AN OVERVIEW OF CLINICAL EPIDEMIOLOGY: METHODOLOGIC AND ANALYTIC CONSIDERATIONS

OSCTR BERD SEMINAR

May 25th , 2018

Tabitha Garwe, PhD MPH

ASSISTANT PROFESSOR

DEPARTMENT OF BIOSTATISTICS AND EPIDEMIOLOGY &

DEPARTMENT OF SURGERY

SCTR Oklahoma Shared Clinical & Translational Resources

Oklahoma Shared Clinical and Translational Resources
http://osctr.ouhsc.edu
NIGMS award U54GM104938

OKLAHOMA Clinical and Translational Science Institute
Located at The University of Oklahoma Health Sciences Center

Seminar Outline

- Theoretical Background – ***Tabitha Garwe***
- Applied Example – **Courtney Montgomery**

SCTR Oklahoma Shared Clinical & Translational Resources

Clinical Epidemiology

- Epidemiology: Cornerstone for evidence based medicine (EBM)
- Clinical Epidemiology – not a different discipline but denotes ***application*** of epidemiologic methods to questions ***relevant to patient care***
- Traditionally, epidemiologic research largely devoted to ***etiologic*** research
- Clinical practice major concerns – adequate diagnosis, prognosis and therapy

OSCTR Oklahoma Shared Clinical & Translational Resources

Clinical Practice: Challenges

- Consider a patient consulting a physician; subsequent action depends on patient profile
  - Clinical profile (symptoms, signs and diagnostic test results)
  - Non-clinical profile (age, gender, socioeconomic status)

D.E.P.TH.

TABLE 1–1 Challenges of Daily Patient Care

Challenge	*Question*	*Needs*
Interpret the clinical profile: predict the presence of the illness	What illness best explains the symptoms and signs of the patient?	Diagnostic knowledge
Explanation of the illness	Why did this illness occur in this patient?	Etiologic knowledge
Predict the course of disease	1. What will the future bring for this patient, assuming no intervention takes place? 2. To what extent may the course of disease be affected by treatment?	Prognostic knowledge (including therapeutic knowledge)
Decision about medical action	Which treatment, if any, should be chosen for this particular patient?	Balancing benefits and risks of available options
Execution of medical action	Initiation of treatment	Skills

Grobbee: Clinical Epidemiology, 2nd Ed

Causal vs Descriptive (Prediction) Research

- Causal – aims to <u>explain</u> a relationship in etiologic terms
  - Does this factor cause the outcome?
  - Questions of treatment efficacy and safety
  - Extraneous determinants (confounders) need to be considered and taken into account in view of validity
- Descriptive – aims to <u>predict</u> rather than to explain
  - Includes diagnostic – determinants typically include elements of the clinical profile (signs, symptoms, test results); outcome is diagnosis of disease that fits the profile
  - And prognostic- determinants similarly include elements of the clinical profile; outcome is prognosis (survival, cure, recurrence)
- Causal and Descriptive – Intervention research
  - Aims to both <u>predict prognosis</u> following the intervention and understand the <u>effect caused by</u> the intervention
  - Typically the causal aspect drives the design (RCT)

Major Types of Clinical Epidemiologic Studies

TABLE 1–2 Major Types of Clinical Epidemiologic Research

Type of Research Question	*Descriptive/ Causal*	*Aim (Clinical Challenge)*	*Relevance*
Diagnostic research	Descriptive	To predict the probability of presence of target disease from clinical and nonclinical profile	Relevance for patient and physician to establish diagnosis and guide management
Etiologic research	Causal	To causally explain occurrence of target disease from determinant	Research relevance, may indicate means of prevention and causal intervention
Prognostic research	Descriptive	To predict the course of disease from clinical and nonclinical profile	Relevance for patient and physician to learn about the future and guide management
Intervention research	Causal and descriptive	1. To causally explain the course of disease as influenced by treatment 2. To predict the course of disease given treatment (options) and clinical and nonclinical profile	1. Relevance for research and drug development/ registration 2. Relevance for patient and physician to decide on optimal management

Grobbee: Clinical Epidemiology, 2nd Ed

PREDICTION RESEARCH

- DIAGNOSTIC
- PROGNOSTIC

8

Diagnostic Research

- Diagnostic Process : multivariable, sequential and probabilistic

- 1. A diagnosis starts with a patient presenting with **a complaint suggestive** of a certain disease to be diagnosed.
- 2. The subsequent work-up is a multivariable process. It involves **multiple diagnostic determinants (tests**) that are applied in a **logical order**
- 3. Setting or ruling out a diagnosis is a **probabilistic** action and the **probability is continuously updated** based on **subsequent** diagnostic test results.
- 4. The **true diagnostic value** of a test is determined by the extent to which it provides diagnostic information **beyond earlier tests**
- 5. The **goal** of the diagnostic process is to eventually **rule in or rule out the disease** with **enough confidence** to take clinical decisions.

Test-Treat Threshold Probabilities

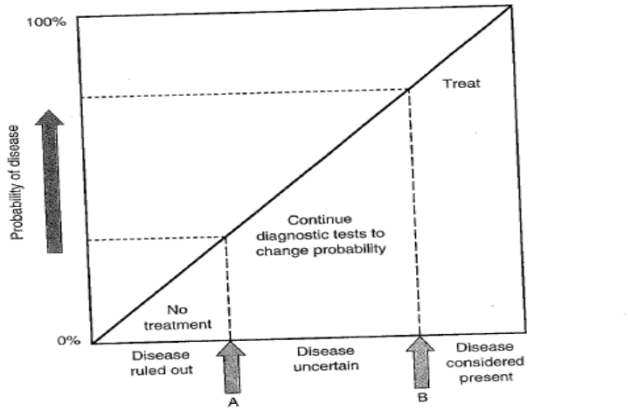
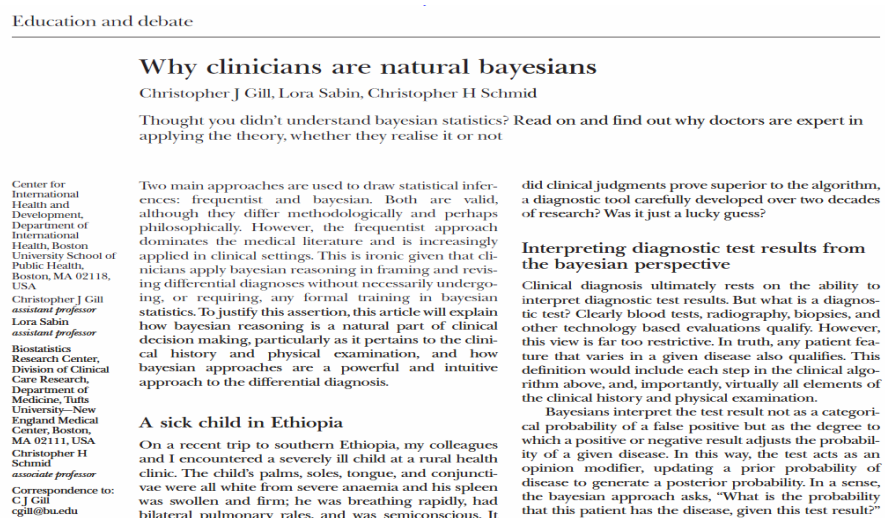


FIGURE 2–1 Diagnostic Testing.

Grobbee



Education and debate

Why clinicians are natural bayesians

Christopher J Gill, Lora Sabin, Christopher H Schmid

Thought you didn't understand bayesian statistics? Read on and find out why doctors are expert in applying the theory, whether they realise it or not

Center for International Health and Development, Department of International Health, Boston University School of Public Health, Boston, MA 02118, USA
Christopher J Gill *assistant professor*
Lora Sabin *assistant professor*
Biostatistics Research Center, Division of Clinical Care Research, Department of Medicine, Tufts University—New England Medical Center, Boston, MA 02111, USA
Christopher H Schmid *associate professor*
Correspondence to: C J Gill cgill@bu.edu

Two main approaches are used to draw statistical inferences: frequentist and bayesian. Both are valid, although they differ methodologically and perhaps philosophically. However, the frequentist approach dominates the medical literature and is increasingly applied in clinical settings. This is ironic given that clinicians apply bayesian reasoning in framing and revising differential diagnoses without necessarily undergoing, or requiring, any formal training in bayesian statistics. To justify this assertion, this article will explain how bayesian reasoning is a natural part of clinical decision making, particularly as it pertains to the clinical history and physical examination, and how bayesian approaches are a powerful and intuitive approach to the differential diagnosis.

A sick child in Ethiopia

On a recent trip to southern Ethiopia, my colleagues and I encountered a severely ill child at a rural health clinic. The child's palms, soles, tongue, and conjunctivae were all white from severe anaemia and his spleen was swollen and firm; he was breathing rapidly, had bilateral pulmonary rales, and was semiconscious. It did clinical judgments prove superior to the algorithm, a diagnostic tool carefully developed over two decades of research? Was it just a lucky guess?

Interpreting diagnostic test results from the bayesian perspective

Clinical diagnosis ultimately rests on the ability to interpret diagnostic test results. But what is a diagnostic test? Clearly blood tests, radiography, biopsies, and other technology based evaluations qualify. However, this view is far too restrictive. In truth, any patient feature that varies in a given disease also qualifies. This definition would include each step in the clinical algorithm above, and, importantly, virtually all elements of the clinical history and physical examination.

Bayesians interpret the test result not as a categorical probability of a false positive but as the degree to which a positive or negative result adjusts the probability of a given disease. In this way, the test acts as an opinion modifier, updating a prior probability of disease to generate a posterior probability. In a sense, the bayesian approach asks, "What is the probability that this patient has the disease, given this test result?"

BMJ 2005;330:1080–3

Evaluating Diagnostic Tests: What do we need to know?

Reliability: repetitions of the test give the same result

Accuracy: test gives the right answer

Usefulness: right answer improves outcome by favorably affecting decisions

Value: expected improvement in health outcomes justifies the risks and costs

12

Reliability Studies

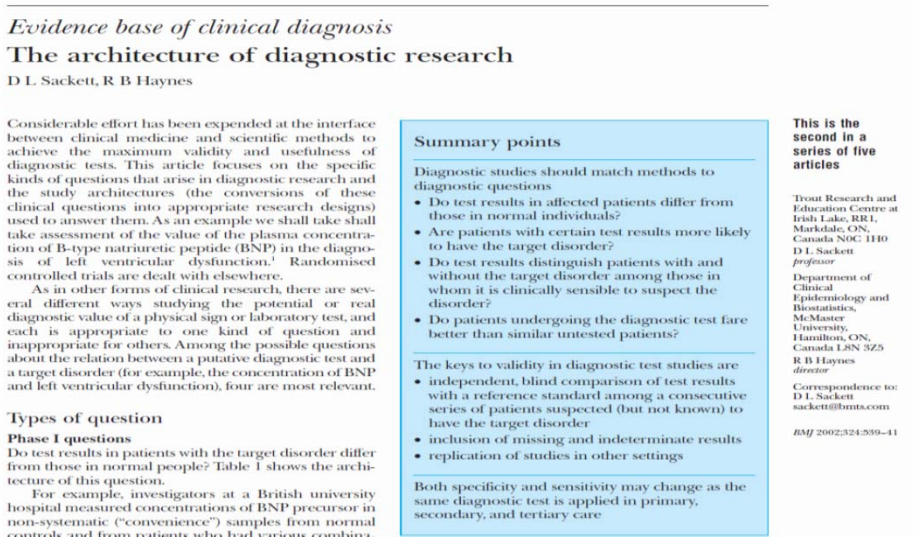
- Inter- and intra-observer agreement among categorical observations
  - Kappa coefficient
- Inter- and intra-observer reliability for continuous measures
  - Within-subject standard deviation (S_w)
    (also called standard error of measurement or SEM)
  - Smallest real difference
  - ICC: Intraclass correlation coefficient
- **Agreement between methods (Bland Altman plots)**
  - Agreement (or bias) between paired continuous measurements
  - Limits of agreement
  - Also applicable to between-rater comparisons

13

Diagnostic Accuracy Studies

- Diagnostic research → improve diagnostic process
  - Occurrence relation → P(D)=f(T1,T2,T3,..Tn)
    - ID combination(s) of tests that have the largest diagnostic yield
  - Does new test provides additional diagnostic value in clinical practice?
  - Is a less burdensome/inexpensive test an alternative?
- Descriptive in nature
  - Excludes diagnostic intervention studies
  - Should be performed in close adherence to daily clinical practice

Phase I – IV Diagnostic Accuracy Studies



Evidence base of clinical diagnosis

The architecture of diagnostic research

D L Sackett, R B Haynes

Considerable effort has been expended at the interface between clinical medicine and scientific methods to achieve the maximum validity and usefulness of diagnostic tests. This article focuses on the specific kinds of questions that arise in diagnostic research and the study architectures (the conversions of these clinical questions into appropriate research designs) used to answer them. As an example we shall take shall take assessment of the value of the plasma concentration of B-type natriuretic peptide (BNP) in the diagnosis of left ventricular dysfunction.[1] Randomised controlled trials are dealt with elsewhere.

As in other forms of clinical research, there are several different ways studying the potential or real diagnostic value of a physical sign or laboratory test, and each is appropriate to one kind of question and inappropriate for others. Among the possible questions about the relation between a putative diagnostic test and a target disorder (for example, the concentration of BNP and left ventricular dysfunction), four are most relevant.

Types of question

Phase I questions

Do test results in patients with the target disorder differ from those in normal people? Table 1 shows the architecture of this question.

For example, investigators at a British university hospital measured concentrations of BNP precursor in non-systematic ("convenience") samples from normal

Summary points

Diagnostic studies should match methods to diagnostic questions

- Do test results in affected patients differ from those in normal individuals?
- Are patients with certain test results more likely to have the target disorder?
- Do test results distinguish patients with and without the target disorder among those in whom it is clinically sensible to suspect the disorder?
- Do patients undergoing the diagnostic test fare better than similar untested patients?

