







La métrologie: une science fondamentale en MPR

DES de Médecine Physique et de Réadaptation

Module 1 – Evaluation en MPR

02/12/2024 – 1h30

Pr David GASQ ^{a, b, c}

^a Université Toulouse 3, Faculté de Santé, Département Médecine, Maïeutique & Paramédicaux

^b CHU de Toulouse, Physiologie Explorations Fonctionnelles, Clinique Universitaire du Mouvement

^c ToNIC, Toulouse NeuroImaging Center, Université de Toulouse, Inserm, UPS





Métrologie et propriétés de mesure

2 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

- Métrologie¹** = science de la mesure
- Propriétés de mesure²** = caractéristiques d'un instrument de mesure qui reflètent la qualité de l'instrument de mesure



DES MPR – Mod. 1

D. GASQ

02/12/2024

¹ « Métrologie – en bref » édition meda 2007. ² Mokkink et al. J Clin Epidemiol. 2010.



3 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Métrologie et propriétés de mesure

- Concepts fondamentaux incontournables → utilité quotidienne en MPR
- Pour interpréter toutes les évaluations et orienter la prise en charge !




"The only statistics you can trust are the ones you have falsified yourself." Winston Churchill 1967



4 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Propriétés de mesure: quel terme pour quoi ?

- Reproducibility
- Reliability
- Repeatability
- Agreement
- Precision
- Variability
- Consistency
- Stability
- Measurement error
- Responsiveness
- Sensitivity to change
- Longitudinal validity
- Test-retest
- Intra/inter-rater
- Within/between-day

- Internal consistency
- Content validity
- Face validity
- Construct validity
- Structural validity
- Hypotheses testing
- Cross-cultural validity
- Criterion validity
- Interpretability

- Standard deviation (SD)
- Coefficient of variation (CV)
- Kappa coefficient
- Intraclass coefficient correlation (ICC)
- Limit of agreement (LOA)
- Small real difference (SRD)
- Minimal clinically important difference (MCID)
- Minimal metrically detectable change (MMDC)
- Minimal detectable change (MDC)
- Standard error measurement (SEM)
- Minimal important change (MIC)
- Minimal important difference (MID)
- Receiver operator characteristic (ROC) curve
- Effect size (ES)



5 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Objectifs d'apprentissage

1. Connaitre la taxonomie des propriétés de mesure

2. Connaitre et comprendre les différents concepts de validité

3. Connaitre et comprendre les effets seuils (plancher et plafond)

4. Connaitre et comprendre les différentes dimensions de la reproductibilité

5. Connaitre et comprendre le concept de sensibilité au changement

6. Savoir comment chercher des informations sur les propriétés de mesure



6 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

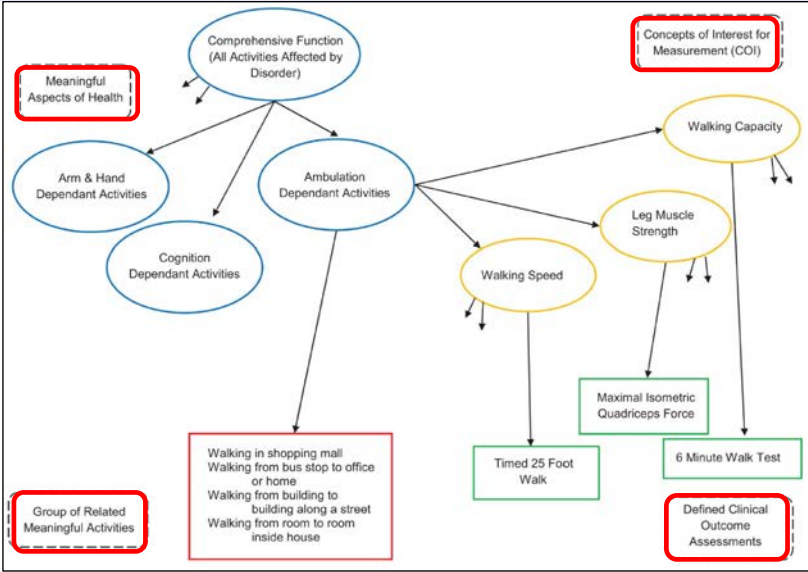
Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Quelle évaluation avec quel instrument de mesure ?



Walton et al. Value Health. 2015.



Les attributs de l'évaluation d'un résultat

7 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1
-
D. GASQ
-
02/12/2024

1. L'évaluation des résultats dépend-elle de la participation active du patient ou du jugement de l'évaluateur ?

- *Clinical outcome assessment* ⇔ évaluation clinique = variabilité potentiellement liée à:
 - Variation de la volonté/motivation/attention du patient
 - Variation du jugement de l'évaluateur selon le moment, son expérience, etc ...
- Biomarqueur ⇔ *mesure en laboratoire*
 - Soumis à peu ou pas d'influence motivationnelle du patient ou de jugement de l'évaluateur
 - Ex. des données biologiques (HbA1c)
 - *Qui de la flexion de genou au cours du cycle de marche ...*

Walton et al. Value Health. 2015; COSMIN Risk of Bias tool to assess the quality of studies on reliability and measurement error of outcome measurement instrument, 2021.



Les attributs de l'évaluation d'un résultat

8 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1
-
D. GASQ
-
02/12/2024

1. L'évaluation des résultats dépend-elle de la participation active du patient ou du jugement de l'évaluateur ?

2. Qui est l'évaluateur ? Quel type d'évaluateur applique son jugement pour former la mesure ?

- Ce sont des **évaluations cliniques**
- Le patient (questionnaire) = **Patient reported outcome measurement (PROM)**
- Clinicien avec une formation professionnelle spécifique = **Clinician reported outcome (ClinROM)**
 - EVA, données cliniques (PSA), échelle basée sur l'interrogatoire du patient (HAD), interprétation imagerie médicale
- Observateur autre que le patient et sans formation spécialisée (parent, accompagnant, auxiliaire de vie) = **Observer reported outcome (ObsRO)**
- C'est la « *performance* » du patient qui est recueillie = **Performance outcomes (PerfOMs)**
 - Le patient est évalué mais aucun jugement de l'évaluateur n'a d'impact sur la mesure.
 - Exécution par le patient d'une tâche définie qui est quantifiée d'une manière spécifiée et ne repose pas sur le jugement pour déterminer la note.
 - Capacité plutôt que performance (cf. CIF)
 - Ex. = test de marche de 6 minutes, Box & Block test

Walton et al. Value Health. 2015; COSMIN Risk of Bias tool to assess the quality of studies on reliability and measurement error of outcome measurement instrument, 2021.



Les attributs de l'évaluation d'un résultat

9 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

1. L'évaluation des résultats dépend-elle de la participation active du patient ou du jugement de l'évaluateur ?
2. Qui est l'évaluateur ? Quel type d'évaluateur applique son jugement pour former la mesure ?
- 3. L'évaluation mesure-t-elle directement ou indirectement un aspect significatif de la façon dont le patient se sent ou fonctionne ?**
 - L'interprétation dépend du lien plus ou moins direct entre la grandeur mesurée et un ressenti pertinent pour le patient
 - *Ex. de la mesure de la force musculaire par dynamométrie isocinétique versus test de 6 minutes de marche*
- 4. Effet du contexte d'utilisation sur les propriétés de mesure**
 - Intérêt selon la phase d'évolution de la pathologie : effets seuils
 - *Ex. des échelles d'évaluation de l'équilibre en post-AVC: PASS versus BBS*

Walton et al. Value Health. 2015; COSMIN Risk of Bias tool to assess the quality of studies on reliability and measurement error of outcome measurement instrument, 2021.



Quel instrument de mesure ?

10 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

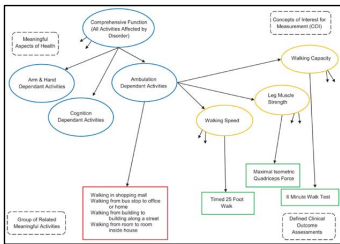
Sensibilité au changement

Guidelines

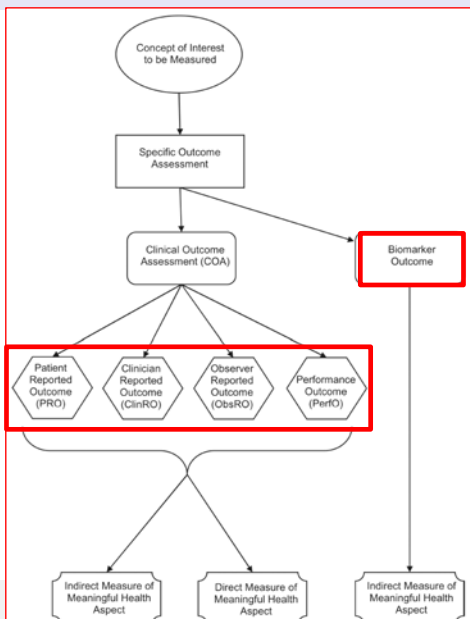
DES MPR – Mod. 1

D. GASQ

02/12/2024



This flowchart illustrates the relationship between measurement concepts and specific instruments. At the top, 'Concepts of Interest to be Measured' leads to 'Specific Outcome Assessment', which then leads to 'Clinical Outcome Assessment (COA)'. 'COA' is further divided into four categories: 'Patient Reported Outcome (PRO)', 'Clinician Reported Outcome (ClinRO)', 'Observer Reported Outcome (ObsRO)', and 'Performance Outcome (PerfO)'. These four categories are grouped under a red box labeled 'Biomarker Outcome'. Below this, the flowchart branches into three types of measures: 'Indirect Measure of Meaningful Health Aspect', 'Direct Measure of Meaningful Health Aspect', and 'Indirect Measure of Meaningful Health Aspect'.



