Critically appraise evidence OMs
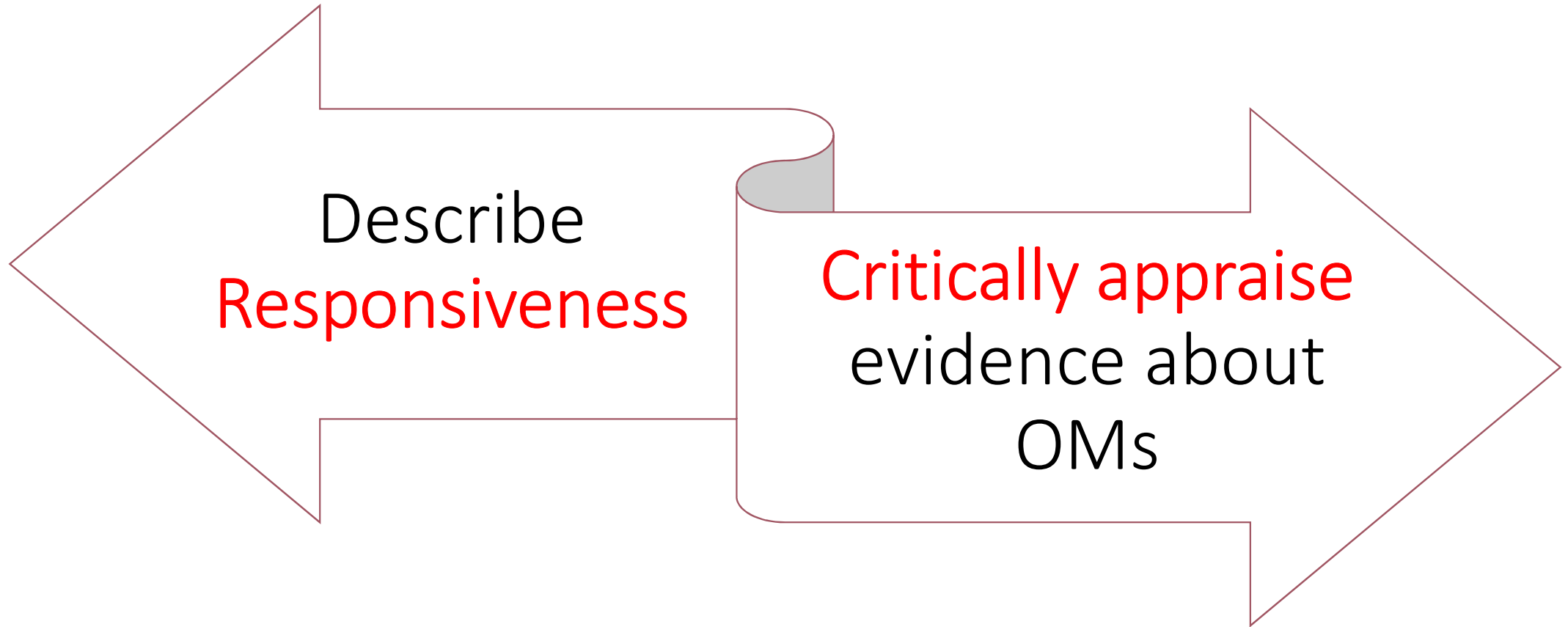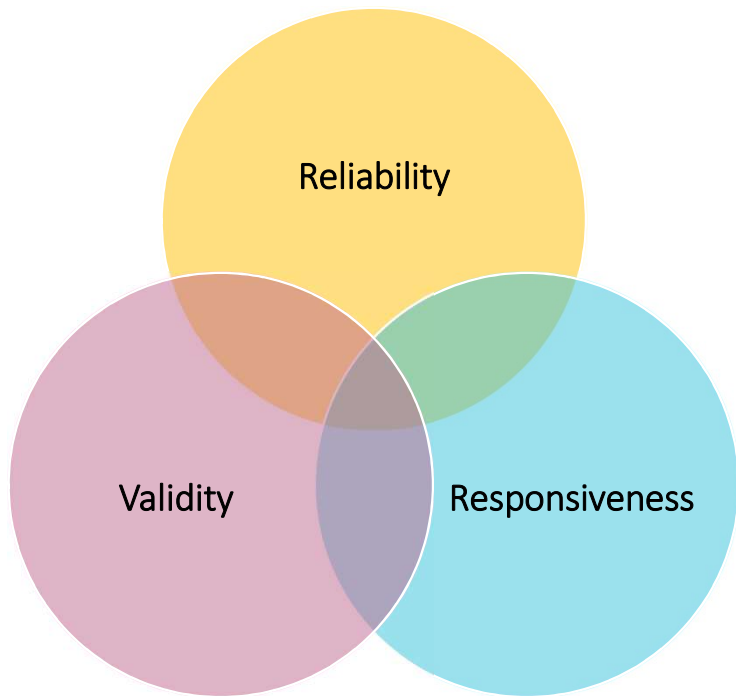
Mohammed TA, Omar Ph.D., PT , CLT

