

Evaluating diagnostics and biomarkers: What should statisticians know?

1

Biological measurement providing information to guide care

- Traditional biomarkers
 - ❖ PSA, BMI, GTT, d-Dimer, pregnancy test
 - ❖ Oncotype DX, Mammaprint, Her-2
- Devices
 - ❖ CT scans, ultrasound, thermography, FOBT, colonoscopy, oral rinse
- Genes
 - ❖ Huntington's, BrCA, Gene expression
- Clinical hx, signs and symptoms!

2

What kind are these used for?

3

Tests are newsworthy

4

 STANFORD
SCHOOL OF MEDICINE

Visit [ISM website](#)

INSIDE STANFORD MEDICINE

Having trouble viewing this email? [Click here](#) to view the hosted version.

July 14, 2014 | Volume 6, No. 13

Microchip could aid diagnosis of type-1 diabetes

Researchers have invented a cheap, portable, microchip-based test that could speed up the diagnosis of type-1 diabetes and enable studies of how the disease develops.

The handheld microchips distinguish between the two main forms of diabetes mellitus, which are both characterized by high blood-sugar levels but have different causes and treatments. Until now, making the distinction has required a slow, expensive test available only in sophisticated health-care settings.



"We would like this to be a technology that satisfies global need," said Brian Feldman, MD, PhD, assistant professor of pediatric endocrinology, who is the senior author of a study describing the new test. [Read story >](#)

STANFORD
KIDS

5

www.bbc.co.uk/news/health-20836082

Man Cll PCORI reCall Weather From Safari Stanford News JHU sites 2.0K Share f t

24 December 2012 Last updated at 19:27 ET

Simple eye scan can reveal extent of Multiple Sclerosis

A simple eye test may offer a fast and easy way to monitor patients with multiple sclerosis (MS), medical experts say in the *Journal of Neurology*.

Optical Coherence Tomography (OCT) is a scan that measures the thickness of the lining at the back of the eye - the retina.

It takes a few minutes per eye and can be performed in a doctor's surgery.

In a trial involving 164 people with MS, those with thinning of their retina had earlier and more active MS.

The team of researchers from the Johns Hopkins University School of Medicine say larger trials with a long follow up are needed to judge how useful the test might be in everyday practice.

The latest study tracked the patients' disease progression over a two-year period.

Unpredictable disease

Multiple sclerosis is an illness that affects the nerves in the brain and spinal cord causing problems with muscle movement, balance and vision. In MS, the protective sheath around nerves, called myelin, becomes damaged which, in turn, leaves the nerves open to damage.

There are different types of MS - most people with the condition have the relapsing remitting type where the symptoms come and go over days, weeks or months.



The retina sits at the back of the eye and houses the cells that provide us with vision

Related Stories

Highest MS rate in Orkney Islands
New MS drug is 'most effective'

 STANFORD UNIVERSITY

6



After learning he had cancer, a man and his wife set out to cross off all the items on their bucket list, but after spending all their money, they learned the cancer diagnosis was a false-positive.

Couple Does Bucket List, Ends Up Broke.
Fueled by false cancer diagnosis, couple sold home, went on vacation.

By: Jena Kehoe
Web2Carz Contributing Writer
Published: June 20th, 2012

1 reddit 1 reddit fail 1 Tweet 1 3

A man named Frank was told in May 2010 that he had cancer. He and his wife, Wilma, then took to their bucket lists, crossing things off one by one. Two years later, they're \$80,000 in debt and blaming it on the New Zealand health care system—the cancer diagnosis turned out not



7

Goal of a Dx/Pred/Scr biomarker?

Not!!! just to distinguish between diseased and non-diseased, or high and low risk patients, or to “predict”.

To improve the overall health outcomes (or reduce cost or suffering), i.e. do more good than harm.



8



10 July 2012 | Volume 305 | Number 6807 | BMJ

GP recruitment scheme falters p127
Making vaccinations compulsory p138
The ethics of the Charlie Gard case p148
How to diagnose IBS in children p162
1.5 GPD hour in the education section

New diagnostic tests: more harm than good?

9

Forbes /

FEB 21, 2012 @ 09:01 PM 1,120 VIEWS

We Spend How Much On Unneeded Tests?

Reuters has a great story on the American College of Physicians efforts to stem medical overuse. Early in the story, it cites a \$250 billion figure for the amount wasted on unneeded tests. But then comes this titan of a figure:



Matthew Herper
FORBES STAFF

I cover science and medicine, and believe this is biology's century.

FOLLOW ON FORBES (2115)

f t r e

FULL BIO >

Options expressed by ...

“I don’t trust professional societies to do it because that’s how they make money – by doing tests and procedures,” said MIT healthcare economist Dr. Jonathan Gruber.

He cites estimates that about \$800 billion – or nearly one-third of all healthcare spending – is wasted in unnecessary diagnostic tests, procedures and extra days in the hospital. Treatment guidelines will help curb overuse, but Gruber and others would prefer the government set them.

via Stemming the tide of overtreatment in U.S. healthcare | Reuters.

10

Example: NeuroPointDx Blood test for autism



11

ST. JOURNAL News Sports Opinion Columns Business Lifestyle Entertainment Back & Side 74°

Madison company's study paving the way for an autism blood test is published in scientific journal

JULY NEUMANN/jneumann@madison.com | JULY 4, 2018



NeuroPointDX CEO Elizabeth Denley left, and Miller Lublig, lab manager, discuss new blood technology at a laboratory in Madison. The company is developing a test to identify chemicals in the blood that can indicate certain types of autism spectrum disorder.

MORE INFORMATION

Facebook Twitter LinkedIn Email Print

NeuroPointDX's ability to be public to its research is limited due to patent issues. The Madison-based company hopes to start offering its test this spring.

A Madison company that aims to "revolutionize" the way autism spectrum disorder is diagnosed and treated is getting a boost from a scientific publication.

Research conducted by Madison-based NeuroPointDX in collaboration with the MIND Institute at the University of California Davis showed subtle differences in blood can identify some children as young as 18 months with autism spectrum disorder.

The findings are being published Thursday in a peer-reviewed

MOST POPULAR

- Dairy industry, rural banks ask Wisconsin Supreme Court to re-vote case
- Downtown Middleton Wine Walk
- Gaylord Fabrica, a commercial roofing entrepreneur, dies
- Arrendahl Pharmaceuticals acquires Milwaukee-based company
- New retail building approved for Towne Center on Junction Road

12

https://madison.com/wsj/business/madison-company-neuropointdx-debuts-blood-test-for-autism/article_c2806431-4b93-5927-96e8-abd457ba3d4e.html

TOPICAL TOP STORY

AUTISM | NEW BLOOD TEST
Madison company NeuroPointDX debuts blood test for autism

JUDY NEWMAN jdnewman@madison.com Nov 1, 2018

Blood samples are loaded into a mass spectrometer to analyze them for signs of metabolic variations that could signal autism spectrum disorder at NeuroPointDX, 504 S. Rosa Road.

NEUROPOINTDX

13

2/27/19

NeuroPoint Dx
Point. Act. Thrive.

[HOME](#) [OUR TEST](#) [PARENTS](#) [PROVIDERS](#) [SEARCH](#)

Helping Children Receive an Earlier Autism Diagnosis

For Parents For Providers

Earlier Diagnosis + Earlier Intervention = Better Outcomes

14

A tool for earlier ASD diagnosis

Earlier diagnosis and intervention can help children with ASD thrive. Behavioral tests can lead to an ASD diagnosis in children as young as 24 months old, but the average age of diagnosis in the United States is more than 4 years! Our NPDX ASD test can help kids receive a diagnosis and begin treatment sooner.

It can be difficult to know when to be concerned because kids develop different skills, like walking and talking, at different times. It can be hard to tell if a child is experiencing delayed development that could signal a condition like ASD or is simply developing at a different pace compared to his or her peers.

This is why a biological test, one that's less susceptible to interpretation, could help doctors diagnose children with ASD at a younger age. The NPDX ASD test was developed for children as young as 18 months old.