This flowchart illustrates the relationship between measurement concepts and specific instruments. At the top, 'Concept of Interest to be Measured' leads to 'Specific Outcome Assessment', which then leads to 'Clinical Outcome Assessment (COA)'. 'COA' is further divided into four categories: 'Patient Reported Outcome (PRO)', 'Clinician Reported Outcome (ClinRO)', 'Observer Reported Outcome (ObsRO)', and 'Performance Outcome (PerfO)'. These four categories are grouped under a red box labeled 'Biomarker Outcome'. Below this, the flowchart branches into three types of measures: 'Indirect Measure of Meaningful Health Aspect', 'Direct Measure of Meaningful Health Aspect', and 'Indirect Measure of Meaningful Health Aspect'.

Walton et al. Value Health. 2015.



11 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Étude des propriétés de mesure

- **CO**nsensus-based Standards for the selection of health **ME**asurement **IN**struments



- COSMIN vise à **améliorer la sélection des instruments de mesure des résultats**, tant dans la recherche que dans la pratique clinique, en développant une **méthodologie et des outils pratiques pour sélectionner l'instrument de mesure** des résultats le plus approprié.

<https://www.cosmin.nl/>



12 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

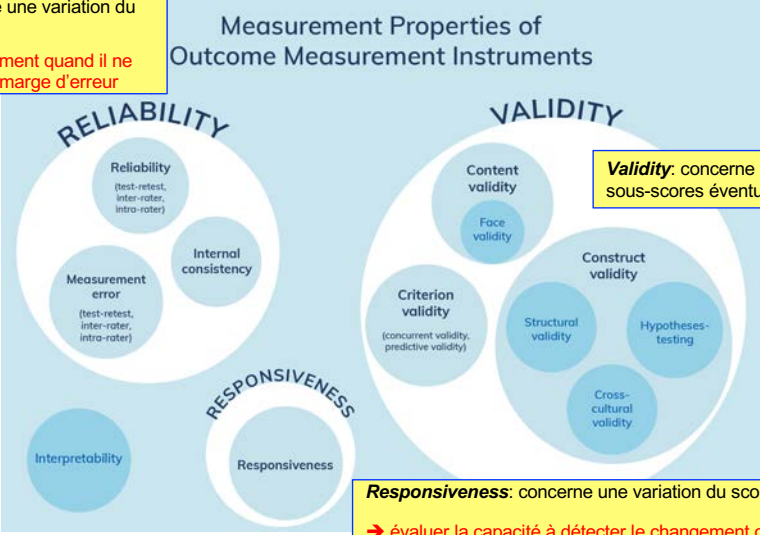
DES MPR – Mod. 1

D. GASQ

02/12/2024



Taxonomie des propriétés de mesure



RELIABILITY

- Reliability (test-retest, inter-rater, intra-rater)
- Internal consistency
- Measurement error (test-retest, inter-rater, intra-rater)
- Interpretability

VALIDITY

- Content validity
- Face validity
- Criterion validity (concurrent validity, predictive validity)
- Construct validity
 - Structural validity
 - Hypotheses-testing
 - Cross-cultural validity

RESPONSIVENESS

- Responsiveness

Reliability: concerne une variation du score test/retest
→ évaluer le changement quand il ne doit pas y en avoir = marge d'erreur

Validity: concerne un seul score (et ses sous-scores éventuels)

Responsiveness: concerne une variation du score pré/post
→ évaluer la capacité à détecter le changement quand il y en a

Mokkink et al. J Clin Epidemiol. 2010. <https://www.cosmin.nl/>



13 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

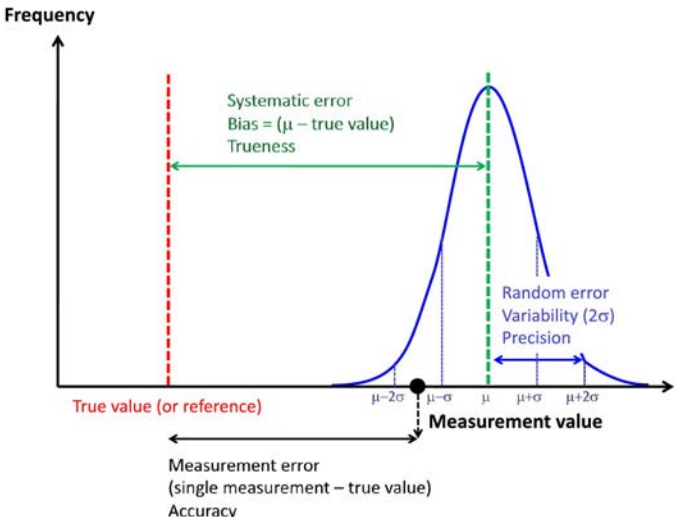
DES MPR – Mod. 1

D. GASQ

02/12/2024

La validité c'est quoi au juste ?

- La vraie valeur ?
- Exactitude (*accuracy*)
- Justesse (*trueness*)
- Précision (*precision*)
- Fiabilité (*reliability*)
- Erreur de mesure
 - Systématique
 - Aléatoire



Frequency

Systematic error
Bias = (μ – true value)
Trueness

Random error
Variability (2σ)
Precision

True value (or reference)

Measurement value

Measurement error
(single measurement – true value)
Accuracy



14 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

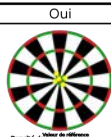
02/12/2024

La validité c'est quoi au juste ?

Précision ?

	Oui	Non
Oui	 Densité de probabilité	 Densité de probabilité
Non	 Densité de probabilité	 Densité de probabilité

Fiabilité ?

	Oui	Non
Oui	 Densité de probabilité	 Densité de probabilité
Non	 Densité de probabilité	 Densité de probabilité

Pour améliorer la validité d'une mesure, il faut souvent multiplier les mesures !!!

https://fr.wikipedia.org/wiki/Exactitude_et_pr%C3%A9cision

15 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Un exemple qui vous parle

The graph plots height (cm) and weight (kg) against age (years) from 0 to 18. It includes curves for +2 DS, TC (Tercile Centrale), and -2 DS. A cartoon elephant is shown on a scale, looking confused with a question mark above its head.

Duclos et Le Bouc 2016.

16 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Validité = quel instrument de mesure ?

- Échelle = 1 chiffre ?
 - Validité = « simple »
- Échelle avec multiples items ?
 - Une / plusieurs sous-parties ?
 - Validité = complexe

The image shows a hand holding an analog scale and a form titled 'Erasmus modified Nottingham Sensory Assessment for the upper limb'. The form includes sections for 'Sensibilité superficielle', 'Discrimination de 3 points', 'Proprioception', and 'Synthèse'.

Mokkink et al. J Clin Epidemiol. 2010. Villepinte et al. 2019.

 Validité			
17 / 79	Domain	Measurement property	Aspect of a measurement property
Taxonomie	Validity		Definition
			The degree to which an HR-PRO instrument measures the construct(s) it purports to measure
		Content validity	The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured
		Face validity	The degree to which (the items of) an HR-PRO instrument indeed looks as though they are an adequate reflection of the construct to be measured
		Construct validity	The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured
		Structural validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
		Hypotheses testing	Idem construct validity
Effets seuils		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
		Criterion validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"
Reproductibilité			
Sensibilité au changement			
Guidelines			
DES MPR – Mod. 1			
D. GASQ			
02/12/2024			
Mokkink et al. J Clin Epidemiol. 2010. Fermanian, Annales de réadaptation et de médecine physique 2005.			

 Validité			
18 / 79	Domain	Measurement property	Aspect of a measurement property
Taxonomie	Validity		Definition
		Internal consistency	The degree of the interrelatedness among the items
		Inclus dans 'Reliability' = Cohérence interne	
		Coefficient alpha de Cronbach	
		Sous-parties de l'échelle → Lié à la validité de construit : prérequis = homogénéité des sous parties	
		Construct validity	The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured
		Structural validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
Effets seuils		Hypotheses testing	Idem construct validity
		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
Reproductibilité		Criterion validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"
Sensibilité au changement			
Guidelines			
DES MPR – Mod. 1			
D. GASQ			
02/12/2024			
Mokkink et al. J Clin Epidemiol. 2010. Fermanian, Annales de réadaptation et de médecine physique 2005.			



Validité

19 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Domain	Measurement property	Aspect of a measurement property	Definition
Validity	Content validity	Face validity	The degree to which an HR-PRO instrument measures the construct(s) it purports to measure
			The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured
			The degree to which (the items of) an HR-PRO instrument indeed looks as though they are an adequate reflection of the construct to be measured
	Construct validity	Structural validity	The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured
			The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
			Idem construct validity
			The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
			The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"
Criterion validity	Hypotheses testing Cross-cultural validity	Adaptation transculturelle si traduction dans une autre langue. Guidelines spécifiques. Concerne validité et reproductibilité	

Mokkink et al. J Clin Epidemiol. 2010. Fermanian, Annales de réadaptation et de médecine physique 2005.