The keys to validity in diagnostic test studies are

- independent, blind comparison of test results with a reference standard among a consecutive series of patients suspected (but not known) to have the target disorder
- inclusion of missing and indeterminate results
- replication of studies in other settings

Both specificity and sensitivity may change as the same diagnostic test is applied in primary, secondary, and tertiary care

This is the second in a series of five articles

Trout Research and Education Centre at Irish Lake, RR1, Markdale, ON, Canada N0C 1H0
D L Sackett
professor

Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada L8N 3Z5
R B Haynes
director

Correspondence to:
D L Sackett
sackett@bmts.com

BMJ 2002;324:539–41

Diagnostic vs Test Research

- Test research – Assess whether a ***single*** diagnostic test adequately can show presence or absence of a particular disease – **often case-control studies**
  - E.g. NT-proBNP in the dx of heart failure
  - Deviates from main principle of *clinically relevant* diagnostic research
  - Diagnostic process involves multiple tests
  - Relevant patient domain often not included – pts presenting with signs and symptoms **suggestive** of target disease
- Test research – relevant and helpful in 1) the developmental phase of a new diagnostic test 2)screening in asymptomatic

Diagnostic Accuracy Studies: Study Designs

- Diagnostic process – cross-sectional by definition (presence/absence of Dz estimated at t=0)
- Cross-sectional study design – most common
- Sometimes characterized as cohort study (t>0)
- Diagnostic case-control study aka ***cross-sectional case-control study***
  - all patients **suspected** cases and a sample of suspected controls
- Diagnostic Intervention Studies – RCTs, longitudinal

Diagnostic Studies: Study Population

- Diagnostic test – goal is to distinguish between those with target disease and those without **in patients <u>suspected</u>** of having a particular disease; Ideally should **exclude**
  - those in whom disease state has already been established
  - High/low disease probability to take action/no action
- Restrict study population to a level of care or setting
- **Consecutive patients**, exclusion criteria should be few
- Signs and symptoms accompanying the disease defines patients 'suspected' of having a particular disorder

Diagnostic Studies: Diagnostic Determinants

- Diagnosis in practice typically made on the basis of multiple dx determinants
- All tests (potentially) used should be considered
  - Logistics and larger sample size required limit the # of tests that can be included
- Assessment of dx determinants should resemble quality in daily practice

Diagnostic Research: Outcomes

- Typically dichotomous
- Gold standard (GS) –used to define disease state
  - In reality, no perfect test exists -> GS=Reference standard (RS)
  - RS-typically the **best** procedure that exists at the time of study initiation - most expensive or invasive
  - Contrasts with the assessment of the diagnostic determinants of interest
  - No single test can constitute the RS → composite reference standard
- Ideally final dx should be established independent of results of dx test under study
  - blinding not guaranteed →***incorporation bias*** → typically leads to overestimation
  - blinding depends on type of reference standard (RS) applied
  - separate RS may not be available, may be infeasible or unethical to apply in all cases
  - **Partial or differential verification bias**

Diagnostic Research: Quantifying Diagnostic Test Accuracy (validity)

- Sensitivity and specificity
- Predictive values
- Likelihood ratios
- Receiver Operating Characteristic (ROC) curve
- Diagnostic Odds Ratio

Pre-test (Prior) & Post-test (Posterior) probability

- Goal of determining a diagnosis for patients is to estimate the probability of disease given the diagnostic test results
- **Pre-test (Prior) probability**
  - Pretest probability is the more general term
  - For screening tests, pretest probability = **prevalence**
  - For diagnostic tests, *pretest probability* ***incorporates history and physical exam items***
- **Post-test probability (Posterior) vs. Predictive value**
  - **Posttest** probability **after a positive** test is the same as **positive predictive value**
  - **Posttest** probability **after a negative (for a diseased person)** test is **1– negative predictive value**

22

Pretest (prior) and post-test (posterior) probability

		Disease status		
		Has disease	No disease	Total
Test Result	Positive	A	B	A + B
	Negative	C	D	C + D
	Total	A + C	B + D	A + B + C + D

Pretest (prior) probability = (A+C)/(A+B+C+D)=**Prevalence**

Posttest probability = A/(A+B) = **Predictive Value Positive (PV+)** or
C/(C+D) = 1- **Predictive Value Negative (PV-)**

23

"Case-control" sampling

		Disease status		
		Has disease	No disease	Total
Test Result	Positive	A	B	A + B
	Negative	C	D	C + D
	Total	A + C	B + D	A + B + C + D

~~PPV=A/(A+B)~~
~~NPV=D/(C+D)~~

24

Diagnostic Odds Ratio (DOR)

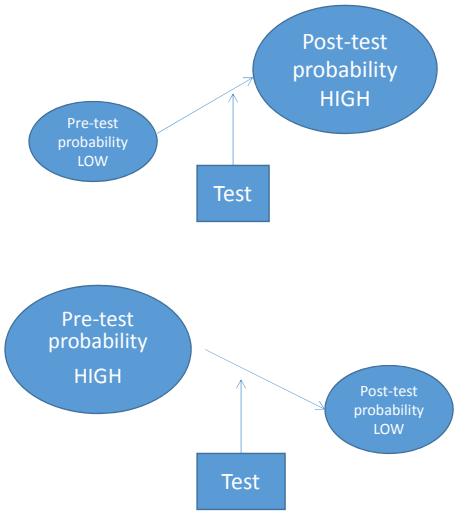
	Disease present	Disease absent
Test positive	True positives (a)	False positives (b)
Test negative	False negative (c)	True negatives (d)

Odds of positive test result in persons with the target condition compared to those without the target condition

DOR = (a/c) / (b/d)
DOR = ad / bc
DOR = Odds of T+|D+ / Odds of T+|D-

Bayesian approach to diagnosis

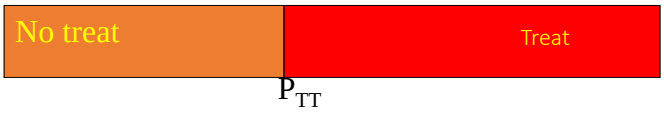
- An accurate test will help reduce uncertainty
- The pre-test probability is revised using test result to get the post-test probability
- Tests that produce the **biggest changes from pretest to post-test probabilities are most useful in clinical practice** [very large or very small likelihood ratios]



Likelihood Ratios [LRs]: AKA 'Bayes Factor'

- Factor by which **odds of disease** either increase or decrease as a result of the test
- LR= $\frac{\text{P(Result) in patients WITH disease}}{\text{P(Result) in patients WITHOUT disease}}$ (can calculate +ve and –ve LRs)
  - PPVs and NPVs can also be used to get **posterior probabilities** (PostTP) from sens, spec, prior probability (prevalence) and test results
  - However, LRs have advantages over predictive values
    - Less likely to change with the prevalence of the disorder
    - can be calculated for several levels of the symptom/sign or test
    - Can be used to combine the results of multiple diagnostic tests
    - Can still be used to calculate the post-test probability for a target disorder.

Treatment Thresholds



- Evaluating diagnostic tests
  - Reproducibility
  - Accuracy
  - **Usefulness: Will it change management?**
  - **Value: Is it worth its risks and costs?**
- The last two may require us to estimate a ***Treatment Threshold***

28

Test Usefulness

- 2 main factors that limit the usefulness of tests
  - 1. They sometimes give wrong answers **(imperfect)**
  - 2. They **have a "cost,"** which includes the financial cost as well as the risks, discomfort, and complications that arise from testing

Quantifying Costs and Benefits

- To calculate the **range of prior probabilities** for which the **expected benefits justify testing**, we need to quantify **three things**:
  - 1) How bad it is to treat someone who does not have the disease? (C)
  - 2) How bad it is not to treat someone who does have the disease? (B)
  - 3) What is the cost of the test? (T)

30

Treatment Threshold Probability (PTT)

- First introduced by Pauker and Kassirer in 1975
  - It is the (posterior) probability of disease at which the **expected costs** of the two types of mistakes we can make (treating people without the disease **(C)** and not treating people with the disease **[B]**) **are balanced**.
  - Expected cost = multiply the cost of these mistakes (C and B) by their probability of occurring.
  - The expected cost **of not treating** is P (the probability of disease) x B = **PB**
  - The expected cost **of treating** is P (the probability of NO disease) x C= **(1 - P) x C = (C- C x P)**

Treatment Threshold

- P_{TT} is the probability of disease at which:

$$P_{TT} \times B = (1 - P_{TT}) \times C$$

And therefore, the treatment threshold odds are given by:

$$\frac{P_{TT}}{(1 - P_{TT})} = \frac{C}{B}$$

and the threshold probability is

$$P_{TT} = \frac{C}{(C + B)}$$

- E.g. treating someone who does not have the disease is half as bad as failing to treat someone who does have the disease – should be willing to treat 2 people without disease to avoid failing to treat one person who has it
  - C=1/2B ; B=2xC; P_{TT} = C/(C + 2C) = C/3C= 1/3 = 0.33

Test/Treat Thresholds: Dichotomous Tests

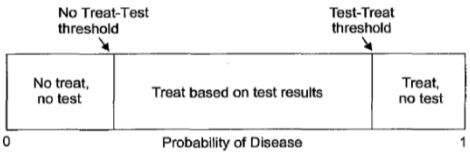


Figure 3.3 The no treat–test and test–treat probability thresholds, between which the test can affect treatment decisions.

- If we know the treatment threshold (P_{TT}), we can use LRs to get testing threshold(zone)

33

Example

Should Malaria Treatment Be Guided by a Point of Care Rapid Test? A Threshold Approach to Malaria Management in Rural Burkina Faso

Zeno Bisoffi[1]*, Halidou Tinto[2], Bienvenu Sodiomon Sirima[3], Federico Gobbi[1], Andrea Angheben[1], Dora Buonfrate[1], Jef Van den Ende[4]

1 Centre for Tropical Diseases, S. Cuore Hospital, Negrar, Verona, Italy, **2** Centre Muraz, Bobo Dioulasso, Burkina Faso, **3** Centre National de Recherche et de Formation sur le Paludisme, Ministry of Health, Ouagadougou, Burkina Faso, **4** Department of Clinical Sciences, Prince Leopold Institute of Tropical Medicine, Antwerp, Belgium

Abstract

Background: In Burkina Faso, rapid diagnostic tests for malaria have been made recently available. Previously, malaria was managed clinically. This study aims at assessing which is the best management option of a febrile patient in a hyperendemic setting. Three alternatives are: treating presumptively, testing, or refraining from both test and treatment. The test threshold is the tradeoff between refraining and testing, the test-treatment threshold is the tradeoff between testing and treating. Only if the disease probability lies between the two should the test be used.

Methods and Findings: Data for this analysis was obtained from previous studies on malaria rapid tests, involving 5220 patients. The thresholds were calculated, based on disease risk, treatment risk and cost, test accuracy and cost. The thresholds were then matched against the disease probability. For a febrile child under 5 in the dry season, the pre-test probability of clinical malaria (3.2%), was just above the test/treatment threshold. In the rainy season, that probability was 63%, largely above the test/treatment threshold. For febrile children >5 years and adults in the dry season, the probability was 1.7%, below the test threshold, while in the rainy season it was higher (25.1%), and situated between the two thresholds (3% and 60.9%), only if costs were not considered. If they were, neither testing nor treating with artemisinin combination treatments (ACT) would be recommended.

Conclusions: A febrile child under 5 should be treated presumptively. In the dry season, the probability of clinical malaria in adults is so low, that neither testing nor treating with any regimen should be recommended. In the rainy season, if costs are considered, a febrile adult should not be tested, nor treated with ACT, but a possible alternative would be a presumptive treatment with amodiaquine plus sulfadoxine-pyrimethamine. If costs were not considered, testing would be recommended.

Citation: Bisoffi Z, Tinto H, Sirima BS, Gobbi F, Angheben A, et al. (2013) Should Malaria Treatment Be Guided by a Point of Care Rapid Test? A Threshold Approach to Malaria Management in Rural Burkina Faso. PLoS ONE 8(3): e58019. doi:10.1371/journal.pone.0058019

Editor: Clive Shiff, Johns Hopkins University, United States of America

Received November 19, 2012; **Accepted** January 29, 2013; **Published** March 5, 2013

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: zeno.bisoffi@sacrocuore.it

Common Biases: What is the right population for a diagnostic accuracy study?

- Those in whom we are **uncertain** of the diagnosis
- Those in whom we will use the test in clinical practice to resolve our uncertainty
- Patients with the disease who suffer from a **wide spectrum** of severity and patients without the disease who have other conditions that are commonly confused with the target disease

Common Biases: Studies of diagnostic test accuracy

- **Incorporation bias (includes review bias)**
  - Classification of disease status partly depends on the results of the index test. The gold standard incorporates the index test. If the gold standard is expert clinical review, this includes failure to blind the expert(s) to the results of the index test → **Sn & Sp falsely** ⬆
- **Partial verification bias (aka verification , referral, ascertainment or work-up)**
  - Patients with positive index tests are more likely to get the gold standard, and only patients who get the gold standard are included in the study → **Sn falsely** ⬆; **Sp falsely** ⬇
- **Differential verification bias (double gold standard bias)**
  - Patients with a positive index test are more likely to receive an immediate, invasive gold standard, whereas patients with a negative index test are more likely to receive clinical follow-up for development of disease → dz resolves spontaneously - **Sn & Sp falsely** ⬆**;** dz only detectable during follow-up period - **Sn & Sp falsely** ⬇
- **Spectrum bias:** Spectrum of disease and non-disease differs from clinical practice. Sn depends on spectrum of diseased and Sp depends on spectrum of non-disease → disease skewed towards 'sickest of the sick' - **Sn falsely** ⬆; non-disease skewed toward 'Wellest of the well' – **Sp falsely** ⬆

Academic Emergency Medicine
Official Journal of the Society for Academic Emergency Medicine

RESEARCH METHODS & STATISTICS

Understanding the Direction of Bias in Studies of Diagnostic Test Accuracy

Michael A. Kohn, MD, MPP, Christopher R. Carpenter, MD, MSc, and Thomas B. Newman, MD, MPH

Abstract

Ordering and interpreting diagnostic tests is a critical part of emergency medicine (EM). In evaluating a study of diagnostic test accuracy, emergency physicians (EPs) need to recognize whether the study uses case–control or cross-sectional sampling and account for common biases. The authors group biases in studies of test accuracy into five categories: incorporation bias, partial verification bias, differential verification bias, imperfect gold standard bias, and spectrum bias. Other named biases are either equivalent to these biases or subtypes within these broader categories. The authors go beyond identifying a bias and predict the direction of its effect on sensitivity and specificity, providing numerical examples from published test accuracy studies. Understanding the direction of a bias may permit useful inferences from even a flawed study of test accuracy.

ACADEMIC EMERGENCY MEDICINE 2013; 20:1194–1206 © 2013 by the Society for Academic Emergency Medicine

Helpful Checklists: STARD & QUADAS

- STARD

Section & Topic	No	Item
TITLE OR ABSTRACT		
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the index test
	4	Study objectives and hypotheses
METHODS		
Study design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants	6	Eligibility criteria
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)

- **The QUADAS tool**

Item		Yes	No	Unclear
1.	Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	( )	( )
2.	Were selection criteria clearly described?	( )	( )	( )
3.	Is the reference standard likely to correctly classify the target condition?	( )	( )	( )
4.	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	( )	( )	( )

PROGNOSTIC RESEARCH: The Motive and Aim of Prognosis

- Many patients expect a statement from their doctor about their prognosis.
  - A *prognosis* refer to all elements of future health
- Prognosis guides subsequent medical actions
  - Many treatments tend to become more cost-effective as the prognosis worsens.
  - Prognostication - "What is the predicted course of the disease in this patient if I do not intervene?"