15

Earlier Diagnosis Metabolic Imbalances Accurate Tool A New Approach

A highly-accurate tool able to identify ~30% of children with ASD

The NPDX ASD test can identify about 30% of children with autism spectrum disorder with an increased risk of an ASD diagnosis. This means that three in 10 kids with autism spectrum disorder could receive an earlier diagnosis, get interventions sooner, and potentially receive more precise treatment suggestions from their doctors, based on information about their own metabolism.

A new approach to thinking about ASD that has been rigorously validated in a large clinical study

Through our clinical study, the Children's Autism Metabolome Project (CAMP), we found differences in the metabolic profiles of certain small molecules in the blood of children with ASD. Our research has identified a connection between metabolism and autism spectrum disorder.

Over 1,100 children enrolled in CAMP, which was conducted at eight clinical sites across the country. NeuroPointDX is grateful to the families who participated and all of the health care professionals who helped make this study a success.

The results of CAMP have been published in a peer-reviewed, highly-regarded journal, *Biological Psychiatry*.

[Read the Publication](#)

16

Biological Psychiatry
Volume 85, Issue 4, 15 February 2019, Pages 345-354

ELSEVIER

Amino Acid Dysregulation Metabotypes: Potential Biomarkers for Diagnosis and Individualized Treatment for Subtypes of Autism Spectrum Disorder

Alan M. Smith, Joseph J. King, Paul R. West, Michael A. Ludwig, Elizabeth L.R. Donley, Robert E. Burns, and David G. Amaral

ABSTRACT

The combination of glutamine, glycine, and ornithine amino acid dysregulation metabotypes identified a dysregulation in AA/BCAA metabolism that is present in 16.7% of the CAMP subjects with ASD and is detectable with a specificity of 96.3% and a positive predictive value of 93.5% within the ASD subject cohort.

Conclusions:
“Identification and utilization of metabotypes of ASD can lead to actionable metabolic tests that support early diagnosis and stratification for targeted therapeutic interventions.”

subject cohort
CAMP subjects. Identification and utilization of metabotypes of ASD can lead to actionable metabolic tests that support early diagnosis and stratification for targeted therapeutic interventions.
Keywords: Amino acids, Autism, Biomarker, Diagnosis, Metabolomics, Metabotype
<https://doi.org/10.1016/j.biopsych.2018.08.018>

17

Background

“ASD is currently diagnosed based on behavioral characteristics exhibited by affected children (12). While specialist clinicians can confidently diagnose children as young as 24 months (13), the average age of diagnosis in the United States is over 4 years (2,14). Families often experience long waits to receive a definitive diagnosis owing to the paucity of trained clinicians able to perform diagnostic assessment.

Early diagnosis is important not only because intensive behavioral therapies are effective in reducing disability in many children with autism (15–17) but also because the benefit of early intervention is greater the earlier the intervention is started.”

18

STANFORD UNIVERSITY

Reference standard

- The Autism Diagnostic Observation Schedule—Second Version (ADOS-2) to confirm an ASD diagnosis in all children.
- Children with developmental delay but without ASD were excluded after recruitment. (n=94)
- Children entering the study as TYP [aka normal] were not routinely administered the ADOS-2.
- The Social Communications Questionnaire was administered to 65 TYP children as a screen for ASD. Of these, 9 children were referred for subsequent ADOS-2 evaluation, with 4 receiving a diagnosis of ASD (and being included in this study) and 5 receiving a diagnosis of TYP.

 STANFORD UNIVERSITY

19

Metric	Training Set	Test Set	Combined Sets
ASD Children	253	263	516
TYP Children	85	79	164
ASD Prevalence (%)	74.9	76.9	75.9
ASD % Male	77.9	79.5	78.7
TYP % Male	64.7	59.5	62.2
ASD Age (Months)	35.9 +/- 7.5	34.5 +/- 7.9	35.2 +/- 7.8
TYP Age (Months)	32.6 +/- 8.5	31.3 +/- 8.8	32 +/- 8.7
Age (range)	18 to 48	18 to 48	18 to 48
DQ ASD	62 +/- 17.8	63.5 +/- 17.7	62.8 +/- 17.8
DQ TYP	98.5 +/- 14.7	101.8 +/- 18.2	100.1 +/- 16.5

Table S3. Subject composition of training and tests. Means +/- Standard Deviation. Abbreviations: TYP, typically developing; ASD, autism spectrum disorder; DQ, developmental quotient.

20

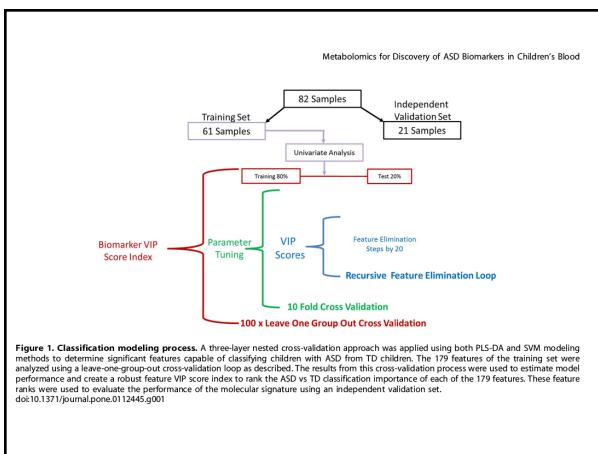


Figure 1. Classification modeling process. A three-layer nested cross-validation approach was applied using both PLS-DA and SVM modeling methods to determine significant features capable of classifying children with ASD from TD children. The 179 features of the training set were analyzed using a leave-one-group-out cross-validation loop as described. The results from this cross-validation process were used to estimate model performance and create a robust feature VIP score index to rank the ASD vs TD classification importance of each of the 179 features. These feature ranks were used to evaluate the performance of the molecular signature using an independent validation set.
doi:10.1371/journal.pone.0112445.g001

21

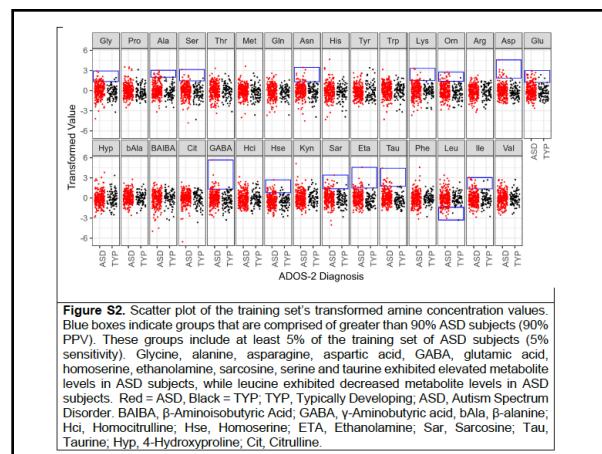
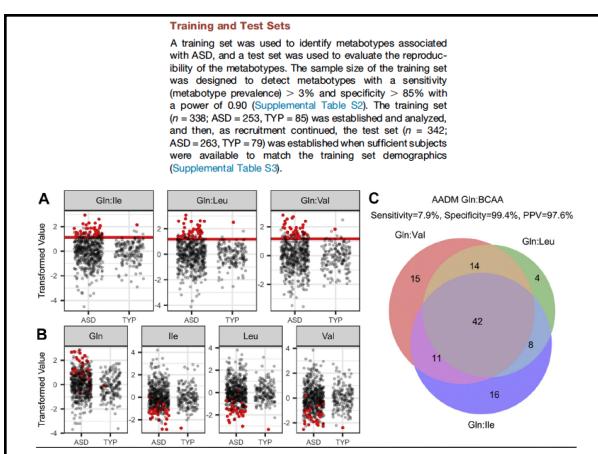
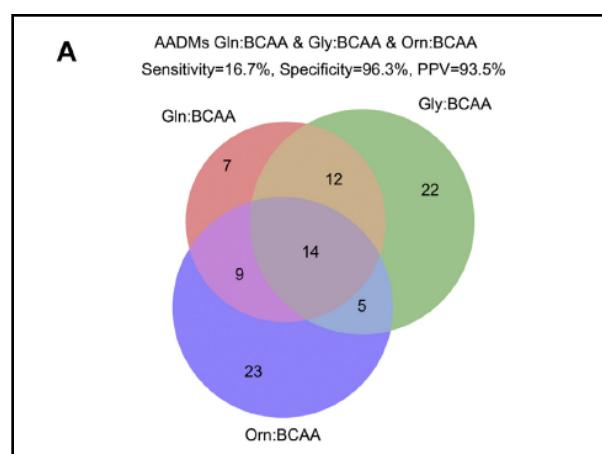


Figure S2. Scatter plot of the training set's transformed amino acid concentration values. Blue boxes indicate groups that are comprised of greater than 90% ASD subjects (90% PPV). These groups include at least 5% of the training set of ASD subjects (5% sensitivity). Glycine, alanine, asparagine, aspartic acid, GABA, glutamic acid, homoserine, ethanolamine, sarcosine, serine and taurine exhibited elevated metabolite levels in ASD subjects, while leucine exhibited decreased metabolite levels in ASD subjects. Red = ASD, Black = TYP. Typically Developing; ASD, Autism Spectrum Disorder; RAIBA, β -Aminobutyric Acid; GABA, γ -Aminobutyric acid; bAla, β -alanine; Hci, Homocitulline; Hse, Homoserine; ETA, Ethanolamine; Sar, Sarcosine; Tau, Taurine; Hyp, 4-Hydroxyproline; Cit, Citrulline.