Validité

20 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Domain	Measurement property	Aspect of a measurement property	Definition
Validity	Content validity	Face validity	The degree to which an HR-PRO instrument measures the construct(s) it purports to measure
			The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured
			The degree to which (the items of) an HR-PRO instrument indeed looks as though they are an adequate reflection of the construct to be measured
	Construct validity	Structural validity	The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured
			The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
			Idem construct validity
	Hypotheses testing	Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
			The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"
Criterion validity		The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"	

Validité contre-critère (vs. gold standard), concurrente/prédictive

Mokkink et al. J Clin Epidemiol. 2010. Fermanian, Annales de réadaptation et de médecine physique 2005.



Exemple d'étude de la validité – Adaptation transculturelle

21 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Table 2
Concurrent-criterion validity assessed by Spearman's rank correlation coefficient between EmNSA-SS UL and the FMA-S UL (n=50).

EmNSA-SS UL	FMA-S UL		
	Light touch subscale	Proprioception subscale	Total score
Tactile sensation subscale	0.75 [*]	–	–
Proprioception subscale	–	0.77 [*]	–
Total score	–	–	0.74 [*]

EmNSA-SS UL: Erasmus modified Nottingham Sensory Assessment Somatosensory component for the upper limb; FMA-S UL: Fugl-Meyer Assessment-Sensory component for the upper limb.
^{*} $P < 0.001$ (two tailed).

Table 3
Internal consistency, floor and ceiling effect of the EmNSA UL and the FMA-S UL (n=50).

	Internal consistency ^a	Floor effect (%)		Ceiling effect (%)	
		EmNSA UL	FMA-S UL	EmNSA UL	FMA-S UL
Total score					
Somatosensory component (n=22)	0.961	0	2	14	34
Stereognosis component (n=10)	0.941	18	–	2	–
Subscales of EmNSA-SS					
Tactile sensation (n=16)	0.954	2	4	36	40
Two-point discrimination (n=2)	0.852	30	–	30	–
Proprioception (n=4)	0.821	4	4	52	54

EmNSA-SS/ST UL: Erasmus modified Nottingham Sensory Assessment Somatosensory/Stereognosis component for the upper limb; FMA-S UL: Fugl-Meyer Assessment-Sensory component for the upper limb.
^a Internal consistency was reported as Cronbach α coefficient.

Villepinte et al. 2019.



Notion de norme (*normative data*)

22 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

- Une norme permet de situer le sujet ⇔ physio vs. patho
 - ! Le score max/min d'une échelle (meilleur score) n'est pas une norme !
 - Effet plafond → Ex. score BBS / PASS
- Norme pour les déficiences et certaines activités +++

Mixed Populations; (McKay et al., 2017; n = 1,000; mean (SD) = 647 (127.0))

Reference values for the 6-Minute Walk Test by age group and sex (meters)

Age Group (years)	Male (mean (SD), n)	Female (mean (SD), n)
3-9	550 (117.6), 70	556 (119.4), 70
10-19	748 (103.9), 80	693 (83.6), 80
20-59	736 (88.3) ^a , 200	674 (76.8), 200
60+	599 (125.4) ^a , 150	550 (115.0), 150

^aSignificant sex differences, $p < 0.01$

23 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Notions d'effet plancher et plafond

- Évalué par le % de sujets ayant le score minimal et maximal (~15%)
 - Variable selon le stade de la pathologie
- Affecte la possibilité de discriminer les sujets et suivre leur évolution
- Surtout avec les échelles multi-items évaluant une fonction

Terwee et al. Journal of Clinical Epidemiology 2007.

24 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Exemple d'une échelle d'étude de la sensibilité

Erasmus modified Nottingham Sensory Assessment for the upper limb
Version validée en français (Villegre et al. 2019)

Form fields: Nom, Prénom, Date de naissance, Date de l'évaluation, Côté pathologique (Droite/Gauche), Examineur.

SENSIBILITÉ SUPERFICIELLE

Segments corporels	Toucher léger		Pression		Piquer		Piquer/Toucher	
	G	D	G	D	G	D	G	D
Douglas								
Main								
Avant bras								
Bras								
Total	0/8	0/8	0/8	0/8	0/8	0/8	0/8	0/8

DISCRIMINATION DE 2 POINTS *

Segments corporels	G	D	mm
Index			
Essence			
Thème			
Total	0/4	0/4	

PROPRIOCEPTION

Articulations	G	D
Prox		
Prox		
Prox		
Total	0/3	0/3

SYNTHÈSE

	G	D
Sensibilité superficielle (/32)	0	0
Discrimination 2 points (/4)	0	0
Proprioception (/8)	0	0
Score somesthésique (/44)	0	0
Score stéréognosie (/20)	0	0

COMMENTAIRES



25 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Exemple de la stabilité posturale

- Berg Balance Scale (BBS)**
 - Échelle générique validée en gériatrie et neurologie (AVC, Parkinson)
 - 14 items, score total de 0 à 56

1. **Passer de la position assise à debout**

INSTRUCTIONS: Veuillez vous lever en essayant de ne pas vous aider avec les mains.

() 4 Peut se lever sans l'aide des mains et garder son équilibre

() 3 Peut se lever seul avec l'aide de ses mains

() 2 Peut se lever en s'aidant de ses mains, après plusieurs essais

() 1 Besoin d'un peu d'aide pour se lever ou garder l'équilibre

() 0 Besoin d'une aide modérée ou importante pour se lever

14. **Se tenir debout sur une jambe (noter la jambe utilisée)**

INSTRUCTIONS: Tenez-vous sur une jambe le plus longtemps possible, sans prendre appui.

() 4 Peut lever une jambe sans aide et tenir plus de 10 secondes**** NOTE : Continuer de chronométrer jusqu'à 60 secondes.

() 3 Peut lever une jambe sans aide et tenir de 5 à 10 secondes

() 2 Peut lever une jambe sans aide et tenir de 3 à 5 secondes

() 1 Essaie de lever une jambe mais ne peut tenir 3 secondes tout en restant debout, sans aide

() 0 Ne peut exécuter l'exercice ou a besoin d'aide pour ne pas tomber

- Effet plancher +++ en post-AVC à la phase subaigue
- Effet plafond (= item 14)

Benaïm et al. 1999; Sawacha et al. 2013



26 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Exemple de la stabilité posturale

- Postural Assessment Scale for Stroke Patients (PASS)**
- Evaluation de la déficience posturale chez le sujet hémiplégique
 - Score global de 0 à 36
- Pas d'effet plancher
- Effet plafond (~idem BBS) +++

	0	1	2	3
Couché				
Se tourne vers le côté sain	impossible	aide importante	modérée	sans aide
Se tourne vers le côté hémiplégique	impossible	aide importante	modérée	sans aide
S'assoit	impossible	aide importante	modérée	sans aide
Assis				
Assis sans support (plan Bobath, pieds au sol)	impossible	Support (1 main)	sans aide au moins 10 sec	sans aide au moins 5 mn
Se couche	impossible	aide importante	modérée	sans aide
Se met debout	impossible	aide importante	modérée	sans aide
Debout				
Avec aide (position libre des pieds)	impossible	soutien de deux personnes au moins 10 s, ou appui uni-podal	Soutien d'une personne au moins 1 min, ou appui +/- asymétrique modérée	Soutien d'une main possibilité de mouvements amples (> épaules)
Sans aide (position libre des pieds)	impossible	aide importante	modérée	sans aide
S'assoit	impossible	aide importante	modérée	sans aide
Ramasse un petit objet (stylo) posé sur le sol	impossible	aide importante	modérée	sans aide
Appui monopodal côté sain	impossible	quelques secondes	au moins 5 s	au moins 10 s
Appui monopodal côté parésié	impossible	quelques secondes	au moins 5 s	au moins 10 s
Sous-total				
Score global				/ 36

Benaïm et al. 1999; Nardone & Schieppati 2010; Sawacha et al. 2013



27 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

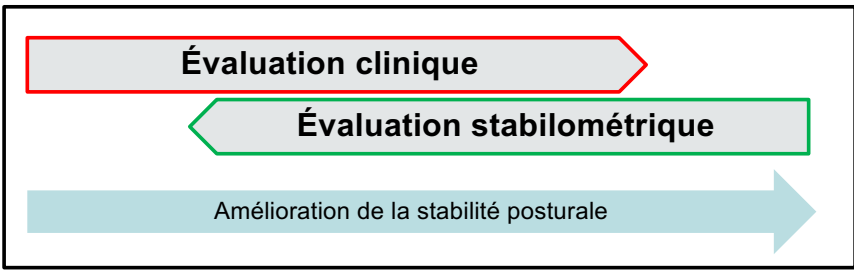
Sensibilité au changement

Guidelines

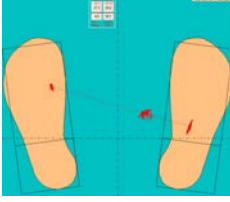
DES MPR – Mod. 1
- D. GASQ
- 02/12/2024

Exemple de la stabilité posturale

- Effet plafond** (+/- plancher) des échelles cliniques d'évaluation de la stabilité posturale
- Effet plancher** de l'évaluation stabilométrique
 - Position orthostatique sans aide technique > 30s
- Outils complémentaires**



The diagram consists of three horizontal arrows. The top arrow is red and points right, labeled 'Évaluation clinique'. The middle arrow is green and points left, labeled 'Évaluation stabilométrique'. The bottom arrow is light blue and points right, labeled 'Amélioration de la stabilité posturale'.



