Study Design and Analysis: Diagnostic Studies 2025

Mary Dunbar MD MSc MSc FRCPC

Assistant Professor of Pediatrics, University of Calgary
Pediatric Neurologist, Alberta Children's Hospital

Slide Credit: some slides taken from Dr. S. Greenaway's lecture from years ago

Objectives

- Understand the issues in the evaluation of a diagnostic test
- Appreciate components of evaluating test performance
- precision and accuracy
- sensitivity and specificity
  - likelihood ratio
  - positive and negative predictive values
  - receiver operating characteristic (ROC) curves
  - additional factors: cost, availability, acceptability, utility

Examples of Diagnostic Tests

- **Biochemical**
  - electrolytes, urea, creatinine
- **Imaging**
  - CXR, MRI
- **Genetic**
  - karyotype, array, WES
- **Microbiological**
  - blood culture
- **Physiological**
  - PFTs, exercise test, GTT
- **Clinical**
  - Lever sign to diagnose ACL tear
- **Patient-reported outcome measures**
  - questionnaire of symptoms to diagnose IBD

Purpose of diagnostic tests

- **Diagnose a disease or condition**
  - TSH
  - echocardiogram
- **Exclude a disease or condition**
  - HbA1C
  - Troponin
- **Estimate prognosis**
  - LDL cholesterol
  - BRCA1 mutation
- **Inform treatment decisions**
  - PSA
  - karyotype

Factors Affecting Diagnostic Test Performance

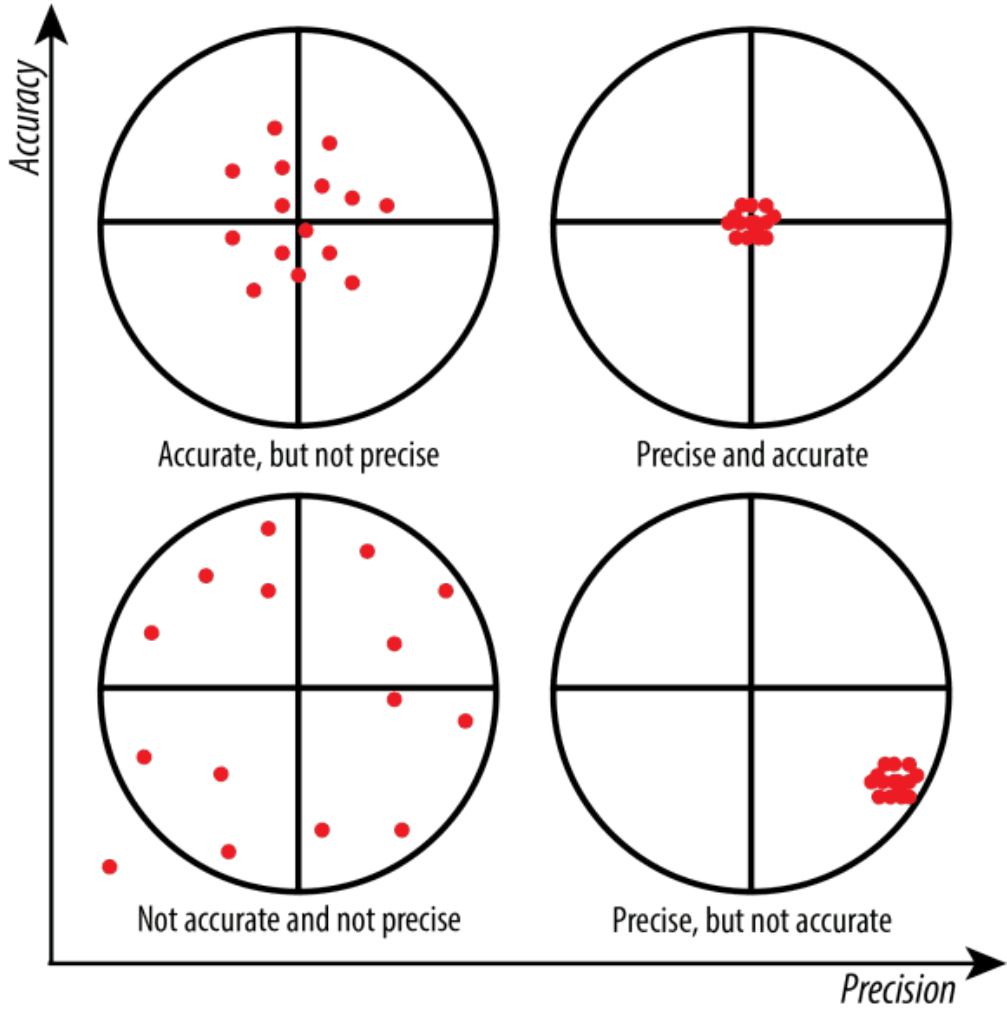
- Prevalence of the disease in the population
- Spectrum of the disease
- Often dependent on other factors
  - part of diagnostic pathway
  - test results may not be independent
  - often depend on prior knowledge
- Gold standard
  - established test which confirms the diagnosis

Types of Studies to Evaluate a Diagnostic Test

- Precision (reproducibility)
  - intra-observer (amount of variation for a single observer)
  - inter-observer (variation between 2 or more observers)
- Accuracy
  - cohort
  - case-control
- Costs, Risks and Acceptability
  - prospective
  - retrospective
- Improvement of clinical outcome
  - RCT
  - case-control