23



24



25

Conclusions

- The AADMs provide one pathway to much earlier diagnosis of a substantial subset of children with ASD. Earlier diagnosis may also provide the opportunity for earlier biological intervention.
- Stratifying ASD based on metabolotypes offers an opportunity to identify efficacious interventions within metabolotypes that can lead to more precise and individualized treatment. The hope is that by combining the established metabolotypes into a more comprehensive diagnostic system, a substantial percentage of children at risk for ASD will be identifiable at a very early age.

26



[Link to NeuropointDx CEO talk.](#)

25.00

28.00



27

Example POSITIVE Test Result Report

Description
Includes the number of amino acids tested and the method of analysis.

Result Summary
This patient's result is **POSITIVE**, indicating the patient has at least one metabolic substrate elevated above average. This section may include additional findings about individual amino acids.

Guidance
For this condition, follow-up with a neurodevelopmental specialist is recommended. We also recommend researching the recommendations for managing a patient with a metabolic specialist for further evaluation.

NeuroPoint^{Dx} Panel Results

NPDX ASD Panel Results

Patient Name: John M. Patient ID: 12345678901234567890 Order Number: 12345678901234567890 Order Date: 01/01/2018 Specimen Received: 01/01/2018 Requested By: Dr. Smith Request Time: 12:00 PM

NPDX ASD Panel Description

This panel includes amino acid analysis for patients with known or suspected autism spectrum disorder (ASD). Amino acid analysis is used to identify metabolic abnormalities associated with autism spectrum disorder (ASD). Specific details of findings are listed below.

Positive patient has metabolic signature associated with ASD

This panel includes amino acid analysis for patients with known or suspected autism spectrum disorder (ASD). Specific details of findings are listed below.

Metabolite 1: An increase reflects the patient's underlying level of amino acids detected. The following amino acids were elevated above average levels: Alanine, Arginine, Aspartic Acid, Asparagine, Cysteine, Glutamic Acid, Glutamine, Histidine, Isoleucine, Leucine, Lysine, Methionine, Phenylalanine, Threonine, Tryptophan, Tyrosine, Valine, and Valproate.

Metabolite 2: A decrease reflects the patient's underlying level of amino acids detected. The following amino acids were decreased below average levels: Alanine, Arginine, Aspartic Acid, Asparagine, Cysteine, Glutamic Acid, Glutamine, Histidine, Isoleucine, Leucine, Lysine, Methionine, Phenylalanine, Threonine, Tryptophan, Tyrosine, Valine, and Valproate.

Additional Findings
Level of individual amino acids is outside of range (See Article Results Table below).

Guidance
Recommends follow up with neurodevelopmental/ASD specialist for further evaluation. Some details include dietary modification may be recommended.

Lab Director: James Fuller PhD, DABCC, FAACC | CLIA NUMBER: 5020122400
NeuropointDx • 505 S. Rosa Road, Mountain View, CA 94039 • 800.854.6004 • www.NeuropointDx.com

Page 1 of 3

28

ARTICLE IN PRESS

Correspondence

Lack of Diagnostic Utility of "Amino Acid Dysregulation Metabolotypes"

To the Editor:

We read with interest the article by Smith et al. [1] on amino acid dysregulation metabolotypes (AADMs). The hypothesis that there may be subtypes of autism spectrum disorder (ASD) marked by measurable metabolic changes is interesting. However, we note that the data presented only demonstrate that the test described has minimal diagnostic utility. Here, we describe five serious problems that led to this conclusion.

ARTICLE IN PRESS

Correspondence

Reply to: Lack of Diagnostic Utility of "Amino Acid Dysregulation Metabolotypes"

To the Editor:

Our article entitled "Amino Acid Dysregulation Metabolotypes for Diagnosis and Individualized Treatment for Subtype of Autism Spectrum Disorder," provides an important contribution to the field of autism spectrum disorder (ASD) [1]. This publication is based on data from the Children's Hospital Los Angeles (CHLA) Autism Diagnostic and Treatment Program, which has enrolled 1100 children with ASD, developmental delay, and typical development. This study had the goal of determining

lower tails; cases simply had more variability than control subjects [see Figure 2 in Smith et al. [1]]. This is noted by the authors, but we note both biological plausibility and generalizability. Many factors, including genetic function, affect this variation. For example, children with ASD have more erratic eating behaviors than normal children, which could lead to metabolic changes. For example, some children with ASD are about 3 months younger than the ASD group, and younger toddlers may have more homogeneous diets.

For a test to be useful for autism spectrum disorder, it must detect cases at earliest, even before behavioral tests. This study provides no evidence that the AADM test can pick up autism sooner than behavioral testing—all cases in the study

ARTICLE IN PRESS

Correspondence

Reply to: Lack of Diagnostic Utility of "Amino Acid Dysregulation Metabolotypes"

To the Editor:

Our article entitled "Amino Acid Dysregulation Metabolotypes for Diagnosis and Individualized Treatment for Subtype of Autism Spectrum Disorder," provides an important contribution to the field of autism spectrum disorder (ASD) [1]. This publication is based on data from the Children's Hospital Los Angeles (CHLA) Autism Diagnostic and Treatment Program, which has enrolled 1100 children with ASD, developmental delay, and typical development. This study had the goal of determining

lower tails; cases simply had more variability than control subjects. Our underlying hypothesis is that subpopulations of ASD exist owing to the heterogeneous underlying biology of ASD. There is a wide range of genetic and environmental factors that contribute to the biological underpinnings of ASD. Accounting these factors with metabolites within the ASD population provides important clues to developing targeted treatments.

Banwait and Goodman [2] suggest that the age of the subject is an important factor in the AADM test. They attempt to identify differences in means within 6-month age bins of the AADM population—evidence that age is not a contributing factor to the AADM test is critical to the validity of the test. Research is currently underway to determine if the identified metabolites are stable over time, defining and characterizing a

30

29

What are the issues? I

- Prevalence of ASD in study population is artificially high at 76%. Prevalence = 1.7% in the general pop, 18% in high risk subs. So PPV(in study) = 76%.**
- Specificity is measured in normal kids, not those with developmental concerns. If Spec drops from 96% to 90%, it would have little value, and if to 83%, the test becomes literally worthless. (Excluded 94 kids with DD.)**
- Average levels of all metabolites were equal between the groups. However, those with ASD or more variable, with more high *and* low values. Weakens plausibility and can be affected by diet.**

ARTICLE IN PRESS

Correspondence

Reply to: Lack of Diagnostic Utility of "Amino Acid Dysregulation Metabolotypes"

To the Editor:

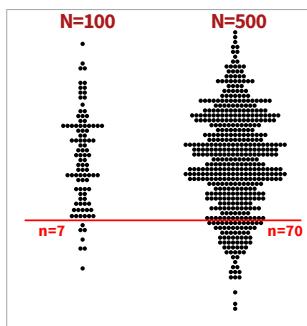
Our article entitled "Amino Acid Dysregulation Metabolotypes for Diagnosis and Individualized Treatment for Subtype of Autism Spectrum Disorder," provides an important contribution to the field of autism spectrum disorder (ASD) [1]. This publication is based on data from the Children's Hospital Los Angeles (CHLA) Autism Diagnostic and Treatment Program, which has enrolled 1100 children with ASD, developmental delay, and typical development. This study had the goal of determining

lower tails; cases simply had more variability than control subjects. Our underlying hypothesis is that subpopulations of ASD exist owing to the heterogeneous underlying biology of ASD. There is a wide range of genetic and environmental factors that contribute to the biological underpinnings of ASD. Accounting these factors with metabolites within the ASD population provides important clues to developing targeted treatments.