27

Visser et al. 2008; Nardone & Schieppati 2010; Sawacha et al. 2013



28 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1
- D. GASQ
- 02/12/2024

Sources de variabilité

- L'outil de mesure**
- Variabilité expérimentale**
 - Le sujet mesuré
 - L'évaluateur
 - ➔ conditions de mesure

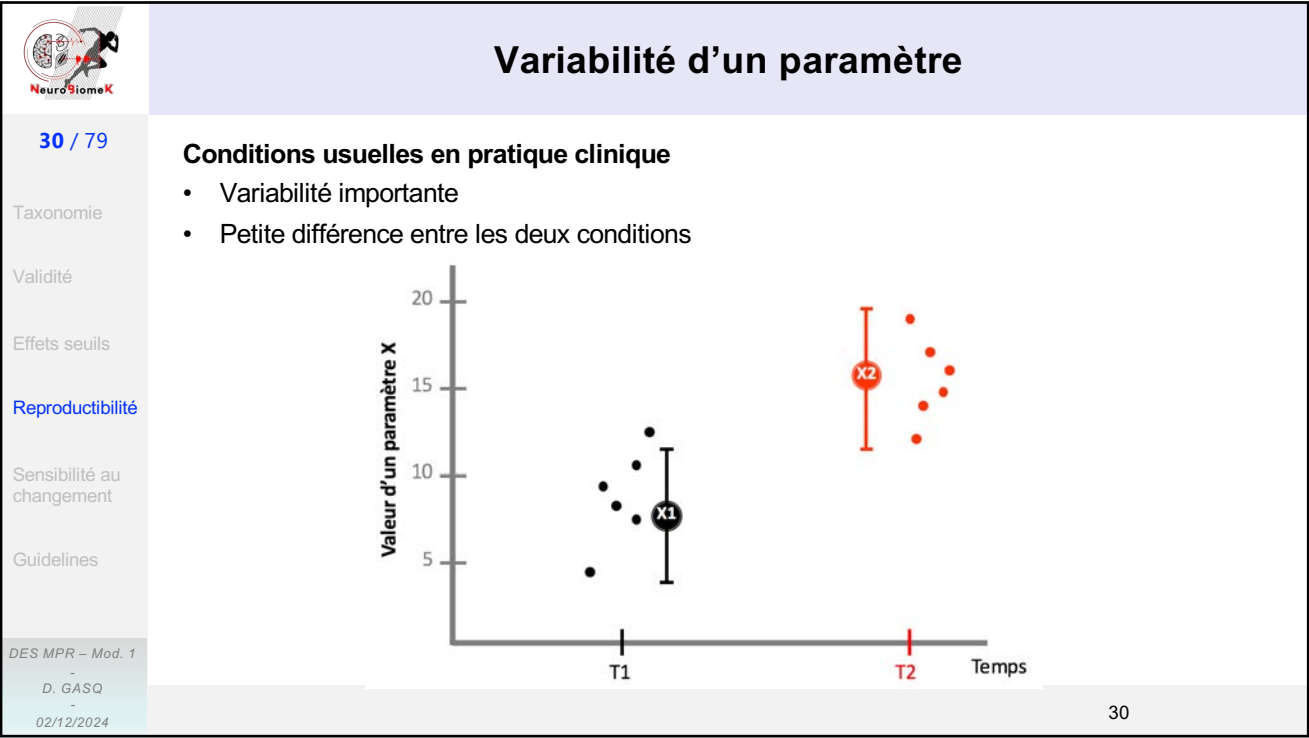
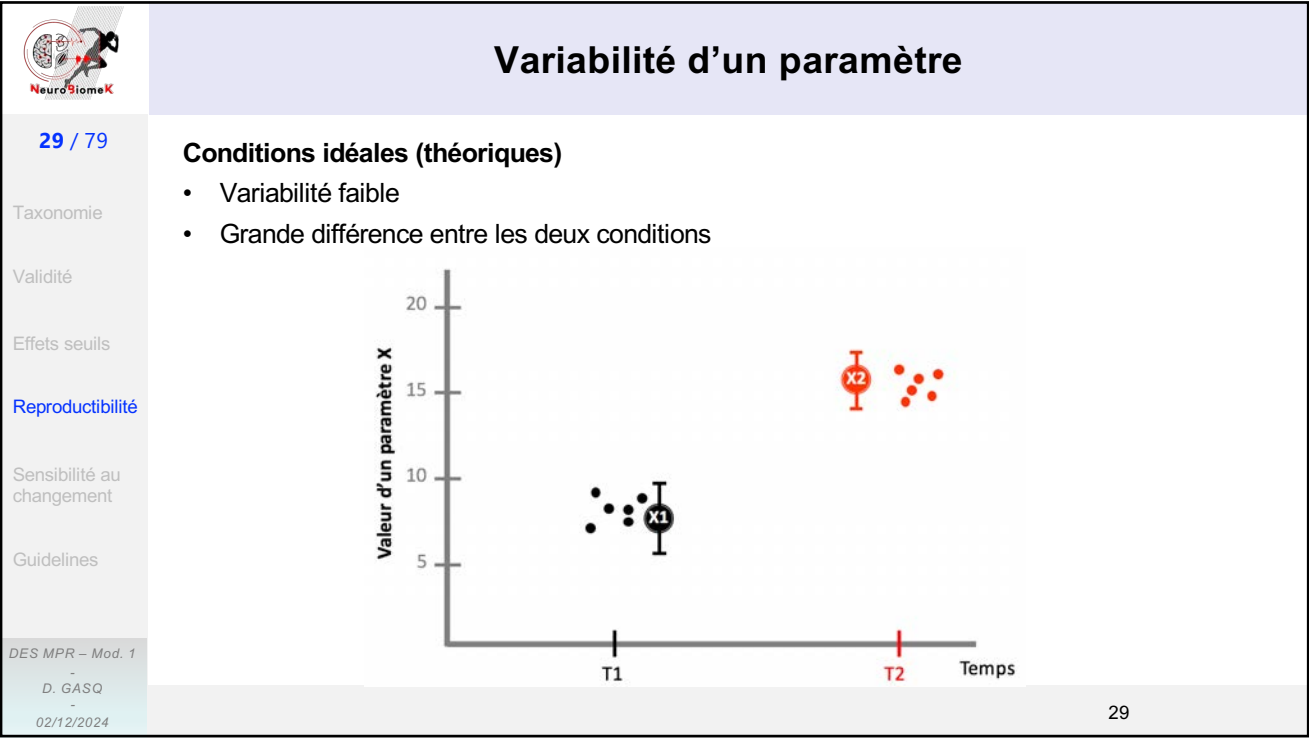


COMMENT VOULEZ-VOUS QUE JE SACHE SI JE SUIS TROP GROS...

JE NE VOIS PAS LE CADRAN DE MA BALANCE

Connaissance de la variabilité de la mesure INDISPENSABLE à la comparaison de 2 examens

28





31 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

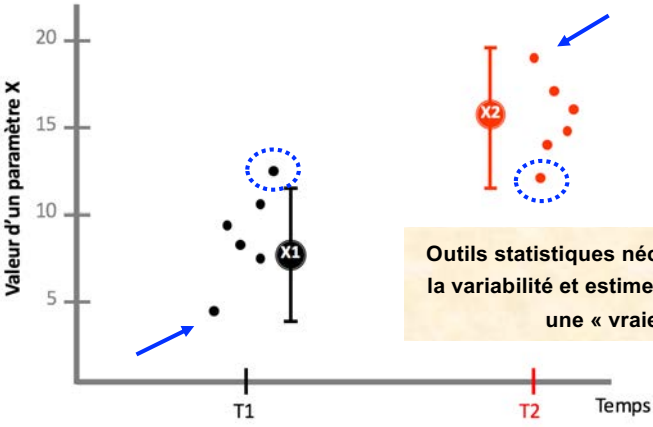
D. GASQ

02/12/2024

Variabilité d'un paramètre

Conditions usuelles en pratique clinique

- Variabilité importante
- Petite différence entre les deux conditions



Outils statistiques nécessaires pour quantifier la variabilité et estimer la probabilité qu'il y ait une « vraie » différence



32 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Reproductibilité

Domain	Measurement property	Aspect of a measurement property	Definition
Reliability	Reliability		The degree to which the measurement is free from measurement error
Reliability (extended definition)			The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: for example, using different sets of items from the same HR-PROs (internal consistency), over time (test–retest) by different persons on the same occasion (interrater) or by the same persons (i.e., raters or responders) on different occasions (intrarater)
	Reliability	Fiabilité / Fidélité	The proportion of the total variance in the measurements which is because of “true” ^a differences among patients
	Measurement error	Marge d'erreur	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured

- Design = test-retest
- Reliability = Fiabilité
 - ICC (coefficient de corrélation intraclasse) si variable continue
 - Coefficient de kappa si variable binaire / nominale
- Marge d'erreur = indicateurs avec différentes appellations : MDC / SDC / SRD / LOA ...



