

# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

## RAPORT TEORETIC PARTEA I



Co-funded by the  
Erasmus+ Programme  
of the European Union

Număr de referință: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

## Noțiuni de bază despre cercetare

Acest raport rezumă câteva aspecte cheie ale cercetării clinice în științele sănătății. O astfel de cercetare poate fi de feluri foarte diferite și poate fi clasificată în sens foarte larg ca cercetare cantitativă, calitativă sau cu metode mixte, în funcție de tipul de metode care sunt utilizate pentru colectarea și analiza datelor. Toate trei au rolul și locul lor definite și niciunul nu este „mai bun” decât altul; alegerea depinde în cele din urmă de obiectivele sau întrebările de cercetare care stau la baza unui anumit studiu.

În esență, cercetarea cantitativă se bazează pe date numerice (care, la rândul lor, pot fi fie calitative, fie cantitative prin natură, la care vom reveni mai târziu). Acesta își propune să descrie și să sintetizeze cantitățile sau dimensiunile diferitelor fenomene și să înțeleagă modul în care diferitele fenomene sunt legate. Acest lucru se face cu intenția de a trage concluzii generalizabile care sunt valabile nu numai în studiul respectiv, ci și în alte situații sau grupuri de oameni și pentru a putea face predicții (de exemplu, cu privire la eficacitatea unei anumite intervenții).

În schimb, cercetarea calitativă nu se bazează pe date numerice, ci pe date audio sau text (poate implica și alte tipuri, de exemplu, observații, imagini sau filme, dar acest lucru este mai puțin obișnuit). Cel mai comun tip de date calitative în științele sănătății sunt transcrierile text ale interviurilor sau povestirilor referitoare la fenomenul pe care se dorește să-l studieze. În general, acestea sunt analizate pentru a rezuma și a obține o înțelegere mai profundă a naturii diferitelor fenomene. Adică, în cercetarea calitativă se dorește să se înțeleagă cum este ceva, ce înseamnă să trăiești într-o anumită condiție, experiențele, opiniile sau gândurile oamenilor cu privire la anumite lucruri sau situații, etc. Există o gamă largă de metode și abordări pentru a efectua cercetări calitative și pentru a analiza calitativ, inclusiv (dar fără a se limita la) analiza de conținut, analiza tematică, teoria fundamentată, hermeneutică și fenomenologică. În cele din urmă, în cazul cercetării cu metode mixte, în cadrul aceluiași studiu este utilizată o combinație de abordări cantitative și calitative.

Acest raport se concentrează pe cercetarea cantitativă și în special pe studiile care intenționează să evalueze eficacitatea sau valoarea unei anumite intervenții sau tratament, adică studii intervenționale. Acest lucru nu se datorează faptului că cercetarea cantitativă este „mai bună” (din nou, „cea mai bună” abordare este guvernată de obiectivele studiului), ci pentru că aceasta este cea mai comună abordare în cercetarea clinică și pentru că de obicei acest tip de studii sunt la baza ghidurilor clinice și a recomandărilor. În plus, există o serie de aspecte de bază ale cercetării cantitative care sunt necesare pentru a înțelege, pentru a înțelege și revizui critic rapoartele de cercetare – unele dintre aceste aspecte nu sunt neglijate sau greșit înțelese, dar sunt importante de înțeles pentru a permite eficient și corect, comunicarea cu pacienții și membrii familiei acestora, precum și cu alți profesioniști din domeniul sănătății (HCPs).

În plus, ne referim la Anexa 1, care oferă scurte explicații ale termenilor și conceptelor obișnuite în cercetarea clinică, dintre care unii pot să nu fie explicați în text.



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



## Practica bazată pe dovezi

În ultimele câteva decenii, principiile practicii bazate pe dovezi (EBP) au câștigat o importanță sporită și, de obicei, stau la baza ghidurilor și recomandărilor clinice. EBP poate fi definită ca integrarea expertizei clinice cu cele mai bune dovezi clinice externe disponibile din cercetarea sistematică (Sacket et al. 1995). În ceea ce privește cele mai bune dovezi clinice disponibile din cercetarea sistematică, un aspect cheie este proiectul cercetării, care se referă la modul în care este efectuat un studiu. Aceasta include modul în care participanții sunt recrutați (sau eșantionați) la studiu, câți oameni care participă la studiu, prin ce mijloace sunt colectate datele, cum arată procedurile studiului (de exemplu, când, cum și de către cine sunt colectate datele și când, cum și de către cine este furnizată intervenția de studiu) și cum sunt analizate datele. Acești toți factori contribuie la măsura în care rezultatele studiului pot fi de încredere, sunt interpretabile, reprezentative și generalizabile.

