Challenges for the Planning and Conduct of Diagnostic Studies With Molecular Biomarkers

Andreas Ziegler

ziegler@imbs.uni-luebeck.de

Institut für Medizinische Biometrie und Statistik
Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein

Berlin, 10.09.2013

Outline

- Levels of Evidence
- Important principles in study design
- Important sources of bias
- Quality of diagnostic studies
  - Example: HIV, malaria, tuberculosis
  - Findings from systematic reviews
  - Verification bias
  - Reasons for selection bias
- Software patents

Levels of Evidence in Diagnostic Studies (GBA)

Level	Criteria
I a	Systematic reviews of studies with evidence level I b
I b	Randomized controlled trials
I c	Other intervention studies
II a	Systematic reviews of diagnostic accuracy studies with evidence level II b
II b	Cross-sectional or cohort study enabling the estimation of measures on diagnostic accuracy (sensitivity, specificity, likelihood ratios, positive and negative predictive value)
III	Other studies enabling the estimation of measures on diagnostic accuracy (sensitivity, specificity, likelihood ratios)
IV	Associations, pathophysiological reflections, descriptive presentations, case reports, and similar; expert opinions not proven by studies, reports of expert committees and consensus conferences

http://www.g-ba.de/downloads/62-492-748/VerfO_2013-04-18.pdf [access data: Aug 12, 2013]

Important Methodological Principles of Diagnostic Accuracy Studies

Principle	Explanation
Two groups of patients	Patients with the disease for estimating sensitivity; group of subjects without disease for estimating specificity
Well-defined patient samples	Independent of ascertainment scheme: description of patient characteristics (e.g., age, gender, disease stage, comorbidities)
Well-defined diagnostic test	Clear definition of diagnostic test Application to all study participants in identical way
Gold standard / reference standard	Determination of true disease status of all study participants by perfect standard or best standard available
Sample of raters	If test requires trained raters, two or more raters required
Blinded investigation	Independent and blind assessment of reference standard and diagnostic test
Standardized reporting of results	Report according to STARD statement, QUADAS, QUADOMICS

Important Sources of Bias in Studies on Diagnostic Accuracy

Bias	Explanation
Spectrum composition bias	Spectrum of patients not representative of the patients who will receive the test in practice
Disease progression bias	Time period between reference standard and index test so long that target condition might have changed
Partial verification bias	Reference standard not applied to all patients
Differential verification bias	Use of different reference standards, use depending on test result
Incorporation bias	Reference standard and index test not independent; special case: Index test part of reference standard
Test review bias	Index test results interpreted with knowledge of results of reference standard
Reference standard review bias	Reference standard results interpreted with knowledge of results of index test
Clinical review bias	Index test interpreted in the light of clinical data that would not be available when test used in practice

Fontela et al. 2009 PLoS One 4:e7759
Ziegler et al. 2013 Dtsch Med Wochenschr 138:e14-e24, Table 4