33 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1
- D. GASQ
- 02/12/2024

Comment étudier la reproductibilité

- Test- retest
 - Between-day / Intra-day ...
- Intra / inter-rater

Patients

First
measurement

Second
measurement

Occasion A

Occasion B

Figure 1 Test-retest design of the assessment of reliability or measurement error of the entire measurement instrument.

Components of measurement instruments

Equipment & preparation

Collection raw data

Data processing & storage

Assignment of the score

Assignment of the score repeated

Technician Z

Technician Z

Technician Z

Rater A

Rater B

Figure 2 Inter-rater design of the assessment of reliability or measurement error of the assignment of the score.

Mokkink et al. Patient Related Outcome Measures 2023.



34 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1
- D. GASQ
- 02/12/2024

Fiabilité avec l'ICC

- Evalue la proportion de variabilité inter-individuelle
 - ICC élevé si variabilité intra-individuelle < variabilité inter-individuelle
 - Intéressant pour la capacité à discriminer les sujets entre eux

$$\text{reliability} = \frac{\text{variability between study objects}}{\text{variability between study objects} + \text{measurement error}}$$

De Vet et al. 2005



35 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Fiabilité avec l'ICC

- Evalue la proportion de variabilité inter-individuelle
 - ICC élevé si variabilité intra-individuelle < variabilité inter-individuelle
 - Intéressant pour la capacité à discriminer les sujets entre eux
 - ICC proche de 1 entre rond et carré / triangle
 - ICC plus faible entre carré & triangles

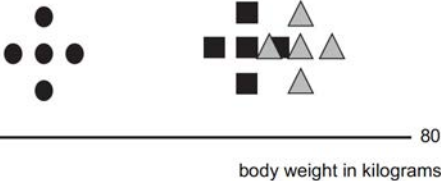


Fig. 1. Five repeated measurements of the body weights of three persons (●, ■, and ▲).

reliability = $\frac{\text{variability between study objects}}{\text{variability between study objects} + \text{measurement error}}$



36 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Fiabilité avec l'ICC

TABLE 5. Intraclass correlation coefficient model summary table.*

Shrout and Fleiss	Computational formula	McGraw and Wong	Model
1,1	$\frac{MS_B - MS_W}{MS_B + (k - 1)MS_W}$	1	1-way random
1,k	$\frac{MS_B - MS_W}{MS_B}$	k	1-way random
	Use 3,1	C,1	2-way random
	Use 3,k	C,k	2-way random
2,1	$\frac{MS_S - MS_E}{MS_S + (k - 1)MS_E + \frac{k(MS_T - MS_E)}{n}}$	A,1	2-way random
2,k	$\frac{MS_S - MS_E}{MS_S + \frac{k(MS_T - MS_E)}{n}}$	A,k	2-way random
3,1	$\frac{MS_S - MS_E}{MS_S + (k - 1)MS_E}$	C,1	2-way fixed
3,k	$\frac{MS_S - MS_E}{MS_S}$	C,k	2-way fixed
	Use 2,1	A,1	2-way fixed
	Use 2,k	A,k	2-way fixed

* Adapted from Shrout and Fleiss (46) and McGraw and Wong (34). Mean square abbreviations are based on the 1-way and 2-way analysis of variance illustrated in Table 2. For McGraw and Wong, A = absolute and C = consistency. MS_B = between-subjects mean square; MS_E = error mean square; MS_S = subjects mean square; MS_T = trials mean square; MS_W = within-subjects mean square.

reliability = $\frac{\text{variability between study objects}}{\text{variability between study objects} + \text{measurement error}}$



37 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Marge d'erreur

- Marge d'erreur test-retest
 - Basée sur la variabilité des mesures répétées au niveau individuel
- Intervalle de confiance de cette variabilité au niveau d'une population
 - Dans l'unité de mesure ou %
 - Plusieurs approches statistiques

“How good is the agreement between repeated measurements?” This concerns the measurement error, and assesses exactly how close the scores for repeated measurements are.

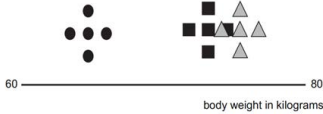


Fig. 1. Five repeated measurements of the body weights of three persons (●, ■, and ▲).

Guyatt et al. 1985; Bland et Altman 1986; De Vet et al. 2005.



38 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Marge d'erreur : un peu d'histoire

- Limit of agreement (LOA)
 - Graphe de Bland & Altman

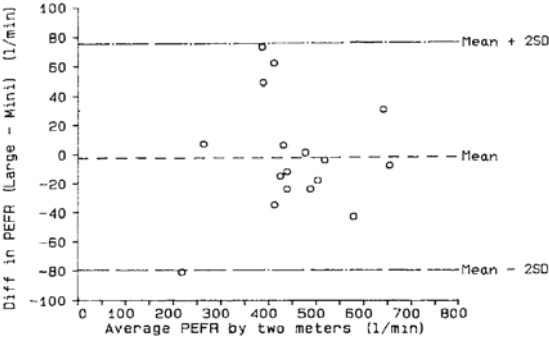


Fig 2—Difference against mean for PEFR data.

THE LANCET, FEBRUARY 8, 1986

Measurement

STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

J. MARTIN BLAND DOUGLAS G. ALTMAN

Bland et Altman 1986.



39 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Marge d'erreur : un peu d'histoire

- **MMDC & MDC**
 - Minimal (metrically) detectable change
 - IC 95% des erreurs
- **Smallest real difference (SRD)**

Quality of Life Research 10: 571-578, 2001.
© 2001 Kluwer Academic Publishers. Printed in the Netherlands.

Smallest real difference, a link between reproducibility and responsiveness

H. Beckerman¹, M.E. Roebroeck², G.J. Lankhorst¹, J.G. Becher¹, P.D. Bezemer³ & A.L.M. Verbeek⁴
¹Department of Rehabilitation Medicine, Universitair Ziekenhuis Vrije Universiteit, Amsterdam (E-mail: H. Beckerman@vrije.univ.nl)

Setting the Minimal Metrically Detectable Change on Disability Rating Scales

Réjean Hébert, MD, David J. Spiegelhalter, PhD, Carol Brayne, MD

ABSTRACT: Hébert R, Spiegelhalter DJ, Brayne C. Setting the minimal metrically detectable change on disability rating scales. Arch Phys Med Rehabil 1997; 78:1305-8.

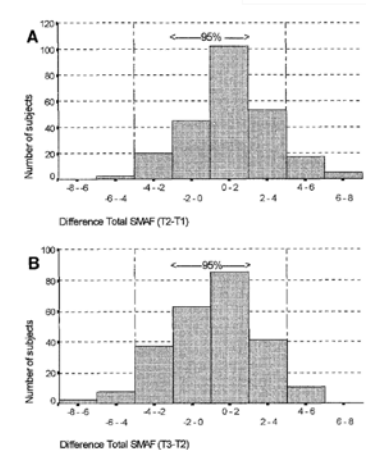


Fig 3. Distribution of the difference in total SMF scores between (A) the second (T2) and baseline (T1) interviews, and (B) the third (T3) and second (T2) interviews.



40 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Recommandations pour la reproductibilité

- La fiabilité = ICC / Kappa
 - Pertinente lorsque l'on souhaite améliorer l'instrument de mesure (par exemple, le protocole de mesure) = recherche
- L'erreur de mesure = marge d'erreur (SDC ou MDC₉₅)
 - Pertinente lorsque l'on utilise l'instrument de mesure = utilisateur de l'outil
 - **MDC₉₅ = marge d'erreur à dépasser pour attester d'un changement au niveau individuel**
- Il est recommandé de toujours évaluer les deux propriétés de mesure
 - Même conception d'étude et mêmes données collectées
 - ICC_{agreement} et SEM_{agreement} avec leur intervalle de confiance à 95%
 - Méthodologies de calcul spécifiques selon la situation
 - MDC₉₅ = 1,96 * $\sqrt{2}$ * SEM

➔ Critique de la sur-utilisation de l'ICC dès 1985 ++

Guyatt 1985; Mokkink et al. Patient Related Outcome Measures 2023. <https://www.cosmin.nl/>

D. Gasq - Toulouse

20



41 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Exemple de calcul de la marge d'erreur et de l'ICC

Parameters	Affected shoulder	Nonaffected shoulder
PT _A : Mean (SD)	69.49° (17.60°)	79.78° (7.60°)
PT _B : Mean (SD)	68.69° (16.25°)	78.88° (8.38°)
Mean _{diff_AB} (SD _{diff_AB})	0.80° (10.00°)	0.90° (9.58°)
PT _A vs. PT _B : % within 5°	43%	43%
PT _A vs. PT _B : % within 10°	72%	72%
Limits of agreement _{AB}	−18.80°–20.40°	−17.88°–19.68°
ICC _{agreement_AB}	0.83	0.28

Abbreviations: ICC: intraclass correlation coefficient; Mean_{diff_AB}: Mean of the differences between PT_A and PT_B; PT_A: physical therapist A; PT_B: physical therapist B; SD: standard deviation; SD_{diff_AB}: standard deviation of the differences between PT_A and PT_B.