Rehabilitation health Science

CAMS-KSU

Objective

Describe **Responsiveness**

Critically appraise evidence about OMs

Responsiveness

- Responsiveness' has been defined as the ability of an instrument to accurately detect change when it has occurred (clinically meaningful change)

- Responsiveness is an important component of the test's validity, especially if the test is to be used to evaluate the effectiveness of interventions.

- Responsiveness is not a constant characteristic of a measure

Responsiveness

- There is little consensus over how responsiveness should be assessed

- Responsiveness is necessary, but additional information is needed to determine the minimally important difference (MID) for a OMs

- MID has been defined as the smallest change in a PRO measure that is perceived by patients as beneficial or that would result in a change in treatment

- MID is used to interpret whether the observed change is important from the patient's or clinician's perspective.

Assessing the responsiveness

What parameter is
to be measured?

Whose perspective
is important:

Responsive to what?

Unit of analysis?

There are a variety of different ways reported when the responsiveness of a measurement is being analyzed. They include:

Effect size (ES)
Standardized response mean (SRM)
Responsiveness statistics (RS)
Standard error of measurement
Receiver operator characteristic curve.
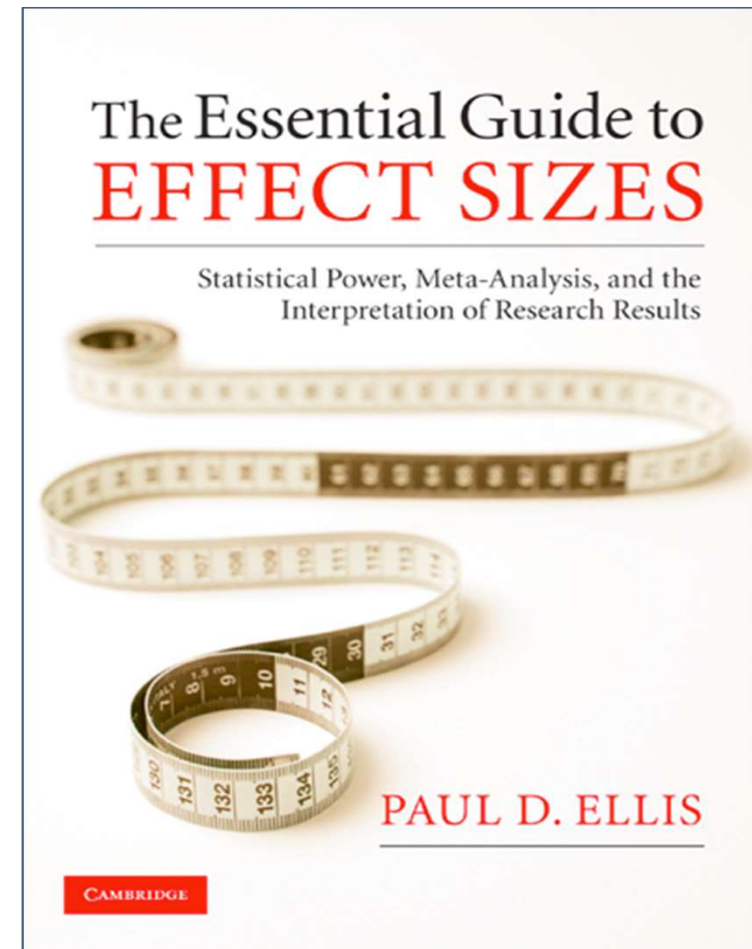
Effect size (ES) "average changes"

Effect size is the magnitude of the difference between groups

❖ Easy to calculate and understand
❖ Used with ratio or interval level of measures
❖ To determine extent of a relationship/difference between variables
❖ Use in case of studying outcome measures and effect of intervention
❖ Enables sample size calculations.
❖ Facilitates comparison between scientific studies

