Evaluating Diagnostic Procedures

Einas Al-Eisa, MSc, PhD

Outline

- What is diagnosis?
- Relevance
- The ideal diagnostic test
- Validity of diagnostic tests
- Sensitivity
- Specificity
- Predictive value

What is Diagnosis?

"The anatomic, biochemical, physiologic, or psychologic derangement"

DIAGNOSIS  **Labeling Pathology**

What about the role of physiotherapists in diagnosis?

What is Diagnosis?

“Diagnosis is the term which names the primary dysfunction toward which the physical therapist directs treatment” (Sahrmann, 1989)

DIAGNOSIS ⟺ **Planning Treatment**

Example

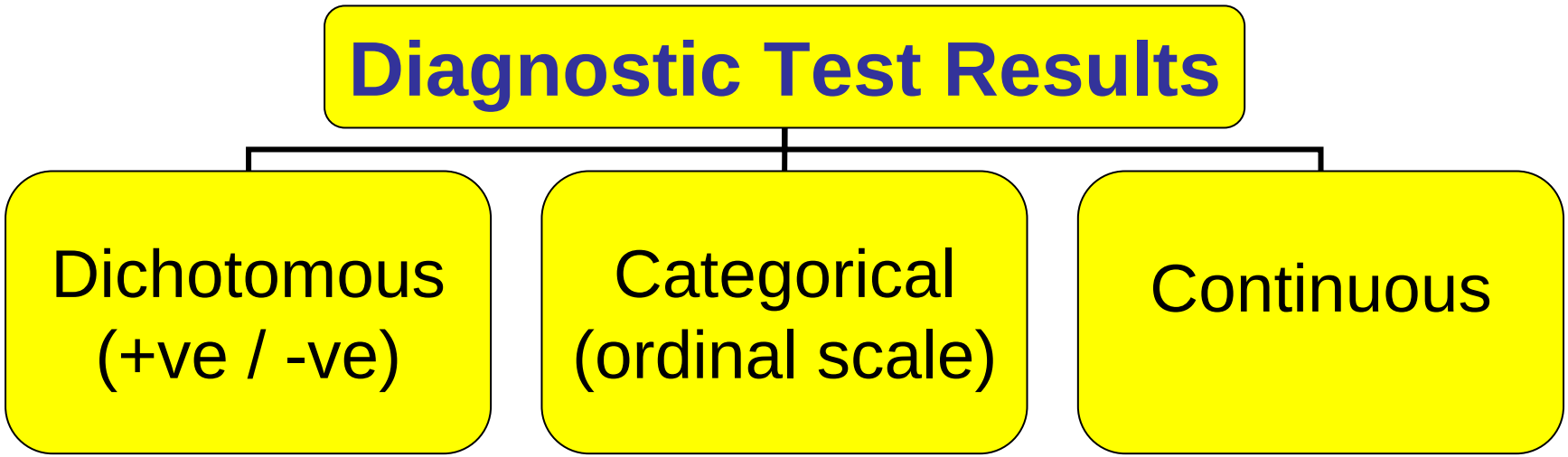
- **Medical Diagnosis:**
  - **Herniated Disc**
- **Physical Therapy Diagnosis:**
  - **Right-sided radiculopathy centralizing with repeated extension**

Example

- **Medical Diagnosis:**
  - **CVA**
- **Physical Therapy Diagnosis:**
  - **Left-sided hemiplegia - Brunnstrom Stage III: all movements in synergy with marked spasticity**

Diagnosis

- Presence or absence of the disease
- Functional deficits
- Identify who would benefit from specific intervention


Diagnostic Test Results
Dichotomous
(+ve / -ve)
Categorical
(ordinal scale)
Continuous

Relevance of diagnostic tests

- Used for clinical decision making
- Involve allocation of resources
- Potential risk to patients

What is the ideal diagnostic test?