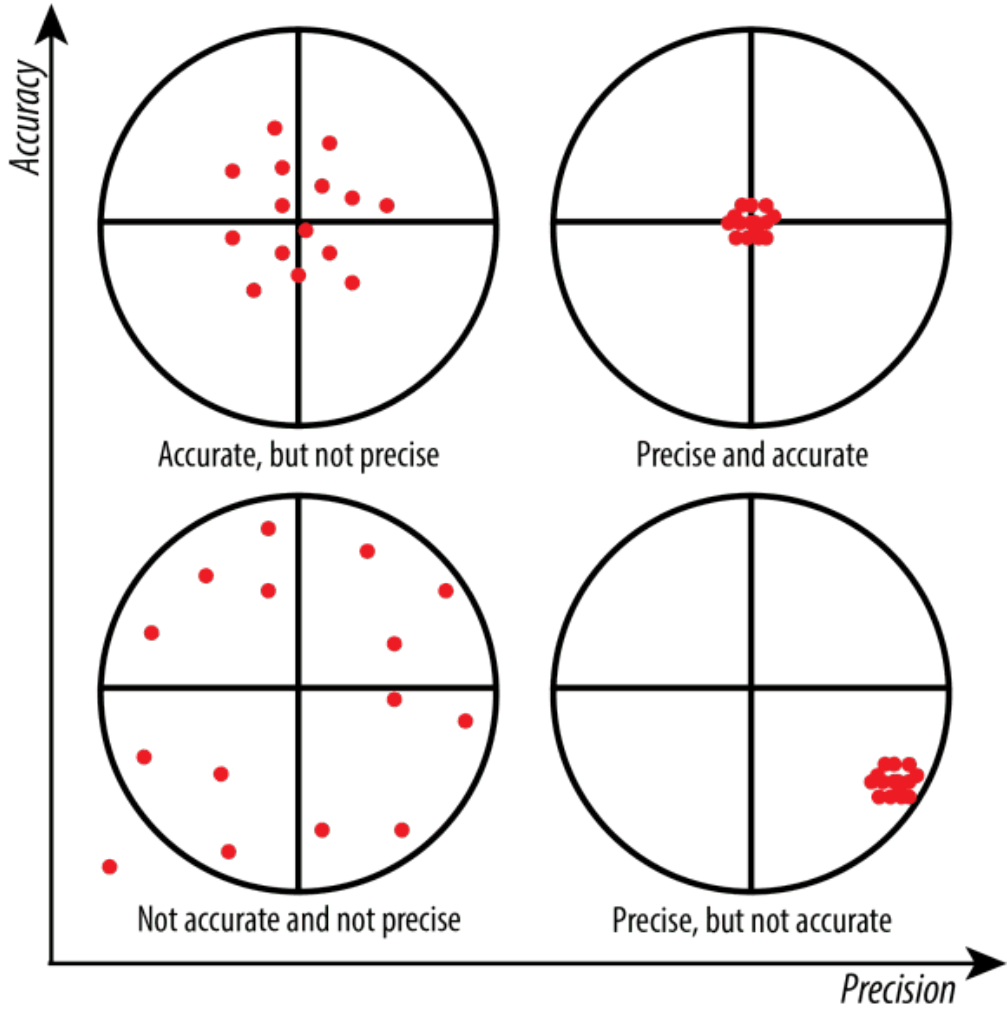
Precision

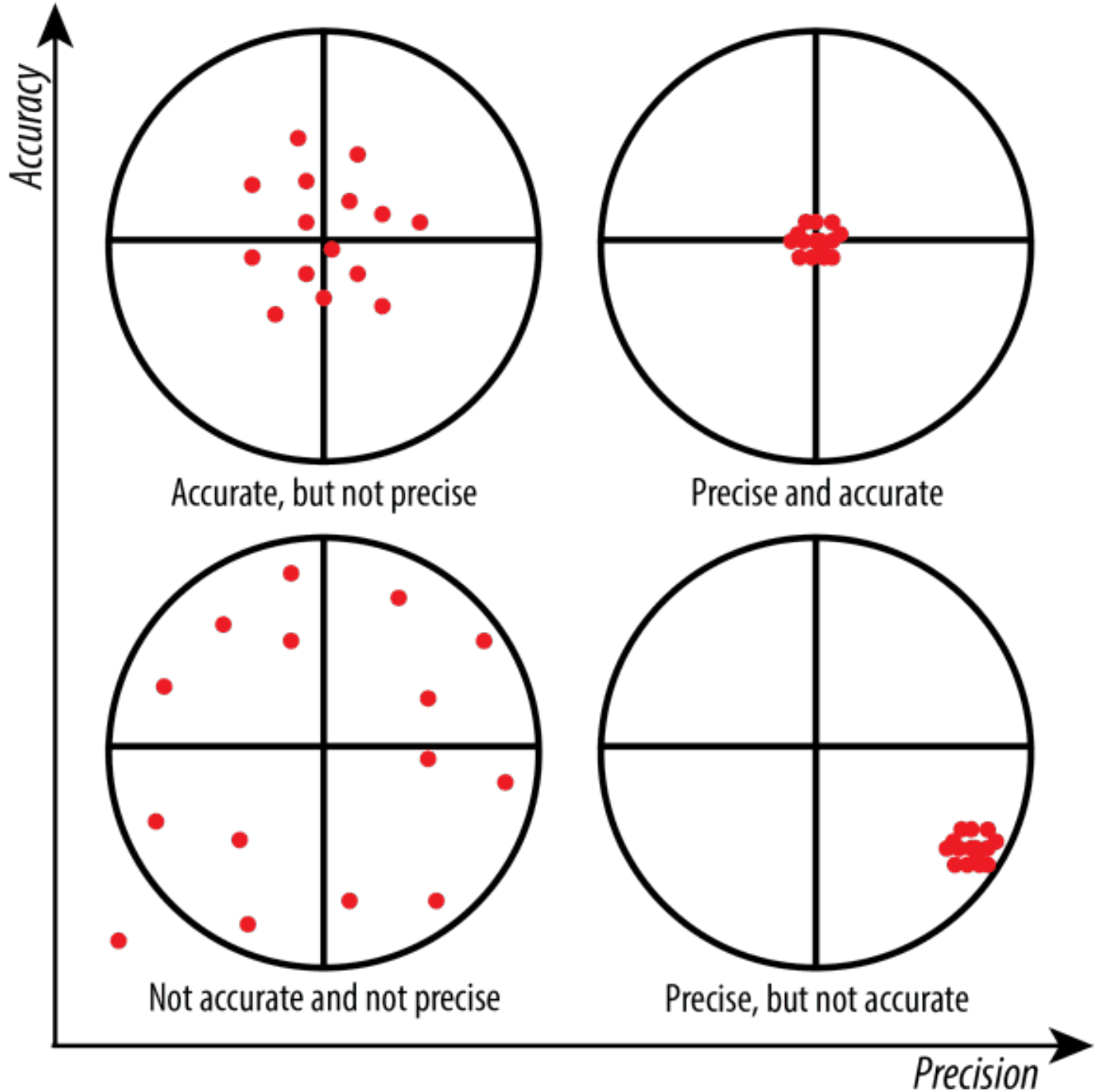
- Reproducibility or repeatability
  - Agreement between repeated measures
- Intra-observer variability
  - agreement with your previous interpretation
- Inter-observer variability
  - agreement between observers



Accuracy

- Closeness of measurements to a specific value
  - To what extent does the test give the right answer
  - Requires a gold standard (definitive assessment)
- Measures of accuracy
  - Sensitivity and specificity
  - positive and negative predictive values
  - receiver operating characteristic (ROC) curve
  - likelihood ratio





https://www.google.com/url?sa=i&url=https%3A%2F%2Fwp.stolaf.edu%2Fit%2Fgis-precision-accuracy%2F&psig=AOvVaw1n_ExUOnO-v4UWNFEReZIr&ust=1667082302398000&source=images&cd=vfe&ved=0CAwQjRxqFwoTCJi63NP7g_sCFQAAAAAdAAAAABAX

Sensitivity & Specificity

- Sensitivity
  - proportion of positive tests out of total disease
  - Given you have the disease, proportion that have a positive test (T+|D+)
  - correctly identified positives
  - true-positive rate
  - The probability that a person with the disease is classified correctly by the test
- Specificity
  - proportion of negative tests out of total non-diseased
  - Given you don’t have the disease, proportion that have a negative test (T-|D-)
  - correctly identified negatives
  - true-negative rate
  - The probability that a person without the disease is classified correctly by the test

Dichotomous Outcome and Test Result 2x2 Contingency Table

	Disease present	Disease absent
Positive test	True positive	False positive
Negative test	False negative	True negative

Calculating Sensitivity and Specificity

	Disease present	Disease absent
Positive test	True positive	False positive
Negative test	False negative	True negative

	Stroke	No stroke	
CT = stroke	56	3	59
CT = no stroke	161	136	297
	217	139	356

Sensitivity: true positives/all stroke = 56/217 = 26%
Specificity: true negatives/all without stroke= 136/139 = 98%

Magnetic resonance imaging and computer tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison." Chalela J, Kidwell CS, Nentwich LM, Luby M, Butman JA, Demchuk AM, Hill MD, Patronas N, Latour L, Warach S. The Lancet, Vol. 369, January 27, 2007, pp. 293-298.

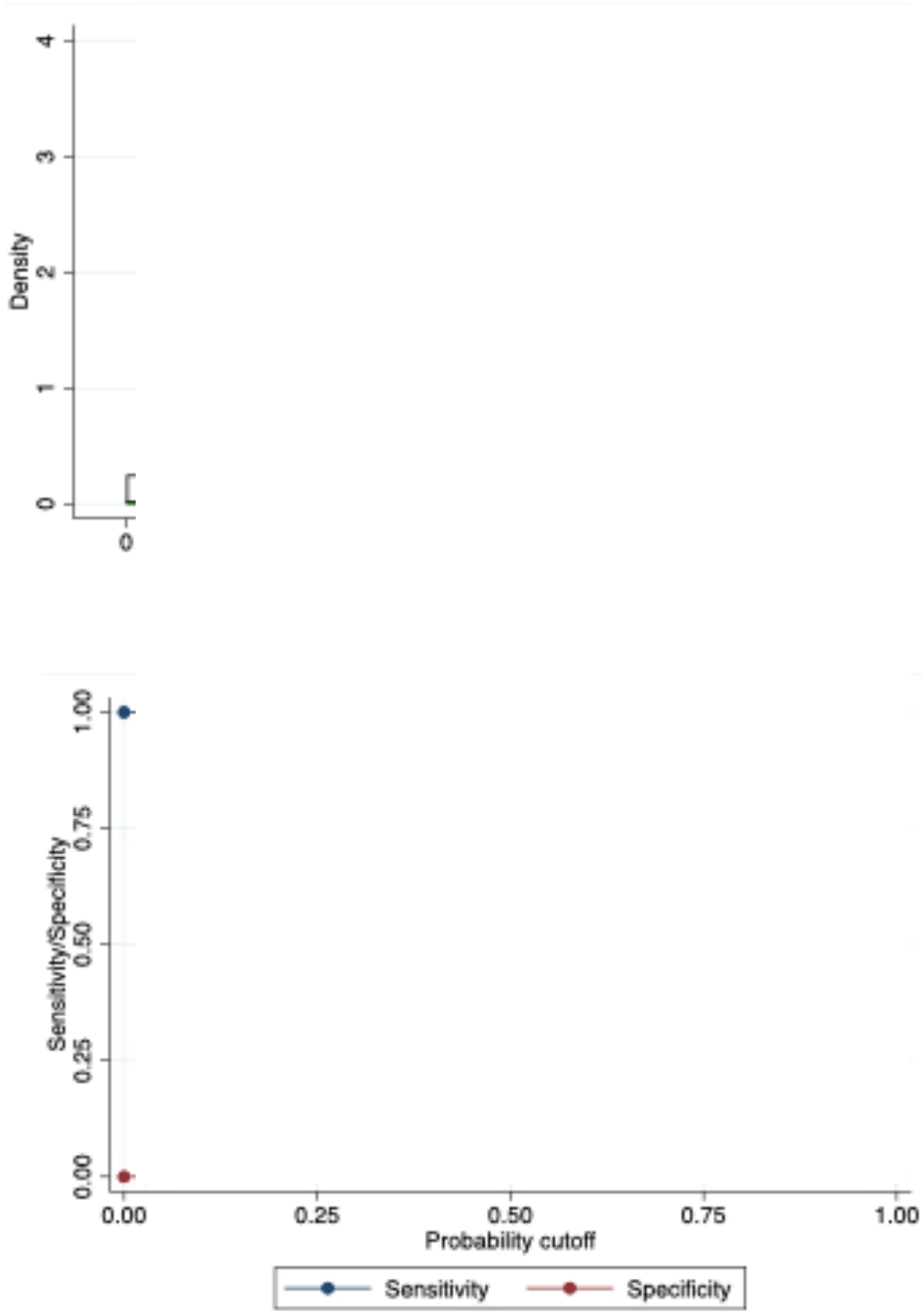
Sensitivity & Specificity: classification