Prognostication: A Multivariable Process

- The aim of prognostication - individual risk prediction
  - Average prognosis – imprecise and clinically of limited value e.g.
    - The prognosis of pancreatic cancer is poor; 5-year survival in osteosarcoma approximates 40%
    - Results should be expressed as absolute risks
  - Typically the prognosis of an individual is determined by a variety of patient characteristics
  - Risk profile - combination of prognostic determinants
    - nonclinical characteristics such as age and gender, and clinical characteristics such as the diagnosis, symptoms, signs, dx tests

Prognostic Research: Design Issues and Conduct

- Research objectives
  - 1) **Which combination** of determinants under study **best predicts** the future outcome.
  - 2) **Additional predictive value beyond** other available predictors.
    - A logical hierarchy of prognostic determinants exists based on everyday practice.
    - Cumbersome or costly prognostic markers (e.g., blood tests and imaging), ideally, should only be used if they have added predictive value
  - 3) May include **comparison of the predictive accuracy** of two (new) markers.

Prognostic Research: Study Designs

- Occurrence relation - Incidence O = f(d1, d2, d3, dn)
- **Study Population:** Domain of a prognostic occurrence relation includes individuals who are **at risk of developing** the outcome of interest
  - usually defined by the presence of a particular condition
  - pts with a 0 or 100% probability of developing the outcome not part of the domain
- Study design most suitable to address prognostic questions is a **cohort study**, preferably a prospective one
  - consecutive patients at risk for developing the outcome
  - Restrict study population to the setting of care
- Sometimes a **case-control** design is used, usually for efficiency reasons
  - Does not allow for an estimation of absolute risks of an outcome unless sampling fraction of controls is known
  - **Case-cohort** design increasingly being used

Experimental or Observational

- Almost all prognostic studies **outside** the realm of ***intervention research*** are observational (cohort)
- Randomized trials can serve as a vehicle for prognostic research
  - Prognostic determinants of interest are just observed and not influenced by the researcher.
  - A prognostic study within a trial bears a greater resemblance to an observational study
    - prognostic study within the **reference** - estimate the prognosis in a patient with a certain condition if no intervention is initiated
    - Prognostic analysis within the **treated** (intervention) - facilitate quantification of the expected course (in terms of absolute risks) in an individual patient following treatment

Prognostic Determinants (Predictors)

- Predictors should preferably be measured using methods applicable-or potentially applicable-to daily practice
- Prognostic determinants - history taking, physical examination, blood tests, imaging, and other test results; may include treatments (current or past)
- Feasibility plays an important role in choosing determinants
- Potential predictors should be **measured and analyzed** with a view to **chronological hierarchy** in practice

Prognostic Outcomes

- Typically dichotomous but may be continuous
- Generally should not study intermediate outcomes
- **Time period of outcome occurrence important**
  - Prediction over a shorter period is commonly less problematic than prediction over a longer time period.
  - Follow-up time may differ- use Kaplan Meier or Cox regression
- Blinding important
  - Less important for hard outcomes like mortality

Bias in Prognostic Research

- Confounding is not an issue in prognostic research, as in all types of prediction research
  - no central determinant for which the relationship to the outcome should be causally isolated from other outcome predictors
- Loss to follow-up
  - Bias due to indeterminates, missing data
    - Multiple imputation
    - Worst-case sensitivity analysis
- Ascertainment bias (Diagnostic Review bias)
- **Overfitting** - internal and external validation

Prognostic vs Diagnostic Research

- Differences
  - Prognostic research (PR) inherently **longitudinal** whereas dx research is **cross-sectional**
  - PR often deals with **continuous** outcomes, such as measures of pain or quality of life, and **multiple** outcomes
  - Prognostic predictions are generally **less accurate** than diagnostic predictions
- Shared characteristics
  - The purely **predictive** aim of prognostic research is shared with diagnostic research and has major implications for the design, conduct, and reporting of research.
    - Both inherently multivariable
    - There is no central factor or determinant whose causal effect must be isolated from the effects of other variables – **confounding not an issue**
    - Study should be performed in and mimic routine clinical practice

Combining tests/predictors and multivariable decision rules

- Diagnostic Research
  - Logistic Regression
  - Recursive Partitioning Analysis(RPA aka CART)
  - Neural Networks[NN]
  - Logistic regression generally accepted statistical method for MV diagnostic studies with a **dichotomous outcome**
  - RPA and NN – criticized for overly optimistic results
- Prognostic Research
  - Logistic regression – dichotomous outcome
  - Cox regression – time to event data
  - Linear regression – continuous outcomes
  - Recursive partitioning analysis
- Added value of test/predictor – ideally easily obtainable tests should be estimated first before costly, burdensome

48

Multivariable Data Analysis in Prediction Research

- Missing data – Impute data to reduce bias
- Model Performance Measures
  - **Calibration** - how well the probability estimated from the test result matches the actual probability
    - Graph of observed versus expected probabilities
    - Goodness of Fit Tests (e.g. Hosmer and Lemeshow)
    - Bland-Altman Calibration Plots with Mean Bias and SD of Errors
  - **Discrimination** - how well the test differentiates between patients more and less likely to have the outcome
    - A commonly used approach to quantifying the discrimination of a prognostic test is the **Area Under the ROC Curve (AUROC)**

Area Under the Receiver Operating Characteristic Curve (AUROC)

- The ROC curve illustrates the tradeoffs between cut points that maximize sensitivity and specificity
  - A plot of the FP probability on the x-axis and the TP probability on the y-axis across several thresholds of a continuous value
  - the probability that, confronted with a pair of randomly chosen patients, one of whom truly has the disease of interest and the other of whom truly does not, the test will accurately identify which of the pair has the disease.
- Each point on the curve represents a Se/Sp pair corresponding to a particular cut-off (decision threshold or criterion value)
- The ROC method - **overall measure** of diagnostic/prognostic performance
  - Can be used to compare the diagnostic/prognostic performance of two or more tests/factors.

Area Under the ROC Curve (AUC)

- Area Under the ROC Curve is **non-parametric**
  - AUC not significantly affected by shapes of underlying populations
  - ROC curve depends only on the ranking of individual measurements (in this case, risk estimates) and not their absolute values.

- Non-informative – AUC=0.5
- Less accurate – 0.5 < AUC < 0.7
- **Moderately accurate – 0.7 < AUC < 0.9**
- **Highly accurate – 0.9 < AUC < 1**
- **Perfect test – AUC=1**

Reclassification Measures

- Discrimination - AUC is the most popular metric but requires **very large 'independent'** associations
- **Net Reclassification Improvement (NRI)**
  - Quantifies the number of individuals that are correctly reclassified into clinically meaningful higher or lower risk categories with the addition of a new predictor, using pre-specified risk groups [Pencina et al., 2008].
- **Integrated Discrimination Improvement (IDI)**
  - In contrast to the NRI, the IDI does not require subjectively predefined risk thresholds.
  - It is the estimated improvement in the average sensitivity of the basic model with addition of the new predictor minus the estimated decrease in the mean specificity, summarized over all possible risk thresholds.
- Newer, statistical methods not yet well developed
  - Careful application is necessary

Multivariable Data Analysis in Prediction Research

- Adequate diagnosis and prognostication requires knowledge about the occurrence of current and future outcomes given ***combinations*** of test/ predictors.
  - Requires studies that follow a **multivariable approach** in **design and analysis** and results in **outcome probabilities** and **predictive tools**
  - Results should be expressed as absolute risks
  - ***Clinical prediction models, predictions rules, prognostic indices, or risk scores***
    - Explicitly transform combinations of values of prognostic determinants documented in an individual patient to an absolute probability of developing the outcome in the future e.g. APACHE score [Knaus et al., 1991]; SAPS [Le Gallet al., 1993]

Multivariable Data Analysis in Prediction Research: Risk Scores

- Simplified risk score (SRS)
  - Commonly done by dividing each regression coefficient by the smallest regression coefficient
  - Must be accompanied by the observed disease frequencies across score categories
  - Some loss in dx accuracy, but minimal and easy to use
  - Grobbee & Hoes suggest reporting both the untransformed model and SRS with AUROCs

54

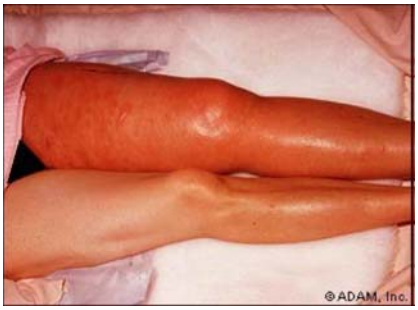
Garwe et al. The Geriatric Field Trauma Triage Risk Score

Table 4 Predictors of serious injury (ISS>=16)

Variable	Serious Injury Odds Ratio (95% CI)	GFTT Risk Score
Male Gender	1.38 (1.25-1.51)	2
Penetrating Injury	4.40 (3.30-5.87)	9
Physiologic Criteria (Initial Scene)		
GCS <=13	2.28 (1.99-2.61)	5
RR <=10 or RR >=24	1.59 (1.32-1.90)	3
SBP < 100	2.08 (1.79 – 2.42)	4
Intubation (or Ventilatory Support)	2.92 (2.32 – 3.69)	6
Anatomic Injury (Physical Exam)		
Skull fracture/intracranial	11.19 (10.08-12.42)	14
Traffic-related Rib fracture	6.22 (5.39-7.17)	10
Traffic-related long bone fracture	2.33 (1.87 – 2.89)	5
Pelvic fracture	2.69 (2.27 – 3.20)	6
Pre-existing Comorbidity		
Cardiac disease	1.19 (1.07-1.32)	1
Coagulopathy	1.74 (1.33-2.28)	3

Hosmer-Lemeshow Goodness-of-Fit, p = 0.059; AUC (95% CI) = 0.8633 (0.8566-0.8699), Bias-corrected AUC = 0.8527

Does D-dimer add value to diagnosing DVT in Primary Care?



56

AUROC Comparison

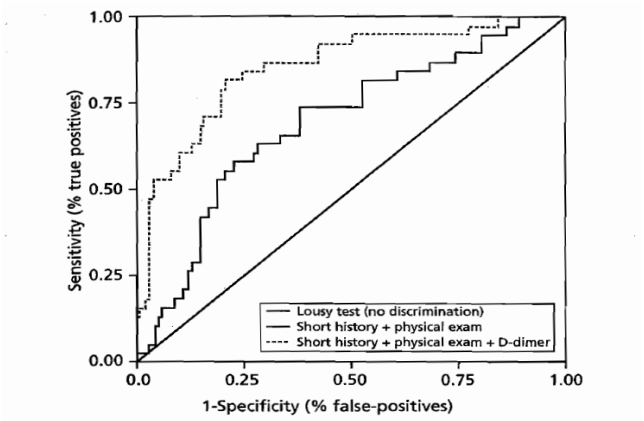


FIGURE 3.3 Example of an ROC curve of the reduced multivariable logistic regression model, including the same six determinants as in Figure 3.2. The ROC area of the "reduced history + physical model" (red) was 0.70 (95% confidence interval [CI], 0.66–0.74) and of the same model added with the D-dimer assay (green) 0.84 (95% CI, 0.80–0.88).

Oudega et al.

57

Diagnostic Simplified Risk Score

*1*male gender + 1*OC use + 1*presence of malignancy + 1*recent surgery + 1*absence of trauma + 1*vein distension + 2*calf difference ≥ 3cm +6*abnormal D-dimer test.*

Table 4: Prevalence of DVT across four score (risk) categories.

Probability or risk Category	number of patients n (%)[1]	DVT present n (%)[2]	DVT absent n (%)[3]
Very low (0–3)	293 (23)	2 (0.7)	291 (99.3)
Low (4–5)	66 (5)	3 (4.5)	63 (95.5)
Moderate (7–9)	663 (51)	144 (21.7)	519 (78.3)
High (10–13)	273 (21)	140 (51.3)	133 (48.7)

1=proportion of all (1295) patients; 2=proportion of presence of DVT within risk category; 3=proportion of absence of DVT within risk category.

58

Recursive Partitioning Analysis

- Same as Classification and Regression Trees (CART)
- **Creates a decision tree**
  - Aim: correctly classify members of the population by splitting it into sub-populations
- Termed **recursive** because each sub-population may in turn be split an indefinite number of times until the splitting process terminates after a particular stopping criterion is reached.
- Explore relation among variables without having a prior model
- **Results in rules or algorithms, not scores**
  - **creates a rule such as 'If a patient has finding a, b, or c, they have an XX probability of disease Y'**

59

Garwe et al. Table 4 Recursive Partitioning Analysis Risk stratification based on the GFTT Score for Injured Older Adults

Outcome Risk Category	Risk score range	n	Outcome Incidence
Serious Injury (ISS >=16)			
Low	< 9	7452	6.4%
Medium	9 -22	4509	49.3%
High	**>=23**	**1314**	**82.5%**

60

Overfitting Bias in Prediction Research

- 'If you torture data sufficiently, it will confess to almost anything' – Fred Menger (Newman, EBD)
- Overfitting - Mainly a problem when a combination of tests/predictors is chosen from many candidate tests to identify a disease or predict a prognosis
  - What might look like a good prediction rule in one sample might perform poorly when applied to other external populations
- **Minimizing overfitting** – internal and external validation

Internal Validation and Shrinkage of the Diagnostic/Prognostic Model

- Initial prediction model applied to the data from which it is derived - usually ***overfitted***
- Amount of overfitting can be estimated and corrected using internal validation methods
  - Split sample and x-validation – one sample used for development and remainder for estimating accuracy
  - **Bootstrapping** – first model is developed on full sample, then multiple random samples are drawn from the full sample
    - Average optimism in discrimination and calibration can be used to adjust original model (i.e. shrink the model)
  - Heuristic shrinkage factor; penalized estimation methods

Internal Validation and Shrinkage of the Prognostic Model

- In general, shrinkage of regression coefficients may improve the performance of a prognostic model substantially.
- When the sample size is large, no shrinkage will be required, in contrast to a small data set, where substantial shrinkage may be needed (Steyerberg & Harrell,2000).
  - If the number of predictors over the number of observations (of least occurring outcome) is less than 1/10, shrinkage is necessary, if this ratio is between 1/10 and 1/20, shrinkage is advisable, and if the ratio is smaller than 1/20, shrinkage is not necessary.