Există o ierarhie generală în ceea ce privește nivelurile de dovezi, adică măsura în care rezultatele cercetării sunt considerate solide și de încredere. Această ierarhie este ilustrată în Figura 1 și se bazează în mare parte pe proiectul de cercetare folosit. Cel mai înalt nivel este reprezentat de recenzii sistematice ale cercetărilor publicate, de preferință cuplate cu o meta-analiză a datelor din studiile originale incluse în revizuire. Termenul de meta-analiză înseamnă că se realizează o nouă analiză statistică a datelor publicate. Rezultatele unei meta-analize sintetizează datele raportate anterior în rezultate noi care reprezintă nivelul actual de cunoștințe privind, de exemplu, eficacitatea sau valoarea unei anumite intervenții (Impellizzeri & Bizzini 2012; Haidich 2010). Dacă lipsesc recenzii sistematice și meta-analize, rezultatele din studiile individuale randomizate controlate (RCT) sunt considerate a doua cea mai bună sursă de dovezi, urmate de studiile nerandomizate, studiile observaționale, seriile de cazuri și, la cel mai scăzut nivel al ierarhiei dovezilor, opiniile experților.

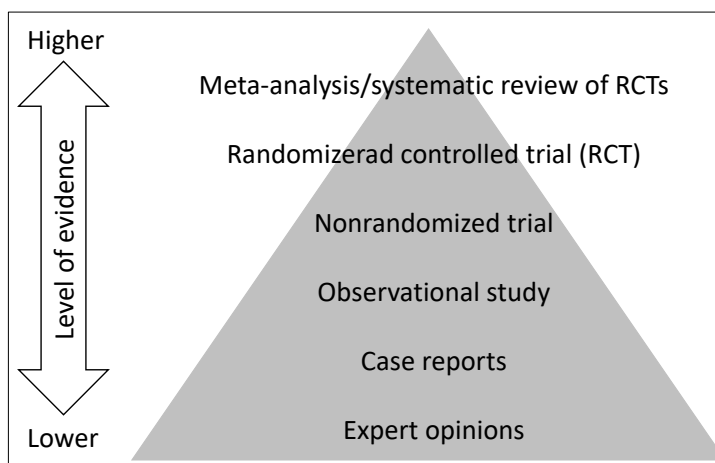


Fig. 1 Ierarhia probelor

## Proiecte de cercetare





# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

După cum este evident din această ierarhie, randomizarea este un aspect cheie de proiectare a cercetării, în special în studiile intervenționale. Cu toate acestea, există câteva alte proiecte de cercetare (Grimes & Schulz 2002; Miller et al. 2020; Noordzij et al. 2009), dintre care unele dintre cele mai comune vor fi discutate mai târziu. Importanța randomizării se datorează faptului că randomizarea participanților la studiu către, de exemplu, grupul de intervenție activ sau grupul de control este cea mai bună și mai eficientă modalitate de a controla orice părtinire a rezultatelor studiului (Kang și colab. 2008). Lăsând șansa să decidă cărui grup va aparține fiecare participant, probabilitatea de a avea grupuri comparabile la începutul unui studiu este maximizată; dacă grupurile diferă în orice aspect potențial important (de exemplu, vârsta), există riscul ca acest lucru să țină seama de unele aspecte ale oricăror diferențe în rezultatele studiului care pot fi observate între grupuri (adică, există o părtinire). Grupul de intervenție (uneori denumit și grupul activ) este format din participanți care primesc intervenția care este studiată, iar grupul de control primește o intervenție alternativă care este utilizată ca comparator în studiu. Aceasta poate fi nici o intervenție, îngrijire obișnuită sau placebo. Placebo este o intervenție inactivă care este cât mai asemănătoare cu cea oferită grupului de intervenție, dar nu are componentele critice despre care se crede că țin seama de rezultat. Studiul clasic controlat cu placebo este un studiu de droguri, în care pilula placebo arată, are gust și se simte exact ca pilula activă, dar nu conține substanța activă farmacologică. Dar există și alte exemple, cum ar fi intervențiile bazate pe informare sau educație, în care condiția placebo nu include aspectele considerate a fi importante pentru obținerea rezultatului dorit. Pentru a consolida și mai mult proiectul, este, de asemenea, de dorit să lăsați participanții, precum și clinicienii să nu știe în ce grup au fost randomizați. Acest lucru este denumit studii unice (participanții la studiu nu știu cărui grup aparțin) și studii dublu-orb (nici participanții și nici clinicienii nu știu cine a fost alocat în ce grup). Din motive etice, este totuși important ca toți participanții să cunoască exact ce înseamnă a fi randomizat în oricare dintre grupurile de studiu și că nu vor fi conștienți de alocarea lor până când studiul nu va fi încheiat. În plus, dacă rezultatele studiului sunt pozitive, participanților din grupul de control ar trebui să li se ofere intervenția, dacă rezultatele sunt favorabile la încheierea studiului (World Medical Association 2018).

