


Hôpitaux de Toulouse | TonIC | Inserm | UNIVERSITÉ TOULOUSE III | NeuroBiomeK

La métrologie: une science fondamentale en MPR

DES de Médecine Physique et de Réadaptation
Module 1 – Evaluation en MPR
04/12/2023 – 1h30
Dr David GASQ ^{a,b}



^a Université Toulouse 3, CHU de Toulouse – Physiologie Explorations Fonctionnelles
^b TonIC, Toulouse NeuroImaging Center, Université de Toulouse, Inserm, UPS

La base de la métrologie = la rigueur !

2 / 65




ET VOUS AUREZ UNE PREUVE DE VOTRE AGRESSION ?
BEN, EUH...

DES MPR – Mod. 1
D. GASQ
04/12/2023

Métrologie et propriétés de mesure

3 / 65

- **Métrologie**¹ = science de la mesure
- **Propriétés de mesure**² = caractéristiques d'un instrument de mesure qui reflètent la qualité de l'instrument de mesure



DES MPR – Mod. 1
D. GASQ
04/12/2023

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

Métrologie et propriétés de mesure

4 / 65

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

Vous pouvez interpréter ça de deux façons différentes.



DES MPR – Mod. 1
D. GASQ
04/12/2023

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

Propriétés de mesure: quel terme pour quoi ?

5 / 65

<ul style="list-style-type: none"> • Reproducibility • Reliability • Repeatability • Agreement • Precision • Variability • Consistency • Stability • Measurement error • Responsiveness • Sensitivity to change • Longitudinal validity • Test-retest • Intra/inter-rater • Within/between-day 	<ul style="list-style-type: none"> • Internal consistency • Content validity • Face validity • Construct validity • Structural validity • Hypotheses testing • Cross-cultural validity • Criterion validity • Interpretability 	<ul style="list-style-type: none"> • 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)
---	---	--

DES MPR – Mod. 1
D. GASQ
04/12/2023

Objectifs d'apprentissage

6 / 65

1. Connaître la taxonomie des propriétés de mesure
2. Connaître et comprendre les différents concepts de validité
3. Connaître et comprendre les effets seuils (plancher et plafond)
4. Connaître et comprendre les différentes dimensions de la reproductibilité
5. Connaître et comprendre le concept de sensibilité au changement
6. Savoir comment chercher des informations sur les propriétés de mesure

DES MPR – Mod. 1
D. GASQ
04/12/2023

Déroulé de l'exposé

7 / 65

1. Taxonomie des propriétés de mesure

- Notions de validité
- Effets seuils
- Reproductibilité
- Sensibilité au changement
- Guidelines

DES MPR - Mod 1
D. GASQ
04/12/2023

Quelle évaluation avec quel instrument de mesure ?

8 / 65

DES MPR - Mod 1
D. GASQ
04/12/2023

Walton et al. Value Health 2015.

Les attributs de l'évaluation d'un résultat

9 / 65

- 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)

DES MPR - Mod 1
D. GASQ
04/12/2023

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

10 / 65

- L'évaluation des résultats dépend-elle de la participation active du patient ou du jugement de l'évaluateur ?
- 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 spécifique et ne repose pas sur le jugement pour déterminer la note.
 - Ex. = test de marche de 6 minutes, Box & Block test

DES MPR - Mod 1
D. GASQ
04/12/2023

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

11 / 65

- L'évaluation des résultats dépend-elle de la participation active du patient ou du jugement de l'évaluateur ?
- Qui est l'évaluateur ? Quel type d'évaluateur applique son jugement pour former la mesure ?
- 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 isocinétiqne versus test de 6 minutes de marche
- 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 BES

DES MPR - Mod 1
D. GASQ
04/12/2023

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 ?

12 / 65


DES MPR - Mod 1
D. GASQ
04/12/2023

Walton et al. Value Health 2015.

Étude des propriétés de mesure

13 / 65

Consensus-based Standards for the selection of health Measurement Instruments



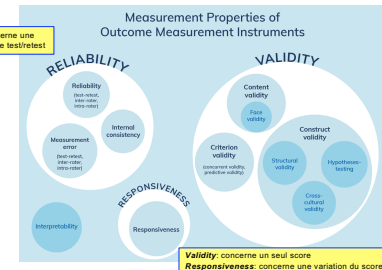
COSMIN vise à améliorer la sélection des instruments de mesure des résultats, tant dans la recherche que dans la pratique clinique, en développant une méthodologie et des outils pratiques pour sélectionner l'instrument de mesure des résultats le plus approprié.