External Validation

- Application and testing of the model in new patients.
  - Generally necessary before a model can be used in practice with confidence
  - Can be performed in patients from the same center but from a later period, patients from other centers or countries
  - Warranted when one aims to apply a model in another setting or in patient subgroups that were not included in the development study

64

External Validation

ELSEVIER

Journal of Clinical Epidemiology 56 (2003) 826–832

Journal of Clinical Epidemiology

External validation is necessary in prediction research: A clinical example

S.E. Bleeker[a,b,c,*], H.A. Moll[a], E.W. Steyerberg[d], A.R.T. Donders[b,e], G. Derksen-Lubsen[c], D.E. Grobbee[b], K.G.M. Moons[b]

[a]*Erasmus Medical Center/ Sophia Children's Hospital Department of Pediatrics, Room Sp 1545 Dr Molewaterplein 60, 3015 GJ Rotterdam, The Netherlands*
[b]*Julius Center for General Practice and Patient Oriented Research, University Medical Center, Utrecht, The Netherlands*
[c]*Juliana Children's Hospital, Emergency Department, The Hague, The Netherlands*
[d]*Center for Clinical Decision Sciences, Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands*
[e]*Center for Biostatistics, Utrecht University, Utrecht, The Netherlands*

Accepted 21 November 2002

Abstract

Background and Objective: Prediction models tend to perform better on data on which the model was constructed than on new data. This difference in performance is an indication of the optimism in the apparent performance in the derivation set. For internal model validation, bootstrapping methods are recommended to provide biascorrected estimates of model performance. Results are often accepted without sufficient regard to the importance of external validation. This report illustrates the limitations of internal validation to determine generalizability of a diagnostic prediction model to future settings.

Methods: A prediction model for the presence of serious bacterial infections in children with fever without source was derived and validated internally using bootstrap resampling techniques. Subsequently, the model was validated externally.

Results: In the derivation set ($n = 376$), nine predictors were identified. The apparent area under the receiver operating characteristic curve (95% confidence interval) of the model was 0.83 (0.78–0.87) and 0.76 (0.67–0.85) after bootstrap correction. In the validation set ($n = 179$) the performance was 0.57 (0.47–0.67).

Conclusion: For relatively small data sets, internal validation of prediction models by bootstrap techniques may not be sufficient and indicative for the model's performance in future patients. External validation is essential before implementing prediction models in clinical practice.

Keywords: Prediction models; Internal validation; Bootstrap; External validation; Logistic Regression

65

TRIPOD STATEMENT

Annals of Internal Medicine RESEARCH AND REPORTING METHODS

Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement

Gary S. Collins, PhD; Johannes B. Reitsma, MD, PhD; Douglas G. Altman, DSc; and Karel G.M. Moons, PhD

Prediction models are developed to aid health care providers in estimating the probability or risk that a specific disease or condition is present (diagnostic models) or that a specific event will occur in the future (prognostic models), to inform their decision making. However, the overwhelming evidence shows that the quality of reporting of prediction model studies is poor. Only with full and clear reporting of information on all aspects of a prediction model can risk of bias and potential usefulness of prediction models be adequately assessed. The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Initiative developed a set of recommendations for the reporting of studies developing, validating, or updating a prediction model, whether for diagnostic or prognostic purposes. This article describes how the TRIPOD Statement was developed. An extensive list of items based on a review of the literature was created, which was reduced after a Web-based survey and revised during a 3-day meeting in June 2011 with methodologists, health care professionals, and journal editors. The list was refined during several meetings of the steering group and in e-mail discussions with the wider group of TRIPOD contributors. The resulting TRIPOD Statement is a checklist of 22 items, deemed essential for transparent reporting of a prediction model study. The TRIPOD Statement aims to improve the transparency of the reporting of a prediction model study regardless of the study methods used. The TRIPOD Statement is best used in conjunction with the TRIPOD explanation and elaboration document. To aid the editorial process and readers of prediction model studies, it is recommended that authors include a completed checklist in their submission (also available at www.tripod-statement.org).

Ann Intern Med. 2015;162:55-63. doi:10.7326/M14-0697 **www.annals.org**
For author affiliations, see end of text.
For contributors to the TRIPOD Statement, see the **Appendix** (available at www.annals.org).

66

CAUSAL RESEARCH

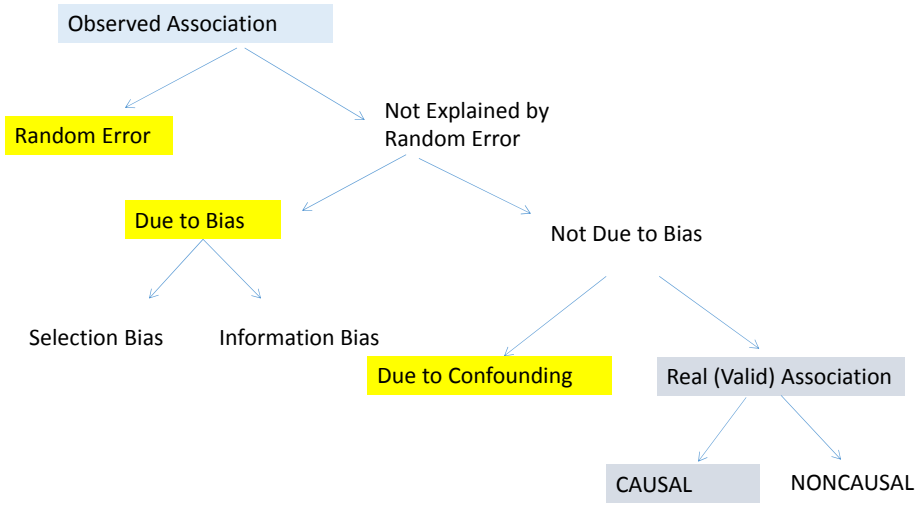
- ETIOLOGIC
- INTERVENTION

67

Etiologic (Causal) Research: A Review

- Etiologic research aims to find causal associations.
- To achieve this goal, alternative explanations for an apparent link between determinant and outcome need to be excluded in the research.
  - Confounding
  - Systematic error (Bias)
  - Random Error
- Typically, focuses on a single determinant at a time.
- Occurrence Relation
  - Causal Occurrence – Outcome=ƒ(D|ED)
- Courtroom Perspective
  - As an investigator (author of the study), you must convince the jury (your peers and readers) that the determinant is causally involved in the occurrence of the disease.

Etiologic Research: Observed Association Between Exposure and Outcome



Intervention Research

- DEPTh Model – Diagnosis, Etiology, Prognosis and **Therapy/Intervention**
- **Intervention** - deliberate action intended to change the prognosis in a patient
  - includes drug treatment, surgery, physiotherapy, lifestyle interventions such as physical exercise, and preventive actions such as vaccination.
  - the intended or main effects of the intervention must be weighed against possible risks (i.e., the unintended or side effects of the intervention)
  - Cost considerations also play a role

Intervention Research

- In intervention research, the principles of causal (etiologic) and descriptive research combine.
  - **Etiologic** - the true effect of the intervention (i.e., caused by the intervention) needs to be estimated **free from confounding variables**.
  - **Prognostic** - it is important to know as precisely as possible both the beneficial and untoward impact the intervention may have on an individual patient's prognosis.
    - E.g. drug X - 1-year mortality may be expected to decrease from 30% to 10% (intended or main effect), while the risk of developing orthostatic hypotension (unintended or side effect) is 10%.
  - Randomized controlled trials (RCTs) play an essential role in IR - role model for causal research

Treatment (Intervention) Effect

- A comparative study is needed to determine the true effect of a treatment (intervention)
- The **treatment effect** and the **three alternative** explanations for the observed treatment response can be illustrated by a simple equation
- Treated (intervention) Group
  - **OE_i (Observed effect) = R_x (**trt effect**)+ NH_i (**natural history**) + EF_i (**extraneous**) + Ob_i (**observation effects**)**
- Not receiving the intervention (reference [r] group)
  - **$OE_r = NH_r + EF_r + Ob_r$**
- **Overall Treatment Effect**

$$OE_i - OE_r = R_x + (NH_i - NH_r) + (EF_i - EF_r) + (OB_i - OB_r)$$

Comparability of Natural History

- Natural history - prognosis of the disease in the absence of treatment.
- Effects of natural history should be the same in all groups compared
  - Matching: carefully selected participants (similar age, proportion of males, severity of the disease, etc)
    - Individual matching on prognostic factors
  - Restriction: restrict the entire study population to a highly homogeneous group of patients
  - Multivariable adjustment: record prognostic indicators in detail
- **The problem**: comprehensive knowledge of all relevant prognostic factors is typically lacking.
  - **complexity of the decision to treat patients accentuates the problem**

Comparability of Natural History

- Setting an indication for treatment
  - many factors considered, some measurable while others are very implicit and neither reflected in the patient file nor measurable
- **Indication for treatment** (i.e., the composite of all reasons to initiate it) is a **very strong prognostic indicator**.
  - Patients indicated for a drug (intervention) typically have a poorer prognosis

Confounding by Indication (Indication Bias)

- **Confounding by indication - t**he effect on natural history of the presence or absence of a pertinent indication in patients with the same disease who are or are not treated [Grobbee & Hoes, 1997].

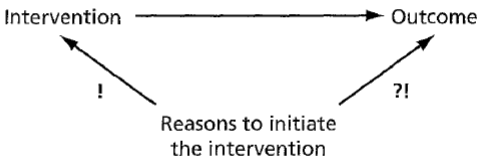


FIGURE 5–6 Reasons underlying the decision to intitiate treatment are important potential confounders.

Academic Emergency Medicine
Official Journal of the Society for Academic Emergency Medicine

RESEARCH METHODS & STATISTICS

Indication Bias Explains Some of the Observed Increased Mortality Associated With Use of Prehospital Intravenous Fluids in a Pediatric Trauma Population

Tabitha Garwe, PhD, Jeremy J. Johnson, MD, and Robert W. Letton, MD

Abstract

Objectives: Traditionally, in both pediatric and adult trauma patients, management of hemorrhage and shock has included early rapid intravenous fluid (IVF) replacement at the scene or during transport to a definitive care facility. Because prehospital resuscitation can be considered as a lifesaving intervention, severely injured patients are more likely to receive IVF. Observational studies not adequately adjusting for this confounding by indication (indication bias) while evaluating the impact of prehospital IVF on mortality in clinically heterogeneous patient populations are likely to find an increased mortality associated with the use of prehospital IVF, an association that may be spurious even after traditional multivariable risk adjustment. Propensity scores can be used to mitigate the impact of this selection bias on the estimated effect. The authors hypothesized that the effect of IVF on mortality will differ based on whether propensity scores (based on a set of prehospital indications for IVF) are adjusted for in a multivariable outcome model.

Confounding by Indication

- Confounding by indication commonly creates nearly insurmountable problems for nonrandomized research on intended effects of treatment
  - Groups of patients with the same indications but different treatments can be compared
  - Residual dissimilarities in characteristics in patients receiving different treatments for the same indications are known, adequately measured, and can be adjusted for.
- **Confounding by contraindication** - reasons to refrain from initiating the intervention may act as confounding variables

Randomization

- Randomization - most effective way to resolve the problem of confounding by indication and other confounding effects of differences in natural history

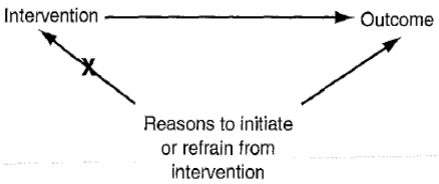


FIGURE 5–7 Major strength of a random allocation of patients to an intervention.

Principles of RCTs: A Summary

- Randomization ensures comparability of natural history (NHi = NHr)
- Blinding and use of placebo ensure comparability of extraneous effects (EFi = EFr)
- Blinding also prevents observer bias due to differential observations or measurements in either group (OBi = OBr)
- Comparability for natural history is always needed for a valid estimation of the treatment effect
- Need for blinding varies according to the objective of the trial and the nature of the outcome
- Limitations
  - cannot always be conducted
  - tend to include highly selected patients

Causal research: Unintended Effects of Interventions

- Main challenge lies in **establishing causality**
- Studies also bear characteristics of prediction (prognostic) research
- Courtroom perspective
  - the researcher has to prove beyond a reasonable doubt that the intervention caused the side effect
  - **Confounding by indication (Indication Bias)**
    - E.g. If COX-2 inhibitors are for some reason preferentially prescribed to patients with an unfavorable cardiovascular risk profile
- RCT limited, case-control design an attractive option

Type A Unintended Effects: Confounding by Indication

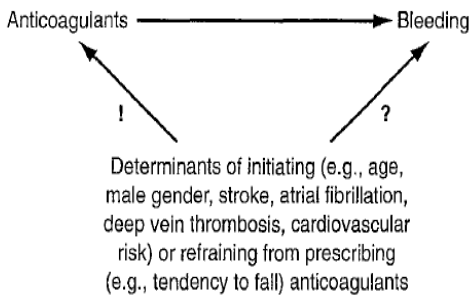


FIGURE 6–2 Potential confounding in the study of type A unintended effects of an intervention with the example of anticoagulants and bleeding.

Type B Unintended Effects and Confounding

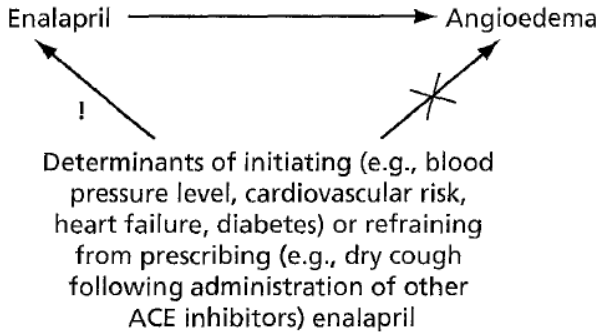


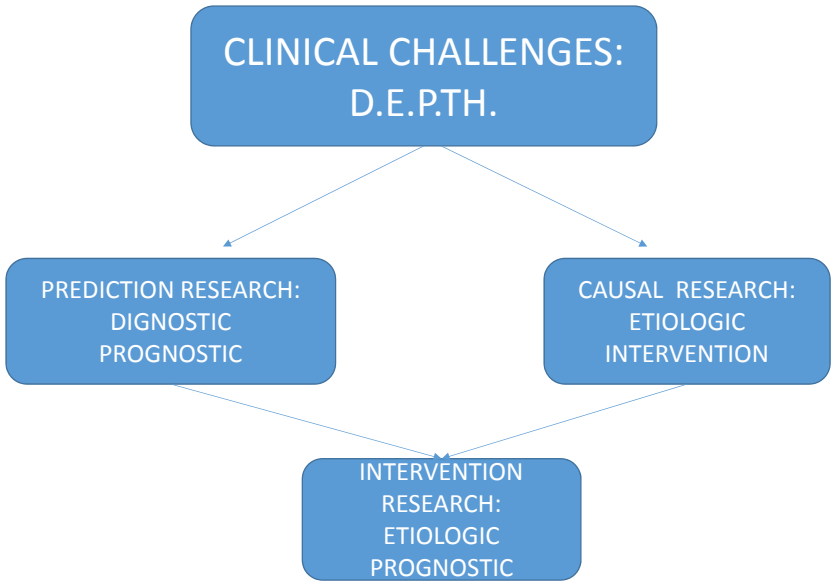
FIGURE 6–3 Potential confounding in the study of type B unintended effects of an intervention with the example of enalapril and angioedema.