În anumite contexte, în special studiile de medicamente farmacologice, studiile sunt clasificate în patru faze (Faza I-IV), reflectând stadiul de dezvoltare a intervenției studiate. Studiile de Fază I se referă la siguranța noii intervenții și se desfășoară cu grupuri relativ mici (până la poate 100) de oameni, de obicei (în cazul unui medicament nou) voluntari din domeniul sănătății, în timp ce studiile de Fază II se concentrează pe determinarea dacă intervenția funcționează așa cum s-a intenționat (adică, pentru a vedea dacă are efectul scontat, ceea ce nu este clar la începutul unui studiu de Fază II). Dacă intervenția trece de Fază II (adică este eficientă), sunt efectuate studii de Fază III pentru a o compara cu alte intervenții (sau placebo/îngrijire standard) și pentru a monitoriza în continuare efectele secundare. Cu intervenții comerciale (de exemplu, medicamente noi), studiile de Fază III sunt baza pentru acordarea de licențe și comercializarea produsului către populația mai largă. În cele din urmă, studiile de Fază IV sunt efectuate după ce intervenția a fost pusă la dispoziție, de obicei pentru a studia „lumea reală” și/sau eficacitatea pe termen lung.



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

În multe cazuri, poate fi practic dificil sau chiar imposibil să randomizați participanții la studiu în grupurile de intervenție activă și de control sau, într-adevăr, să aveți un grup de control rezonabil comparabil, în același timp, asigurați-vă că nu există nicio interacțiune sau „efecte de propagare” între grupuri. Câteva exemple de intervenții care sunt dificil de controlat într-o astfel de manieră includ intervenții educaționale, psihosociale, de îngrijire și de autogestionare. O altă situație poate fi atunci când o intervenție este testată mai întâi, iar investigatorii nu își pot permite timpul, efortul sau banii pentru a efectua un RCT la scară largă. Indiferent de motiv(e), proiectele de studii care nu includ randomizarea sunt denumite studii nerandomizate sau cvasi-experimentale. Un studiu care include un grup de control, dar nu are randomizare este uneori denumit și studiu clinic controlat (CCT), dar este încă un studiu cvasi-experimental. În esență, un proiect cvasi-experimental se referă la orice tip de studiu căruia îi lipsește randomizarea, un grup de control sau ambele. Acest lucru contrastează cu modelele experimentale adevărate, cum ar fi RCT, care utilizează randomizarea și un grup de control.

Studiile observaționale diferă de tipurile de studii menționate anterior prin aceea că investigatorul nu intervine și nu face nimic în afară de colectarea datelor. Există trei tipuri de bază de modele de studii observaționale: studii de cohortă (un grup de persoane este urmărit prospectiv în timp), studii caz-control (un grup de persoane cu o anumită caracteristică sunt studiate retrospectiv, de obicei împreună cu un grup de control care este foarte similar cu grupul de studiu, cu excepția unei anumite caracteristici), și studii transversale (datele sunt colectate la un moment dat). Studiile de cohortă și de caz-control sunt utilizate în mod obișnuit în cercetarea epidemiologică, unde este de interes să se studieze expuneri sau caracteristici care pot prezice evenimente viitoare (studii de cohortă) sau expuneri sau caracteristici din trecut pot fi asociate cu dezvoltarea a ceva (studii de caz-control). În plus, studiile de cohortă sunt folosite pentru a studia istoria naturală a diferitelor condiții. Deoarece studiile transversale se bazează pe colectarea datelor la un moment dat, le lipsește perspectiva temporală și oferă doar un „instantaneu” a variabilelor studiate. Ca atare, utilizarea lor este limitată la scopuri descriptive și pentru a explora diverse asocieri.

Similar cu rolul important al alocării aleatorii pe grupuri de studiu în studiile intervenționale, selecția aleatorie a participanților la studiu este cea mai bună și mai eficientă modalitate de a asigura eșantioane reprezentative și rezultate reprezentative în studiile non-intervenționale, cum ar fi anchetele. Acest lucru se datorează faptului că lăsarea întâmplării să decidă cine să fie luat în considerare pentru participarea la un studiu maximizează probabilitatea ca eșantionul (adică, persoanele selectate pentru participare) să oglindească populația (adică, tipul de persoane despre care studiul intenționează să învețe) cu privire la caracteristici precum vârsta, sexul și alte variabile relevante. Acest lucru este important deoarece cu cât eșantionul de studiu este mai reprezentativ pentru populația sa, cu atât este mai probabil ca rezultatele studiului să fie generalizabile. Aici trebuie menționat că investigatorul este cel care definește care este populația de studiu. Acest lucru se realizează de obicei prin intermediul unor criterii de includere și excludere, care clarifică astfel la ce fel de persoane se referă studiul, ce încearcă să învețe și despre care să tragă concluzii și, prin



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

urmare, cărora le generalizează. În schimb, grupul de persoane care participă efectiv la un studiu este denumit eșantion.