De Vet et al. 2005



42 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Exemple de calcul de la marge d'erreur et de l'ICC

Table 4. Intra (n = 48) and inter-rater (n = 46) reliability of the EmNSA UL scale and subscales^a

		ICC (95% CI)	MDC95
EmNSA-SS total score (/44)	Intra-rater	0.92 (0.86 to 0.95)	12.3
	Inter-rater	0.92 (0.87 to 0.96)	14.6
Tactile sensation (/32)	Intra-rater	0.88 (0.79 to 0.93)	8.6
	Inter-rater	0.90 (0.83 to 0.95)	14.1
Discrimination of two points (/4)	Intra-rater	0.89 (0.82 to 0.94)	1.9
	Inter-rater	0.71 (0.53 to 0.83)	2.5
Proprioception (/8)	Intra-rater	0.86 (0.77 to 0.92)	4.4
	Inter-rater	0.71 (0.53 to 0.83)	3.5
EmNSA-ST (/20)	Intra-rater	0.92 (0.76 to 0.97)	23.1
	Inter-rater	0.84 (0.74 to 0.91)	27.5

- **ICC**: indicateur de variabilité inter-individuelle
- **MDC** = marge d'erreur utilisable en pratique clinique



43 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Marge d'erreur et nombre d'essais moyennés ?

Données stabilométriques chez 7 sujets sains de 62 à 73 ans

for the 120-Second Trial Duration

COP Variables	Direction	Mean ± SD	Trial*	MMDC
Sway area (mm ²)	—	135.8±46.3	1	±68.4
			13	±27.2
COP range (mm)	AP	26.3±5.6	1	±7.4
	ML	14.1±4.3	8	±3.5
RMS (mm)	AP	3.7±1.0	1	±5.1
	ML	2.3±0.9	6	±2.7
COP velocity (mm/s)	AP	7.9±1.7	1	±1.2
	ML	2.3±0.9	7	±0.9
	AP	7.9±1.7	1	±1.2
	ML	5.2±1.0	8	±0.6

*Number of trial averaged for ICC ≥.90.

Connaitre les conditions d'obtention de la marge d'erreur que vous allez utiliser

Lafond et al. 2004



44 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Notion de biais systématique

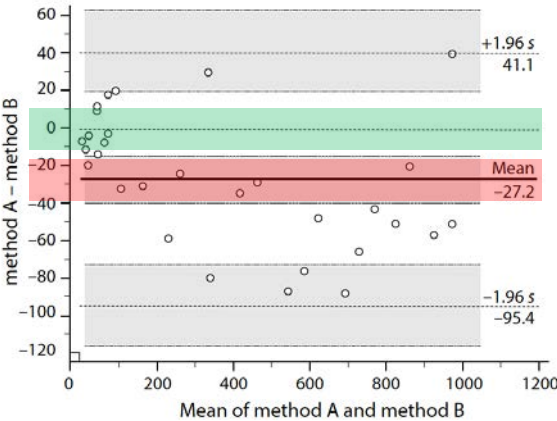


FIGURE 6. Same plot as Figure 2, with the representation of confidence interval limits for mean and agreement limits (shaded areas, data from table 2).

THE LANCET, FEBRUARY 8, 1996

Measurement

STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

J. MARTIN BLAND DOUGLAS G. ALTMAN

Department of Clinical Epidemiology and Social Medicine, St George's Hospital Medical School, London SW17; and Division of Medical Statistics, MRC Clinical Research Centre, Northwick Park Hospital, Harrow, Middlesex

Giavarina 2015



45 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Notion de biais systématique

- Biais = erreur systématique
 - Si progrès = apprentissage/habitude ?
 - Si dégradation = fatigue ?

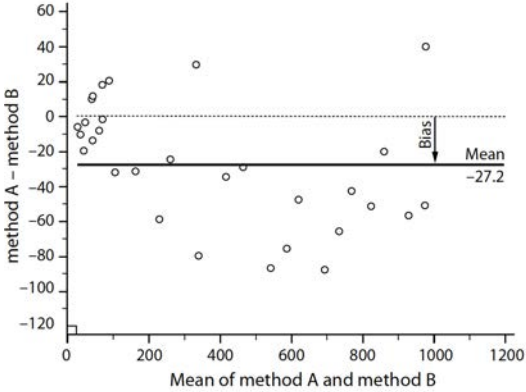


FIGURE 2. Plot of differences between method A and method B vs. the mean of the two measurements (data from table 1). The bias of -27.2 units is represented by the gap between the X axis, corresponding to a zero differences, and the parallel line to the X axis at -27.2 units.

THE LANCET, FEBRUARY 8, 1996

Measurement

STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

J. MARTIN BLAND DOUGLAS G. ALTMAN

Department of Clinical Epidemiology and Social Medicine, St George's Hospital Medical School, London SW17; and Division of Medical Statistics, MRC Clinical Research Centre, Northwick Park Hospital, Harrow, Middlesex



46 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Exemple de biais systématique

- Tenir compte ou pas du biais systématique ?

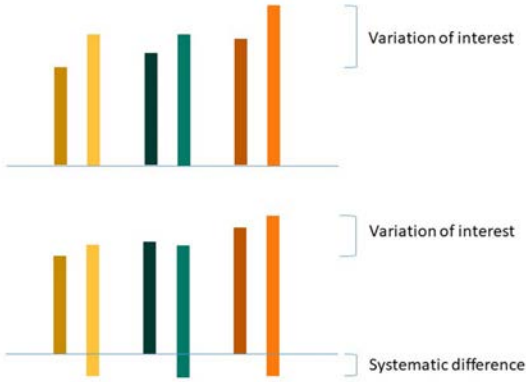


Figure 6 Variation of interest for agreement parameters (upper panel) and variation of interest for consistency parameters (lower panel): patients: yellow, green, orange; raters: dark, light.

- Méthodologie de calcul adaptée
- $ICC_{\text{agreement}}$ et $SEM_{\text{agreement}}$ avec leur intervalle de confiance à 95%

Mokkink et al. Patient Related Outcome Measures 2023.



47 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Exemple de calcul de la marge d'erreur

Impact de la méthode statistique

Table 4. Intra (n = 48) and inter-rater (n = 46) reliability of the EmNSA UL scale and subscales^a

		mDiff (95% CI)	MDC95	MDC95 w/o bias correction
EmNSA-SS total score	Intra-rater	0.65 (–0.63–1.93)	12.3	
	(/44) Inter-rater	0.91 (–0.31–2.14)	14.6	8.1
Tactile sensation	Intra-rater	0.27 (–0.89–1.43)	8.6	
	(/32) Inter-rater	0.93 (–0.07–1.94)	14.1	6.6
Discrimination of two points (/4)	Intra-rater	0.08 (–1.14–0.31)	1.9	
	Inter-rater	–0.07 (–0.42–0.29)	2.5	2.4
Proprioception (/8)	Intra-rater	0.29 (–0.01–0.60)	4.4	
	Inter-rater	0.04 (–0.48–0.57)	3.5	3.5
EmNSA-ST (/20)	Intra-rater	1.67 (0.97–2.37) ^d	23.1	
	Inter-rater	2.00 (0.94–3.06) ^d	27.5	7.0

D'après Villepinte et al. 2019.

"The only statistics you can trust are the ones you have falsified yourself." Winston Churchill 1967



48 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

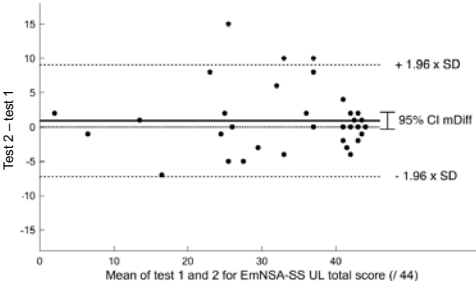
D. GASQ

02/12/2024

Exemple de biais systématique

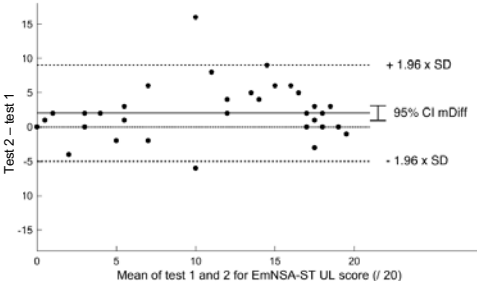
- Où se situe le biais (C ou D) ?
 - Que pourrait-il signifier ?

C



Score somesthésie

D



Score stéréognosie

Villepinte et al. 2019.