Quality of Diagnostic Studies
Example: Tuberculosis, HIV, Malaria

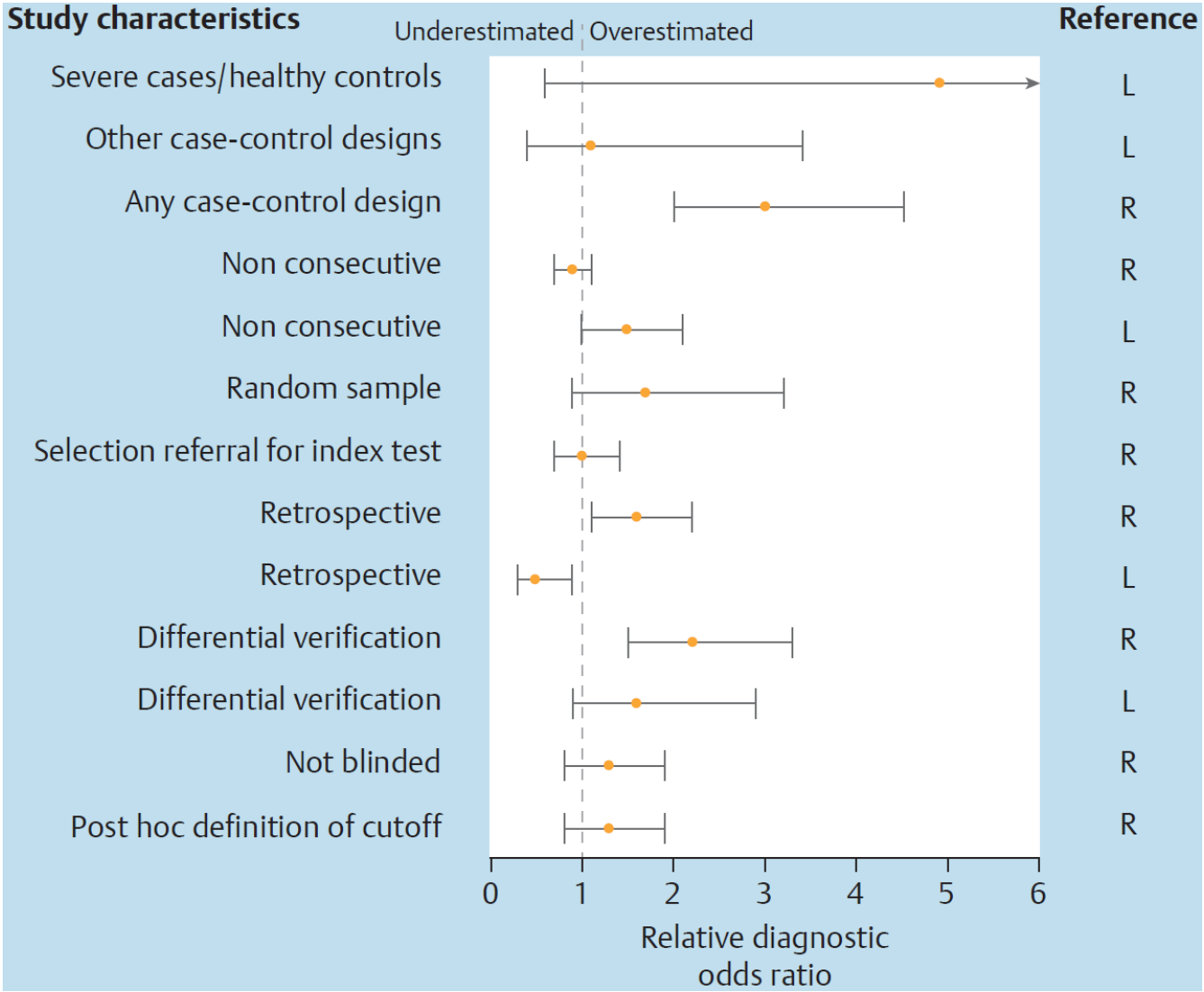
Inclusion criteria

- Commercially available test kit
- Tuberculosis, HIV, malaria
- Original publication
- Reporting of
  - Methods
  - Sensitivity
  - Specifity
- Publications 2004 – 2006
- Languages: English, French, Spanish, Portuguese

Quality of Diagnostic Studies
Example: Tuberculosis, HIV, Malaria

- 238 articles fulfilled inclusion criteria
- ALL studies with design issues
- 10% of studies with adequate description of reference standard
- 19% without index test observation bias
- 24% without reference standard observation bias
- < 25% of studies with 9 STARD indicators
  - 0% reported methods for calculation and estimation of reproducibility
  - 14% reported number of eligible subjects not included
  - 16% with blinded observer

Quality of Diagnostic Studies
Design dependent bias



Lijmer et al. 1999 J Am Med Assoc 282:1061-1066 • Rutjes et al. 2006 Can Med Assoc J 174:469-476 • Ziegler et al. 2013 Dtsch Med Wochenschr 138:e14-e24, Fig. 1

Quality of Diagnostic Studies
Design dependent bias

- "We found that studies that used retrospective data collection or that routinely collected clinical data were associated with an overestimation of the DOR by 60%."
- "Case–control studies had much higher estimates of diagnostic accuracy …"

- "Verification is a key issue in any diagnostic accuracy study. Studies that relied on 2 or more reference standards to verify the results of the index test reported odds ratios that were on average 60% higher than the odds ratios in studies that used a single reference standard."

Lijmer et al. 1999 J Am Med Assoc 282:1061-1066
Rutjes et al. 2006 Can Med Assoc J 174:469-476

Verification Bias: Example

- Study population: 6691 patients with suspected prostate cancer
- Index test: PSA (4.1ng/ml)
- Reference standard: biopsy
- 705 patients PSA + biopsy

Age	< 60 year	≥ 60 years
Sensitivity:	43%	35%
Specifity:	77%	88%

- 5986 patients only PSA (correction by Begg-Greenes)

Age	< 60 years	≥ 60 years
Adjusted sensitivity	18%	19%
Adjusted specifity	98%	94%

- Recommendation: decrease PSA threshold for biopsy

Verification Bias: Explanation

- Synonymous: work-up bias, referral bias

	Reference standard	
Index test	Positive	Negative
Positive	80 **a**	10 **b**
Negative	20 **c**	40 **d**