- Sensitivity and Specificity tell you about misclassification errors
- Studies that display results as sensitivity and specificity are Validation Studies
  - Step 1: obtain a sample of people with and without a disease
  - Step 2: administer a test or procedure to classify them
  - Step 3: compare the results of the classification to a “gold standard” and construct a 4x4 table

Sensitivity and Specificity - Challenges

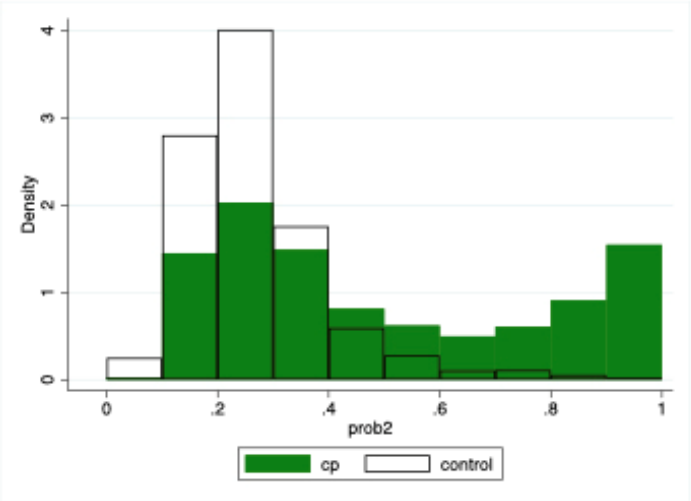
- Never consider these two parameters separately
  - Trade off between sensitivity and specificity
  - As one increases, the other decreases
  - e.g. higher cutoff leads to increased specificity but decreased sensitivity
- A highly sensitive test is prone to false-positives
  - incorrectly label someone as having the disease
- A highly specific test is prone to false-negatives
  - fail to identify disease
- What is **important** to you?
  - Avoid missing someone or avoid incorrectly labelling someone?

Trade off Between Sensitivity and Specificity

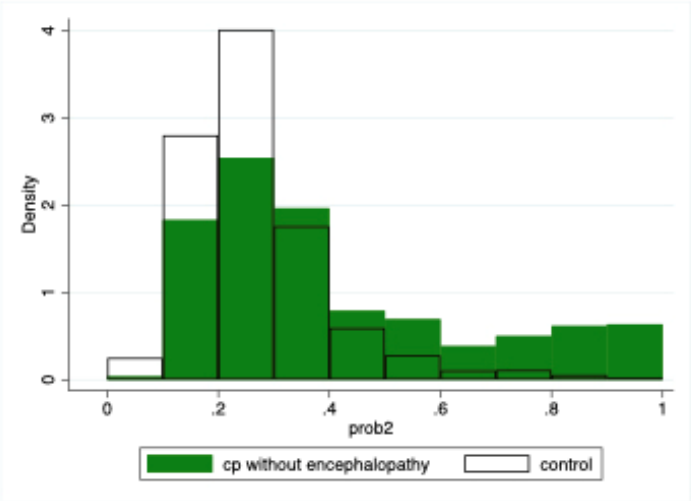


Trade off Between Sensitivity and Specificity

Ai



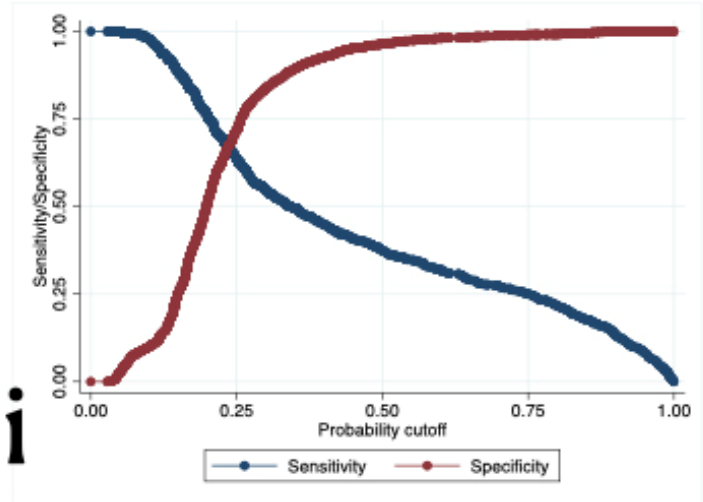
Bi



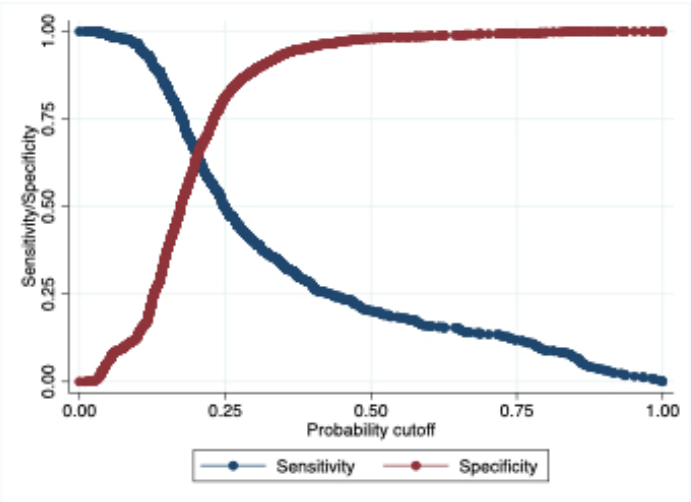
Aii

Bii

Aiii



Biii



Sensitivity and Specificity - Challenges

- Affected by severity of disease
  - results from a CXR for detection of lung cancer will depend on severity of illness and stage of the disease, size of the tumour etc.
- Sensitivity and specificity describe how well a **test** performs
  - Don’t convey significance of the test result for an individual **patient**

Likelihood Ratio (Positive)

- Assesses potential **utility** of a diagnostic test
  - Assesses how likely the patient with a positive test has the disease
- Probability of positive test given disease relative to probability of positive test given no disease (**true positives/false positives**)
- Answers question: How much more likely is a positive test result in the presence of disease compared with absence of disease?
- LR = sensitivity/(1-specificity)
- Answer is an odds

Negative Likelihood Ratio

- Probability that a person with the disease tested negative/ probability that a person without the disease tested negative
- 1-Sensitivity (false negative rate)/Specificity (true negative rate)

Likelihood Ratio

- Has predictive value and stable with changes in prevalence
- Ranges from zero to infinity
- The higher the value, the more likely the patient has the condition
  - 0 - 1 = decreased evidence for disease
  - 1 = no diagnostic value
  - >1 = increased evidence for disease

Likelihood Ratio Example

Serum Ferritin (mg/dL)	LR (of iron deficiency anemia)
<15	51.8
15-24	8.8
25-34	2.5
45-100	0.5
>100	0.08

Sloane 2008

Liklihood ratio example

	Stroke	No stroke	
CT = stroke	56	3	59
CT = no stroke	161	136	297
	217	139	356

True positives/false positives = (56/217)/(3/139) = 0.258/0.0216 = 12
Sensitivity/(1-specificity) = (56/217)/(1-(136/139) = 0.258/(1-0.978) = 12

Prediction

- Predictive values
- Ability of a diagnostic test **to make a diagnosis** in the future
- Positive predictive value (PPV)
  - proportion of diseased with positive test result
  - proportion of **people with a positive test who have the disease**
- Negative predictive value (NPV)
  - proportion of healthy individuals with a negative test result
  - proportion of people with a **negative test who are free of disease**

Prediction

- A test with a high positive predictive value makes the disease quite likely in a subject with a positive test
- A test with a high negative predictive value makes the disease quite unlikely in a subject with a negative test

Positive predictive value (PPV) = true positive tests/all positive tests
Negative predictive value (NPV) = true negative tests/all negative tests

Prediction

	Disease present	Disease absent
Positive test	True positive	False positive
Negative test	False negative	True negative

	Stroke	No stroke	
CT = stroke	56	3	59
CT = no stroke	161	136	297
	217	139	356

Positive predictive value (PPV) = true positive tests/all positive tests
Negative predictive value (NPV) = true negative tests/all negative tests

PPV (true positives/all positives)= 56/59 = 95%
NPV (true negatives/all negatives)= 136/297 = 46%

Magnetic resonance imaging and computer tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison." Chalela J, Kidwell CS, Nentwich LM, Luby M, Butman JA, Demchuk AM, Hill MD, Patronas N, Latour L, Warach S. The Lancet, Vol. 369, January 27, 2007, pp. 293-298.