Procesul de identificare și selectare a persoanelor pentru participarea la un studiu este denumit eșantionare (Elfil & Negida 2017; Shorten & Moorley 2014). Există multe tipuri de metode de eșantionare disponibile, dar cele mai multe dintre ele pot fi clasificate fie ca metode de eșantionare cu probabilitate (adică, aleatorii) sau nonprobabilitate (adică, non-aleatorii). În esență, eșantionarea probabilă înseamnă că fiecare persoană din populație are o probabilitate egală de a fi selectată pentru eșantionul de studiu. După cum s-a menționat mai devreme, aceasta este, în general, de preferat eșantionării nonprobabile, deoarece maximizează reprezentativitatea populației. Cu toate acestea, în multe cazuri poate fi dificil sau chiar imposibil să se folosească acest tip de eșantionare și se va folosi în schimb un fel de eșantionare non-probabilă. Metodele obișnuite de eșantionare probabilă includ eșantionarea aleatorie simplă (selectarea aleatorie a unui număr  $X$  de persoane dintr-o listă a tuturor persoanelor din populație), eșantionarea aleatorie sistematică (selectarea aleatorie a unei persoane urmată de selecția, de exemplu, a fiecărei a zecea persoană din grupul randomizat inițial), eșantionare aleatorie stratificată (selecție aleatorie din diferite subgrupuri relevante predefinite de persoane) și eșantionare în grup (selectare aleatorie a unităților, de exemplu, spitale sau zone geografice, mai degrabă decât persoane, urmată de selecția persoanelor din acele unități). Metodele obișnuite de eșantionare non-probabilă includ eșantionarea convenabilă (selectarea participanților în funcție de disponibilitatea și accesibilitatea lor, de exemplu, toate persoanele care frecventează o anumită clinică și îndeplinesc criteriile de includere/excludere), eșantionarea normată (similar cu eșantionarea stratificată, dar fără selecție aleatorie) și eșantionare *Snowball*-bulgăre de zăpadă, (se cere persoanelor selectate inițial să identifice alte persoane de același fel; utilizată de obicei în studii care implică persoane greu accesibile, de exemplu, persoane fără adăpost).

După cum s-a menționat mai devreme, proiectul cercetării se referă la modul în care este efectuat un studiu și susține măsura în care rezultatele studiului pot fi de încredere, sunt interpretabile, reprezentative și generalizabile. Aceste aspecte revin la ceea ce păstrează validitatea unui studiu, care se referă la cât de încrezători putem fi în ceea ce privește rezultatele unui studiu, concluziile și implicațiile aferente. Există mai multe tipuri de validitate descrise în literatură, dintre care trei tipuri principale sunt luate în considerare aici: validitatea internă, validitatea externă și validitatea concluziei statistice. Validitatea internă se referă la măsura în care putem avea încredere în rezultatele unui studiu, în timp ce validitatea externă se referă la măsura în care acele rezultate pot fi generalizate la alte situații dincolo de studiul în cauză. În mod evident, validitatea externă necesită astfel validitate internă, care, la rândul său, se bazează pe probleme de proiectare a cercetării, cum ar fi cele abordate mai sus, adică randomizarea, grupurile de control, eșantionarea și așa mai departe. Aspectele suplimentare includ metode și proceduri legate de colectarea datelor. De exemplu, dacă rezultatele studiului sunt relevante din punct de vedere clinic, dacă datele au fost colectate folosind instrumente și protocoale care sunt valide și de încredere și dacă rezultatele sunt interpretabile și semnificative. Aceste aspecte vor fi abordate mai detaliat mai târziu. Al treilea aspect de validitate, validitatea concluziei statistice, se referă la dacă numărul de



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

participanți la studiu a fost adecvat pentru a îndeplini obiectivele studiului și în legătură cu metodele utilizate pentru analiza datelor, precum și dacă acele metode sunt adecvate având în vedere obiectivele studiului. și tipul de date din studiu. Astfel de aspecte vor fi abordate și ulterior.