Effect sizes calculation

Statistical significance is the least interesting thing about the results. You should describe the results in terms of measures of magnitude –not just, does a treatment affect people, but how much does it affect them

- *Cohen's d*

- *Pearson's r*

The Essential Guide to
EFFECT SIZES

Statistical Power, Meta-Analysis, and the
Interpretation of Research Results

PAUL D. ELLIS

CAMBRIDGE

Effect sizes

Interpretation

Cohen's d

Effect size based on means or distances between/among means

$$\text{Effect size (ES)} = \frac{average\ changes}{SD\ of\ inital\ scores}$$

0.2 = small

0.5= moderate

0.8 = large

1.3=very large

Interpretation

Pearson's r correlation

Small ±0.2
Medium ±0.5
Large ± 0.8

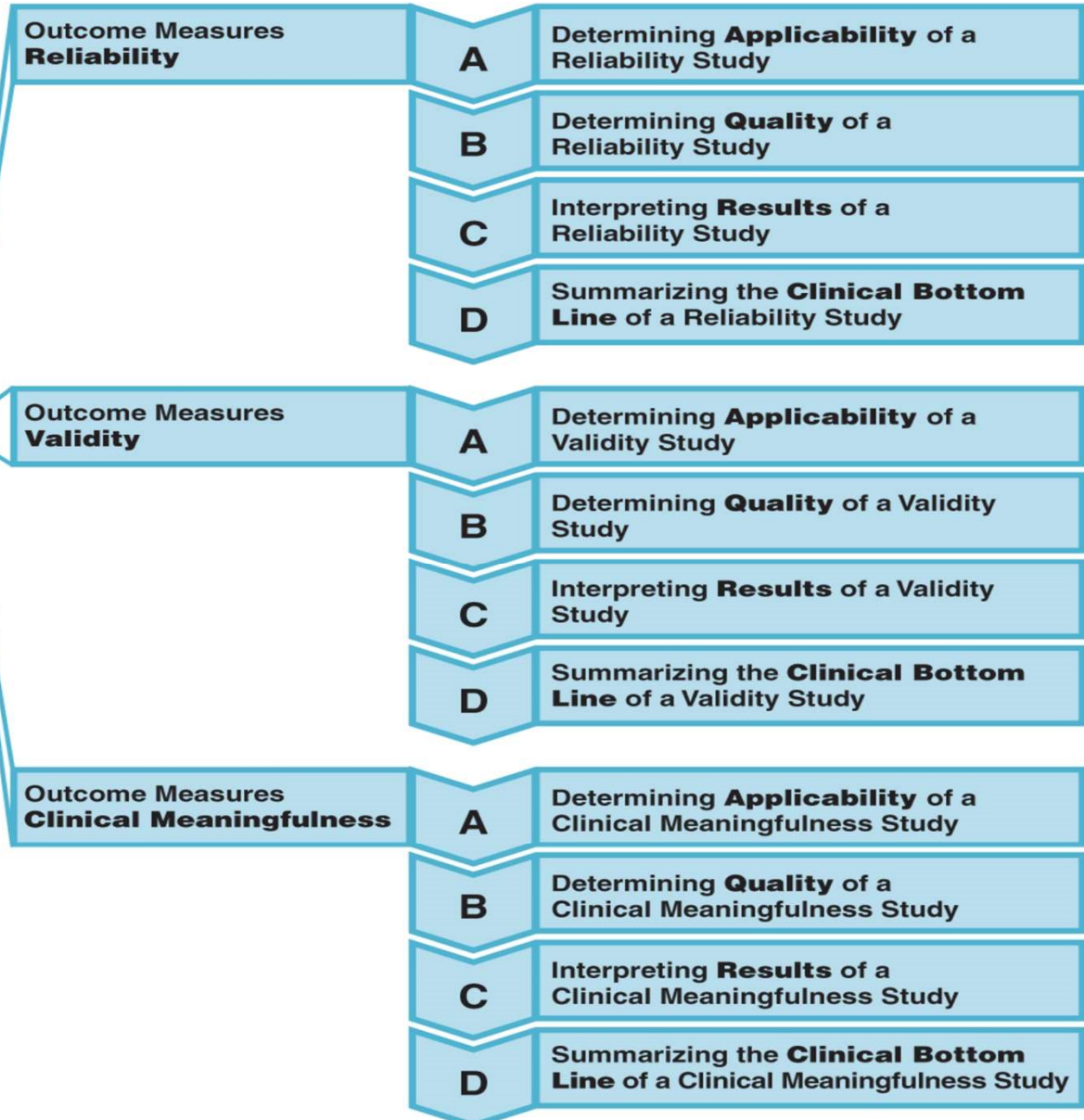# Evidence-Based Approach to Choosing Outcome Measures
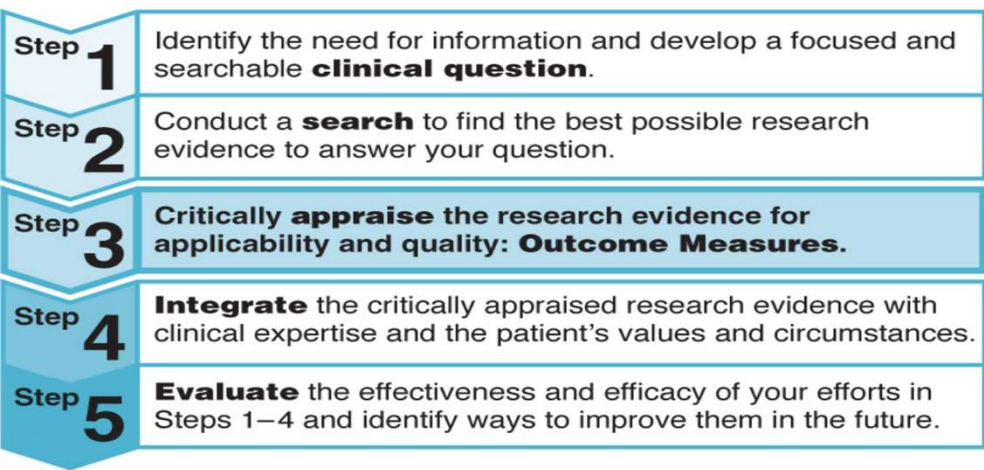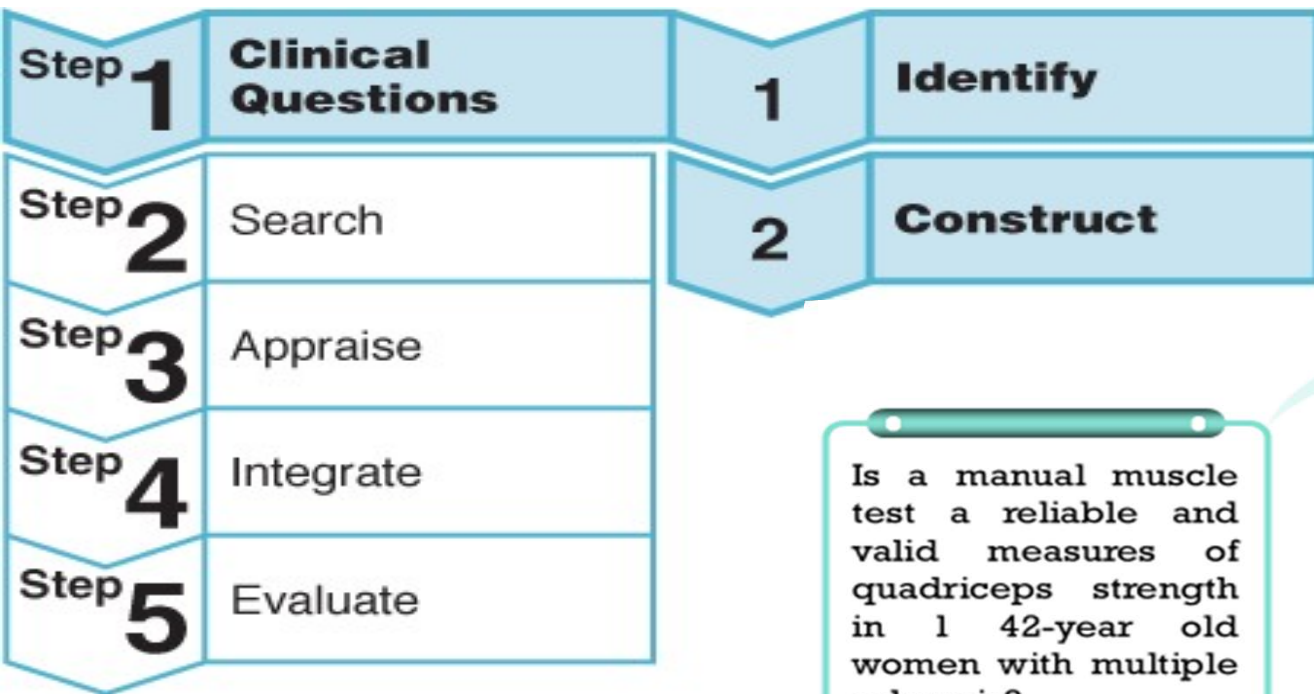
Evidence-Based Approach to Choosing OMs