Calculating Sensitivity and Specificity

	Disease present	Disease absent
Positive test	True positive	False positive
Negative test	False negative	True negative

	Stroke	No stroke	
CT = stroke	56	3	59
CT = no stroke	161	136	297
	217	139	356

Sensitivity: true positives/all stroke = 56/217 = 26%
Specificity: true negatives/all without stroke= 136/139 = 98%

Magnetic resonance imaging and computer tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison." Chalela J, Kidwell CS, Nentwich LM, Luby M, Butman JA, Demchuk AM, Hill MD, Patronas N, Latour L, Warach S. The Lancet, Vol. 369, January 27, 2007, pp. 293-298.

SPin & SNout

- SPecific tests that are POSITIVE rule IN disease
  - Low rate of false positives (true negative rate is high)
- SeNsitive tests that are NEGATIVE rule OUT disease
  - Low rate of false negatives

	Stroke	No stroke	
CT = stroke	56	3	59
CT = no stroke	161	136	297
	217	139	356

Sensitivity: true positives/all stroke = 56/217 = 26%
Specificity: true negatives/all without stroke= 136/139 = 98%

PPV (true positives/all positives)= 56/59 = 95%
NPV (true negatives/all negatives)= 136/297 = 46%

Predictive values - Challenges

- Cannot be used in case-control studies
  - used for random samples or cohorts where **observed prevalence is equivalent to true prevalence**
- *Affected by prevalence* (proportion of subjects with disease)
  - high prevalence
    - PPV increases and NPV decreases
  - low prevalence
    - PPV decreases, NPV increases
- Less portable from population to population
  - due to effect of prevalence

What are all these terms again?

Test	Numerator	Denominator	Goal
Sensitivity	Positive tests in those with disease (true positives)	All with disease	In those with disease, what proportion will test positive?
Specificity	Negative tests in those without disease (true negatives)	All without disease	In those without disease, what proportion test negative?
Likelihood ratio	Sensitivity (true pos rate)	1-specificity (false pos rate)	How much more likely is disease if test is positive?
Positive predictive value	Positive tests in those with disease (true positives)	All positive tests	What proportion with a positive test actually have the disease?
Negative predictive value	Negative tests in those without disease	All negative tests	What proportion with a negative test don’t have the disease?

Effect of prevalence

	Stroke	No stroke	
CT = stroke	56	3	59
CT = no stroke	161	136	297
	217	139	356

Prevalence = 217/356 = 61%

Sensitivity (true positives/all stroke) = 56/217 = 26%

Specificity (true negatives/all without stroke)= 136/139 = 98%

PPV (true positives/all positives)= 56/59 = 95%

NPV (true negatives/all negatives)= 136/297 = 46%

	Stroke	No stroke	
CT = stroke	6	7	13
CT = no stroke	16	327	343
	22	334	356

Prevalence = 22/356 = 6%

Sensitivity (true positives/all stroke) = 6/22 = 26%

Specificity (true negatives/all without stroke)= 327/334 = 98%

PPV (true positives/all positives)= 6/13 = 46%

NPV (true negatives/all negatives)= 327/343 = 95%

	Stroke	No stroke	
CT = stroke	83	1	84
CT = no stroke	236	35	271
	320	36	356

Prevalence = 320/356 = 90%

Sensitivity (true positives/all stroke) = 83/320 = 26%

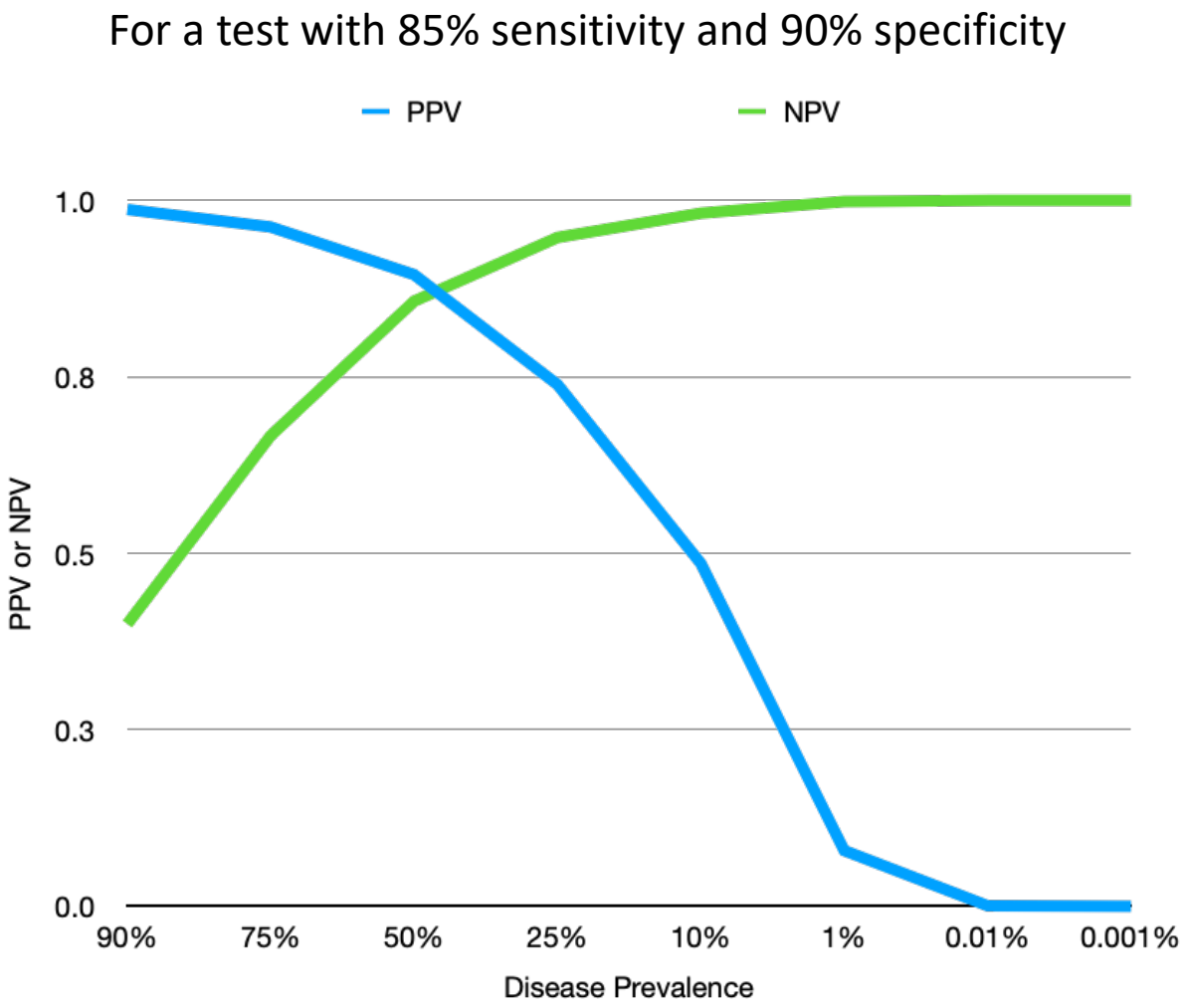
Specificity (true negatives/all without stroke)= 35/36 = 98%

PPV (true positives/all positives)= 83/84 = 99%

NPV (true negatives/all negatives)= 35/271 = 13%

Effect of Prevalence on PPV and NPV

Prevalence	Sensitivity	Specificity	PPV	NPV
90%	0.85	0.9	0.987097	0.400000
75%	0.85	0.9	0.962264	0.666667
50%	0.85	0.9	0.894737	0.857143
25%	0.85	0.9	0.739130	0.947368
10%	0.85	0.9	0.485714	0.981818
1%	0.85	0.9	0.079070	0.998319
0.01%	0.85	0.9	0.000849	0.999983
0.001%	0.85	0.9	0.000085	0.999998



Effect of prevalence

- Example of newborn screening for congenital hypothyroidism
- Amazing test
- But low prevalence = low PPV

	Cord sampling	Heel-stick sampling
Sensitivity	100%	100%
Specificity	99.6%	98.3%
Recall rate	0.04%	1.7%
Positive predictive value	7.95%	2.30%

https://www.researchgate.net/publication/336248581_Cord_blood_versus_heel-stick_sampling_for_measuring_thyroid_stimulating_hormone_for_newborn_screening_of_congenital_hypothyroidism/figures?lo=1

Prevalence and Diagnostic Tests

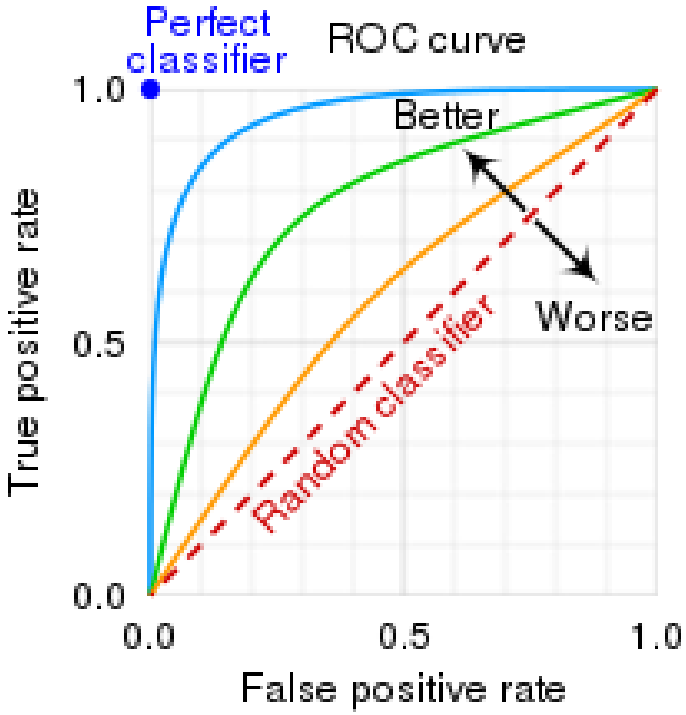
- Diagnostic tests function best when prevalence is between 40-60%
  - Chose the right population to test
- **Function poorly at extremes of prevalence**
- “When you are already pretty sure that the patient either does or does not have the diagnosis in question, additional testing may not alter that probability very much”
- e.g. ECHO for endocarditis or chest CT for pulmonary embolus

Summary of terms

- Sensitivity and specificity
  - How good is the test compared to gold standard?
- Likelihood ratio
  - How much more likely is a positive test result in the presence of disease compared with absence of disease? (true positives/false positives)
- Predictive value
  - Given a test result, what is the probability of actually having the disease?