Problemele de validitate devin deosebit de pertinente în studiile intervenționale și alte studii care încearcă să abordeze cauzalitatea, adică dacă un anumit efect sau rezultat este cauzat de un anumit lucru (de exemplu, o intervenție). Pentru a susține o relație cauzală, există câteva cerințe de bază care trebuie îndeplinite (a se vedea, de exemplu, Greenhalgh 1997 b). Acestea sunt mai multe și vom include doar câteva aici. În primul rând, ar trebui să existe o asociere între cauză și efect. Cu toate acestea, acest lucru nu este suficient, deoarece poate fi văzută o asociere din diferite motive. De exemplu, dacă există o asociere puternică (de exemplu, o corelație) între X și Y, aceasta nu spune în sine dacă X este cauza lui Y sau invers. În plus, ar putea exista și altceva (să-i spunem Z) care are o relație cauzală atât cu X, cât și cu Y, ceea ce face să pară că X și Y sunt puternic asociate atunci când de fapt nu sunt (în acest caz Z ar fi considerat un factor de confuzie). În al doilea rând, trebuie să existe o relație temporală între cauză și efect. Mai exact, cauza trebuie să preceadă efectului, altfel nu poate fi responsabilă pentru efect. O implicație a acestui lucru este că este imposibil să se tragă concluzii cauzale din datele transversale. În al treilea rând, ar trebui să fie posibilă excluderea unor explicații alternative pentru efect. Acesta este un motiv major pentru care este important să existe un grup de control (pentru a exclude, de exemplu, că efectul nu s-ar produce doar lăsând timpul să treacă) și să se folosească repartizarea aleatorie a participanților (pentru a se asigura că grupurile sunt comparabile).

## Cifre: Măsurare și evaluare

După cum s-a menționat inițial, cercetarea cantitativă se bazează pe date numerice. Prin urmare, este relevant să luăm în considerare semnificația numerelor, ceea ce pot reprezenta și ceea ce ar putea să nu reprezinte. În principiu, numerele în sine sunt abstracte și lipsite de sens și vor necesita informații din exterior pentru a avea sens. Luați în considerare, de exemplu, numărul 39. Fără alte informații, tot ce putem spune este, în esență, că este un număr între 38 și 40. Pentru a-i da sens, ar trebui cel puțin să știm ce variabilă reprezintă (de exemplu, timpul, temperatura, distanța) și unitatea pe care o exprimă (de exemplu, ani, grade C, km). În plus, vor fi necesare cunoștințe cu privire la intervalul și contextul posibil (de exemplu, vârsta unei persoane, temperatura corpului sau distanța până la cabinetul local de medicină generală). După cum am menționat anterior, acest tip de informații trebuie să provină din afara datelor, iar orice calcule sau analize statistice vor rămâne abstracte și lipsite de sens fără ele.

### *Sensul numerelor*



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



În timp ce cercetarea cantitativă se referă la date numerice, este important să fim conștienți de faptul că datele numerice pot fi fie cantitative, fie calitative (Sevens 1946). Această distincție se referă la ceea ce reprezintă numerele (Tabelul 1).

**Tabelul 1.** Tipuri și niveluri de date numerice

Tip	Nivelul de date	Caracteristici	Informații furnizate	Exemplu
Calitativ	Nominal	<ul style="list-style-type: none"><li>• Categorii/etichete numerice</li><li>• Fără semnificație cantitativă</li></ul>	<ul style="list-style-type: none"><li>• Clasificare</li></ul>	Evaluări (de exemplu, sex, etnie, grupă de sânge) 0 = Femeie; 1 = Bărbat
	Ordinal	<ul style="list-style-type: none"><li>• Categoria ordonată</li><li>• Intervale necunoscute (inegale) între numere</li></ul>	<ul style="list-style-type: none"><li>• Ordinea clasamentului</li></ul>	Evaluări, scale de evaluare (de exemplu, bunăstare, severitatea simptomelor) 0 1 2 3 4 5
Cantitativ	Interval	<ul style="list-style-type: none"><li>• Linear</li><li>• Intervale egale între numere</li><li>• Zero arbitrar (fără absolut)</li></ul>	<ul style="list-style-type: none"><li>• Ordinea clasamentului</li><li>• Diferența cantitativă</li></ul>	Măsurătorile (de exemplu, temperatură/Celsius) ... -1 0 1 2 3 4...
	Raport	<ul style="list-style-type: none"><li>• Linear</li><li>• Intervale egale între numere</li><li>• Zero absolut</li></ul>	<ul style="list-style-type: none"><li>• Ordinea clasamentului</li><li>• Diferența cantitativă</li><li>• Raporturi</li></ul>	Măsurători (de exemplu, lungime, biomarkeri, tensiune arterială) 0 1 2 3 4 5...