- An evidence-based approach to selecting outcome measures involves making judgements about the quality of the validity and reliability studies, interpreting the findings and deciding whether they are applicable to one's own specific practice.

Psychometric properties	Relevant Study result	Evidence to patient/client

Step 1 — Identify the need for information and develop a focused and searchable **clinical question**.

Step 2 — Conduct a **search** to find the best possible research evidence to answer your question.

Step 3 — Critically **appraise** the research evidence for applicability and quality: **Outcome Measures**.

Step 4 — **Integrate** the critically appraised research evidence with clinical expertise and the patient's values and circumstances.

Step 5 — **Evaluate** the effectiveness and efficacy of your efforts in Steps 1–4 and identify ways to improve them in the future.

Outcome Measures Reliability

- **A** — Determining **Applicability** of a Reliability Study
- **B** — Determining **Quality** of a Reliability Study
- **C** — Interpreting **Results** of a Reliability Study
- **D** — Summarizing the **Clinical Bottom Line** of a Reliability Study

Outcome Measures Validity

- **A** — Determining **Applicability** of a Validity Study
- **B** — Determining **Quality** of a Validity Study
- **C** — Interpreting **Results** of a Validity Study
- **D** — Summarizing the **Clinical Bottom Line** of a Validity Study

Outcome Measures Clinical Meaningfulness

- **A** — Determining **Applicability** of a Clinical Meaningfulness Study
- **B** — Determining **Quality** of a Clinical Meaningfulness Study
- **C** — Interpreting **Results** of a Clinical Meaningfulness Study
- **D** — Summarizing the **Clinical Bottom Line** of a Clinical Meaningfulness Study

Asking a Question

Step **1**	**Clinical Questions**
Step **2**	Search
Step **3**	Appraise
Step **4**	Integrate
Step **5**	Evaluate

1	**Identify**
2	**Construct**

Is a manual muscle test a reliable and valid measures of quadriceps strength in 1 42-year old women with multiple sclerosis?

Clinical OMs

PICO model

Is a manual muscle test as reliable and valid as hand held dynamometer for measuring quadriceps strength in 1 42-year old women with multiple sclerosis?

Will the Postural Assessment Scale for Stroke Patients (PASS) detect change in balance following rehabilitation in an 82 years old women with hemiplegia?
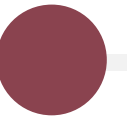
Self-reported outcome measure

PICO model

Is the Postural Assessment Scale for Stroke Patients (PASS) more sensitive to change than berg balance scale following rehabilitation in an 82 years old women with hemiplegia