Naïve correction for verification bias →

	Reference standard	
Index test	Positive	Negative
Positive	80 **a**	10 **b**
Negative	20 + **60 = 80** **c**	40 + **120 = 160** **d**

$$sens = \frac{a}{a+c} = \frac{80}{100} = 80\%$$

$$spec = \frac{d}{b+d} = \frac{40}{50} = 80\%$$

$$sens = \frac{a}{a+c} = \frac{80}{160} = 50\%$$

$$spec = \frac{d}{b+d} = \frac{160}{170} = 94.11\%$$

Ziegler et al. 2013 Dtsch Med Wochenschr 138:e14-e24, Fig. 3

Reasons for Selection Bias

- Hypernormal controls
- Severe cases (selection for symptoms)
- Differences in ascertainment of patients (patient spectrum in emergency room compared to day clinic)
- Referral for index test bias: inclusion into study depending on index test result
- Loss to follow-up bias: healthy study participants missing at follow-up visits
- Participation bias (selection-bias): selected part of subjects participates in study
- Limited challenge bias: only inclusion in presence of previous test(s)
- Increased challenge bias: only inclusion in presence of specific diagnosis
- Study examination bias: only inclusion if kept up at specific investigation or if specific investigation possible

Phases of Diagnostic Biomarker Studies

Phase	Description	Aim of study
I a	Discovery	Identification of promising biomarkers
I b	Assay development, assay validation	Define and optimize analytical process into robust, reproducible, and valid device
I c	Retrospective validation	Clinical assay detects disease; development of first algorithm for combination test
II	Retrospective refinement	Validation of early detection properties of biomarker (set); development and/or refinement of algorithm(s) for combination tests
III	Prospective investigation	Determination of diagnostic accuracy (sensitivity, specificity) in the situation of clinical routine
IV a	Randomized controlled trial	Quantification of effect of making the biomarker information available to the doctor to reduce disease burden
IV b	Health economics study	Quantification of cost-effectiveness

Ziegler et al. 2012 Hum Genet 131:1627-38

Effect of Precision
Sample Size

- Simplified standard sample size formula

$$n = 2 \cdot 10 \cdot \frac{1}{\Delta^2}$$

Simplification for $\alpha = 0.05$, $\beta = 0.2$:

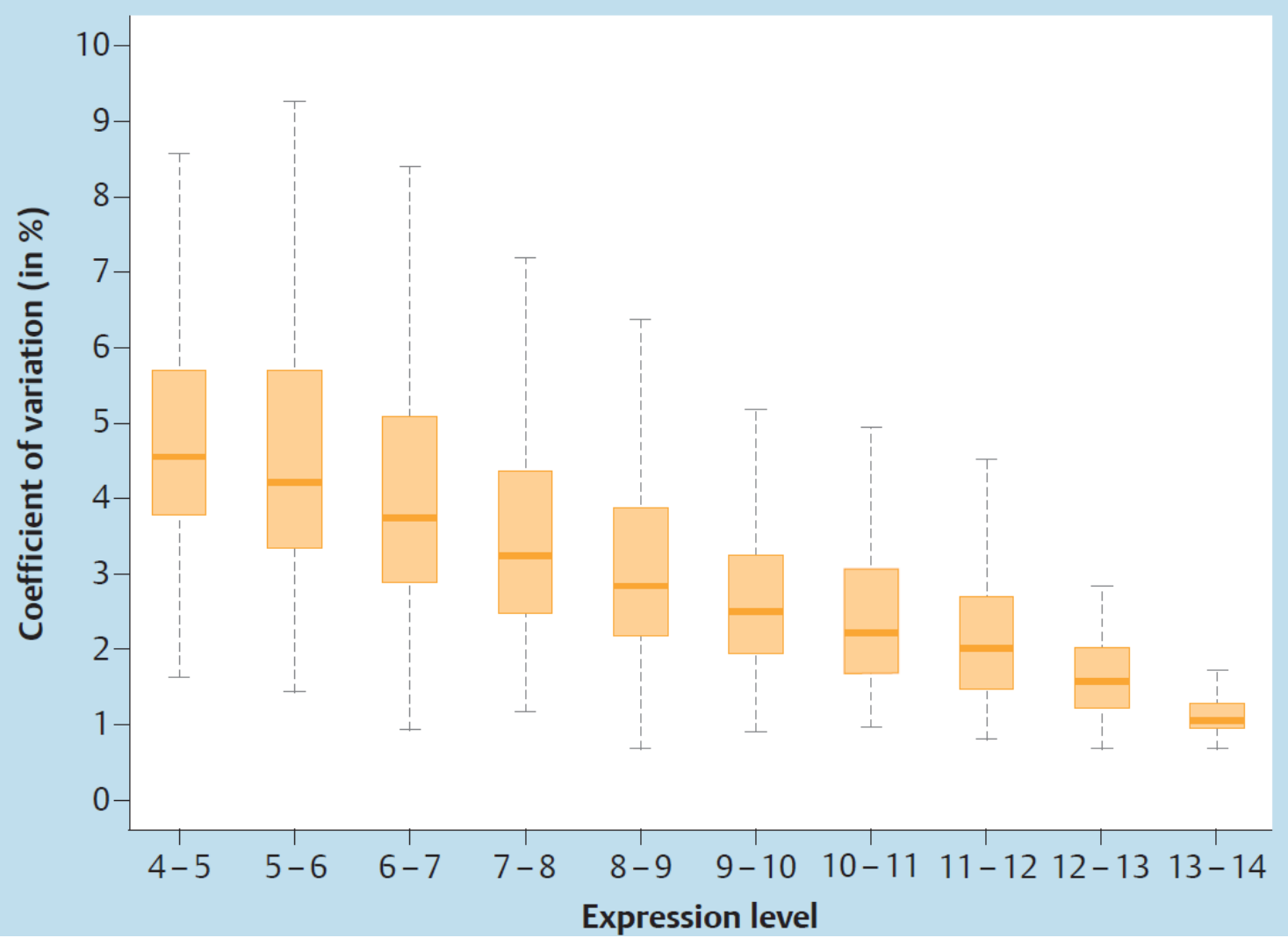
$$10 \approx (z_{1-\alpha/2} + z_{1-\beta})^2$$