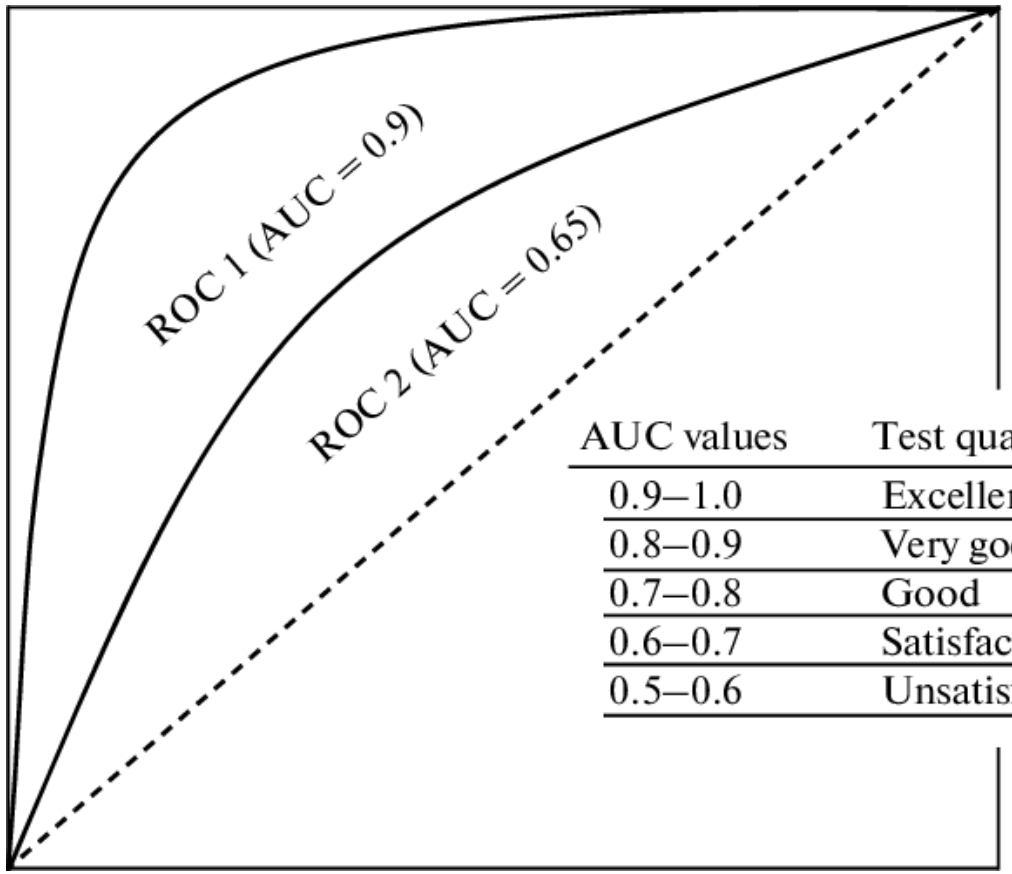
Receiver Operating Characteristic (ROC) Curves

- Test result is not simply positive or negative
- Continuous test results
- Potentially multiple cutoffs
- Sensitivity (Y-axis) vs. 1-specificity (X-axis)
- Best cut-off maximizes sensitivity and specificity
  - 1 = perfect test
  - 0.5 = useless test (equivalent to random chance)
- Quantifies information gain for a test
- Provides summary **estimate of the accuracy** of the test



Area Under the ROC Curve (AUC)

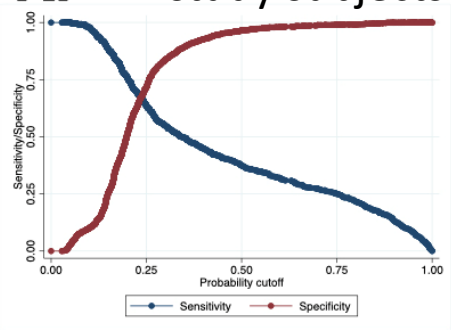
- Values between 0.0 and 1.0
  - perfectly inaccurate to perfectly accurate
  - 0.5 = useless test



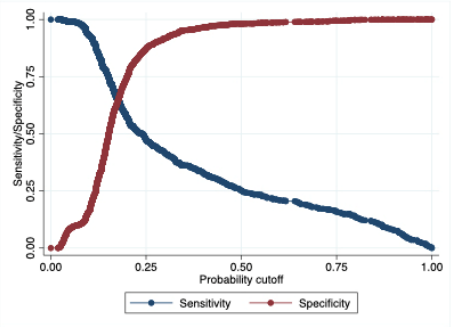
AUC values	Test quality
0.9–1.0	Excellent
0.8–0.9	Very good
0.7–0.8	Good
0.6–0.7	Satisfactory
0.5–0.6	Unsatisfactory

Examples of ROC Curves

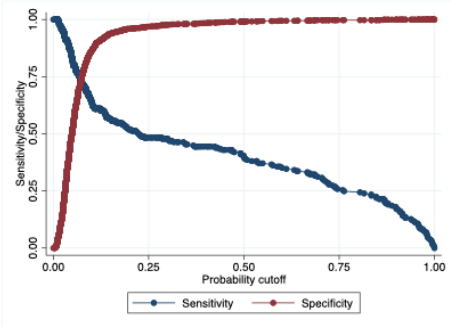
Ai All study subjects



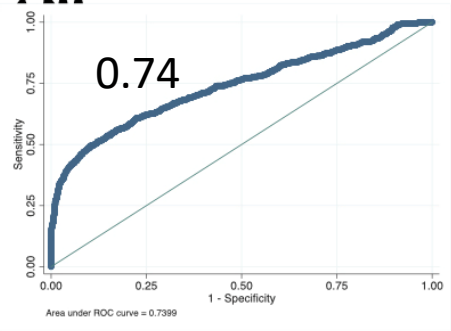
Bi Controls vs. mild CP



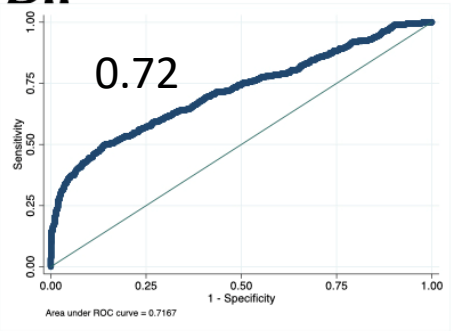
Ci Controls vs. severe CP



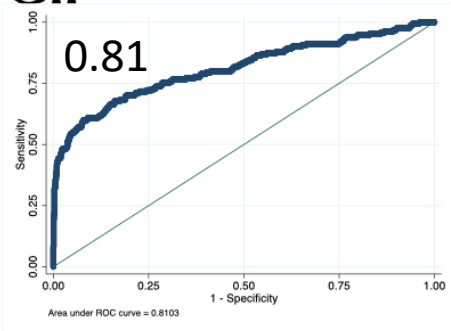
Aii



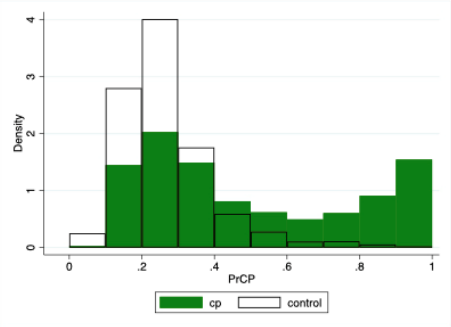
Bii



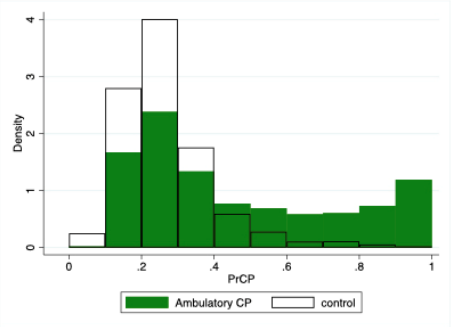
Cii



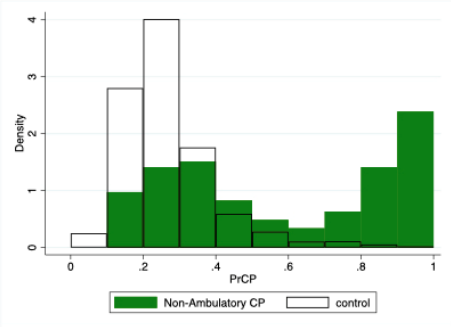
Aiii



Biii



Ciii



Additional Considerations

- Cost
- Availability
- Acceptability
  - i.e. invasive test with potentially serious complications
- Clinical utility
  - ideally assessed using a RCT
    - assess outcomes
    - document adverse events
    - assess impact on decision-making
    - assess patient satisfaction and cost-effectiveness