Indication Bias – A Major Concern in Clinical Research

- Traditional methods used to mitigate this bias
- Randomization – most effective way
- **Alternatives to randomization**
  - **Propensity Scores**
    - Find patients with same indication for treatment but received different treatments
  - **Instrumental Variables**
    - Find a variable strongly associated with treatment assignment but not the outcome
    - Adjusts for unmeasured confounding

83

SUMMARY



84

SUMMARY

- **Why should we differentiate between causal and prediction research in clinical epidemiology?**
  - Analytic implications
    - Confounding a non-issue in prediction research, a potentially confounding variable is simply another predictor in the model
    - Need to internally and externally validate predictive models
  - Peer-review
    - Clinical journals increasingly asking authors to identify type of study i.e. prognostic, diagnostic, therapeutic
    - Minimize author-reviewer misunderstanding
      - E.g. a prognostic study reviewed as an etiologic study or vice-versa
      - My personal experience

85

Acknowledgements

- BSE 6193 'Methods in Clinical Epidemiology' Contributors
  - Julie Stoner, PhD
  - David Thompson, PhD, PT
  - Sarah Vesely, PhD

86

THANK YOU!

CLINICAL EPIDEMIOLOGY: RESOURCES AND SELECTED REFERENCES

Courses and workshops

The *Methods in Clinical Epidemiology* course, BSE 6193 will be offered in the fall of 2019, Dept of Biostatistics and Epidemiology, OUHSC College of Public Health

Look out for topic-specific workshops offered through BERD in the future

Recommended Textbooks

i) *Evidence-Based Diagnosis. 2009. TB Newman and MA Kohn, Cambridge University Press.*

ii) *Clinical Epidemiology. Principles, Methods, and Applications for Clinical Research (2nd Edition) 2015. DE Grobbee & AW Hoes. Jones and Bartlett Publishers.*

Helpful Checklists

i) *STARD – Diagnostic Accuracy Studies*
ii) *QUADAS – Diagnostic Accuracy Studies included in Systematic Reviews*
iii) *TRIPOD - Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis*
iv) *STROBE – Observational Epidemiologic Studies*
v) *PRISMA – Systematic Reviews of RCTs/Intervention Studies*

Selected References

<u>Diagnostic Research</u>

Sim, J.; Wright, C. C. (2005). "The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements". *Physical Therapy*. **85** (3): 257–268.

Banerjee, M.; Capozzoli, Michelle; McSweeney, Laura; Sinha, Debajyoti (1999). "Beyond Kappa: A Review of Interrater Agreement Measures". *The Canadian Journal of Statistics*. **27** (1): 3–23.

Bland JM, Altman DG (1986). "Statistical methods for assessing agreement between two methods of clinical measurement" . *Lancet*. **327** (8476): 307–10

Bland JM, Altman DG (1999). "Measuring agreement in method comparison studies". *Statistical Methods in Medical Research*. 8 (2): 135–60

Brian J. Manning, Thorarinn Kristmundsson, Björn Sonesson, Timothy Resch. Abdominal aortic aneurysm diameter: a comparison of ultrasound measurements with those from standard and three-dimensional computed tomography reconstruction. *J Vasc Surg. 2009 Aug; 50(2): 263–268.*

Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ. 2002 Mar 2; 324(7336):539-41.*

Gill Christopher J, Sabin Lora, Schmid Christopher H. Why clinicians are natural Bayesians. *BMJ 2005; 330:1080*

David A Grimes, Kenneth F Schulz. Refining clinical diagnosis with likelihood ratios. *Lancet 2005; 365: 1500–05*

Pauker, S. G., and J. P. Kassirer (1975). "Therapeutic decision making: A cost-benefit analysis." *N Engl J Med 293(5): 229-34.*

Pauker, S. G., and J. P. Kassirer (1980). "The threshold approach to clinical decision making." *N EnglJ Med 302(20): 1109-17.*

Bisoffi, Z., Tinto, H., Sirima, B. S., Gobbi, F., Angheben, A., Buonfrate, D., & Van den Ende, J. (2013). Should Malaria Treatment Be Guided by a Point of Care Rapid Test? A Threshold Approach to Malaria Management in Rural Burkina Faso. PLoS ONE, 8(3).

Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. Acad Emerg Med. 2013 Nov; 20(11):1194-206

Prediction Research

Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.* New York, NY: Springer Science+Business Media; 2009

Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med. 2004; 23: 1631Y1660.*

Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol. 2003 May; 56(5):441-7.*

Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med. 2000 Apr 30; 19(8):1059-79.*

Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med. 2004 Aug 30; 23(16):2567-86.*

Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: A review and illustration. *Eur J Clin Invest. 2012;42: 216-228.*

Intervention Research

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika 1983; 70: 41–55.*

Newgard CD, Hedges JR, Arthur M, Mullins RJ. Advanced statistics: the propensity score–a method for estimating treatment effect in observational research. *Acad Emerg Med 2004; 11: 953–61.*

Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health 2000; 21: 121–45.*

D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med 1998;17:2265–81.*

Clinical Epidemiology: Examples from Lupus, Sjogren's and Sarcoidosis

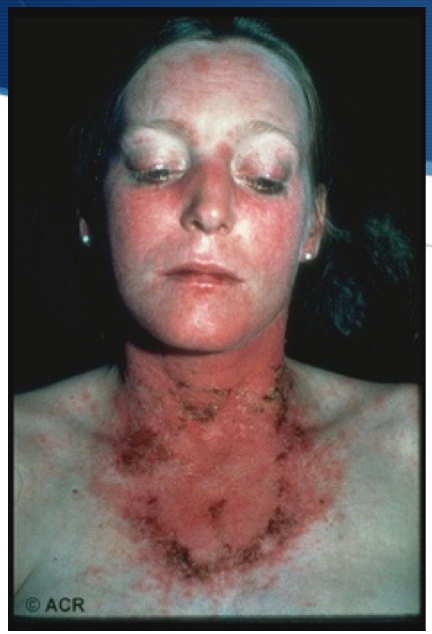
Dr. Courtney Montgomery
Associate Member
Division of Genomic and Data Sciences
Arthritis and Clinical Immunology Research Program
Oklahoma Medical Research Foundation
Oklahoma City, OK

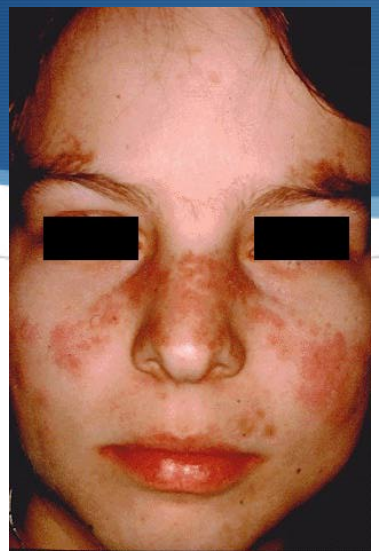
OMRF

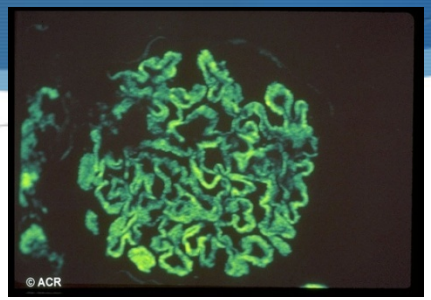
Severity Indices in SLE

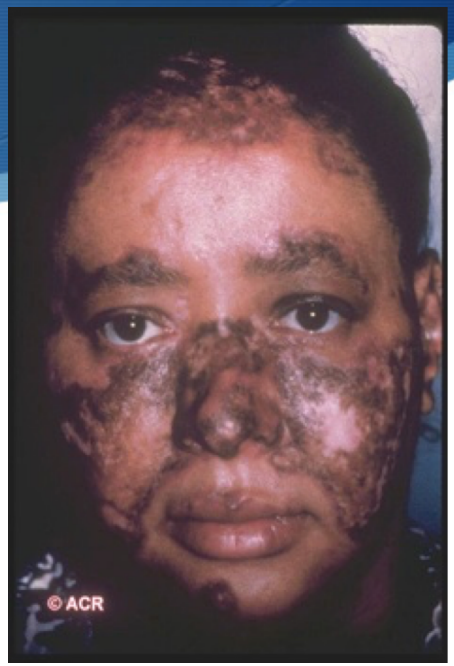
ACR Classification Criteria for SLE

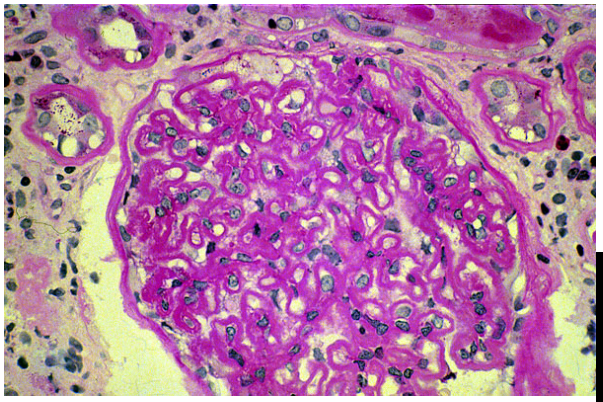
- Malar Rash
- Discoid Rash
- Photosensitivity
- Oral Ulcers
- Arthritis
- Serositis
  - Pleurisy
  - Pericarditis
- Kidney Involvement
- Neurological Disorder
  - Seizures
  - Psychosis
- Hematological Disorder
  - Hemolytic Anemia
  - Leukopenia
  - Lymphopenia
  - Thrombocytopenia
- Serological Disorder
  - Anti-dsDNA
  - Anti-Sm
  - Antiphospholipid Antibodies
    - IgG
    - IgM
    - Lupus Anticoagulant
    - False-+VDRL
- Anti-Nuclear Antibody (ANA)

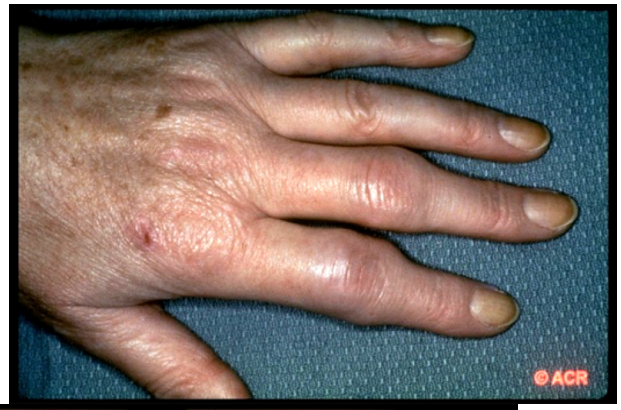

© ACR

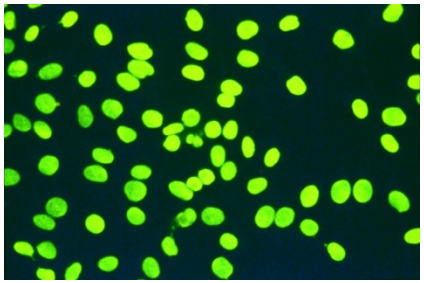


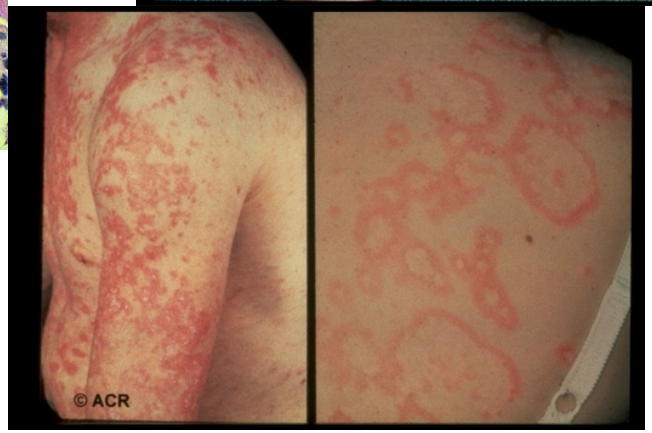

© ACR

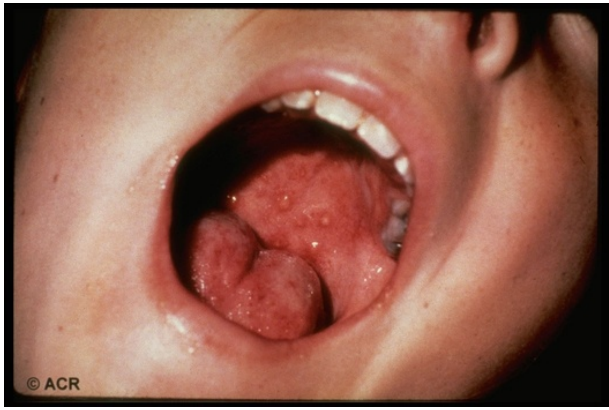

© ACR




© ACR




© ACR


© ACR

Background: Severity in SLE

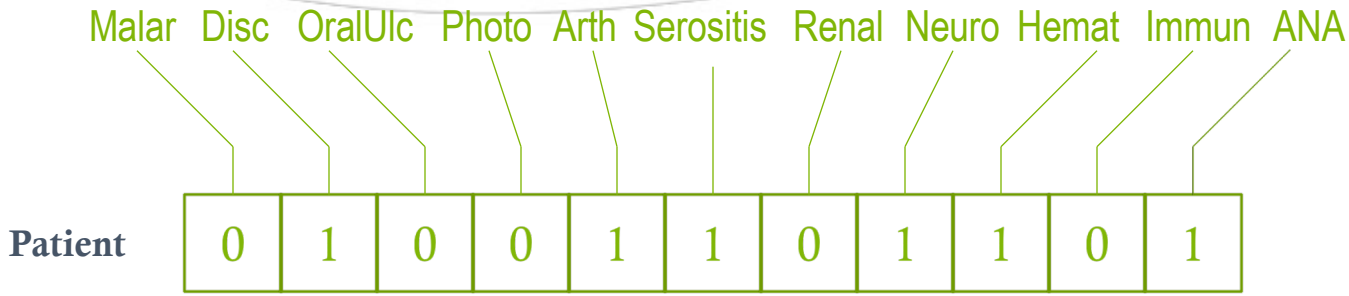
- SLE has a large and complex set of manifestations, which vary widely in frequency, severity and cumulative organ damage.
- Severity in SLE varies along demographic lines:
  - Sex
  - Race/ethnicity
  - Age at onset
- SLE severity is strongly associated with prognosis and life expectancy.
- There is need to advance our understanding of the genetic and molecular mechanisms underlying predisposition to severe SLE