- Sample size formula for coefficient of variation
  - Fold change f example: $f = 2$ $\mu_2/\mu_1 = 2 \Rightarrow \mu_2 = 2 \cdot \mu_1$
  - Coefficient of variation (relative variability) $cv = \sigma/\mu$

$$n = 2 \cdot 10 \cdot \frac{1}{\Delta^2} = 2 \cdot 10 \cdot \frac{v^2}{(1-f)^2}$$

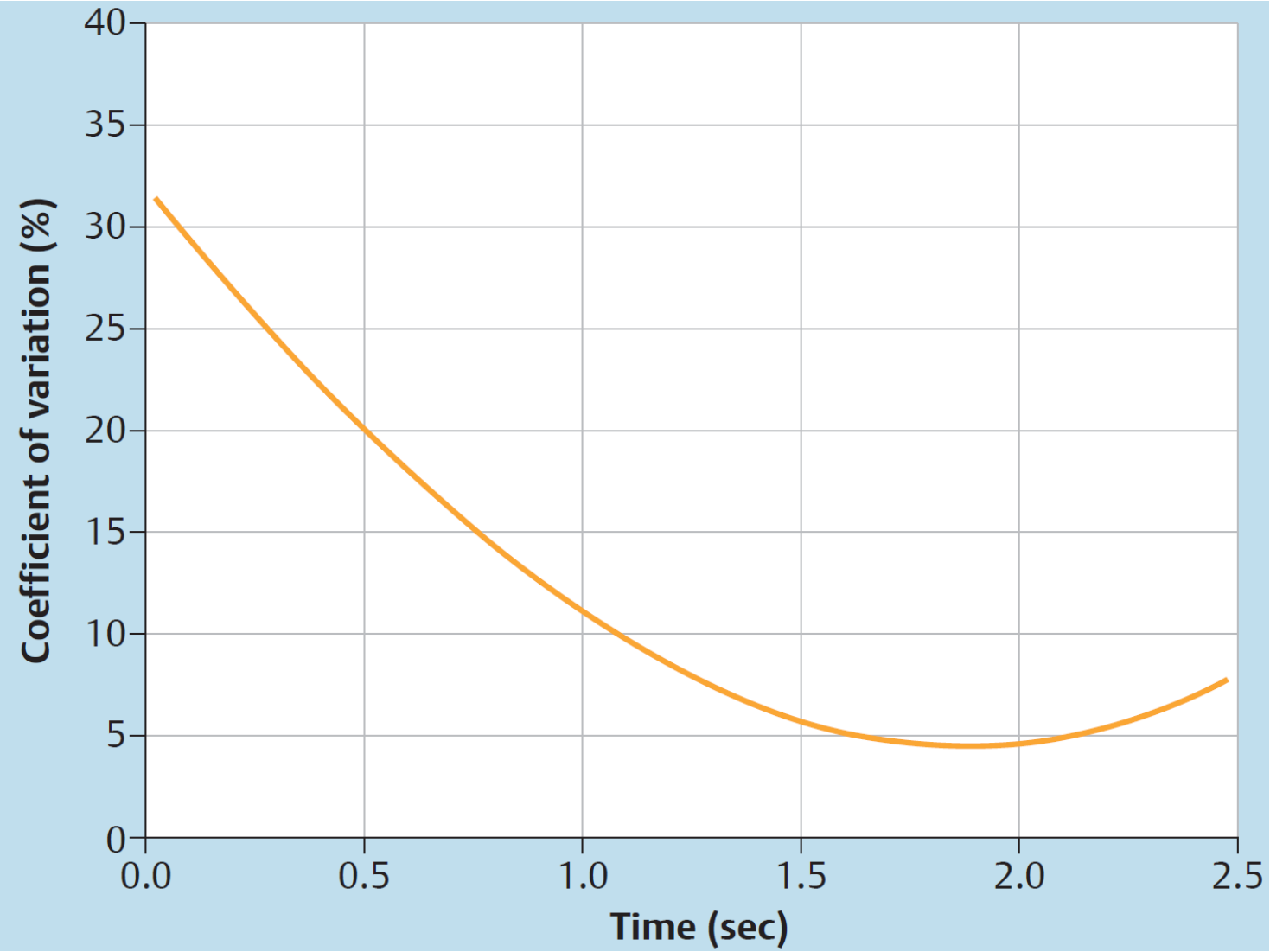
- **Squared increasey of sample size with coefficient of variation**