Banwait and Goodman [2] suggest that the age of the subject is an important factor in the AADM test. They attempt to identify differences in means within 6-month age bins of the AADM population—evidence that age is not a contributing factor to the AADM test is critical to the validity of the test. Research is currently underway to determine if the identified metabolites are stable over time, defining and characterizing a

31

How to create more outliers....



What are the issues?

4. No evidence that this test detects ASD before clinical assessment can.
5. Neither test result eliminates the need for a clinical assessment; the definition of a clinically useless test.
6. All authors work for the company.

32

33

Reply 1 “...based on a lack of understanding.”

1. “...these criticisms are based on a lack of understanding of the state of biomarker research in ASD and specifically on the iterative, multistep process that research on biomarkers must take to establish a diagnostic test.”
2. Pointed to a disclaimer about PPV and Spec not in the online preprint.

34

ACCEPTED MANUSCRIPT

Original

subjects with 7-10% sensitivity and 92-98% PPVs. Taken together, all AADMs identified an altered metabolic phenotype of imbalanced BCAA metabolism in 16.7% of CAMP ASD subjects with a specificity of 96.3% and PPV of 93.5%.

Published

10% sensitivity and 92% to 98% PPVs. Taken together, all AADMs identified an altered metabolic phenotype of imbalanced BCAA metabolism in 16.7% of CAMP subjects with ASD with a specificity of 96.3% and a PPV of 93.5%. We wish to note that our use of PPV was not adjusted for prevalence of ASD in the general population or for populations at greater risk of ASD. The PPV reported is based on the CAMP study population and was used as one facet of the criteria to define metabotypes. We do not imply that a general population screen using the AADMs would achieve a similar specificity or PPV.

35

Reply 1, cont. “...based on a lack of understanding.”

3. “Diagnostic values were based on the CAMP study population to compare biomarker performance between a training set and a test set. The focus of our study was to demonstrate the existence of specific metabotypes in the CAMP study population.”
4. **On lack of mean differences:** “Our metabotyping approach identifies subpopulations of ASD subjects with values not observed in the majority of typically developing subjects without regard to mean values. Our underlying hypothesis is that subpopulations of ASD exist owing to the heterogeneous underlying biology of ASD.”

36

Reply 2 You have even less understanding

5. “The .. point that a biomarker-based test would be useful only if it could detect cases at earlier ages than behavioral tests, demonstrates a lack of understanding of the current state of ASD diagnosis and treatment. Currently, behavioral testing can reliably diagnose children beginning at 24 months of age (8), but the median age of diagnosis is 4.3 years (9). The CAMP study offers the opportunity to assess children as young as 18 months of age.”

(But behavioral testing was the reference standard and applied to all ASD subjects. The DQ was also calculated and dramatically different.)

37

Reply 2, cont.**You keep getting dumber**

6. "The point that the test does not satisfy the first criterion for a decision tool because it does not affect subsequent actions ignores the usefulness of earlier referral and treatment. The metabotype-based test panels are not intended to replace the standard of care, but rather to identify children at risk for ASD who can be referred to a specialist for further evaluation. Earlier intervention is well known to improve outcome, so the suggestion that a diagnostic test is worthless because it does not replace the next step in the standard of care is without merit."

But...every ASD patient in this study was suspected without this test. The "next step" was the reference standard. Since everyone suspected of ASD will get the reference standard regardless of this test result – negative or positive - that is why it is useless.

STANFORD UNIVERSITY

38

Reply 3**You have offended our honor.**

7. "The final comment made by Sainani and Goodman demeans the authors of the letter. They imply that ethical, unbiased research cannot be done through public–private collaborations."

STANFORD UNIVERSITY

39

BishopBlog

Ramblings on academic-related matters. For information on my research see <https://www.psy.ox.ac.uk/research/oxford-study-of-children-s-communication-impairments>. Twin analysis blog: <http://dttemp.blogspot.com/>, ERP time-frequency analysis blog: bishoptechbits.blogspot.com/. For tweets, follow @deevybee.

Saturday, 12 January 2019
NeuroPointDX's blood test for Autism Spectrum Disorder: a critical evaluation
 NeuroPointDX (NPDX), a Madison-based biomedical company, is developing blood tests for early diagnosis of Autism Spectrum Disorder (ASD). According to their [Facebook page](#), the NPDX ASD test is available in 45 US states. [It does not appear to require FDA approval](#). On the [Payments tab](#) of the website, we learn that the test is currently self-pay (not covered by insurance), but for those who have private insurance, a payment plan is available, whereby the test is conducted after a down payment is received, but the results are not disclosed to the referring physician until two further payments have been made.

So what does the test achieve, and what is the evidence behind it?

Claims made for the test
 On their website, NPDX describe their test as a 'tool for earlier ASD diagnosis'. Specifically they say:

'It can be difficult to know when to be concerned because kids develop different skills, like walking and talking, at different

Search This Blog

Search

Subscribe To

Posts

Comments

Prizewinning blog



The CAMP study is [registered](#) on ClinicalTrials.gov, where it is described as follows:

*'The purpose of this study is to identify a metabolite signature in blood plasma and/or urine using a panel of biomarker metabolites that differentiate children with autism spectrum disorder (ASD) from **children with delayed development (DD)** and/or typical development (TD), to develop an algorithm that maximizes sensitivity and specificity of the biomarker profile, and to evaluate the algorithm as a diagnostic tool.'* (My emphasis)

The study is also included on the NIH Project Reporter portfolio, where the description includes the following information:

*'Stemina seeks funding to enroll 1500 patients in a well-defined clinical study to develop a biomarker-based diagnostic test capable of classifying ASD relative to other developmental delays at greater than 80% accuracy. In addition, we propose to identify metabolic subtypes present within the ASD spectrum that can be used for personalized treatment. The study will include ASD, DD and TD children between 18 and 48 months of age. **Inclusion of DD patients is a novel and important aspect of this proposed study from the perspective of a commercially available diagnostic test.**'* (My emphasis)

40

41

Key questions

- Is the intended use clear?
 - Rule in? Rule out? Avoid/replace the reference std? For whom? When? Decide on intervention?
- Does the population tested mirror the population for the intended use? Do we know how and why they participated?
- Are study test logistics the same as in practice? (Including fasting, etc.)
- Does the evaluation take into account everything else we know at the time of testing? Incremental value?
- Will the clinical/decision pathway with the test lead to better outcomes than without it? (Are interventions known to be effective?)
- Performance in completely independent population?

42

STANFORD UNIVERSITY

thebmj

BMJ 2019;364:l640 doi: 10.1136/bmj.l640 (Published 19 February 2019)

Page 1 of 5

ANALYSIS**Dangerous diagnostics? Regulatory reform in the genomic era**

Molecular diagnostics are expanding rapidly yet many have not been externally evaluated. Kelly Holloway and colleagues identify failings in the regulatory system and report on recent efforts at reform

Kelly Holloway *postdoctoral fellow*¹, Fiona A Miller *professor*¹, Alberto Gutierrez *partner*², Stuart Hogarth *lecturer*³

¹Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada; ²NDA Partners, Virginia, USA; ³Department of Sociology, University of Cambridge, Cambridge, UK

43

Issues in dx biomarker study design

44

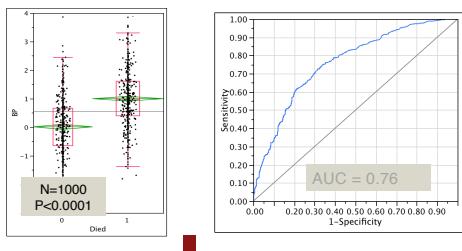


© Cartoontank.com

47

We will care about...