DES MPR - Mod 1
D. GASQ
04/12/2023

Taxonomie des propriétés de mesure

14 / 65

Reliability: concerne une variation du score les/à l'instant



Validity: concerne un seul score

Responsiveness: concerne une variation du score prépost

DES MPR - Mod 1
D. GASQ
04/12/2023

Déroulé de l'exposé

15 / 65

1. Taxonomie des propriétés de mesure
2. Notions de validité
3. Effets seuils
4. Reproductibilité
5. Sensibilité au changement
6. Guidelines

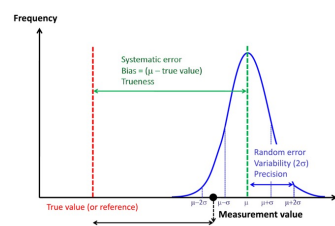


DES MPR - Mod 1
D. GASQ
04/12/2023

La validité c'est quoi au juste ?

16 / 65

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

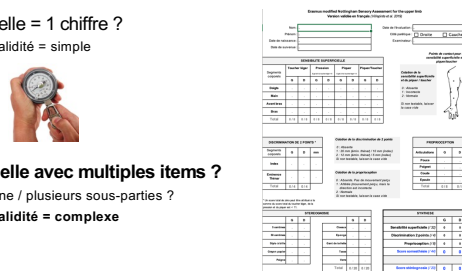


DES MPR - Mod 1
D. GASQ
04/12/2023

Validité = quel instrument de mesure ?

17 / 65

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



DES MPR - Mod 1
D. GASQ
04/12/2023

Validité

18 / 65

Domain	Measurement property	Aspect of a measurement property	Definition
Taxonomie	Validity	Content validity	The degree to which an HR-PRO instrument measures the construct(s) it purports to measure
		Face validity	The degree to which the content of an HR-PRO instrument is 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
		Criterion validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
Effets seuils	Reproducibility	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	Item construct validity
Sensibilité au changement	Guidelines	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"

DES MPR - Mod 1
D. GASQ
04/12/2023

Validité			
Domain	Measurement property	Aspect of a measurement property	Definition
Validité	Internal consistency	The degree of the interrelatedness among the items	
	Inclus dans Reliabilité = Cohérence interne Coefficient alpha de Cronbach Sous-parties de l'échelle → lié à la validité de construit : pré-requis = 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	
	Validité de construit = pour des échelles multi-items Hypothèses = validité convergente / divergente Utilisation de corrélations		
	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	
	Analyses factorielles, rasch...		
	Hypotheses testing	Item 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	
	Criterion validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"	
Mokkink et al. J Clin Epidemiol. 2010; Ferriman, Annales de réadaptation et de médecine physique 2005.			

Validité			
Domain	Measurement property	Aspect of a measurement property	Definition
Validité	Content validity	The degree to which an HR-PRO instrument measures the construct(s) it purports to measure	
	Face validity	The degree to which the content of an HR-PRO instrument is 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	Item 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	
	Criterion validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"	
Mokkink et al. J Clin Epidemiol. 2010; Ferriman, Annales de réadaptation et de médecine physique 2005.			

Validité			
Domain	Measurement property	Aspect of a measurement property	Definition
Validité	Content validity	The degree to which an HR-PRO instrument measures the construct(s) it purports to measure	
	Face validity	The degree to which the content of an HR-PRO instrument is 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	Item 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	
	Criterion validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"	
Mokkink et al. J Clin Epidemiol. 2010; Ferriman, Annales de réadaptation et de médecine physique 2005.			

Exemple d'étude de la validité – Adaptation transculturelle			
Domain	Measurement property	Aspect of a measurement property	Definition
Validité	Content validity	The degree to which an HR-PRO instrument measures the construct(s) it purports to measure	
	Face validity	The degree to which the content of an HR-PRO instrument is 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	Item 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	
	Criterion validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"	
Mokkink et al. J Clin Epidemiol. 2010; Ferriman, Annales de réadaptation et de médecine physique 2005.			