Example

JAMA Pediatrics | Original Investigation

Development of a Bedside Tool to Predict the Diagnosis of Cerebral Palsy in Term-Born Neonates

Amira Rouabhi; Nafisa Husein, MSc; Deborah Dewey, PhD; Nicole Letourneau, PhD, RN; Thierry Daboval, MD;
Maryam Oskoui, MDCM; Adam Kirton, MD; Michael Shevell, MDCM; Mary J. Dunbar, MD;
for the Canadian Cerebral Palsy Registry

Why a new test?

- Cerebral palsy is an impairment of motor development due to a static abnormality of the CNS that occurs before the age of 1 (ie, in development)
- Affects ~1/500 children
- CP is a clinical diagnosis
- CP takes time to become apparent due to maturation of the CNS
- Early interventions improve outcomes
- How can we identify children at risk?
  - Term infants with encephalopathy at birth ~12% develop CP

Classic risk factors

- Prematurity (~40%)
- Bad delivery (~10-20%)
- These children are easy to identify and follow
- But these account for a minority of CP cases (~50%)
- What about the rest?

Study

- Canadian Cerebral Palsy registry = cases = 1265
- APrON (Alberta Pregnancy Outcomes and Nutrition) = controls = 1985
- Look a common elements and try to find ones specific to CP

		Controls (n=1985)		CP (n=1265)		Univariable			Multivariable (45 multiple imputations)		
		Number (%) or median (IQR)	Missing (%)	Number (%) or median (IQR)	Missing (%)	OR	95% CI	P-value	OR	95% CI	Standardized dominance statistic (ranking)
Pregnancy and maternal characteristics	Maternal age (years)	31 (29-34) n=1936	47 (2.4%)	30 (26-33) n= 1246	19 (1.5%)	N/A			N/A		
	Number of pregnancies	2 (1-3) n=1985	0 (0%)	2 (1-3) n=1245	20 (1.6%)	1.2	1.2-1.3	<0.0001	1.4	1.3-1.5	0.041 (6)
	History of miscarriage	464/1985 (23.4%)	0 (0%)	308/1232 (25%)	33 (2.5%)	1.1	0.92-1.3	0.3	Not significant		
	Number of miscarriages	0 (0-0) n=1985	0 (0%)	0 (0-0.5) n=1232	33 (2.6%)	1.1	0.97-1.2	0.25	0.75	0.64-0.87	0.0075 (13)
	Tobacco use	110/1835 (6.0%)	150 (7.6%)	202/1128 (17.9%)	137 (10.8%)	3.1	2.4-4.0	<0.0001	2.3	1.7-3.0	0.078 (4)
	Alcohol use	130/1802 (7.2%)	183 (9.3%)	143/1120 (12.8%)	145 (11.5%)	1.7	1.32-2.2	<0.0001	Not significant		
	Drug use	14/1843 (0.8%)	142 (7.2%)	132/1224 (10.8%)	41 (3.2%)	15.8	9.0-29.8	<0.0001	10.4	6.1-18.0	0.15 (3)
	Diabetes	104/1983 (5.2%)	2 (0.1%)	74/12330 (6.0%)	35 (2.8%)	2.4	1.7-3.3	<0.0001	2.1	1.5-3.0	0.039 (7)
	Pre-eclampsia	13/1983 (0.7%)	2 (0.1%)	46/1173 (3.9%)	92 (7.3%)	6.2	3.3-12.6	<0.0001	4.0	2.0-8.0	0.037 (9)
Labor and Delivery characteristics	Prolonged rupture of membranes (>18hrs)	234/1976 (11.8%)	9 (0.5%)	90/1172 (7.7%)	93 (7.4%)	0.62	0.48-0.80	0.0002	0.5	0.37-0.69	0.053 (5)
	Chorioamnionitis	8/1985 (0.4%)	0 (0%)	79/808 (9.8%)	457 (36.1%)	26.8	12.9-64.4	<0.0001	15.4	6.9-39.1	0.21 (2)
	5-minute Apgar Score	9 (9-9) n=1981	4 (0.2%)	9 (7-9) n=1159	106 (8.4%)	0.63	0.59-0.67	<0.0001	0.64	0.60-0.70	0.31 (1)
	Cord pH	7.26 (7.21-7.3) n=1691	294 (14.8%)	7.25 (7.14-7.3) n= 810	455 (36.0%)	0.03	0.014-0.61	0.0001	Not significant		
	Maternal fever in labor	77/1985 (3.9%)	0 (0%)	87/1049 (8.3%)	216 (17.1%)	2.2	1.6-3.1	<0.0001	Not significant		
	Emergency Caesarian section	244/1985 (12.3%)	0 (0%)	300/461 (65%)	804 (63.6%)	13.3	10.5-16.9	<0.0001	N/A		
Infant characteristics	Male sex	1033/1985 (52%)	0 (0%)	719/1265 (56.8%)	0 (0%)	1.2	1.05-1.4	0.007	1.2	1.0-1.5	0.011 (12)
	Gestational age (weeks)	39.4 (38.7-40.1) n=1936	49 (2.4%)	39 (38-40) n=1228	37 (2.9%)	0.82	0.78-0.87	<0.0001	0.89	0.83-0.97	0.037 (8)
	Birth weight (kg)	3.4 (3.1-3.69) n=1983	2 (0.1%)	3.3 (2.98-3.65) n=1219	46 (3.6%)	0.66	0.57-0.76	<0.0001	0.11	0.02-0.59	0.17 (10)
	Birth weight (kg^2)								1.3	1.04-1.66	0.014 (11)
	Small for Gestational Age	95/1934 (4.9%)	51 (2.6%)	126/1227 (10.3%)	38 (3.0%)	2.2	1.7-3.0	<0.0001	Not significant		
	Encephalopathy	0/1985 (0%)	0 (0%)	335/1184 (28.3%)	81 (6.4%)	*786.4	139-31160	<0.0001	N/A		

Figure 3. Proposed Prognostic Tool Interface and Interpretation for a Range of Predicted Cerebral Palsy (CP) Probability Results

A Proposed prognostic tool interface

Prediction of cerebral palsy diagnosis in term infants

Diabetes during pregnancy	No	Yes
Preeclampsia	No	Yes
Chorioamnionitis	No	Yes
Prolonged rupture of membranes	No	Yes
Tobacco use during pregnancy	No	Yes
Illicit drug use during pregnancy	No	Yes
Male sex	No	Yes
5-minute Apgar score	10	
Gestational age at delivery (wk)	39	
Birthweight (kg)	3.3	
Number of pregnancies	3	
Number of miscarriages	1	
Probability of CP	0.21	
Result indicates	Lower risk than baseline	

B Interpretation for a range of predicted CP probability results

Prediction ranges (probability of CP)	Relative risk of CP (95% CI)	Interpretation/recommendation	Expected % of all neonates in this range	Expected cases of CP in this range
0-0.3	0.50 (0.45-0.54)	Lower risk compared with baseline, screen for parent concern, unable to sit at 9 mo or hand asymmetry <12 mo	70.6	1/1005
0.3-0.5	0.98 (0.86-1.1)	Equal risk compared with baseline, but may benefit from motor screening	23.6	1/511
0.5-0.7	2.9 (2.2-3.7)	Slightly higher risk compared with baseline, should have motor screening	3.9	1/174
0.7-0.9	9 (6.3-12.7)	Higher risk compared with baseline, should have motor screening; consider additional tests	1.7	1/55
>0.9	62 (26.0-149.9)	Much higher risk compared with baseline, should have motor screening; consider additional tests	0.25	1/8

A, Example of possible prognostic screening tool, with an example patient and result. B, Relative risk, expected percentage, and expected cases calculated using model sensitivity and specificity at each range applied to a hypothetical population of neonates with a birth prevalence of CP of 1 in 500 (eTable 6 in Supplement 1).