Individual classification, not group differences



49

REVIEWS

Pragmatic issues in biomarker evaluation for targeted therapies in cancer

Armand de Gramont, Sarah Watson, Lee M. Ellis, José-Angel Jiménez-Torremón, Alimrey de Gramont and others
De Gramont et al.
Nature Reviews (2015),
12: 197-212.

<https://doi.org/10.1038/nrcancer.2014.20>

Introduction

The development of cancer therapies is increasingly informed by our understanding of tumour biology, and biomarkers – especially predictive biomarkers – are crucial tools for identifying patients who are likely to benefit from targeted therapies. More effective patient selection can improve the success rate of new therapies, which are often expensive and have significant side effects that limit their cost-effectiveness, thus demanding reliable predictive biomarkers.

One commonly used definition of a biomarker is a substance or characteristic measured objectively and reproducibly either a normal biological state, or the response to a pathophysiological state, or a therapeutic intervention, and used for various purposes to predict survival (progression-free or overall), monitor disease course, target engagement and the immediate consequence on treatment. Biomarkers are used to identify patients who are most likely to benefit from a targeted therapy, to predict the response to therapy (surrogate biomarker); and to identify patients who are most likely to benefit from a targeted therapy (confirmatory biomarker).

Identification and validation of biomarkers are key to the best possible therapeutic strategies, thereby avoiding unnecessary treatments, reducing costs, and eventually reducing total health care expenditure.

In general, biomarkers are developed and evaluated in unselected patient populations, after only limited clinical trials. As an example, in phase III trials for first-line chemotherapy, 2.3 months with a median duration of 10–12 months, and a median overall duration of more than 20 months. Therefore, the cost-effectiveness of a new targeted therapy is potentially marginal in oncology.¹ Thus, biomarker validation is a major challenge for oncologists who are most likely to benefit from the addition of a new targeted therapy. In addition, the number of patients who are likely to benefit more effectively which would be equally important for patient and health providers, is often very small. This is particularly true for targeted drug development programmes, from which the number of patients who benefit is often small, robust measurements and easy validation for analyses of group differences are difficult to achieve. Biomarkers that emerge will probably be limited, and their use will be restricted to specific subgroups of patients. Identification and validation of biomarkers are key to the best possible therapeutic strategies, thereby avoiding unnecessary treatments, reducing costs, and eventually reducing total health care expenditure.

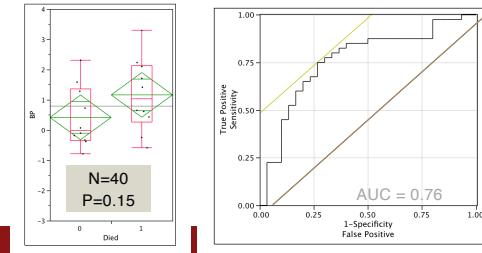
Most cancer therapies, especially those developed in unselected patient populations, offer only limited clinical benefit. As an example, in phase III trials for first-line chemotherapy and 2.3 months with a median duration of 10–12 months, and a median overall duration of more than 20 months. Therefore, the cost-effectiveness of a new targeted therapy is potentially marginal in oncology.¹ Thus, biomarker validation is a major challenge for oncologists who are most likely to benefit from the addition of a new targeted therapy. In addition, the number of patients who are likely to benefit more effectively which would be equally important for patient and health providers, is often very small. This is particularly true for targeted drug development programmes, from which the number of patients who benefit is often small, robust measurements and easy validation for analyses of group differences are difficult to achieve. Biomarkers that emerge will probably be limited, and their use will be restricted to specific subgroups of patients. Identification and validation of biomarkers are key to the best possible therapeutic strategies, thereby avoiding unnecessary treatments, reducing costs, and eventually reducing total health care expenditure.

In general, biomarkers are developed and evaluated in unselected patient populations, after only limited clinical trials. As an example, in phase III trials for first-line chemotherapy, 2.3 months with a median duration of 10–12 months, and a median overall duration of more than 20 months. Therefore, the cost-effectiveness of a new targeted therapy is potentially marginal in oncology.¹ Thus, biomarker validation is a major challenge for oncologists who are most likely to benefit from the addition of a new targeted therapy. In addition, the number of patients who are likely to benefit more effectively which would be equally important for patient and health providers, is often very small. This is particularly true for targeted drug development programmes, from which the number of patients who benefit is often small, robust measurements and easy validation for analyses of group differences are difficult to achieve. Biomarkers that emerge will probably be limited, and their use will be restricted to specific subgroups of patients. Identification and validation of biomarkers are key to the best possible therapeutic strategies, thereby avoiding unnecessary treatments, reducing costs, and eventually reducing total health care expenditure.

45

We will care about...

Individual classification, not group differences



48

We care about....

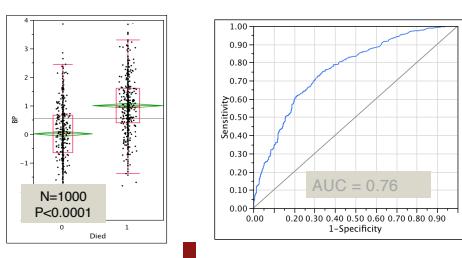
- ✓ Who is being tested
- ✓ Where they are being tested
- ✓ Why they are being tested
- ✓ How they are being tested
- ✓ Actions based on testing
- ✓ Consequences of those actions
 - To the one tested
 - To society

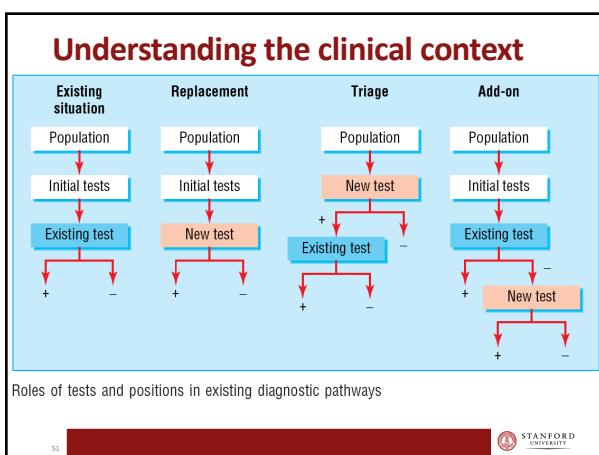
50



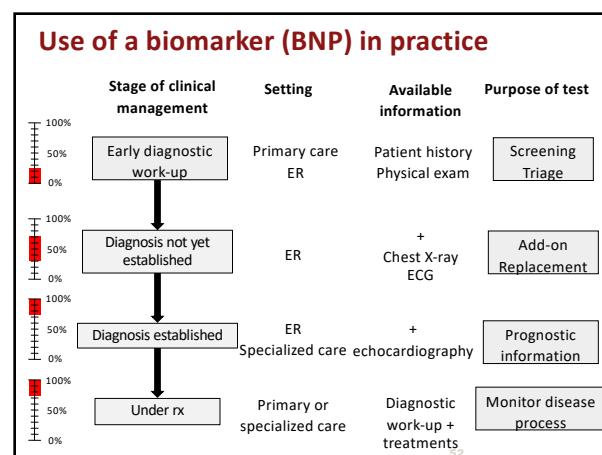
We will care about...

Individual classification, not group differences





51



52

PCORI Methodology Standards for Diagnostic Test Comparative Effectiveness Studies

DT-1 Specify clinical context and key elements of diagnostic test study design

A comparative evaluation of diagnostic tests should specify each of the following items and provide rationale in support of the particular choices:

- 1) the intended use of the test and the corresponding clinical context, including referral for additional testing, referral for additional treatments, and modification of current treatment and target populations;
- 2) the goal of the comparison;
- 3) the technical specifications of the tests as implemented in the study;
- 4) the approach to test interpretation;
- 5) the sources and process for obtaining reference standard information
- 6) the procedures for obtaining follow-up information and determining patient outcomes

53

PCORI Methodology Standards for Diagnostic Test Comparative Effectiveness Studies

DT-2 Study design should be informed by investigations of the clinical context of testing

Design of comparative effectiveness studies should outline clinical pathways involving the tests and the anticipated implications of test use on downstream processes of care and patient outcomes. In the written research methods and study protocol, investigators should give examples of clinical pathways to demonstrate thorough understanding of the clinical context.