Datele numerice calitative reprezintă diferite categorii ale unei variabile. Există două tipuri de date numerice calitative, denumite nominale și ordinale. Datele la nivel ordinal sunt coduri numerice fără nicio semnificație cantitativă (adică nu există „mai mult” sau „mai puțin” implicit), de exemplu, sexul poate fi clasificat ca feminin sau masculin și etichetat cu cifre. Datele ordinale sunt, de asemenea, coduri numerice, dar în acest caz este implicată o ordine de la mai puțin la mai mult. Acest tip de date este foarte comun în practica clinică și cercetare și este, de exemplu, ceea ce vedem cu diverse scale de evaluare (de exemplu, 0=niciunul, 1=ușor, 2=moderat, 3=sever). În timp ce datele ordinale sunt ordonate de la mai puțin la mai mult, nu spune cât de mult mai puțin sau mai mult, ci doar că, de exemplu, 2 este mai mult decât 1 și mai puțin decât 3. Acest lucru se datorează faptului că diferențele reale dintre categoriile ordonate sunt necunoscute și, în general, inegale (de exemplu, diferența dintre 1 și 2 poate fi de două ori mai mare decât cea dintre 0 și 1).

Datele cantitative, pe de altă parte, reprezintă măsurători liniare. Aceasta înseamnă că intervalele (sau diferențele) dintre două valori numerice sunt aceleași în intervalul de valori posibile, ceea ce înseamnă, la rândul său, că calculele precum adunarea și scăderea sunt semnificative și că diferențele pot fi exprimate. Există două tipuri de date numerice cantitative, denumite date de interval și raport. Diferența dintre acestea două este că, cu datele la nivel de interval, nu există zero absolut. În practică, datele de interval nu sunt foarte





# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

comune și unul dintre puținele exemple comune este temperatura măsurată în Celsius sau Fahrenheit (în timp ce temperatura conform Kelvin este pe o scară de nivel de raport). Datele la nivel de raport, pe de altă parte, sunt foarte frecvente în practica clinică, precum și în viața de zi cu zi (de exemplu, lungimea, greutatea, tensiunea arterială, diferite concentrații de biomarkeri în probele de sânge și așa mai departe). În practică, diferența dintre datele de nivel de interval și raport este marginală. Ambele permit același tip de calcule matematice, cu excepția faptului că rapoartele și procentele nu pot fi exprimate în mod valid pe baza datelor brute de interval. De exemplu, dacă temperatura exterioară este de +8 °C azi și a fost de +2 °C ieri, nu putem spune că astăzi este de 4 ori mai cald ( $8/2=4$ ), deoarece 0 nu este zero absolut. Acest lucru este clar dacă se consideră aceeași relație, dar cu temperaturi mai scăzute cu 3 grade, adică +5°C și -1°C ( $5/-1=-5$ ). Dacă, pe de altă parte, luăm în considerare două diferențe pe o scară de interval (să spunem diferența dintre +5°C și -1°C, adică 6 °C, și cea dintre +8 °C și +4°C, adică , 4 °C), rapoartele și procentele pot fi calculate și exprimate într-un mod semnificativ (în acest exemplu,  $6/4=1,5$ , adică prima diferență este cu 50% mai mare decât a doua).

Distincția dintre datele numerice calitative și cantitative are o istorie lungă. De exemplu, în manualul său din 1920 *Physics: The Elements*, fizicianul Norman Campbell a făcut distincția între numere și cifre, unde numerele reprezintă date cantitative care permit calcule aritmetice, în timp ce cifrele reprezintă date calitative care nu permit (Campbell 1920). Implicațiile sunt importante, dar nu rareori uitate.

În general, și din perspectiva cercetării clinice, datele cantitative implică măsuri și măsurători, în timp ce datele calitative sunt aprecieri și evaluări. Deși această distincție este importantă în sens practic, acestea nu sunt în mod inerent diferite sau separate unele de altele la un nivel mai fundamental. Acest lucru se datorează faptului că toate măsurătorile încep cu un fel de observație sistematică și pe baza acesteia se face o evaluare a cantității observate (de la mai puțin la mai mult). Revenind la taxonomia analizată mai sus, aceste aprecieri și evaluări sunt calitative prin natură și reprezintă toate măsurătorile, chiar și măsurătorile fizice, cum ar fi lungimea și greutatea, pe care tindem să le considerăm de la sine înțeles, dar la un moment dat în istorie au fost inventate și concepute prin acest mod/tip de proces. Transformarea observațiilor și aprecierilor în cuantificare și măsurători ale cantității necesită un proces de măsurare care, printre altele, presupune definirea și exprimarea unei unități care este stabilă și liniară. Exemple de astfel de unități includ metrul, **kilogramul și al doilea**. Într-adevăr, deși există mai multe definiții ale măsurării disponibile, definiția internațională afirmă că măsurarea este un proces de obținere a valorilor cantității (adică, unități) care pot fi atribuite în mod rezonabil unei cantități (Joint Committee for Guides in Metrology 2012), care este în conformitatea definiției de măsurare obiectivă oferită de Institutul de Măsurare Obiectivă (Institutul de Metrologie): „repetarea unei cantități unitare care își menține dimensiunea, într-un interval admisibil de eroare, indiferent de instrumentul utilizat și indiferent cine sau ce persoană sau lucru relevant este măsoarat” (Comitetul de program al Institutului de Măsurare Obiectivă 2000). Acest lucru clarifică faptul că este important să se facă distincția între numerele care reprezintă evaluări și cele care reprezintă măsurători, deoarece semnificațiile lor sunt fundamental diferite.