- The one accurate in discriminating between those with and without the disease

- Always +ve in someone with the disease
- Always –ve in someone with no disease

Gold Standard

Gold standard

- Concurrent test:
  - X-rays
  - Blood test
- Obtained at a future time:
  - Autopsy
- No gold standard:
  - Long term outcome (e.g., need of further hospitalization, length of stay)

Validity of a test is based on four proportions:

- Sensitivity
- Specificity

- Positive predictive value
- Negative predictive value

		Condition (disease)	
		Yes	No
Test	Positive	True positive (a)	False positive (b)
	Negative	False negative (c)	True negative (d)

	Condition (disease)	
	Yes	**No**
Test Positive	True positive (a)	False positive (b)
Test Negative	False negative (c)	True negative (d)
	Sensitivity= a/(a+c)	**Specificity**= d/(b+d)

Sensitivity

- Proportion of patients with the condition who have a positive test result
- Tests with high sensitivity have few false negatives
- A negative result ***rules out*** the condition

Specificity

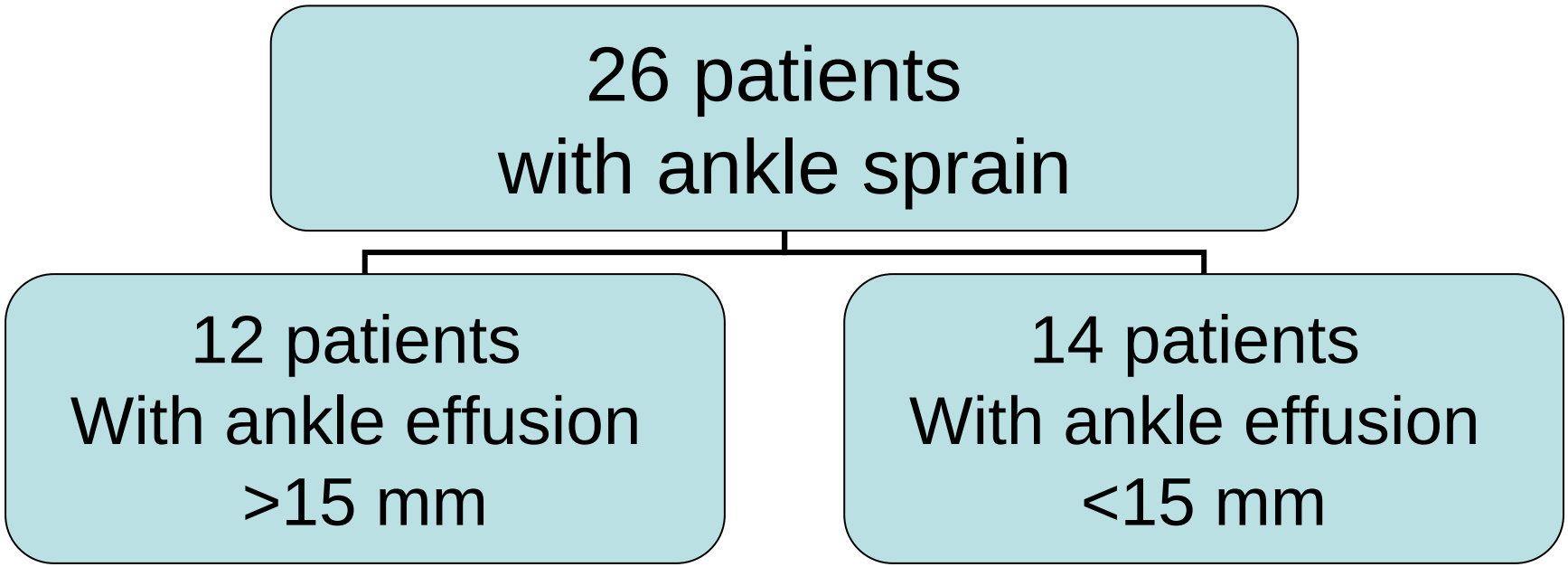
- Proportion of patients without the condition who have a negative test result
- Tests with high specificity have few false positives
- A positive result ***rules in*** the condition

Example

- Clark et al (1996). Improving the detection of radiographically occult ankle fractures: positive predictive value of an ankle joint effusion. ***Clinical Radiol*** ;51:632-636.