Figure 1. Using the Prognostic Tool Developed Using Multiple Imputation on 3250 Participants on a Sample of 2509 With Complete Data to Determine Potential Thresholds and Assess Model Performance

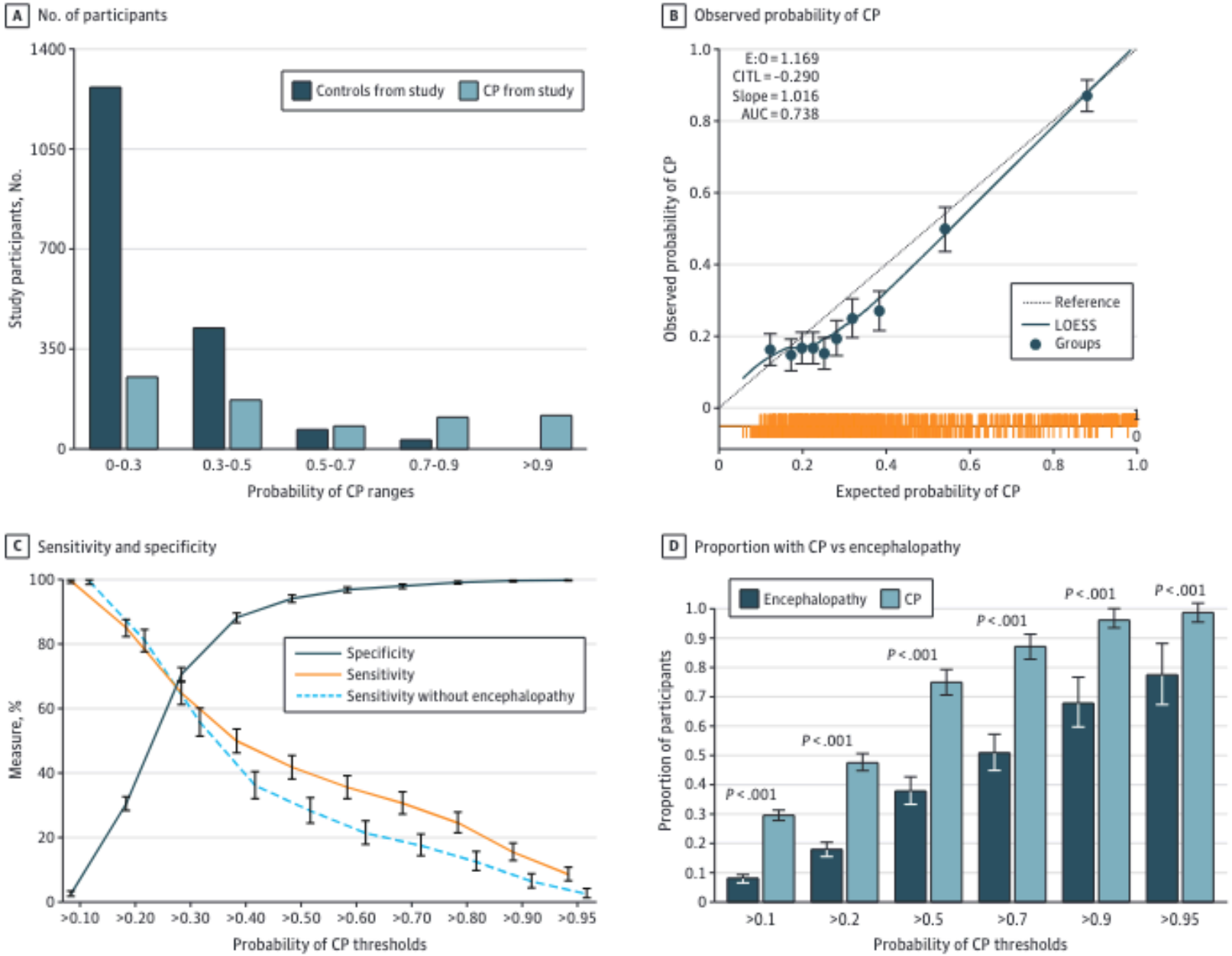


Figure 3. Proposed Prognostic Tool Interface and Interpretation for a Range of Predicted Cerebral Palsy (CP) Probability Results

A Proposed prognostic tool interface

Prediction of cerebral palsy diagnosis in term infants

Diabetes during pregnancy	No	Yes
Preeclampsia	No	Yes
Chorioamnionitis	No	Yes
Prolonged rupture of membranes	No	Yes
Tobacco use during pregnancy	No	Yes
Illicit drug use during pregnancy	No	Yes
Male sex	No	Yes
5-minute Apgar score	10	
Gestational age at delivery (wk)	39	
Birthweight (kg)	3.3	
Number of pregnancies	3	
Number of miscarriages	1	
Probability of CP	0.21	
Result indicates	Lower risk than baseline	

B Interpretation for a range of predicted CP probability results

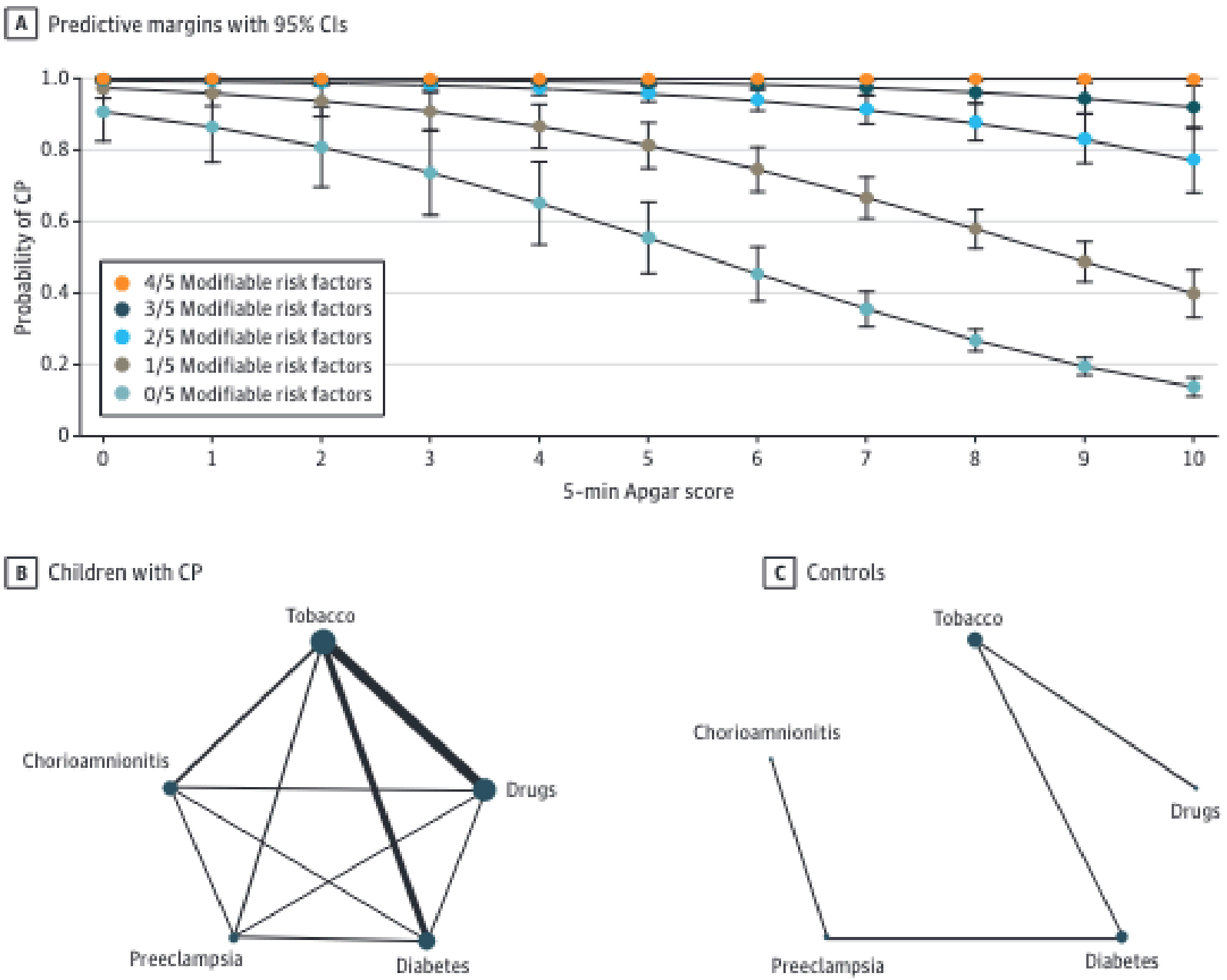
Prediction ranges (probability of CP)	Relative risk of CP (95% CI)	Interpretation/recommendation	Expected % of all neonates in this range	Expected cases of CP in this range
0-0.3	0.50 (0.45-0.54)	Lower risk compared with baseline, screen for parent concern, unable to sit at 9 mo or hand asymmetry <12 mo	70.6	1/1005
0.3-0.5	0.98 (0.86-1.1)	Equal risk compared with baseline, but may benefit from motor screening	23.6	1/511
0.5-0.7	2.9 (2.2-3.7)	Slightly higher risk compared with baseline, should have motor screening	3.9	1/174
0.7-0.9	9 (6.3-12.7)	Higher risk compared with baseline, should have motor screening; consider additional tests	1.7	1/55
>0.9	62 (26.0-149.9)	Much higher risk compared with baseline, should have motor screening; consider additional tests	0.25	1/8

A, Example of possible prognostic screening tool, with an example patient and result. B, Relative risk, expected percentage, and expected cases calculated using model sensitivity and specificity at each range applied to a hypothetical population of neonates with a birth prevalence of CP of 1 in 500 (eTable 6 in Supplement 1).