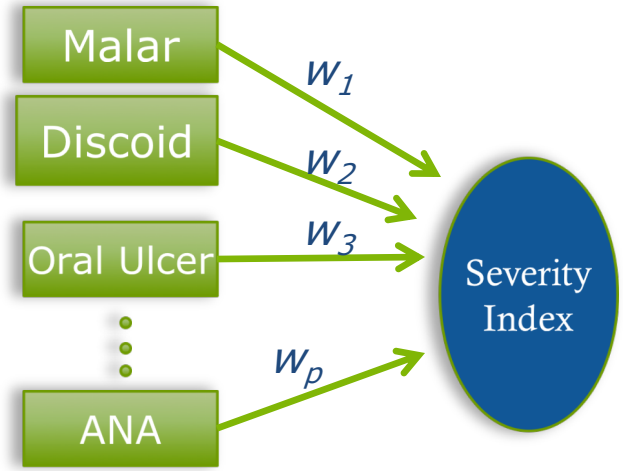
New ongoing project: Development of novel, simple severity index for SLE

- Current instruments for quantifying lupus severity:
  1) SLICC/ACR Damage Index (SDI):
     - 42 items, 12 organ systems (score range: 0-47)
     - Scoring: Damage occurring since onset of lupus, ascertained by clinical assessment and present for at least 6 months unless otherwise stated.
     - Time-consuming, typically completed by physician
     - Not practical for some research settings
  2) Lupus Damage Index Questionnaire (LDIQ):
     - Translation of SDI items into "lay language"
     - 55 yes/no questions
     - Questionnaire completed by patient or through interview
     - Not ideal for large studies
  3) Brief Index of Lupus Damage (BILD): Shorter version of LDIQ
- These instruments require considerable time and expense to compute, and require participation of physicians, patients, or interviewers.
- Project Aim: To develop simple severity index for lupus that can be computed using easily accessible data elements: ACR criteria and sub-criteria

ACR criteria-based index

- Index constructed from ACR criteria:
  - Often the most readily available data on lupus patients
- Each criteria/sub-criteria treated as separate item on index
- Weighting of items is data-driven





- Prescription history used as surrogate for severity
- Group 1 – Ever prescribed at least one of:
  - Cyclophosphamide (Cytoxan), Nitrogen mustard
  - Mycophenolate Mofetil (Cellcept)
  - Cyclosporine (SandImmune, Gengraf, Neoral)
  - Rituximab (Rituxan, Zytux)
- Group 2 – Never been prescribed any of the above

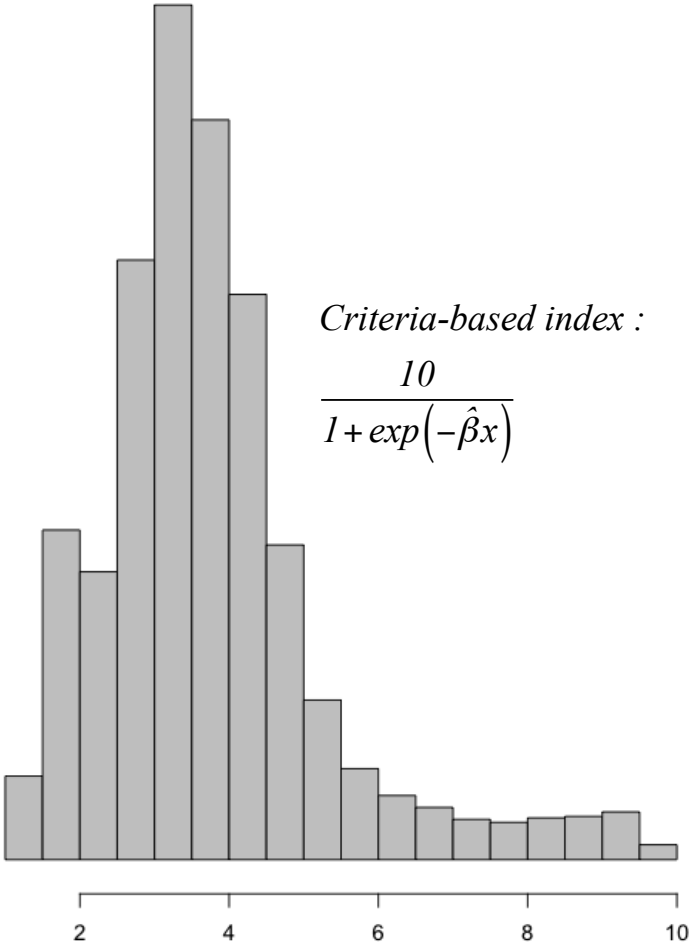
Weights computed using L_1/L_2-penalized GLM

- Training set (*tr*): $n = 1612$
- Test set (*te*): $n = 805$
- AUC_{test}: 0.722

y : Severe medication prescription (Group 1 vs. 2)

x : ACR criteria/sub-criteria

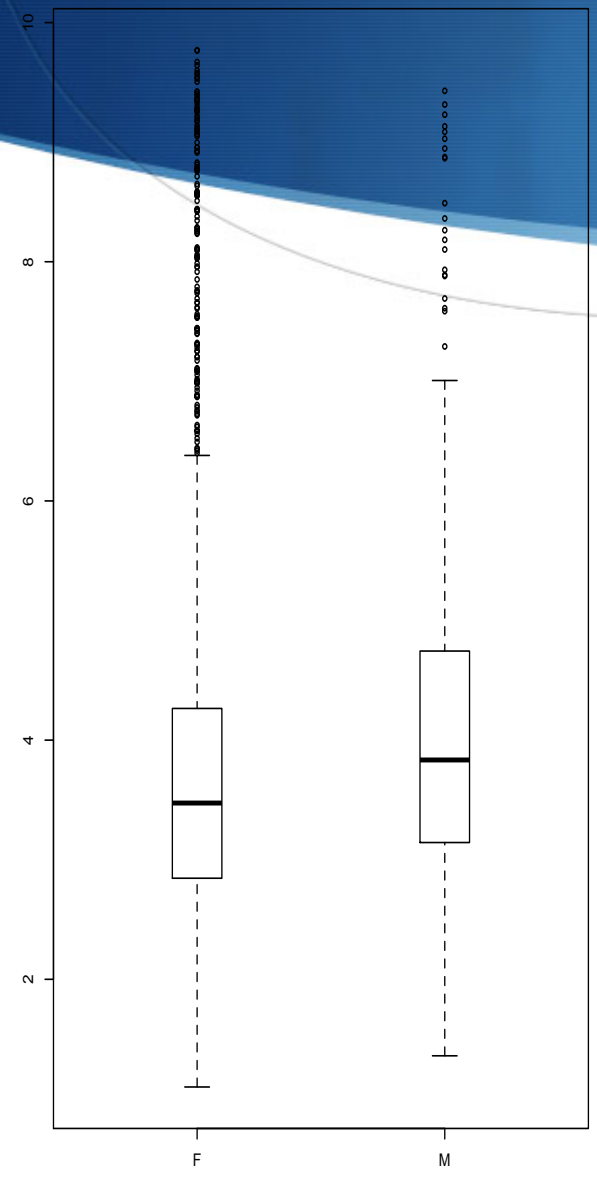
$$\frac{10}{1+exp\left(-\hat{\beta}x\right)}$$ *: Criteria-based severity index*

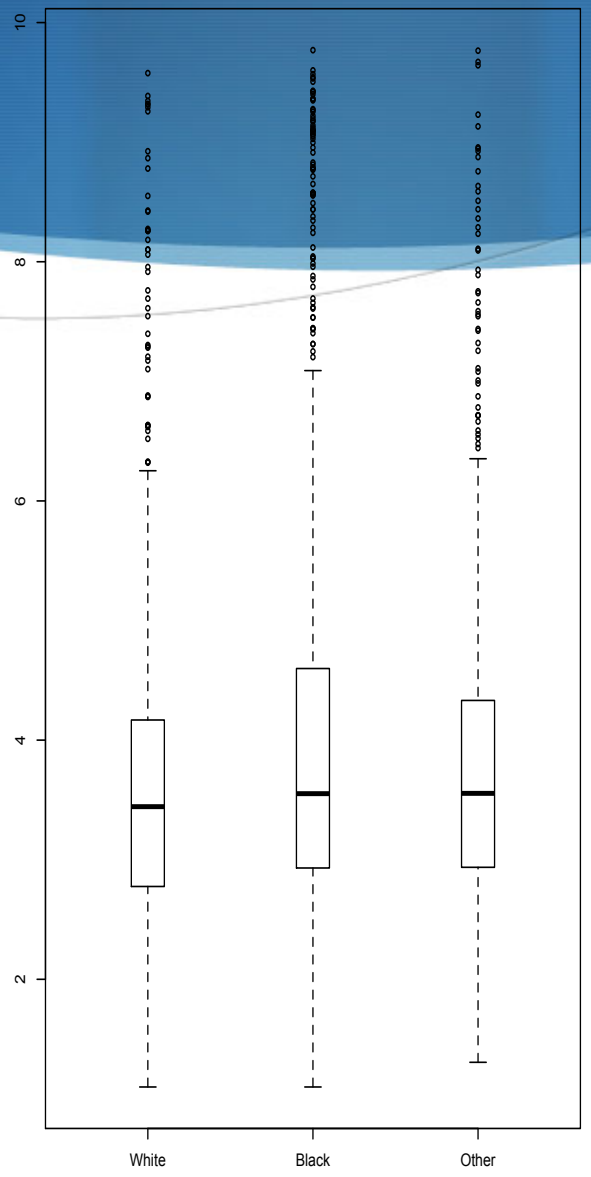


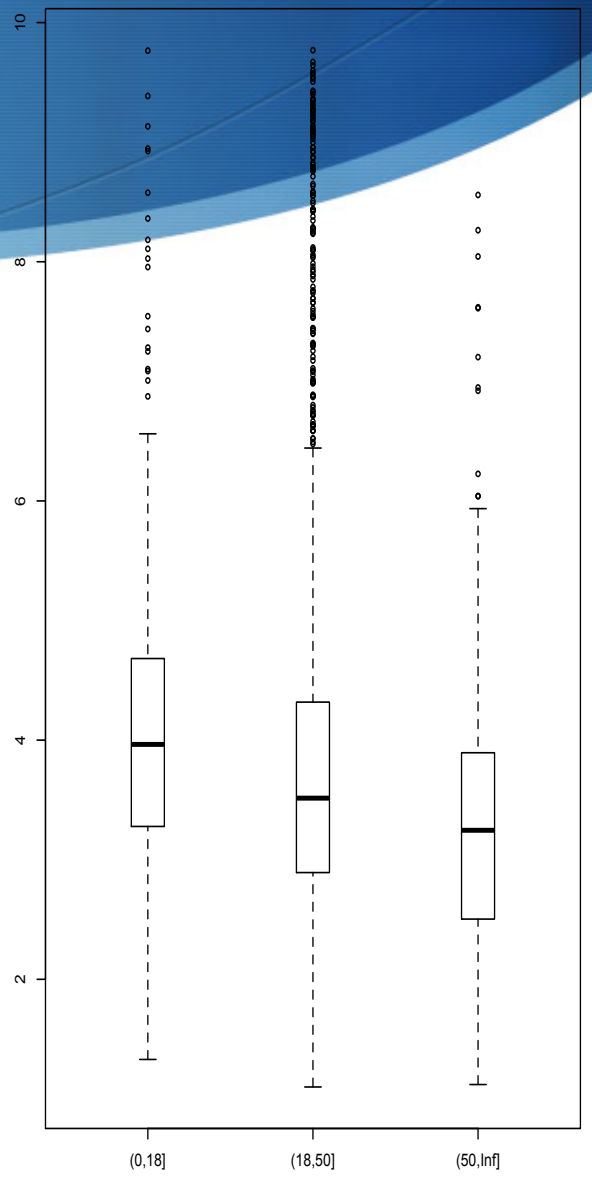
Criteria	Sub-criteria	Coeff(β)
Malar Rash		0.301
Discoid Rash		-0.216
Photosensitivity		-0.138
Oral Ulcer		.
Arthritis		-0.471
Serositis	Pericarditis	0.272
	Pleuritis	0.256
Hematological		0.081
Creatinine*		0.208
Neurological	Seizures	0.410
	Psychosis	0.15
Immunological		0.796
ANA		-0.226

*Creatinine allowable max: 25 (if value>25, then value=25)

Demographics







ACR Criteria-based Severity Index

Relationship with Katz Severity Index (Linear model)

Coefficients	Estimate	p-value
Age (@ met criteria)	-0.012004	0.000511 ***
Sex_M	-0.033931	0.806892
race_Black	0.418570	3.08e-05 ***
race_Other	-0.163964	0.151076
Severity (ACR Based)	1.761993	<2e-16 ***

Bello et al, Lupus Sci Med. 2016

Survival Prediction

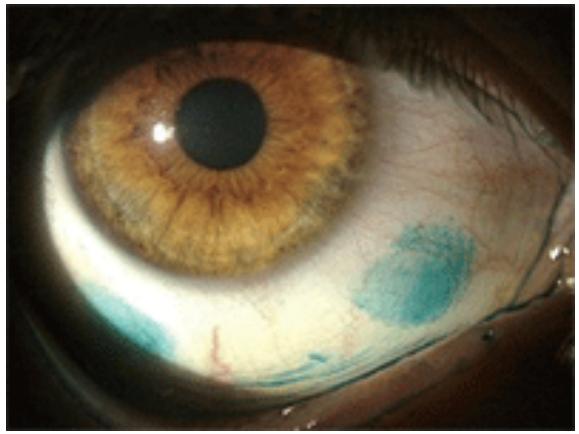
Relationship with Mortality (Cox Prop. Hazards model)