49 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

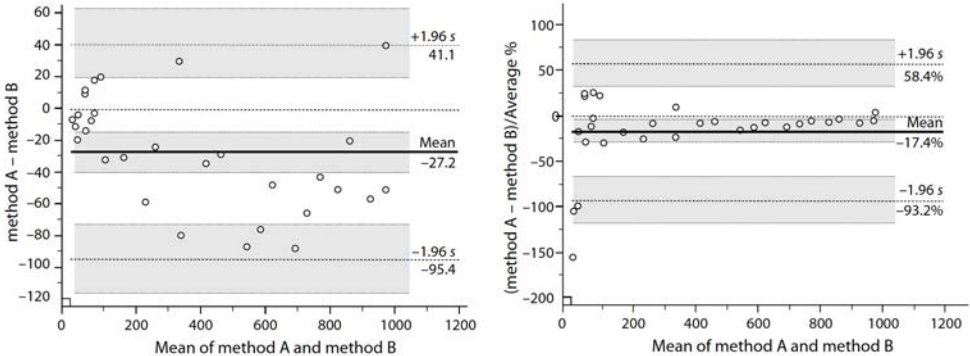
D. GASQ

02/12/2024

Marge d'erreur en synthèse

Vous avez compris l'intérêt de connaître la marge d'erreur d'une échelle de mesure que vous utilisez pour interpréter le changement

→ L'erreur de mesure doit-elle être exprimée dans l'unité de la mesure ou en pourcentage ?



Giavarina 2011



50 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

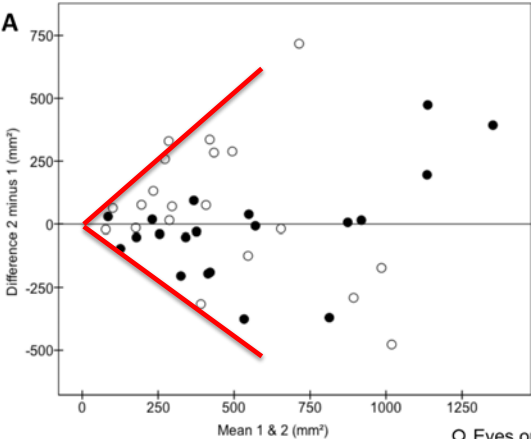
Guidelines

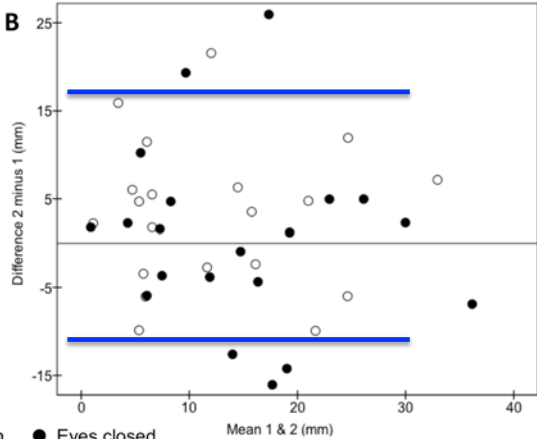
DES MPR – Mod. 1

D. GASQ

02/12/2024

Distribution des écarts à la moyenne ?





Gasq et al. 2014



51 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1
- D. GASQ
- 02/12/2024

Personal data

Calcul marge d'erreur = adaptée à la distribution des erreurs

- 30 sujets hémiplegiques
- Reprod inter-jour (7 jours) des paramètres spatio-temporels de marche
- Absence hétéroscédasticité
 - MDC = MDC en unité originelle

	Condition	MDC	MDCp
Vitesse (m/s)	VS	0.31	36.23
	VM	0.22	21.19
EA (m/s)		0.29	60.20
EAp (%)		72.95	93.24

IM	VS	2.19	21.86
	VM	2.08	19.67
IAS	VS	20.86	21.49
	VM	13.23	16.63
IAT	VS	7.70	9.23
	VM	9.93	12.43



52 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1
- D. GASQ
- 02/12/2024

Personal data

Calcul marge d'erreur = adaptée à la distribution des erreurs

- 30 sujets hémiplegiques
- Reprod inter-jour (7 jours) des paramètres spatio-temporels de marche
- Absence hétéroscédasticité
 - MDC = MDC en unité originelle
- Hétéroscédasticité
 - MDCp = MDC en %

	Condition	MDC	MDCp
Vitesse (m/s)	VS	0.31	36.23
	VM	0.22	21.19
EA (m/s)		0.29	60.20
EAp (%)		72.95	93.24

IM	VS	2.19	21.86
	VM	2.08	19.67
IAS	VS	20.86	21.49
	VM	13.23	16.63
IAT	VS	7.70	9.23
	VM	9.93	12.43



53 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

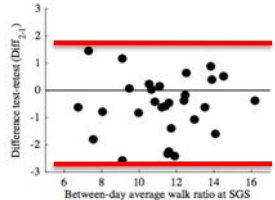
D. GASQ

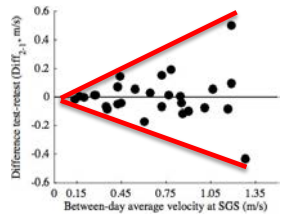
02/12/2024

Comment allez vous faire en pratique ?

Vous étudiez la reproductibilité d'un biomarqueur

1. Choix de l'ICC adéquat
2. Graphe de Bland-Altman : biais systématique ? Outliers ?
3. Distribution des différences test-retest : hétéroscédasticité ?
4. Calcul du MDC₉₅ en intégrant le biais systématique
 - Si hétéroscédasticité = expression en % (ou transformation logarithmique)
 - Sinon = valeur dans l'unité originelle





Bland et Altman 1986; Hopkins 2000; Lexell et al. 2005; de Vet et al. 2005; Mokkink et al. Patient Related Outcome Measures 2023.



Sensibilité au changement

54 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Domain	Measurement property	Aspect of a measurement property	Definition
Responsiveness			The ability of an HR-PRO instrument to detect change over time in the construct to be measured
	Responsiveness	= <i>sensitivity to change</i>	Idem responsiveness
Interpretability ^b			The degree to which one can assign qualitative meaning—that is, clinical or commonly understood connotations—to an instrument's quantitative scores or change in scores.

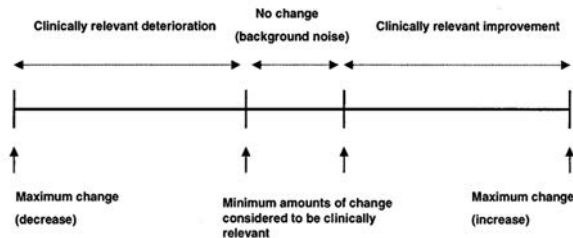


Figure 1. Available options for change over time in disease severity.

De Vet et al. 2001

Terwee et al. 2007; Mokkink et al. J Clin Epidemiol. 2010; Feramanian, Annales de réadaptation et de médecine physique 2005.



55 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Sensibilité au changement

Domain	Measurement property	Aspect of a measurement property	Definition
Responsiveness			The ability of an HR-PRO instrument to detect change over time in the construct to be measured
	Responsiveness	= sensitivity to change	Idem responsiveness
Interpretability ^b			The degree to which one can assign qualitative meaning—that is, clinical or commonly understood connotations—to an instrument's quantitative scores or change in scores.

- Mon pèse personne a-t-il détecté un changement de poids avant que je ne change de taille de vêtements ?



Terwee et al. 2007; Mokkink et al. J Clin Epidemiol. 2010; Fermanian, Annales de réadaptation et de médecine physique 2005.



56 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

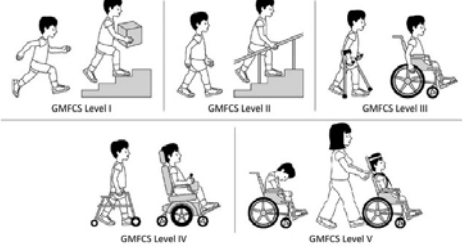
DES MPR – Mod. 1

D. GASQ

02/12/2024

Sensibilité au changement

- Propriété fondamentale si outil d'évaluation pour le suivi
 - Notion de changement cliniquement pertinent ⇔ du sens pour le patient
 - Clinical important difference = CID
 - Minimal (clinically) important difference = MID / MCID
 - Difficile à quantifier et à étudier
 - Pas simplement une différence pré/post-thérapeutique !!
- Non indispensable si outil de discrimination



Mokkink et al. Journal of Clinical Epidemiology 2010.



57 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Quantifier la sensibilité au changement ?

- Notion statistique
 - Taille d'effet et *standardized response mean*
 - Souvent utilisée → vérifier la pertinence clinique ?
- Référence à une notion clinique indispensable
 - Changement cliniquement pertinent pour le patient
- Evaluation pertinente si $MDC \leq MCID$

Outcome measures

Hedges's g [95% CI]

MAL AOU	1.04 [0.80, 1.28]
MAL OOM	1.00 [0.78, 1.23]
FMA-UE	0.65 [0.39, 0.90]
Smoothness	0.64 [0.38, 0.90]
WMFT Time	0.48 [0.26, 0.71]
Movement duration	0.47 [0.30, 0.85]
WMFT Quality	0.43 [0.18, 0.67]
Peak velocity	0.41 [0.23, 0.60]
Shoulder ROM	0.34 [0.12, 0.55]
Control strategy	0.30 [0.10, 0.51]
Movement efficiency	0.30 [0.01, 0.60]
Elbow ROM	0.28 [0.10, 0.46]
Trunk ROM	0.27 [-0.39, 0.94]

Standardised mean change

Villepinte et al. 2021



58 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Quantifier la sensibilité au changement ?