CAVEAT

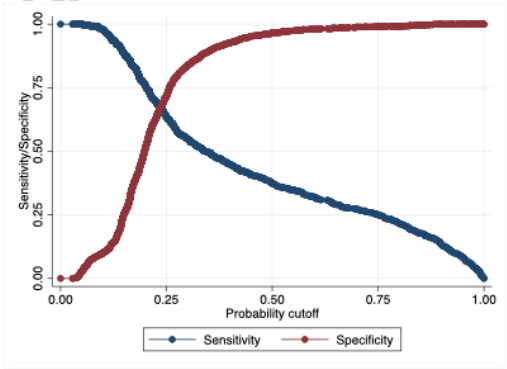
- The ”prevalence” of CP in our study is high! 38%
- This means PPV and NPV are very misleading if we look at the general population! (~0.2%)
  - Recall the PPV and NPV should not be used in a case control study
  - (doesn’t stop the reviewers from asking for it)

Figure 2. Cerebral Palsy (CP) Risk Factors Are Additive

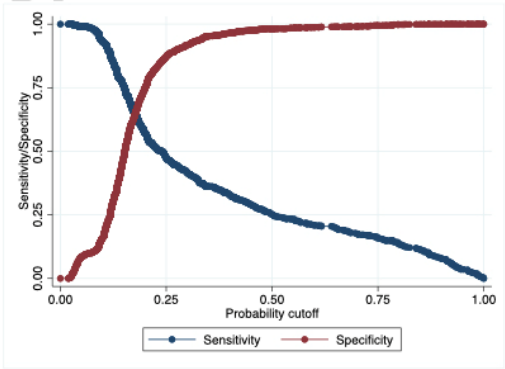


Dose-response

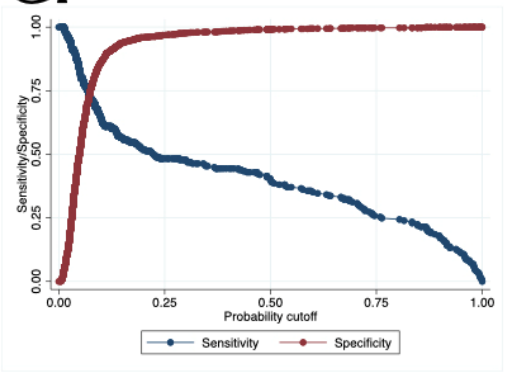
Ai



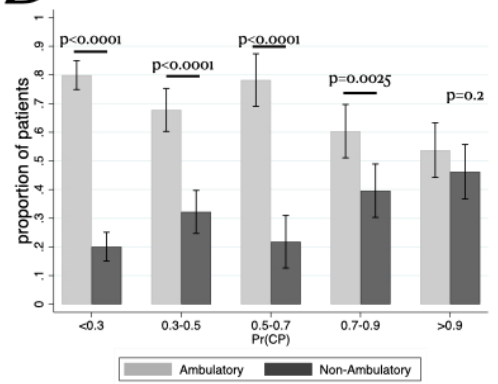
Bi



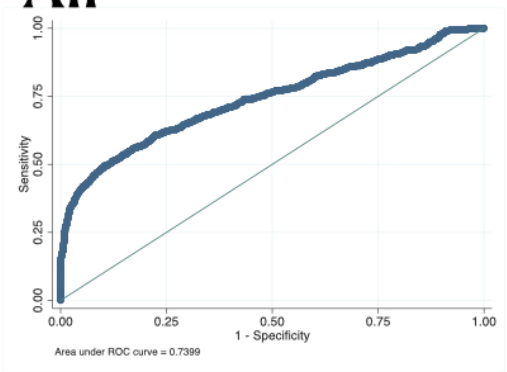
Ci



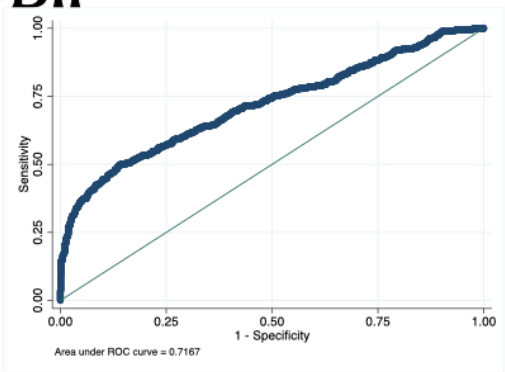
D



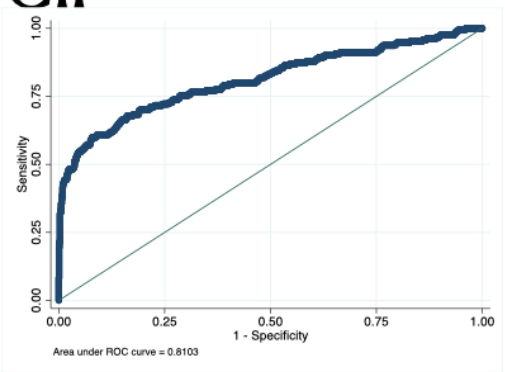
Aii



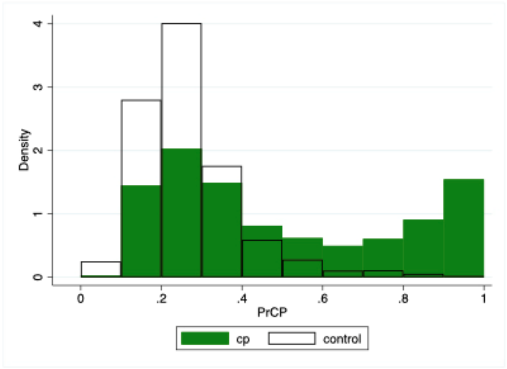
Bii



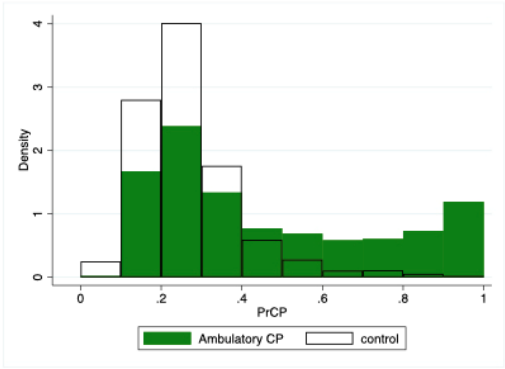
Cii



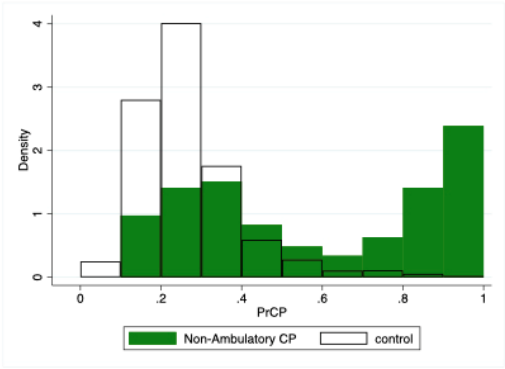
Aiii



Biii



Ciii



Is this acceptable??

- Screening test – want high sensitivity, low specificity
- But low specificity = worried parents, unnecessary tests
- **Acceptability**: Screening is non-invasive (no blood, etc)
- **Availability**: can be done by anyone, most variables will be known
- **Utility**: does this actually identify additional cases of CP???
- **Cost**: tool is free, but requires time; next level screening requires resources
- Next level screening non-invasive (well baby check)
- Tiny subset referred for more intensive screening such as Hammersmith Infant Neurological Examination, General Movements Assessment (can be administered by PTs)

Summary

- Multiple metrics to evaluate a diagnostic test
- Test performance
  - precision (reproducibility) and accuracy
  - sensitivity or specificity
  - likelihood ratio
- Positive and negative predictive values
  - affected by disease prevalence
  - function poorly at the extremes
- ROC curves
  - estimate accuracy of the test for different cutoff values
  - summarized with AUC
- Impact and non-clinical factors

What you want to do	Test(s)	Numerator	Denominator	Sensitive to prevalence?
Evaluate a diagnostic test against another (given a disease, what proportion are correctly diagnosed)	Sensitivity Specificity	Cases correctly identified by test	All with (or without) disease	No
Know what a positive test result means for the patient	Likelihood ratio (positive)	Sensitivity (true positive rate)	1-specificity (false positive rate)	No
Know what a negative test result means for the patient	Likelihood ratio (negative)	1-sensitivity (false negative rate)	Specificity (true negative rate)	No
Given a test result, what proportion are correct?	Positive predictive value Negative predictive value	Cases correctly identified by test	All with positive (or negative) test	Yes
Accuracy of a test with continuous (not binary) results	Receiver operative characteristic	Sensitivity (true positive rate)	1-specificity (fale positive rate)	No

Thanks!

Mary.Dunbar@ahs.ca