal cutoff for CFQ score according to ROC01, SeEqualSp and maxEfficiency (OptimalCutpoints package).
What are your conclusions?

Note

When the diagnostic tools comes in a form of probability (e.g., ML applications), it is usually to use 0.5 as the cutoff for the probability. Accuracy is then estimated.

This cut-off might not be the most optimal for accuracy.