54

Phases of Dx Biomarker Development

STANFORD UNIVERSITY

55

Diagnostic studies have phases just like therapeutic studies

Clinical Trials

- Safety (maximum tolerated dose)
- Pharmacokinetics
- Prelim. Efficacy
- Dosage response
- Clinically relevant effects
- Efficacy
- Safety surveillance

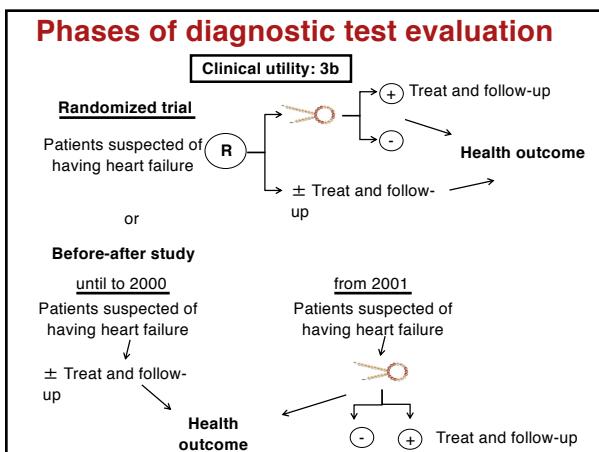
Diagnostic studies

19 models have been proposed, albeit all with similar ordinal sequence.

Lijmer et al. *Med Decis Making* 2009; 29: E13

STANFORD UNIVERSITY

56



63

Key takeaways

- Diagnostic biomarker eval only starts with sens and spec
- The research question must take the context of the measurement into consideration, including the decisions to be affected and the consequences of those decisions.
- Study design varies greatly across the phases of biomarker evaluation and by intended use.
- One of the most important skills is to recognize the phase of the study, with its concomitant purpose, appropriate conclusions and next steps.*

STANFORD UNIVERSITY

64

Stages and Levels of evidence (Simon et al., 2009, JNCI)

65

Table 1. Elements of tumor marker studies that constitute Levels of Evidence determination*				
Category	A	B	C	D
Element	Prospective	Prospective using archived samples	Prospective/observational	Retrospective/observational
Clinical trial	PCT designed to address tumor marker	Prospective trial not designed to address tumor marker, but design accommodates tumor marker utility	Prospective observational registry, treatment and follow-up not dictated	No prospective aspect to study
Patients and patient data	Prospectively enrolled, treated, and followed in PCT	Prospectively enrolled, treated, and followed in PCT, but especially if a predictive marker is considered, a PRCT addressing the treatment of interest	Prospectively enrolled observational registry, but treatment and follow-up standard of care	No prospective enrollment of treatment or follow-up; patient data collected by retrospective chart review
Specimen collection, processing, and archival	Specimens collected, processed, and assayed for specific marker in real time	Specimens collected, processed, and assayed prospectively using generic SOPs. Assayed after trial completion	Specimens collected, processed, and assayed prospectively using generic SOPs. Assayed after trial completion	Specimens collected, processed and archived with no prospective SOPs
Statistical design and analysis	Study powered to address tumor marker question	Study powered to address therapeutic question and underpowered to address tumor marker question	Study not prospectively powered at all. Retrospective study design confounded by selection of specimens for study	Study not prospectively powered at all. Retrospective study design confounded by selection of specimens for study
Validation	Result unlikely to be play of chance Although preferred, validation not required	Focused analysis plan for marker question developed before doing assays Requires one or more validation studies	Focused analysis plan for marker question developed before doing assays Requires subsequent validation studies	No focused analysis for marker question developed before doing assays Requires subsequent validation

* PCT = prospective controlled trial; PRCT = prospective randomized controlled trial; SOPs = standard operating practices.

3 R.D.
Vol. 101, Issue 21 | November 4, 2009

66

Table 2. Revised determination of Levels of Evidence using elements of tumor marker studies*

Level of evidence	Category from Table 1	Validation studies available
I	A	None required
I	B	One or more with consistent results
II	B	None or inconsistent results
II	C	2 or more with consistent results
III	C	None or 1 with consistent results or inconsistent results
IV-V	D	NAT

* Levels of Evidence (LOEs) revised from those originally proposed by Hayes et al. (3).

† NA = not applicable because LOE IV and V studies will never be satisfactory for determination of medical utility.

67

Mathematics of Diagnostic testing

STANFORD UNIVERSITY

68

Test accuracy

How the test performs when we know the diagnosis.

Sensitivity: Proportion of persons *with* disorder who will test *positive*

Specificity: Proportion of persons *without* disorder who will test *negative*

"Accuracy": Proportion of total test results that are right. Not usually helpful or used.

ROC: Display of Sensitivity versus 1-Specificity for all possible definitions of test positivity.



69

Test information

What a known test result tells us about the unknown diagnosis

Post-test probability: Probability of disease after test

Predictive Value: Probability of being right after a positive test or negative test.

Likelihood Ratio: How much the test changes the disease odds (from before to after test).



70

Bayes Theorem

$$\text{Pre-test disease odds} \times \text{Likelihood ratio} = \text{Post-test disease odds}$$



71

Bayes Theorem

$$\frac{\Pr(H_0 | \text{Data})}{\Pr(H_1 | \text{Data})} = \underbrace{\frac{\Pr(H_0)}{\Pr(H_1)}}_{\text{Post-test Odds}} \times \underbrace{\frac{\Pr(\text{Data} | H_0)}{\Pr(\text{Data} | H_1)}}_{\text{Likelihood Ratio}}$$

OR

$$\text{Likelihood ratio} = \frac{\text{Post-test odds}}{\text{Pre-test odds}} = \frac{\overbrace{\Pr(H_0 | \text{Data})}^{\text{Post-test Odds}}}{\overbrace{\Pr(H_1 | \text{Data})}^{\text{Pre-test Odds}}}$$

72

Diagnostic Test mnemonics

$$LR(+) = \text{Sens}/(1-\text{Spec})$$

$$LR(-) = (1-\text{Sens})/\text{Spec}$$

*SnNout

* A highly **Sensitive** test, that when **negative** rules **out** the disorder. (i.e. $LR(-)$ small])

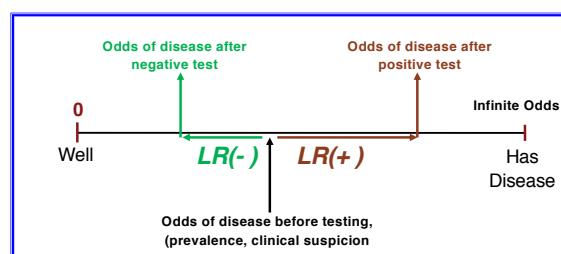
*SpPin

* A highly **specific** test, that when **positive** rules **in** the disorder. (i.e. $LR(+)$ large)

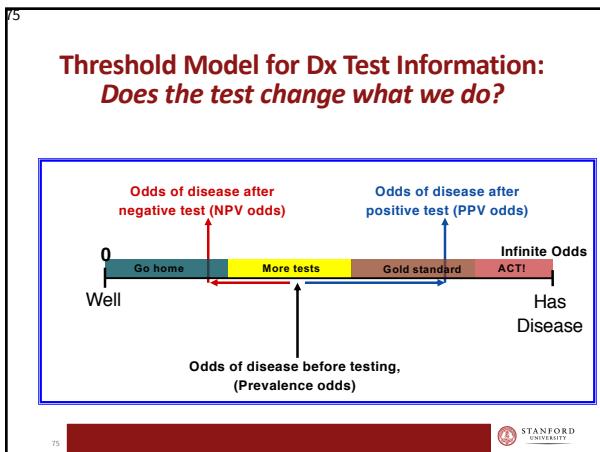


73

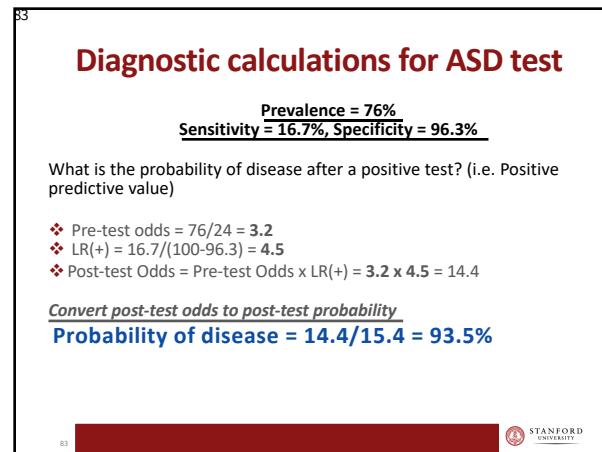
Bayesian Model for Analyzing Diagnostic Test Information