Coefficients	Estimate	p-value
Age (@ met criteria)	0.03034	0.0119 *
Sex_M	0.31803	0.4422
race_Black	-0.23879	0.4805
race_Other	-0.53029	0.2517
Severity (ACR Based)	0.69101	1.6e-07 ***

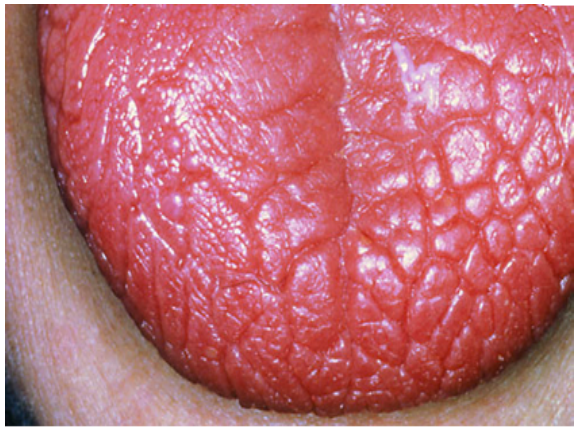
Prediction of Sjogren's organ involvement

About Sjögren's Syndrome (SS)

- Chronic autoimmune disease
  - 9x more prevalent in women than in men
  - Targets moisture-producing glands



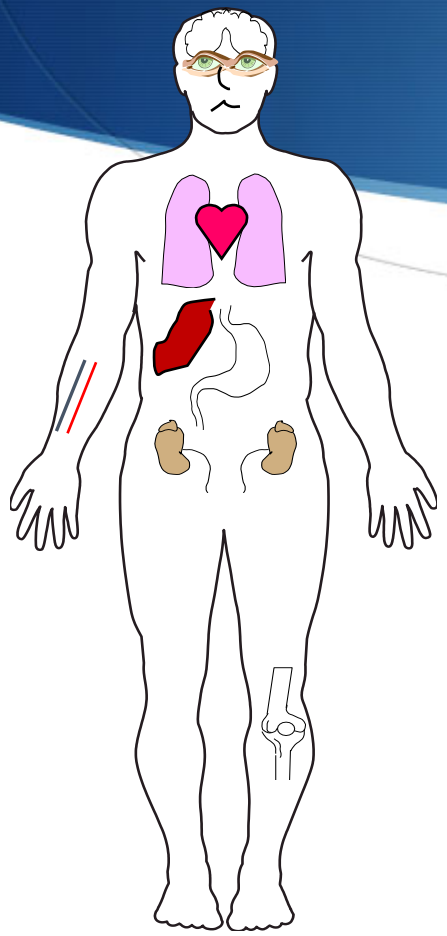
Dry eyes



Dry mouth

  - Diverse array of systemic signs and symptoms

Systemic Signs & Symptoms

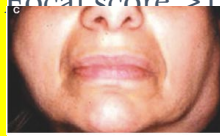
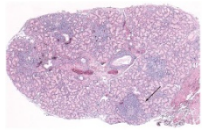


- Rheumatologic:
- Neurologic:
- GI/Hepatobiliary:
- Dermatologic:
- Pulmonary:
- Endocrine:
- Genitourinary:
- Hematologic:
- Constitutional:

Classification Criteria

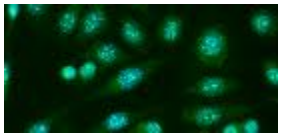
Objective

Biopsy:
Focal lymphocytic sialoadenitis
Focal score >1

Nat Clin Pract Rheumatol 2: 262–269

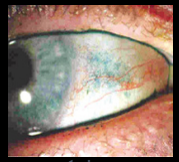
Anti-Ro or Anti-La



aklides.com

Objective

Schirmer's Test or
Lissamine Green
Staining

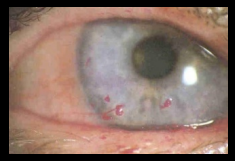
gulfmd.com www.salujaeyecare.com

Whole unstimulated
salivary flow
<1.5 ml in 15 min

www.intelligentdental.com

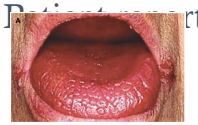
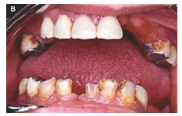
Subjective

Dry Eyes
Patient report



www.salujaeyecare.com

Dry Mouth/PG
swelling
Patient report

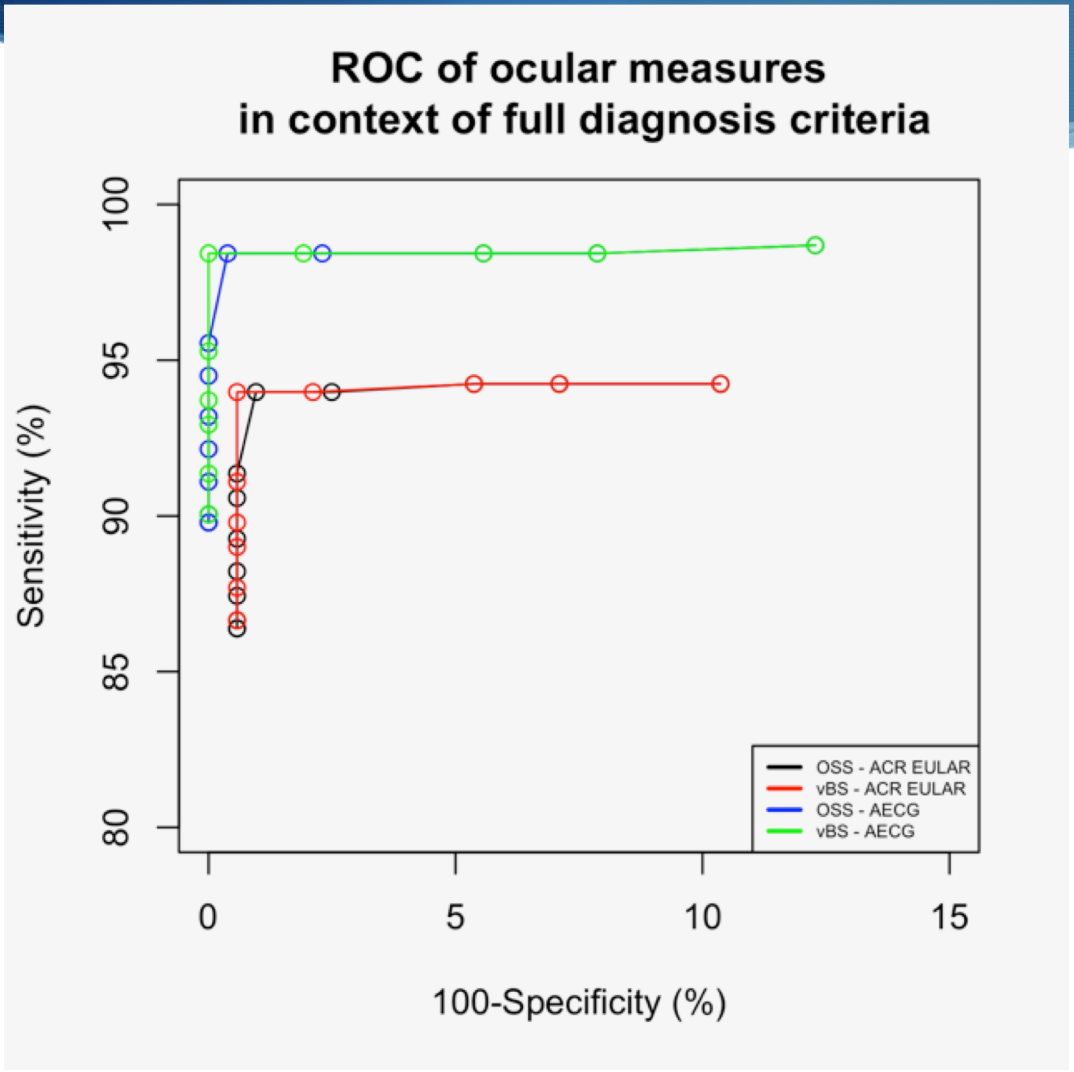



Nat Clin Pract Rheumatol 2: 262–269

Diagnosis may take up to 10 years

Revised criteria for diagnosis of Sjögren's syndrome; Ann. Rheum. Dis. 61 (6):554-8.

Ocular involvement at predictor of overall organ involvement in SS

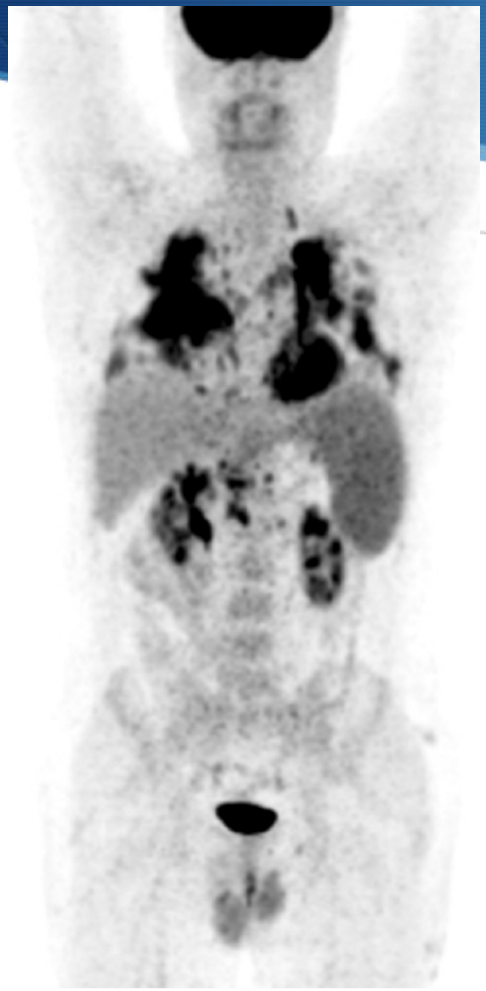


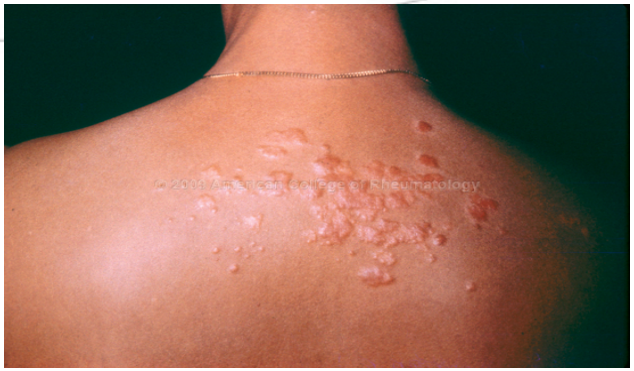
The project was to analyze the efficacy of a new ocular staining method in the context of Sjogren's syndrome classification. The proposed staining method was compared to the previously established method as a stand alone variable as well as in conjunction with the other SS diagnostic criteria.

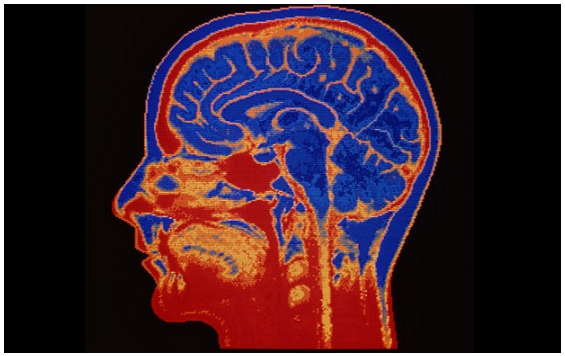
Rasmussen et al, 2018

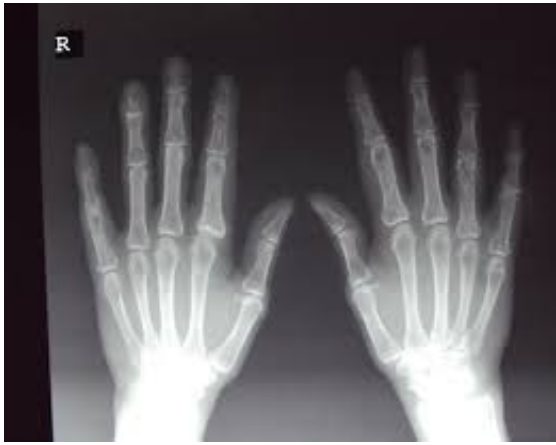
Genetic and Clinical Profiling in Sarcoidosis