EnNSA-SL UL	FMA-5 UL	Light touch subscale	Proprioception subscale	Total score
Tactile sensation subscale	0.75	-	-	-
Proprioception subscale	-	0.77	-	-
Total score	-	-	0.74	-

EnNSA-SL UL: Erasmus modified Nottingham Sensory Assessment Somatosensory component for the upper limb; FMA-5 UL: Fugl-Meyer Assessment- Sensory component for the upper limb. * P < 0.001 (two tailed).

Internal consistency, floor and ceiling effect of the EnNSA UL and the FMA-5 UL (n=50).	Internal consistency ^a
Total score	EnNSA UL
Somatosensory component (n=22)	0.85
Stereognosis component (n=10)	0.94
Subscale of EnNSA-UL	
Tactile sensation (n=16)	0.94
Two-point discrimination (n=2)	0.82
Proprioception (n=4)	0.82

EnNSA-SLST UL: Erasmus modified Nottingham Sensory Assessment Somatosensory/Stereognosis component for the upper limb; FMA-5 UL: Fugl-Meyer Assessment- Sensory component for the upper limb. ^a Internal consistency was reported as Cronbach α coefficient.


Notion de norme (normative data)																
Validité	<ul style="list-style-type: none"> Une norme permet de situer le sujet: physio vs. patho <ul style="list-style-type: none"> 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 +++ 															
Reproductibilité	Mixed Populations: McKay et al., 2007; n = 1,000; mean (SD) = 6.7 (12.0) Reference values for the 6-Minute Walk Test by age group and sex (metres)															
Sensibilité au changement	<table border="1"> <thead> <tr> <th>Age Group (years)</th> <th>Male (mean (SD), n)</th> <th>Female (mean (SD), n)</th> </tr> </thead> <tbody> <tr> <td>3-9</td> <td>550 (117.6), 70</td> <td>556 (119.4), 70</td> </tr> <tr> <td>10-19</td> <td>748 (103.6), 80</td> <td>643 (83.6), 80</td> </tr> <tr> <td>20-29</td> <td>738 (88.3)^a, 200</td> <td>674 (76.8), 200</td> </tr> <tr> <td>60+</td> <td>599 (125.4)^a, 150</td> <td>550 (125.0), 150</td> </tr> </tbody> </table>	Age Group (years)	Male (mean (SD), n)	Female (mean (SD), n)	3-9	550 (117.6), 70	556 (119.4), 70	10-19	748 (103.6), 80	643 (83.6), 80	20-29	738 (88.3) ^a , 200	674 (76.8), 200	60+	599 (125.4) ^a , 150	550 (125.0), 150
Age Group (years)	Male (mean (SD), n)	Female (mean (SD), n)														
3-9	550 (117.6), 70	556 (119.4), 70														
10-19	748 (103.6), 80	643 (83.6), 80														
20-29	738 (88.3) ^a , 200	674 (76.8), 200														
60+	599 (125.4) ^a , 150	550 (125.0), 150														
Guidelines	^a Significant sex differences, p < 0.05															

Déroulé de l'exposé	
7. Taxonomie	1. Taxonomie des propriétés de mesure
7. Validité	2. Notions de validité
7. Effets seuils	3. Effets seuils
7. Reproductibilité	4. Reproductibilité
7. Sensibilité au changement	5. Sensibilité au changement
7. Guidelines	6. Guidelines

Notions d'effet plancher et plafond

25 / 65

- 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

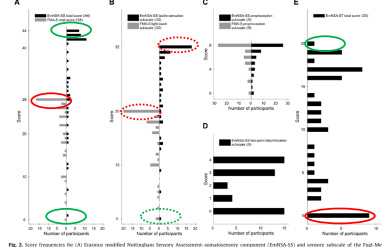


On a dit : "les 2 pieds sur la balance"

Tenwee et al. Journal of Clinical Epidemiology 2007.

Exemple d'une échelle d'étude de la sensibilité

26 / 65



Vilepinte et al. 2019.