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

## *Scale de evaluare*

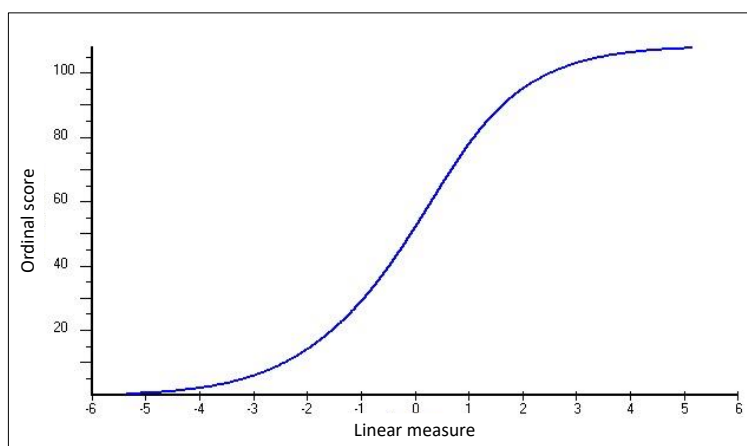
Scalele de evaluare sunt utilizate în mod obișnuit în practica clinică, precum și pentru a evalua rezultatele în cercetarea clinică. Termenul scală de evaluare este adesea folosit într-un sens foarte larg pentru a reprezenta evaluări structurate bazate pe observații sau auto-rapoarte. Utilizarea pe scară largă a scalelor de evaluare înseamnă că acestea joacă un rol central în deciziile cu mize mari privind, de exemplu, utilitatea terapiilor. Cu toate acestea, calitatea rezultatelor studiului clinic depinde direct de calitatea instrumentelor utilizate pentru colectarea datelor. Prin urmare, este relevant să se ia în considerare scalele de evaluare și utilizarea lor în practica clinică și cercetare (Cano & Hobart 2011).

Metoda sau instrumentul folosit pentru a evalua sau măsura o anumită variabilă trebuie să prezinte anumite criterii de calitate pentru ca rezultatele să fie demne de încredere și interpretabile. Acest lucru este adesea luat ca de la sine înțeles. De exemplu, atunci când măsurăm temperatura a ceva, pur și simplu o facem și folosim rezultatul fără a mai lua în considerare procesele de asigurare a calității (de exemplu, calibrarea termometrului la un standard de referință și stabilirea incertitudinii sale de măsurare) care ne asigură că putem face acest lucru. În măsurarea entităților fizice și chimice, acest tip de control al calității este încorporat în sistemul de dezvoltare a instrumentelor și, în general, nu este de interes pentru utilizatorii finali, cum ar fi HCPs. În timp ce criteriile de calitate se aplică și scalelor de evaluare și altor metode și instrumente de evaluare, acestea nu pot fi considerate de la sine înțelese în același mod, deoarece procesele de control al calității precum cele menționate mai sus pentru măsurile fizice nu sunt disponibile. În plus, scalele de evaluare sunt asociate cu unele complexități, deoarece de obicei sunt destinate să vizeze așa-numitele variabile latente care nu sunt direct observabile (de exemplu, severitatea simptomelor, calitatea vieții, sprijin, dizabilitate).

Scalele de evaluare se bazează pe manifestări observabile ale variabilei pe care intenționează să o țintească sau să o reprezinte. Aceste manifestări sunt exprimate ca itemi (întrebări, afirmații, comportament observat sau performanță etc.) care alcătuiesc instrumentul, iar variațiile răspunsurilor la itemi sunt presupuse că reflectă variații ale variabilei pe care scala de evaluare intenționează să o reprezinte (Hagell 2019). Practic, utilizarea unei scale de evaluare presupune prezentarea itemilor individului (fie pentru auto-raportare, fie prin observare), căruia i se cere să afirme una dintre cele două sau mai multe categorii de răspuns ordonate care descriu fiecare un anumit nivel. Fiecărei categorii de răspuns descriptiv (de exemplu, niciunul – ușoară – moderată – severă) i se atribuie un număr integral (de exemplu, 0 – 1 – 2 – 3) ca mijloc de împărțire a continuumului latent subiacent în diferite cantități. Răspunsurile la itemi sunt apoi însumate în mod obișnuit într-un scor total menit să situeze respondentul pe un continuum de la mai puțin la mai mult. Aceste scoruri totale brute sunt apoi folosite pentru a descrie nivelul de indivizi sau grupuri de oameni, pentru a face comparații și pentru a evalua schimbările.