Effect of Precision
Example: Gene Expression Data



Ziegler et al. 2013 Dtsch Med Wochenschr 138:e14-e24, Fig. 7

Effect of Precision
Example: HPLC



Ziegler et al. 2013 Dtsch Med Wochenschr 138:e14-e24, Fig. 8

Note on Software Patents

	PAT. NO.	Title
1	8,095,483	Support vector machine--recursive feature elimination (SVM-RFE)
2	8,005,293	Gradient based training method for a support vector machine
3	7,937,351	Support vector machine using iterative chunking and violators
4	7,873,189	Face recognition by dividing an image and evaluating a similarity vector with a support vector machine
5	7,836,000	System and method for training a multi-class support vector machine to select a common subset of features for classifying objects
6	7,805,388	Method for feature selection in a support vector machine using feature ranking
7	7,797,257	System for providing data analysis services using a support vector machine for processing data received from a remote source
8	7,707,129	
9	7,685,077	
10	7,599,897	
11	7,567,352	
12	7,542,959	
13	7,533,076	
14	7,511,835	
15	7,483,809	
16	7,478,074	
17	7,475,048	
18	7,472,063	
19	7,421,417	
20	7,406,450	Spread kernel support vector machine
21	7,395,253	Lagrangian support vector machine
22	7,386,527	Effective multi-class support vector machine classification
23	7,372,583	Controlling a fabrication tool using support vector machine
24	7,356,187	Method and apparatus for adjusting the model threshold of a support vector machine for text classification and filtering
25	7,293,001	Hybrid neural network and support vector machine method for optimization
26	7,283,984	Method and apparatus for optimizing support vector machine kernel parameters
27	7,240,042	System and method for biological data analysis using a bayesian network combined with a support vector machine
28	7,054,847	System and method for on-line training of a support vector machine
29	7,020,642	System and method for pre-processing input data to a support vector machine
30	6,961,719	Hybrid neural network and support vector machine method for optimization
31	6,941,301	Pre-processing input data with outlier values for a support vector machine
32	6,889,129	Vehicle seat occupant classifying method and apparatus based on a support vector machine
33	6,327,581	Methods and apparatus for building a support vector machine classifier

Searching US Patents Text Collection...

Results of Search in US Patents Text Collection db for:
TTL/"random forest": 0 patents.

No patents have matched your query

Refine Search TTL/"random forest"

http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.htm&r=0&p=1&f=S&l=50&Query=TTL%2F%22support+vector+machine%22&d=PTXT

Summary

- Most important sources of bias
  - Inappropriate selection of subjects (cases / controls)
  - Verification bias
    - Different reference standard
    - Reference standard only in portion of patients
- Overestimation of diagnostic accuracy in case of inappropriate study design
- Validation challenging
- Coefficient of variation of assay important
- Multimarker rules need attention
- Be aware of software patents!

German: Ziegler et al. 2013 Dtsch Med Wochenschr 138:e2-e13
English: Ziegler et al. 2013 Dtsch Med Wochenschr 138:e14-e24