- **Gold standard** for identifying ankle fractures was CT of the ankle

- The new test involved measuring the extent of ankle **joint effusion** on the plain radiographs


26 patients
with ankle sprain
12 patients
With ankle effusion
>15 mm
14 patients
With ankle effusion
<15 mm

		Fracture found with CT	
		Yes	No
Ankle effusion found on x-ray	⩾15mm	10 (a)	2 (b)
	<15mm	2 (c)	12 (d)

		Fracture found with CT		
		Yes	No	
Ankle effusion found on x-ray	⩾15mm	10 (a)	2 (b)	12
	<15mm	2 (c)	12 (d)	14
		12	14	

Sensitivity = a/(a+c) = 10/12 = 0.833 = **83.3%**

Specificity = d/(b+d) = 12/14 = 0.857 = **85.7%**

But, what happens if we change the cut-off criteria?

Cut-off ⩾12 mm

- Sensitivity = 100% (all 12 patients with fractures visualized by CT had an effusion of 12mm or more)
- Specificity = 64.3% (because of many false positives)

Cut-off >18 mm

- Sensitivity = 58.3% (because of many false negatives)
- Specificity = 100% (all 14 patients without fractures had an effusion of less than 18 mm)

What’s the best cut-off value?

- cost versus benefits
- What is worst: to predict a storm that does not come (false positive) or fail to predict a storm that does occur (false negative)

What’s the impact of making a mistake?

- False positive : needlessly worrying the healthy
- False negative : falsely reassuring the ill

If the disease being screened for is serious, but treatable in the early stages

need **high sensitivity**

(to lower the probability of false negatives)

If the disease being screened for is less serious, and can be effectively treated even at later stages

need **high specificity**

(to lower the probability of false positives)

Example

- **Balance test** used to predict those at risk of falling
- Individuals with high scores are referred to a balance exercise program
- Would you choose a lower or higher cut-off scores?

Example

- **Balance test** to predict those at risk of falling

↓

Set the cut-off score **low** to avoid false negatives

↓

High sensitivity

What if the test is used to determine the presence of a condition that requires life threatening surgery?


RESCUE

Predictive value

- **Feasibility** = a test must demonstrate that it is an **efficient** use of time and resources and that it yields sufficient number of **accurate** responses to be clinically useful

Positive predictive value

- Estimates the likelihood that a person who tests positive actually have the disease

Negative predictive value

- Indicates the probability that a person who tests negative is actually disease free

	Disease	No disease
Test positive	True positive A	False positive B
Test negative	False negative C	True negative D

(PV+) = a / (a+b)

(PV-) = d / (c+d)

Example

- Amendt et al. (1990). Validity and reliability testing of the scoliometer. ***Physical Therapy***;70:108-117.

Methods

- Trunk angle measured by the scoliometer was used to screen for the presence or absence of scoliosis
- **Gold standard:** radiographs
- **Cut-off:** 5 degrees
- N=34

		Scoliosis with x-ray		
		Yes	No	
Trunk angle	⩾ 5º	15 (a)	13 (b)	28
	<5º	1 (c)	5 (d)	6
		16	18	

Sensitivity = a/(a+c) = 15/16 = **94%**

Specificity = d/(b+d) = 5/18 = **28%**

		Scoliosis with x-ray		
		Yes	No	
Trunk angle	⩾ 5º	15 (a)	13 (b)	**28**
	<5º	1 (c)	5 (d)	**6**
		16	**18**	

PV+ = a/(a+b) = 15/28 = **54%**

PV- = d/(c+d) = 5/6 = **83%**

High sensitivity

↓

Positive cases are identified easily

↓

We will not miss many true cases

↓

It is less likely that a person with
a negative test will have the disease

↓

High negative predictive value

High specificity

↓

Negative cases are identified easily

↓

It is less likely that a person with a positive test will be normal

↓

High positive predictive value


DALLAS
30
يابدوو
yabdoo.com

The Gross Motor Function Classification System for Cerebral Palsy: a study of reliability and stability over time

Ellen Wood* MD MSc FRCP(C), Assistant Professor, Department of Pediatrics, Dalhousie University, Halifax, Nova Scotia;
Peter Rosenbaum MD FRCP(C), Professor, Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada.