Searching for evidence
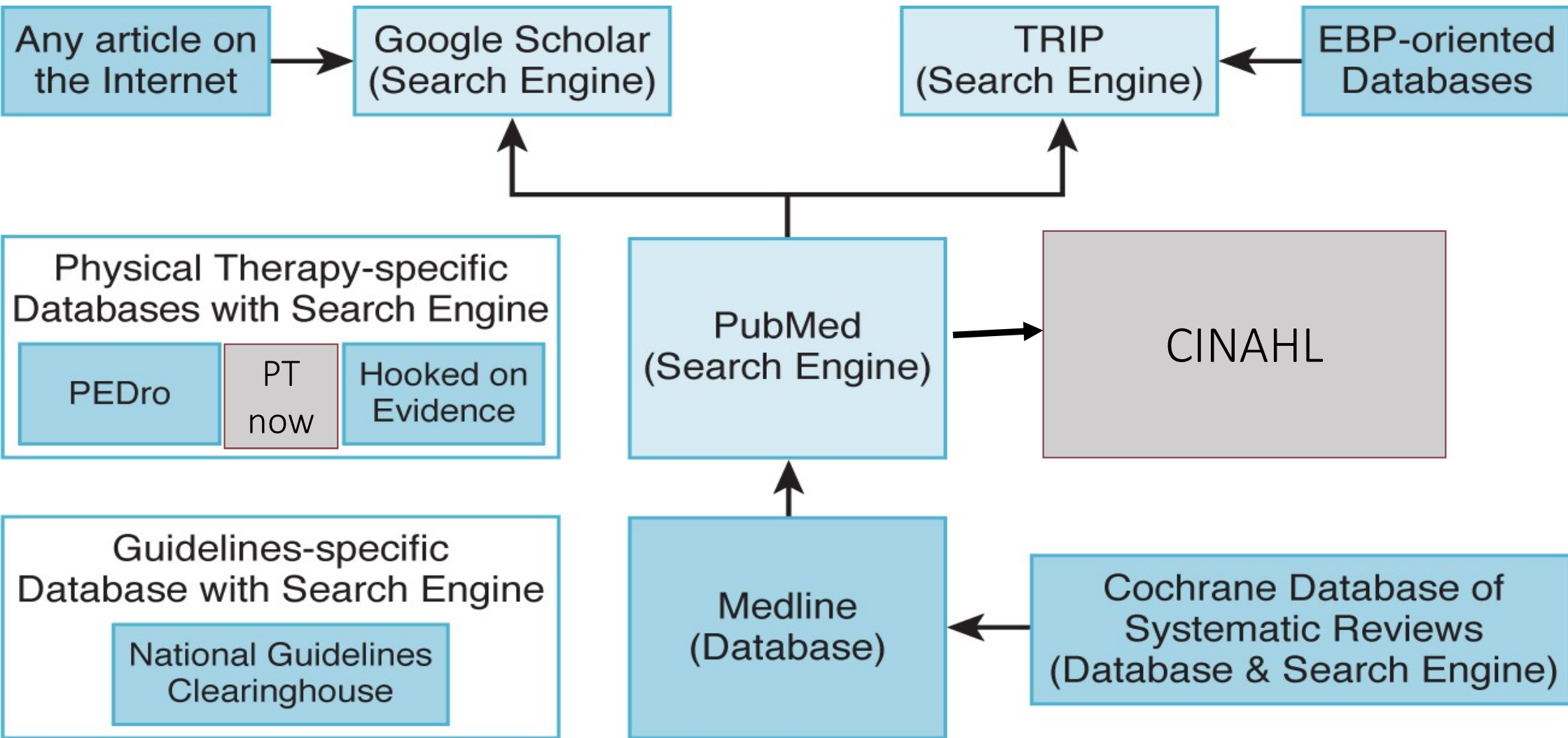
Appraisal Tools for Psychometric Studies

Appraisal Tools for Psychometric Studies

Several checklists for measurement properties exist

- SACMOT: Scientific advisory committee of the medical outcome trust in 2002

- COSMIN: COnsensus-based Standards for the selection of health Measurement Instruments.(www.cosmin.).

- EMPRO: Evaluating the Measurement of Patient-Reported Outcomes;

- Christina Jerosch-Herold"s Checklist for critical appraisal of OMs

No consensus on terminilogy and defintions
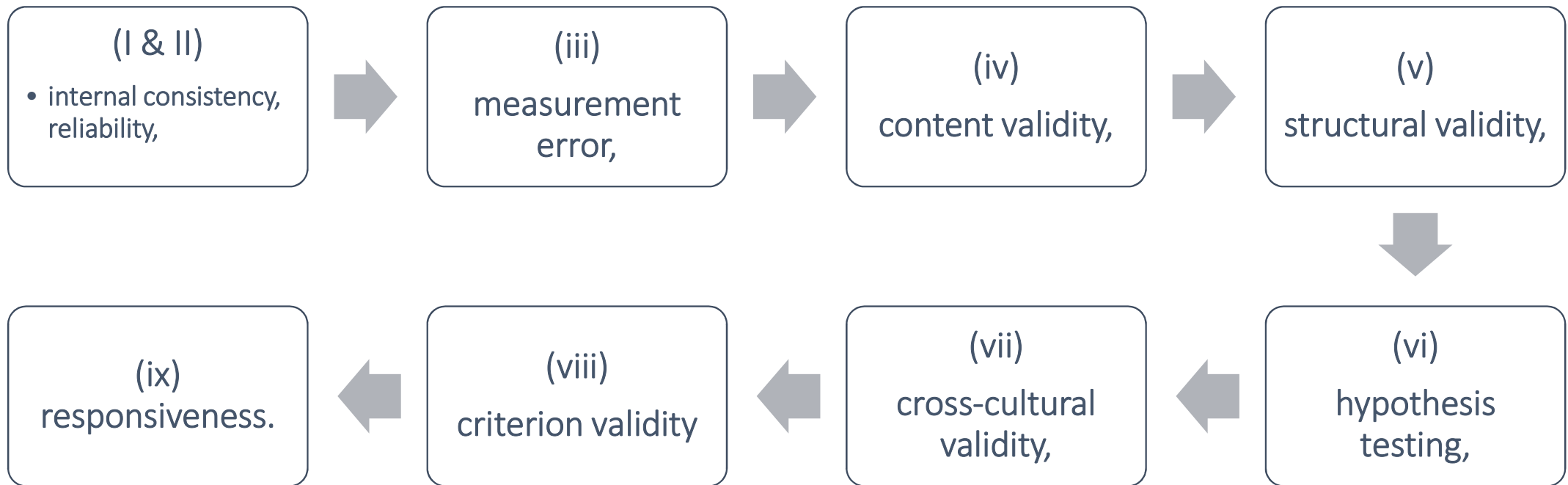None is generally accepted and widely used