Exemple de la stabilité posturale

27 / 65

- Berg Balance Scale (BBS)**
 - Échelle générique validée en gériatrie et neurologie (AVC, Parkinson)
 - 14 items, score total de 0 à 56
- Effet plancher +++ en post-AVC à la phase subaigüe

1. Passer de la position assise à debout	14. Se tenir debout sur une jambe (noter la jambe utilisée)
INSTRUCTIONS: Veuillez vous lever en essayant de ne pas vous aider avec les mains.	INSTRUCTIONS: Tenez-vous sur une jambe le plus longtemps possible, sans prendre appui.
(14 Peut se lever sans l'aide des mains et garder son équilibre	(14 Peut lever une jambe sans aide et tenir de 5 à 10 secondes
(13 Peut se lever seul avec l'aide de ses mains	(13 Peut lever une jambe sans aide et tenir de 3 à 5 secondes
(12 Peut se lever en s'aidant de ses mains, après plusieurs essais	(12 Essai de lever une jambe mais ne peut tenir 3 secondes tout en restant debout, sans aide
(11 Besoin d'un peu d'aide pour se lever ou garder l'équilibre	(10 Ne peut effectuer l'exercice ou a besoin d'aide pour ne pas tomber
(10 Besoin d'une aide modérée ou importante pour se lever	

Bernheim et al. 1989; Sawacha et al. 2013

Exemple de la stabilité posturale

28 / 65

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

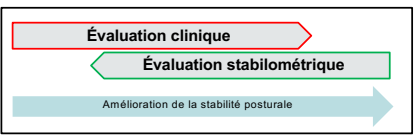
Couche	0	1	2	3
Se tenir vers le côté sans	impossible	aide importante	modérée	sans aide
Se tenir vers le côté hémiplegique	impossible	aide importante	modérée	sans aide
Se tenir	impossible	aide importante	modérée	sans aide
Assis	0	1	2	3
Assis sans support (plan horizontal, pieds au sol)	impossible	Support (1 main)	sans aide au moins 10 s	sans aide au moins 10 s
Se couche	impossible	aide importante	modérée	sans aide
Se met debout	impossible	aide importante	modérée	sans aide
Debout	0	1	2	3
Avec aide (personne libre des pieds)	impossible	sources de deux personnes au moins 10 s ou appui cas-impédal	Sources d'une personne au moins 10 s ou appui as-impédal	Sources d'une main possibilité de mouvements amples (> 45°)
Sans aide (position libre des pieds)	impossible	impossible	modérée	sans aide
S'assoit	impossible	aide importante	modérée	sans aide
Rattrape un petit objet (cylindre) 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 atteint	impossible	quelques secondes	au moins 5 s	au moins 10 s
Sous-total				
Score global	/ 36			

Bernheim et al. 1989; Nardone & Schieppati 2010; Sawacha et al. 2013

Exemple de la stabilité posturale

29 / 65

- 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

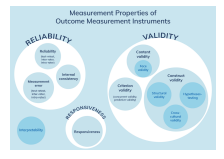


Vissier et al. 2006; Nardone & Schieppati 2010; Sawacha et al. 2013

Déroulé de l'exposé

30 / 65

- Taxonomie des propriétés de mesure
- Notions de validité
- Effets seuils
- Reproductibilité**
- Sensibilité au changement
- Guidelines



D. Gasq

Sources de variabilité

31 / 65

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

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

DES MPR - Mod. 1
D. GASQ
04/12/2023

Variabilité d'un paramètre

32 / 65

Conditions idéales

- Variabilité faible
- Grande différence entre les deux conditions

DES MPR - Mod. 1
D. GASQ
04/12/2023

Variabilité d'un paramètre

33 / 65

Conditions usuelles

- Variabilité importante
- Petite différence entre les 2 conditions

DES MPR - Mod. 1
D. GASQ
04/12/2023

Variabilité d'un paramètre

34 / 65

Conditions usuelles

- Variabilité importante
- Petite différence entre les 2 conditions

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

DES MPR - Mod. 1
D. GASQ
04/12/2023

Reproductibilité

35 / 65

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)	Reliability		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 IIR-FROs (internal consistency), over time (test-retest) by different persons on the same occasion (intrater) or by the same persons (i.e., raters or responders) on different occasions (intrater)
Reliability	Fidélité / Fidélité		The proportion of the total variance in the measurements which is because of "true" 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

- Reliability
 - ICC (coefficient de corrélation intraclass) si variable continue
 - Coefficient de kappa si variable binaire / nominale
- Marge d'erreur = minimal/small detectable change (MDC / SDC)
- Design = test-retest

DES MPR - Mod. 1
D. GASQ
04/12/2023

Comment étudier la reproductibilité

36 / 65

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

DES MPR - Mod. 1
D. GASQ
04/12/2023

Fiabilité avec l'ICC

37 / 65

- 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

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

DES MPR - Mod 1
D. GASQ
04/12/2023
De Vet et al. 2005

Fiabilité avec l'ICC

38 / 65

- 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 ▲).