Systemic

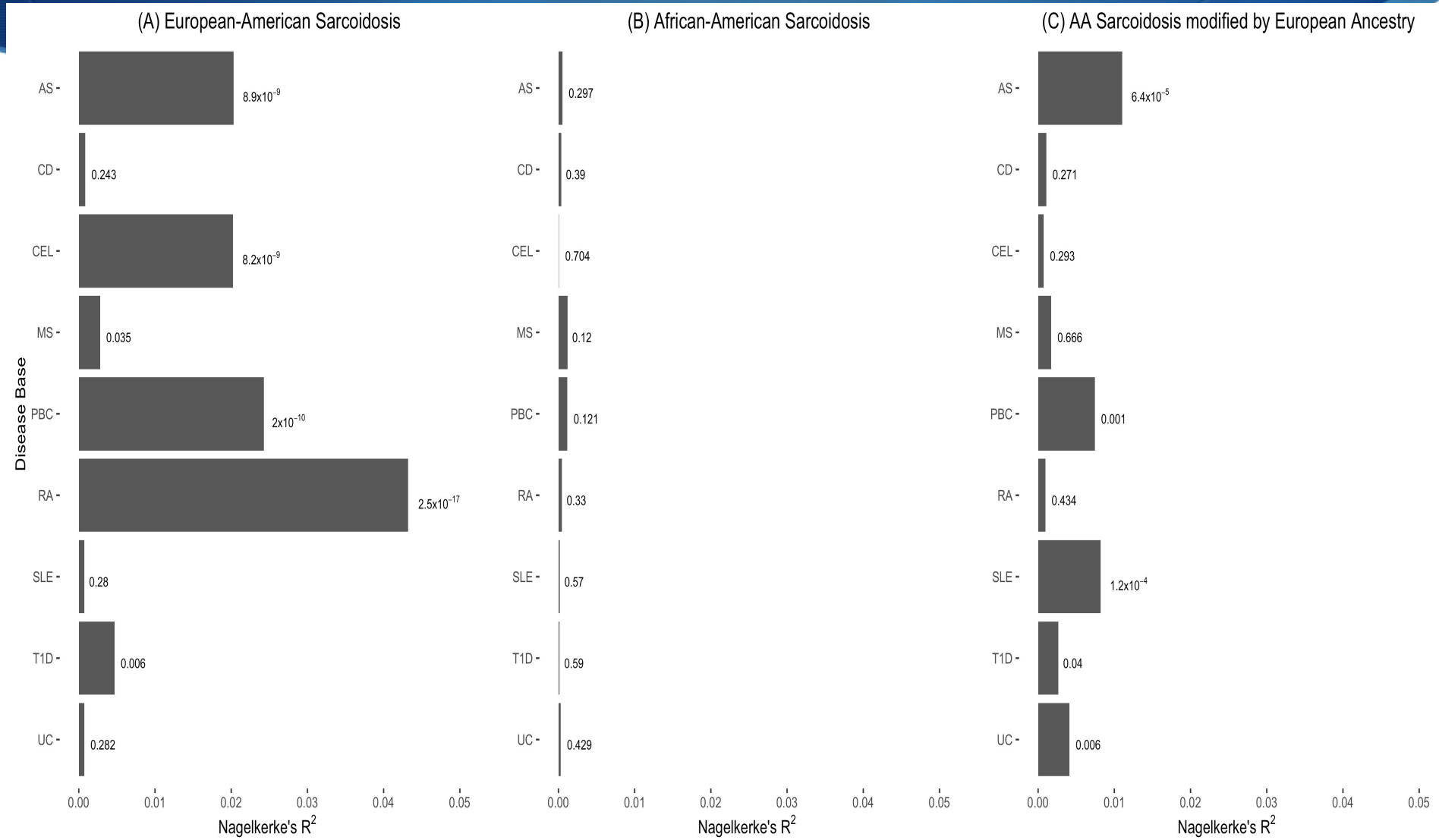






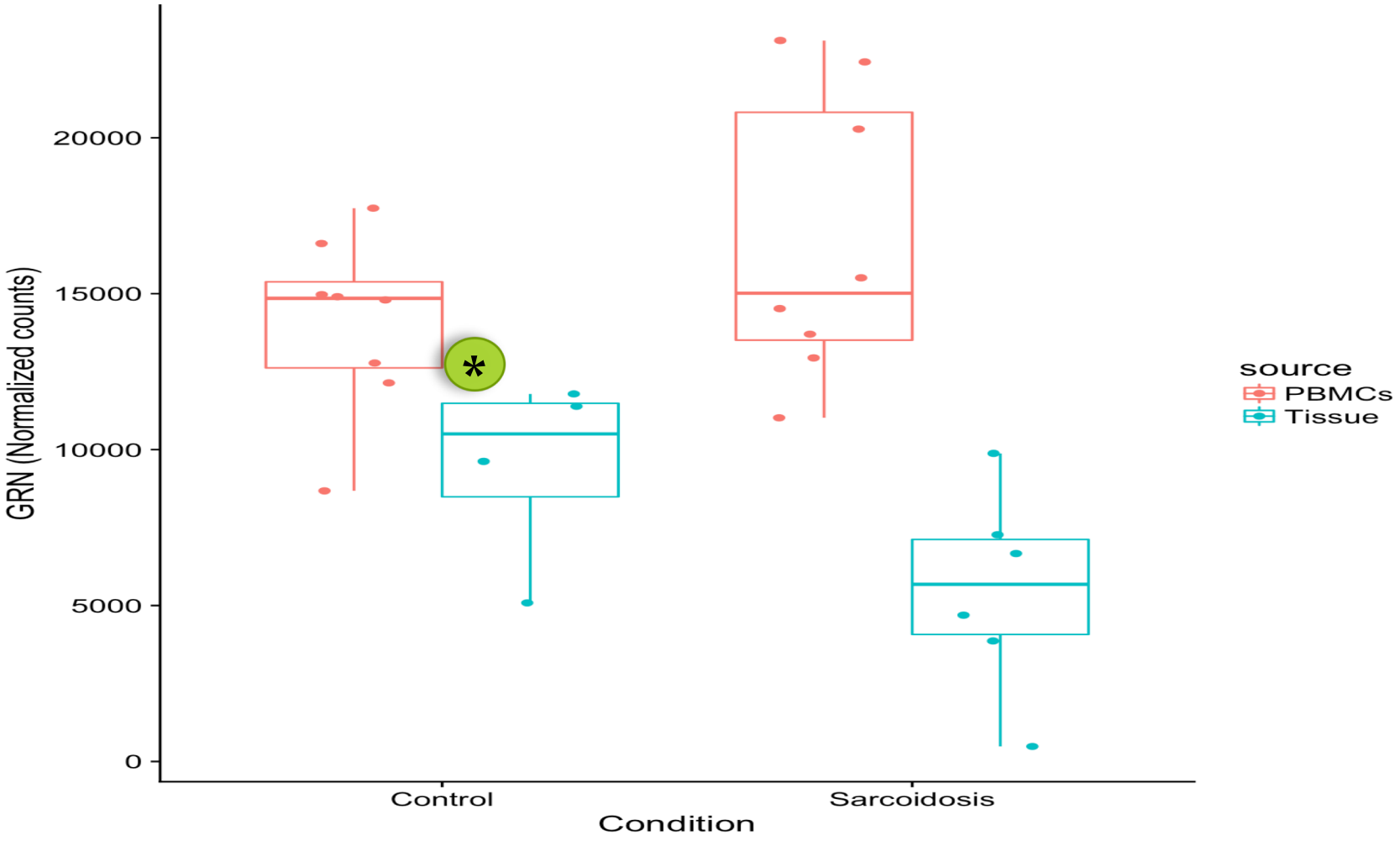


Looks Autoimmune

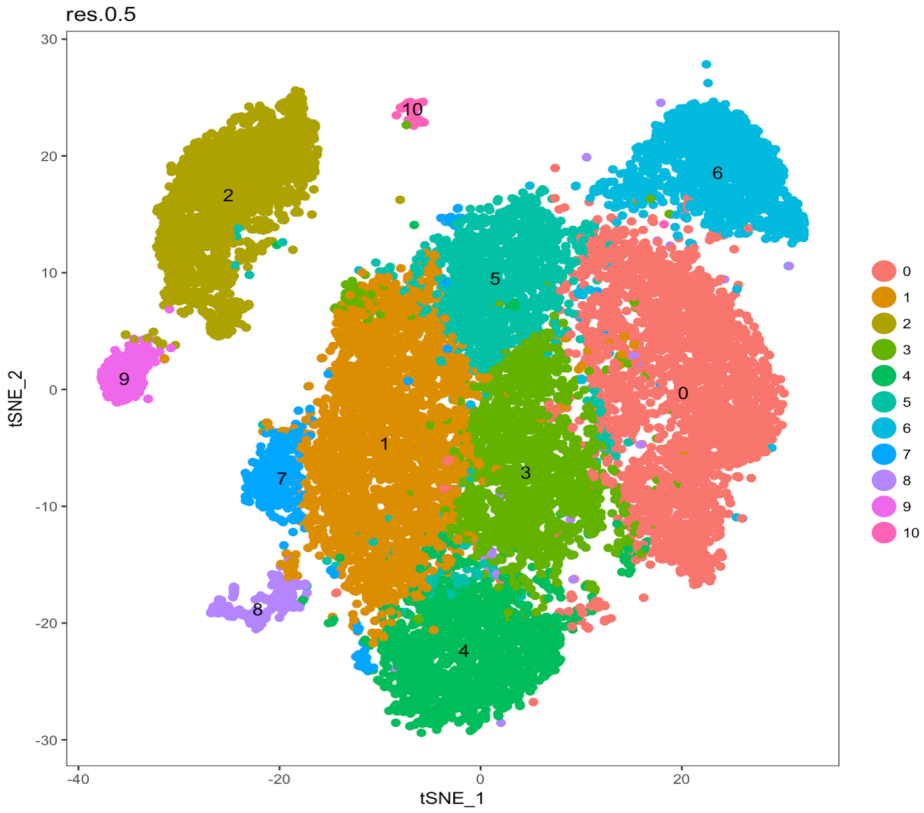


Lareau et al, Gene Immun. 2017

Tissue compared to blood



Clustering Single Cells



Inactivated monocytes:
High expression of:

- FCeR1G (basophils)

Low expression of:

- Monocyte markers (FcgR3A, MNDA)
- Activation markers: S100s

Activated Monocytes:
High expression of:

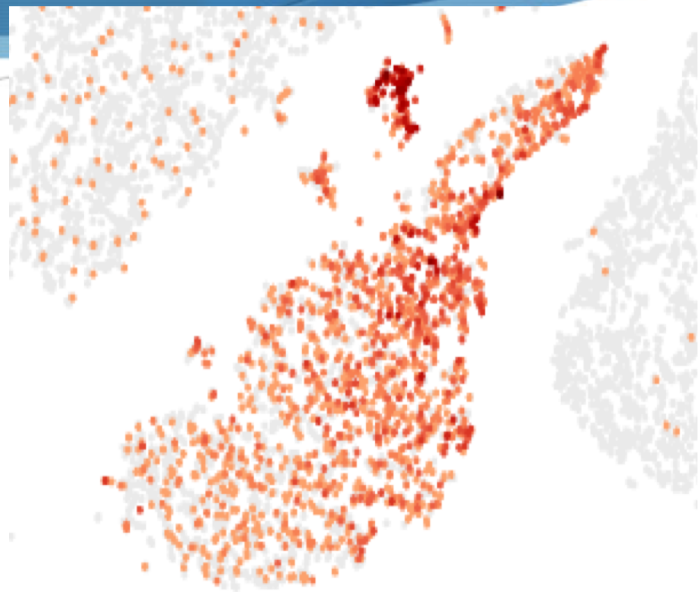
- CTSS (Cathepsin S)
- HLA-DRA/-DPA1 (MHC II)
- MNDA
- IFI30

Low expression of:

- MMD (Monocyte to Mac)

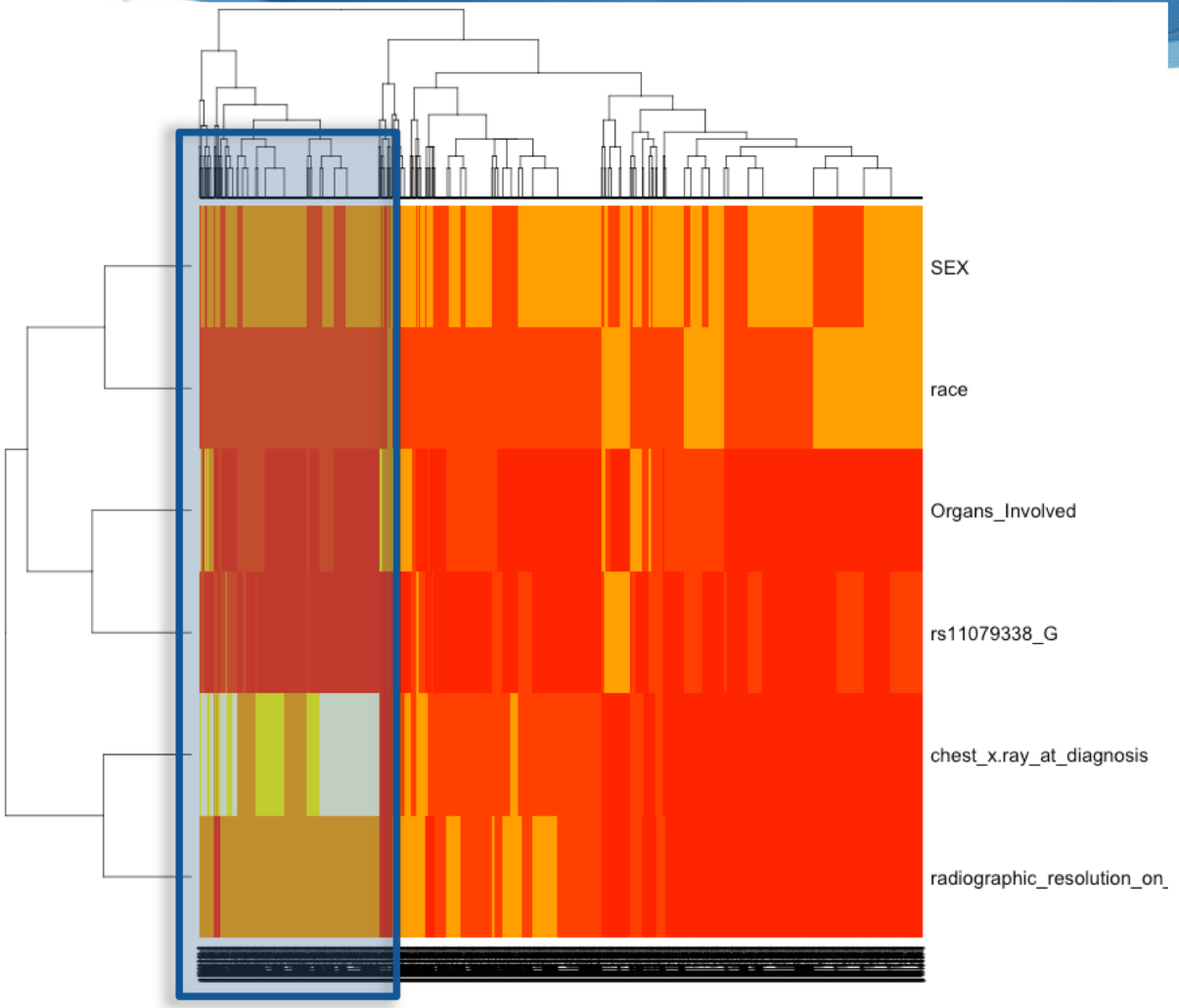
Progranulin (PGRN)

Literature states it is expressed in most cell types in the blood.
In our analyses we saw highest expression in this cluster of cells



Monocytes and monocytes to macrophages

eQTL



Those with two copies of the alternate allele were:

- Female (orange)
- EA (as was our pilot cohort)
- Had both lung and extrathoracic organ involvement (mixed yellow, orange, red)
- Had higher scadding stage at diagnosis (yellow to white)
- Had persistent disease (orange)

Acknowledgments

- OMRF
  - Montgomery Lab
    - Lori Garman
    - Richard Pelikan
    - Nathan Pezant
    - Kaity White
    - Allshine Chen
    - Caleb Lareau
    - Ambra Pastori
    - Stuart Glenn
  - Patrick M. Gaffney
    - Graham B. Wiley
- Vanderbilt University
  - Wonder Drake
  - Ozioma Chioma
- Northwell New York
  - Michael C. Iannuzzi
- Henry Ford Health System
  - Benjamin A. Rybicki
  - Albert M. Levin
  - Indra Adrianto*
- Funding Sources
  - NIH-NHLBI: R01, R01S
  - NIH-NIGMS: U54
  - Chapman Trust