	SACMOT	COSMIN
Number of items	Not in item format; 8 attributes	5–18 items across 9 measurement properties
Availability of tool and manual	Tool published in peer-reviewed journal article	freely accessible/, free of charge
Development	Expert panel	Delphi procedure, expert consensus
Scoring system	No	Yes, 4-point (excellent, good, fair, poor);
Psychometrics	No information	% agreement: 68% of assessments above 80% kappa: 61% below 0.40; 6% above 0.75 (low kappa may be due to skewed item score distribution)
Instructions for synthesis of design and statistical outcome evaluation	No information	Quality score per measurement property

The COSMIN Appraisal Tools

The COSMIN checklist (Mokkink et al.,2010a)

- It is a carefully developed tool, using Delphi consensus-based procedures.

- It has a number of objectives, including, to identify methodologically sound instruments, design and report psychometric studies, and inform the peer-review process.

- In total, it contains 114 items, 96 of which relate to psychometric properties, which are presented in the checklist as Boxes.

The COSMIN Appraisal Tools

(I & II)
- internal consistency, reliability,

(iii)
measurement error,

(iv)
content validity,

(v)
structural validity,

(vi)
hypothesis testing,

(vii)
cross-cultural validity,

(viii)
criterion validity

(ix)
responsiveness.

The COSMIN checklist for internal consistency

Box A. Internal consistency

		yes	no	?
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model?	☐	☐	☐

Design requirements

		yes	no	?
2	Was the percentage of missing items given?	☐	☐	
3	Was there a description of how missing items were handled?	☐	☐	
4	Was the sample size included in the internal consistency analysis adequate?	☐	☐	☐
5	Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	☐	☐	
6	Was the sample size included in the unidimensionality analysis adequate?	☐	☐	☐
7	Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	☐	☐	☐
8	Were there any important flaws in the design or methods of the study?	☐	☐	

Statistical methods

		yes	no	NA
9	for Classical Test Theory (CTT): Was Cronbach's alpha calculated?	☐	☐	☐
10	for dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	☐	☐	☐
11	for IRT: Was a goodness of fit statistic at a global level calculated? e.g. χ^2, reliability coefficient of estimated latent trait value (index of (subject or item) separation)	☐	☐	☐

The COSMIN checklist for content validity

Box D. Content validity (including face validity)

General requirements	yes	no	?
1 Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	☐	☐	☐
2 Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	☐	☐	☐
3 Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	☐	☐	☐
4 Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	☐	☐	☐
5 Were there any important flaws in the design or methods of the study?	☐	☐	

Relevant study results (findings)

Results reported that are specific to your clinical questions

- Cronbach's alpha (α)
- Correlation coefficient (s)
- Effect size (s)
- Standardized response mean (SRM)
- Others
- Statistical significant and/precision of the relevant study results

P-values foe each relevant statistics reported by authors

Confidence interval for each relevant statistics reported by authors

Application of evidence to patient/client

Are there clinically meaningful differences between the subjects in the study and your patients/client?

Yes/no/insufficient

Can you administer the OMs of interest appropriately in your clinical setting with your current resources ?

Yes/no/ insufficient

Does the OMs of interest fit within the patient/client expressed values and preference?

Yes/no

Will you use the PROMs of interest for this patient/client?

yes/no