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

DES MPR - Mod 1
D. GASQ
04/12/2023
De Vet et al. 2005

Fiabilité avec l'ICC

39 / 65

Shrout and Fleiss	Computational formula	McGraw and Wong	Model
1.1	$MS_b = k - 1MS_a$	1	1-way random
1.k	$\frac{MS_b}{MS_b + MS_e}$	k	1-way random
2.1	$\frac{MS_b - MS_c}{MS_b + (k-1)MS_c}$	C.1	2-way random
2.k	$\frac{MS_b - MS_c}{kMS_b - (k-1)MS_c}$	A.k	2-way random
3.1	$\frac{MS_b}{MS_b + (k-1)MS_c}$	C.1	2-way fixed
3.k	$\frac{MS_b}{MS_b + MS_c}$	C.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_c = subjects mean square; MS_w = within-subjects mean square.

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

DES MPR - Mod 1
D. GASQ
04/12/2023
Weir 2005

Marge d'erreur

40 / 65

- 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.

DES MPR - Mod 1
D. GASQ
04/12/2023
Guyatt et al. 1985, Bland et Altman 1986, De Vet et al. 2005

Marge d'erreur : un peu d'histoire

41 / 65

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

Fig 2—Difference against mean for PEFR data.

Bland et Altman 1986.

THE LANCET, FEBRUARY 8, 1986

Measurement

J. MARTIN BLAND DOUGLAS G. ALTMAN

STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

DES MPR - Mod 1
D. GASQ
04/12/2023

Marge d'erreur : un peu d'histoire

42 / 65

- MMDC & MDC
 - Minimal (metrically) detectable change
 - IC 95% des erreurs
- Smallest real difference (SRD)
 - Smallest real difference, a link between reproducibility and responsiveness

Fig 4. Distribution of the difference in total MMDC scores between (A) and (B) for 100 and 1000 subjects (100 differences, and 1000 differences).

Hibbert et al. 1997, Beckerman et al. 2001.

Setting the Minimal Metrically Detectable Change on Disability Rating Scales

Rosenstock, MD, Dettl, J, Fong, MD, PhD, C, et al.

H. Beckman, M.E. Rothman, G.J. Laskowski, J.G. Butler, P.D. Rozema, & A.L.M. Verbeek

DES MPR - Mod 1
D. GASQ
04/12/2023

Recommandations pour la reproductibilité

43 / 65

- 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 * √2 * SEM

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

DES MPR - Mod 1
D. GASQ
04/12/2023
Gayath 1995; Mokkink et al. Patient Related Outcome Measures 2023. <https://www.cochrane.org/>

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

44 / 65

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.

DES MPR - Mod 1
D. GASQ
04/12/2023
De Vet et al. 2005

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

45 / 65

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

		ICC (95% CI)	MDC ₉₅
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: montre le niveau de l'erreur de mesure
- MDC = utilisable en pratique clinique

DES MPR - Mod 1
D. GASQ
04/12/2023
Vilreppé et al. 2019

Marge d'erreur et nombre d'essais moyennés ?

46 / 65

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	
			8 ± 3.5	
RMS (mm)	ML	14.1 ± 4.3	1 ± 5.1	
			6 ± 2.7	
COP velocity (mm/s)	AP	3.7 ± 1.0	1 ± 1.2	
			7 ± 0.9	
COP velocity (mm/s)	ML	2.3 ± 0.9	1 ± 1.2	
			8 ± 0.6	
COP velocity (mm/s)	AP	7.9 ± 1.7	2 ± 1.0	
	ML		2 ± 0.6	

*Number of trial averaged for ICC ≥ 0.90.

DES MPR - Mod 1
D. GASQ
04/12/2023
Lafont et al. 2004

Notion de biais systématique

47 / 65

- 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 difference, and the parallel line to the X axis at -27.2 units.

DES MPR - Mod 1
D. GASQ
04/12/2023
Gavettini 2015

Notion de biais systématique

48 / 65

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).

DES MPR - Mod 1
D. GASQ
04/12/2023
Gavettini 2015

Exemple de biais systématique

49 / 65

- Tenir compte ou pas du biais systématique ?

Figures 4 Variation of interest for agreement parameters (upper panel) and variation of interest for consistency parameters (lower panel); positive yellow bars, negative bars, blue.