ELSEVIER

Journal of Clinical Epidemiology 63 (2010) 37–45

Three ways to quantify uncertainty in individually applied “minimally important change” values

Henrica C.W. de Vet^{a,b,*}, Berend Terluin^a, Dirk L. Kool^{a,b}, Leo D. Roorda^d, Lidwine B. Mokkink^{a,b}, Raymond W.J.G. Oude^{a,b,c}, Erik J.M. Hendriks^c, Lex M. Bouter^d, Caroline B. Terwee^{a,b}

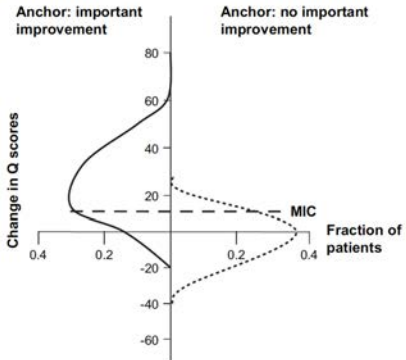


Fig. 1. Anchor-based minimally important change (MIC) distribution of the questionnaire Q with indication of the receiver operating characteristic cutoff point at MIC = 10.5. At this point, sensitivity = 0.74 and specificity = 0.91. For each curve, the relative frequency is presented. Solid curve represents importantly improved patients and dotted curve represents not importantly improved patients.

De Vet et al. 2010.



59 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1
- D. GASQ
- 02/12/2024

Quantifier la sensibilité au changement ?

ELSEVIER

Journal of Clinical Epidemiology 61 (2008) 102–109

REVIEW ARTICLES

Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes

Dennis Revicki^{a,*}, Ron D. Hays^b, David Cella^c, Jeff Sloan^d

What is new?

- Recommend that the minimal important difference (MID) be based primarily on appropriate patient-based and clinical anchors that are correlated at ≥ 0.30 with the patient-reported outcome (PRO), with clinical trial experience used to further inform understanding of MID.
- MID may vary by population and context, and thus a single MID may be insufficient for all study applications involving a PRO instrument.

- Estimation of MID for a specific PRO measure should be based on multiple approaches and triangulation of methods.
- Various methods for estimating MIDs often converge, and generalizability of MID estimates for similar applications is supported.
- Recommend basing the final selection of MID values on systematic review and evaluation process such as a modified Delphi method.

De Vet et al. 2001; Terwee et al. 2007



60 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1
- D. GASQ
- 02/12/2024

Quantifier la sensibilité au changement ?

- MDC ~ MCID pour la vitesse de marche

SYSTEMATIC REVIEW

Minimal clinically important difference for change in comfortable gait speed of adults with pathology: a systematic review

Richard W. Bohannon PT DPT EdD NCS^a and Susan S. Glasgow PT DPT GCSP^b

^aProfessor, ^bClinical Assistant Professor, Program in Physical Therapy, Department of Knowledge, Yang School of Education, University of Connecticut, Storrs, CT, USA

Abstract

Rationale, aims and objectives: Walking speed is an important performance variable, but information on the minimal clinically important difference (MCID) for the measure has not been consolidated. In this review, we aimed to summarize information on the MCID for change in comfortable gait speed measurements for patients with pathology.

Methods: Relevant literature was identified by searches of four databases (PubMed, Web of Knowledge, CINAHL, and Scopus), hand searches and consultation with an expert. Inclusion required that articles reported a MCID for comfortable gait speed measurements. Articles were excluded if the MCID was not determined using receiver operating characteristic (ROC) curve analysis. Articles were abstracted for information on participants, interventions, gait speed documentation and the determination of MCID. Quality was assessed using a hybrid 9-item (0–18 point) instrument.

Results: Seven articles were selected based on inclusion and exclusion criteria. The populations studied included stroke ($n = 3$), hip fracture ($n = 2$), multiple sclerosis ($n = 1$) and mixed ($n = 1$). Using 13 different anchors the studies reported MCIDs of 0.06–0.26 m s⁻¹. All but three of these MCIDs were between 0.10 and 0.20 m s⁻¹. All MCIDs for which the area under the ROC curve exceeded 0.70 were between 0.10 and 0.17 m s⁻¹.

Conclusions: Changes in gait speed of 0.10 to 0.20 m s⁻¹ may be important across multiple patient groups.

Bohannon et al. 2014.



61 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

DES MPR – Mod. 1

D. GASQ

02/12/2024

Liens utiles

- Trouver les propriétés psychométriques des outils
 - <https://strokengine.ca/en/>
 - <https://www.sralab.org/>
 - <https://www.rehab-scales.org/>
- Recherche bibliographique
 - Mesh Pubmed
 - Reproducibility of Results
 - Suggestion de termes sur COSMIN

Reproducibility of Results

The statistical reproducibility of measurements (often in a clinical context), including the testing of instrumentation or techniques to obtain reproducible results. The concept includes reproducibility of physiological measurements, which may be used to develop rules to assess probability or prognosis, or response to a stimulus; reproducibility of occurrence of a condition; and reproducibility of experimental results.

Year introduced: 1989

PubMed search builder options

☐ Restrict to MeSH Major Topic.

☐ Do not include MeSH terms found below this term in the MeSH hierarchy.

Tree Number(s): E05.318.370.725, E05.337.851, N05.715.360.325.685, N06.850.520.445.725

MeSH Unique ID: D015203

Entry Terms:

- Reproducibility of Findings
- Reproducibility Of Result
- Of Result, Reproducibility
- Of Results, Reproducibility
- Result, Reproducibility Of
- Results, Reproducibility Of
- Reproducibility of Finding
- Finding Reproducibilities
- Finding Reproducibility
- Reliability of Results
- Reliability of Result
- Result Reliabilities
- Result Reliability
- Reliability (Epidemiology)
- Validity (Epidemiology)
- Validity of Results
- Validity of Result
- Result Validities
- Result Validity
- Face Validity
- Validity, Face
- Reliability and Validity
- Validity and Reliability
- Test-Retest Reliability
- Reliabilities, Test-Retest
- Reliability, Test-Retest
- Test Retest Reliability



62 / 79

Taxonomie

Validité

Effets seuils

Reproductibilité

Sensibilité au changement

Guidelines

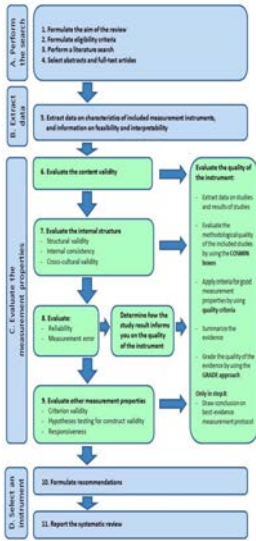
DES MPR – Mod. 1

D. GASQ

02/12/2024

Liens utiles

- Guidelines et outils pour faire une étude / une revue de littérature
 - <https://www.cosmin.nl/>
- Publications pour approfondir le sujet = liens sur le site de COSMIN
 - Ex. Mokkink et al. Studies on Reliability and Measurement Error of Measurements in Medicine - From Design to Statistics Explained for Medical Researchers. Patient Relat Outcome Meas. 2023;14:193-212.



The flowchart illustrates the 11 steps of the COSMIN methodology for systematic reviews of PROMs. It is organized into four main phases: 1. Formulate the aim of the review, 2. Perform a literature search, 3. Extract data on characteristics of included measurement instruments, and 4. Evaluate the quality of the measurement properties. Steps 1-4 are grouped under 'Formulate the aim of the review'. Steps 5-9 are grouped under 'Perform a literature search'. Steps 10-11 are grouped under 'Extract data on characteristics of included measurement instruments'. Steps 12-13 are grouped under 'Evaluate the quality of the measurement properties'. The flowchart includes decision points for 'Evaluate the content validity', 'Evaluate the internal structure', 'Evaluate the reliability', and 'Evaluate the measurement error'. It also includes a box for 'Evaluate the quality of the instrument' and a box for 'Evaluate the measurement properties'.

4.1 The eleven-step procedure for conducting a systematic review of PROMs, PerFORMs, or laboratory values

Below, a summary is given for the eleven-step procedure. In the user manual of the COSMIN methodology for systematic reviews of PROMs (45) a thorough explanation of each step is provided. Only the steps that are different for a review of outcome measurement instruments other than PROMs are described here in detail. Please note that the number of the step are changed.

The methodology of a systematic review of outcome measurement instruments is subdivided into three parts (A, B, and C) (5).

Step 1-4: Perform the literature search

Step 5: Data extraction

Steps 6-9: Evaluate the measurement properties

In step 6 the content validity is assessed. In step 7 the internal structure (structural validity, internal consistency and cross-cultural validity/measurement invariance) is assessed. As the assessment of reliability and measurement error requires an additional step (i.e. understanding how the results of a study inform you about the reliability or measurement error of an outcome measurement instrument), these two measurement properties are now assessed in a separate step, i.e. step 8, apart from the assessment of the measurement properties criterion validity, hypotheses testing for construct validity, and responsiveness (i.e. step 9).

Steps 10-11: Select the outcome measurement instrument

The steps 10 and 11 concerns the formulating recommendations (step 10) and the reporting of the systematic review (step 11).

D. Gasq - Toulouse

31