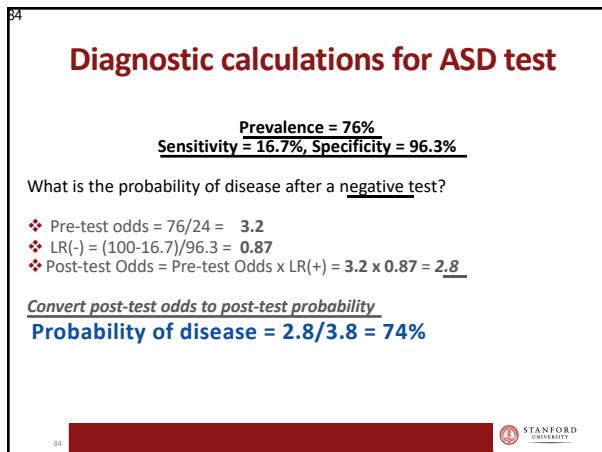
74



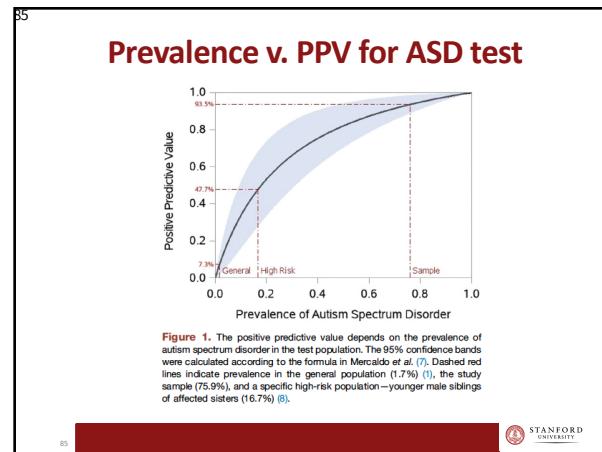
75



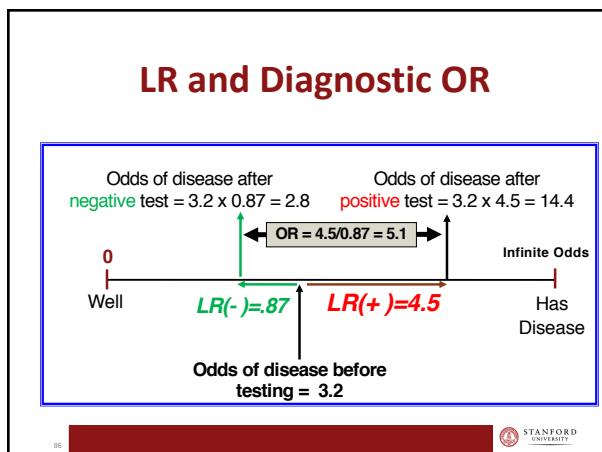
83



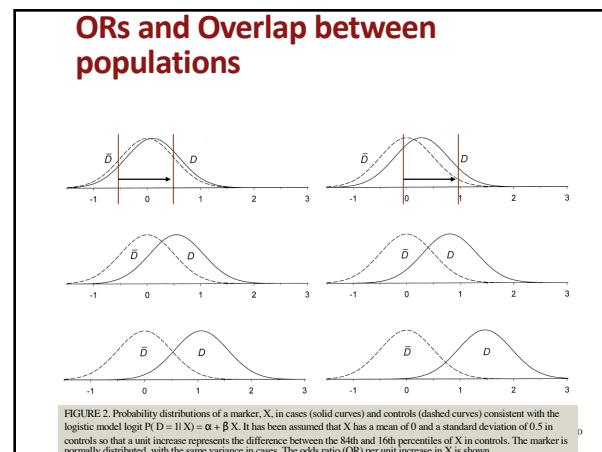
84



85



86



ROCs for constant OR

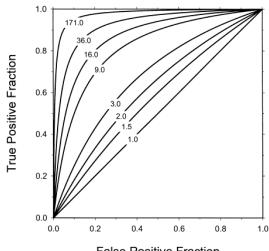


FIGURE 1. Correspondence between the true-positive fraction (TPF) and the false-positive fraction (FPF) of a binary marker and the odds ratio. Values of (TPF, FPF) that yield the same odds ratio are connected.

From Pepe, et. al, *Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker* "AJE, 2004



88

What is the best cutoff point?

Many are taught
the ROC point
closest to (1,0).

That is rarely the
right answer.

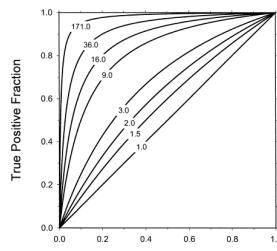


FIGURE 1. Correspondence between the true-positive fraction (TPF) and the false-positive fraction (FPF) of a binary marker and the odds ratio. Values of (TPF, FPF) that yield the same odds ratio are connected.



89

Choosing optimal cutpoint

- ✓ The optimal Dx test threshold is one that **is the point at which the test does more good than harm**.
- ✓ **That is not the point that minimizes the number of errors.**
- ✓ It is the one that achieves a target PV at that exact value of the test, defined by the relative harm of a FP vs. a FN.
- ✓ **Therefore, the cutoff cannot be determined w/o the prevalence and an estimate of relative harms, which are not on the ROC!**
 - The cost of a false positive (FP) = **Harm of (unnecessary) rx.**
 - The cost of a false negative (FN) = **{Treatment benefit}**

$$\text{Target PV Odds} = \frac{FP \text{ Harm}}{FN \text{ Harm}} = \frac{\text{Harm of unneeded tx or action}}{\text{Harm of missed dx}}$$

50



90

Decision curve

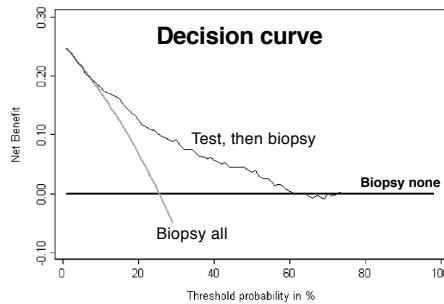


Figure 3. Decision curve analysis for free-to-total PSA ratio in men with elevated PSA. Gray line: biopsy all men. Thin black line: use free-to-total PSA ratio to determine who to biopsy. Thick black line: biopsy no man.

91

Assessing incremental value

51



93

Exercise Tomographic Thallium-201 Imaging in Patients with Severe Coronary Artery Disease and Normal Electrocardiograms -- Christian et al., Annals of Internal Medicine, December 1994, Volume 121 issue 11 | Pages 825-832

Annals of Internal Medicine
Established in 1927 by the American College of Physicians

Home | Current Issue | In the Clinic | ACP Journal Club | CME | Collections | Patient Information

Institution: Welch Med Libr JHU | Log In via User Name/Password

ARTICLE

Exercise Tomographic Thallium-201 Imaging in Patients with Severe Coronary Artery Disease and Normal Electrocardiograms

Timothy F. Christian; Todd D. Miller; Kent R. Bailey; and Raymond J. Gibbons

1 December 1994 | Volume 121 issue 11 | Pages 825-832

Objective: To assess the incremental value and cost-effectiveness of exercise tomographic thallium-201 imaging compared with clinical and exercise electrocardiographic variables for detecting three-vessel or left main coronary artery disease in patients with normal at-rest electrocardiograms.

Design: Prospective cohort study.

Participants: 411 patients (77 [19%] had three-vessel or left main disease) with normal at-rest electrocardiograms who underwent exercise tomographic thallium-201 studies and subsequently had coronary angiography.

Measurements: Clinical, exercise, and thallium-201 variables: Univariate followed by multivariate logistic regression.