**Correspondence to first author at* IWK Grace Health Centre, 5850/5980 University Avenue, PO Box 3070, Halifax, Nova Scotia, Canada B3J 3G9.
E-mail: ewood@is.dal.ca

Cerebral palsy (CP) refers to a group of non-progressive disorders of the development of motor function affecting movement and posture (Bax 1964). CP is caused either by a developmental abnormality of, or an injury to, the immature brain. The incidence of CP is 1.5 to 2.5 per 1000 live births (Aicardi 1992). Although this is a chronic disorder, little is known about the patterns of motor development in children with CP. Many interventions are recommended to the child and their family by many different health professionals, yet there is an absence of objective data to demonstrate that ultimate motor function is improved by these interventions. Without a clear understanding of the natural history of motor development in CP, it is difficult to assess the impact of interventions beyond that improvement in motor function which would have occurred due to normal growth and development; however, the amount of 'natural' change is not well understood.

Many authors have suggested prognostication systems based on a constellation of clinical features to predict eventual motor function, especially independent ambulation. Bleck (1975) and Capute (1979) looked at the presence or absence of seven primitive reflexes to diagnose CP, predict independent walking, and plan interventions. However, neither of these authors reported any reliability or validity data for their criteria. Other authors have examined whether independent sitting by age 2 years would predict later walking ability. Molnar and Gordon (1974) found it was a poor predictor, whereas Watt et al. (1989) reported that independent floor

Purpose

- Measure the inter-rater reliability of the GMFCS
- Assess the stability of a child's GMFCS over time
- Determine the predictive validity and likelihood ratios of the GMFCS in predicting walking

Methods

- Retrospective chart review
- N= 85 children with CP (7 had missing data + 78 had complete data)

Table IV: Time 1 versus Time 4 GMFCS level

GMFCS at Time 1	*GMFCS at Time 4*				
	I	*II*	*III*	*IV*	*V*
I	4	1	1	1	–
II	5	7	9	2	–
III	2	1	5	7	2
IV	2	–	4	9	9
V	–	–	1	1	5

Table IV: Time 1 versus Time 4 GMFCS level

GMFCS at Time 1	*GMFCS at Time 4*				
	I	*II*	*III*	*IV*	*V*
I	4	1	1	1	–
II	5	7	9	2	–
III	2	1	5	7	2
IV	2	–	4	9	9
V	–	–	1	1	5

35 (a)	12 (b)
7 (c)	24 (d)

PV+ = a/(a+b) = 35/47 = **0.74**

PV- = d/(c+d) = 24/31 = **0.77**

Table IV: Time 1 versus Time 4 GMFCS level

GMFCS at Time 1	*GMFCS at Time 4*				
	I	*II*	*III*	*IV*	*V*
I	4	1	1	1	–
II	5	7	9	2	–
III	2	1	5	7	2
IV	2	–	4	9	9
V	–	–	1	1	5

17 (a)	13 (b)
5 (c)	43 (d)

PV+ = a/(a+b) = 17/30 = **0.57**

PV- = d/(c+d) = 43/48 = **0.90**

Table VII: Positive and negative predictive value of GMFCS

Time periods	*Level III combined with*	
	I and II	*IV and V*
Time 1 to 4		
Positive predictive value	0.74	0.57
Negative predictive value	0.77	0.90
Time 2 to 4		
Positive predictive value	0.87	0.62
Negative predictive value	0.94	0.92
Time 3 to 4		
Positive predictive value	0.91	0.80
Negative predictive value	0.89	0.93