Exemple de biais systématique

50 / 65

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

DES MPR - Mod. 1
D. GASQ
04/12/2023
Vilepinte et al. 2019

Exemple de calcul de la marge d'erreur Impact de la méthode statistique

51 / 65

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

		mDiff (95% CI)	MDC95	MDC95 with bias correction
EmNSA-SS total score (44)	Intra-rater	0.65 (-0.63-1.93)	12.3	8.9
	Inter-rater	0.91 (-0.31-2.14)	14.6	8.9
Tactile sensation (32)	Intra-rater	0.27 (-0.89-1.43)	8.6	8.6
	Inter-rater	0.93 (-0.07-1.94)	14.1	8.6
Discrimination of two points (4)	Intra-rater	0.08 (-1.14-0.31)	1.9	2.4
	Inter-rater	-0.07 (-0.42-0.29)	3.5	2.4
Proprioception (8)	Intra-rater	0.29 (-0.01-0.60)	4.4	3.9
	Inter-rater	0.04 (-0.48-0.57)	3.5	3.9
EmNSA-ST (20)	Intra-rater	1.67 (0.97-2.37)*	23.1	7.0
	Inter-rater	2.00 (0.94-3.06)*	27.5	7.0

DES MPR - Mod. 1
D. GASQ
04/12/2023
D'après Vilepinte et al. 2019. *The only statistics you can trust are the ones you have labelled yourself! Winston Churchill 1967

Marge d'erreur en synthèse

52 / 65

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 ?

DES MPR - Mod. 1
D. GASQ
04/12/2023
Gavettini 2011

Distribution des écarts à la moyenne ?

53 / 65

Distribution hétérosélastique
+, proportionnalité marge erreur & valeur de mesure

Distribution homosélastique

Gasq et al. 2014

Comment allez vous faire en pratique ?

54 / 65

- Choix de l'ICC adéquat
- Graph de Bland-Altman : biais systématique ? Outliers ?
- Distribution des différences test-retest : hétérosélasticité ?
- Calcul du MDCs :
 - Si hétérosélasticité = expression en % (ou transformation logarithmique)
 - Si non = valeur dans l'unité originelle

DES MPR - Mod. 1
D. GASQ
04/12/2023
Bland et Altman 1996; Hopkins 2000; Lisse et al. 2005; de Vet et al. 2005; Mokkink et al. Patient Related Outcome Measures 2023.

Déroulé de l'exposé

1. Taxonomie des propriétés de mesure
2. Notions de validité
3. Effets seuils
4. Reproductibilité
5. **Sensibilité au changement**
6. Guidelines

DES MPR - Mod. 1
D. GASQ
04/12/2023

Sensibilité au changement

Domain	Measurement property	Aspect of a measurement property	Definition
Taxonomie	Responsiveness		The ability of an HR-PRO instrument to detect change over time in the construct to be measured
Validité	Interpretability*	Responsiveness = <i>sensitivity to change</i>	Item responsiveness 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.

DES MPR - Mod. 1
D. GASQ
04/12/2023

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

Sensibilité au changement

Domain	Measurement property	Aspect of a measurement property	Definition
Taxonomie	Responsiveness		The ability of an HR-PRO instrument to detect change over time in the construct to be measured
Validité	Interpretability*	Responsiveness = <i>sensitivity to change</i>	Item responsiveness 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 père pesé personne a-t-il détecté un changement de poids avant que je ne change de taille de vêtements ?

DES MPR - Mod. 1
D. GASQ
04/12/2023

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

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

DES MPR - Mod. 1
D. GASQ
04/12/2023

Mokkink et al. Journal of Clinical Epidemiology 2010.

Quantifier la sensibilité au changement ?

• Notion statistique

- Taille d'effet et *standardized response mean*
- Souvent utilisé → vérifier la pertinence clinique ?

• Référence à une notion clinique indispensable

- Changement cliniquement pertinent pour le patient

• Evaluation valide si MDC < MCID

DES MPR - Mod. 1
D. GASQ
04/12/2023

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

Quantifier la sensibilité au changement ?

REVIEW ARTICLES

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

Dennis Revicki^{1,6}, Ron D. Hays², David Cella³, Jeff Sloan⁴

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.

What is new?

- Estimation of MID for a specific PRO measure should be based on multiple approaches and triangulation of methods.
- Various methods for estimating MID's 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.

DES MPR - Mod. 1
D. GASQ
04/12/2023

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