Services

- Send comment/rapid response letter
- Notify a friend about this article
- Add to Personal Archive
- Download to Citation Manager
- ACP Search
- Get Permissions
- Google Scholar**
- Search for Related Content
- PubMed**
- Articles in PubMed by Author

STANFORD UNIVERSITY

94

Univariate Predictors of Three-Vessel or Left Main Coronary Artery Disease*

Variable	Left Main or Three-Vessel Disease		<i>P</i> Value
	Present (n = 77)	Absent (n = 334)	
Clinical			
Age, y	62 ± 8	59 ± 10	0.02
Male sex, n (%)	69 (90)	263 (79)	0.03
Diabetes, n (%)	9 (12)	20 (6)	0.03
Insulin-dependent	9 (12)	19 (6)	
Non-insulin-dependent	42 (55)	114 (34)	0.001
Typical angina, n (%)			
Exercise			
Chest pain during exercise, n (%)	36 (47)	99 (30)	<0.001
Peak heart rate, beats per minute	125 ± 21	135 ± 23	<0.001
Change in systolic blood pressure, mm Hg	23 ± 28	40 ± 26	0.005
Peak systolic blood pressure × peak heart rate	19 600 ± 5400	23 600 ± 6200	<0.001
Duration estimated metabolic equivalents	7.3 ± 1.9	8.0	0.005
Magnitude of ST depression, n (%)			
≥1 mm	58 (75)	166 (50)	<0.001
≥2 mm	36 (47)	71 (21)	<0.001
Thallium-201			
Global TI-201 score after exercise	43 ± 9	48 ± 8	<0.001
Global TI-201 score (delayed – after exercise)	10 ± 9	6 ± 6	<0.001
Abnormal segments (immediate), n	7 ± 4	5 ± 4	<0.001
Abnormal segments after delayed imaging, n	3 ± 2	2 ± 2	0.007
Abnormal segments (delayed – after exercise), n	4 ± 4	3 ± 3	0.001
Increased pulmonary uptake, n (%)	7 (9)	22 (7)	>0.2

* Means are expressed ± SD.

Christian, T. F. et al. Ann Intern Med 1994;121:825-832

Annals of Internal Medicine

95

Multivariate Analysis: Prediction of Three-Vessel or Left Main (Coronary Artery) Disease

Model	Direction	Odds Ratio (95% CI)	<i>P</i> Value
Clinical			
Diabetes mellitus	Present	2.0 (1.3 to 3.1)	0.001
Typical angina	Present	2.3 (1.4 to 3.9)	0.001
Sex	Male	3.1 (1.4 to 4.0)	0.007
Age*	Older	1.4 (1.1 to 1.9)	0.01
Chi-square = 31.3			
Clinical and exercise			
Diabetes mellitus	Present	1.9 (1.2 to 3.0)	0.005
Typical angina	Present	1.9 (1.1 to 3.3)	0.02
Sex	Male	2.3 (0.9 to 5.3)	0.07
Age*	Older	1.2 (0.9 to 1.7)	0.16
Magnitude of ST depression	More	1.5 (1.3 to 1.8)	<0.001
Peak heart rate × peak systolic blood pressure†	Lower	0.9 (0.86 to 0.95)	<0.001
Chi-square = 65.0			
Clinical, exercise, and thallium-201			
Diabetes mellitus	Present	1.9 (1.2 to 3.0)	0.004
Typical angina	Present	1.8 (1.1 to 3.2)	0.03
Sex	Male	2.2 (0.9 to 5.3)	0.07
Age*	Older	1.2 (0.9 to 1.7)	0.17
Peak heart rate × peak systolic blood pressure†	Lower	0.9 (0.4 to 0.95)	<0.001
Magnitude of ST depression	More	1.4 (1.2 to 1.7)	0.001
Global TI-201 score (delayed – after exercise)	Higher	1.1 (1.0 to 1.1)	0.02
Chi-square = 70.4			

* Increments of 10 years (each 10-year increase in age increases the odds of several disease 1.4-fold).

† Increments of 1000 units.

Christian, T. F. et al. Ann Intern Med 1994;121:825-832

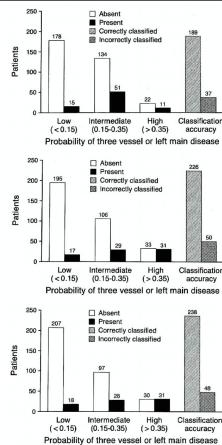
Annals of Internal Medicine

The anatomical results of patients classified as having a low, intermediate, or high probability of developing three-vessel or left main coronary artery disease by the use of multivariate models

Top: Clinical variables only (diabetes, history of typical angina, sex, and age).

Middle: Clinical and exercise variables (heart rate-blood pressure product and the magnitude of exercise ST-segment depression were added independently).

Bottom: Clinical, exercise, and thallium-201 variables (the change in global score was added independently).



97

Incremental Value

Change in % Correctly Classified

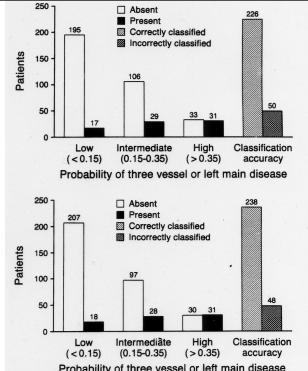
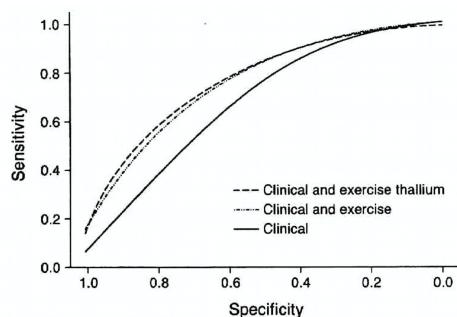


Figure 1. The anatomical results of patients classified as having a low, intermediate, or high probability of developing three-vessel or left main coronary artery disease by the use of multivariate models. Top: Clinical variables only (diabetes, history of

The receiver-operator characteristic curves of the three incremental multivariate models



Christian, T. F. et al. Ann Intern Med 1994;121:825-832

Annals of Internal Medicine

99

Results

We classified patients into probability groups (low, <0.15; intermediate, 0.15 to 0.35; high, >0.35) by each of the three logistic regression models (Figure 1). When the clinical model was used, 189 patients (46%) were correctly classified (low probability with absence of severe coronary disease or high probability with presence of severe coronary disease), and 37 patients (9%) were incorrectly classified The clinical and exercise model increased the number of patients correctly classified by 37 but at the expense of 13 additional patients incorrectly classified, for a net of 24 additional correct classifications (6% of the study group; SE = 3%). The clinical, exercise, and thallium-201 model led to 12 additional correct classifications and 2 fewer incorrect classifications, for a net increase of 14 correct classifications (3% of the study group; SE = 2%). All of these reclassified patients moved from the intermediate- to low-probability group.



100

Cost analysis

When we used only clinical variables, we correctly predicted the presence or absence of three-vessel or left main coronary artery disease in 189 patients (46%). When we used the Minnesota Medicare reimbursement fee for exercise treadmill testing of \$89, the cost per additional correct classification with exercise testing was \$1524 per patient correctly classified ($411 \times \$89$ per 24 additional correct classifications). ...When we added thallium scintigraphy to exercise and clinical data ...14 additional patients were correctly classified. When we used the Minnesota Medicare reimbursement for an exercise thallium-201 study of \$700 (includes treadmill charge of \$89), the cost was **\$20 550** per additional correct classification ($[411 \times \$700]/14$). The cost for thallium scintigraphy was considerably higher when we used the unbiased classifications from the cross-validation analysis: **\$143 880** per additional correct classification ($[411 \times \$700]/2$).

101



101

Questions I consider

- ✓ What is the situation are we trying to improve with this biomarker?
- ✓ How close is the experimental context to the real world, in terms of procedures, population, decisions and consequences?
- ✓ What is the specific decision/action that the biomarker will change, i.e. what happens in the absence of the biomarker, and in its presence?
- ✓ What is the evidence that that action will occur, and will improve net patient outcomes?
- ✓ What is the net cost to the care system and society of using the biomarker?

103



103

Goal of a Dx/Pred/Screening test?

To improve overall health outcomes
i.e. do more good than harm.

110



110

111



111