Cu toate acestea, după cum sa discutat mai sus, scorurile obținute din scalele de evaluare reprezintă evaluări care sunt calitative și ordinale prin natură, mai degrabă decât măsuri. Printre altele, acest lucru înseamnă că o anumită diferență în scorurile totale brute poate reprezenta lucruri neînțelese, în funcție de locul în care se observă de-a lungul intervalului posibil de scoruri. De exemplu, diferența dintre un scor de 1 și 2 nu poate fi mai mică decât cea dintre scorurile de 5 și 8, dacă acele scoruri ar fi fost transformate în măsurare liniară. Acest lucru se datorează faptului că scorurile totale ale scalei de evaluare brute sunt ordinale, adică sunt ordonate prin rang, dar intervalele sau distanțele dintre ele sunt inegale (Hobart & Cano 2009). Relația generală dintre scorurile totale ale scalei de rating ordinale și măsurile liniare pe care le reprezintă este ilustrată în Figura 2.



**Fig. 2** Relația dintre scorurile ordinale și măsurile liniare

Transformarea scorurilor totale ale scalei de evaluare brute ordinale în măsuri liniare este posibilă cu condiția ca scala de evaluare să fie de o calitate suficientă și să îndeplinească criteriile de măsurare. Tehnica pentru a realiza acest lucru este denumită Teoria Măsurării Rasch (RMT) și se bazează pe un model de măsurare dezvoltat de Georg Rasch (Rasch 1960). RMT a fost aplicat în științele sănătății de mai bine de 30 de ani (Wright & Linacre 1989) și utilizarea sa a crescut considerabil în ultimele două decenii. Cu toate acestea, aplicarea sa a fost în primul rând testarea calității scalelor de evaluare pentru a vedea dacă acestea permit măsurarea liniară, în timp ce utilizarea sa pentru a obține măsurători liniare a fost mai puțin obișnuită. Indiferent, dacă s-a constatat că o scală de evaluare funcționează conform intenției și este de bună calitate atunci când este evaluată conform RMT, acesta este, în general, considerat un indicator bun că scala de rating este de bună calitate. Cu toate acestea, chiar dacă o scală îndeplinește criteriile riguroase RMT, scorurile sale totale brute sunt încă de natură ordinală. Nu vom lua în considerare alte detalii cu privire la RMT și utilizarea sa, dar ne vom referi la literatura disponibilă pe acest subiect (a se vedea, de exemplu, sAndrich & Marais 2019; Cano & Hobart 2011; Hagell 2019; Hobart & Cano 2009; Hobart 2003; Tennant et al. 2004; Wright & Linacre 1989; Wright & Stone 1979; Wright & Masters 1982).

Atunci când se apreciază calitatea scalelor de evaluare, există anumite criterii care de obicei sunt luate în considerare. În esență, acestea sunt aceleași cu cele folosite pentru măsurile fizice tradiționale (Pendrill 2019), dar diferă în ceea ce privește rigoarea prin care sunt testate,

iar terminologiile și tradițiile științifice diferă. De exemplu, domeniul asigurării calității măsurilor fizice este denumit metrologie (Howarth & Redgrave 2008), în timp ce cel al scalelor de evaluare este denumit psihometrie (Hobart & Cano 2009). Deși RMT reprezintă un salt înainte substanțial atunci când vine vorba de măsurarea pe scară de rating asigurată de calitate, metodele tradiționale (deseori denumite Teoria clasică a testelor, CTT) sunt încă aplicate mai frecvent (Andrich & Marais 2019; Cano & Hobart 2011; Hagell 2019; Hobart & Cano 2009; Hobart 2003). Cu toate acestea, indiferent de metodele utilizate, există două aspecte principale de calitate ale scalelor de evaluare, denumite în mod obișnuit proprietăți psihometrice: validitate și fiabilitate (Alrubaiby et al. 2014; Hobart & Cano 2009; Hobart et al. 1996; Hobart 2003; Streiner). & Norman 2014; FDA 2009).

Dacă scorurile scalei de evaluare cu încredere pot fi interpretate ca fiind reprezentări bune ale variabilei pe care intenționează să o reprezinte, este o chestiune de validitate. Pentru ca o scală de evaluare să fie validă, aceasta trebuie să se bazeze pe o definiție clară a variabilei pe care intenționează să o vizeze, iar itemii săi trebuie să reprezinte diferite aspecte ale aceleiași variabile. În consecință, și similar ca în cazul măsurării fizice (Kuhn 1961), munca teoretică și calitativă aprofundată este de obicei o condiție prealabilă pentru dezvoltarea cu succes a scalelor de evaluare de înaltă calitate și, de asemenea, un aspect important atunci când se judecă relevanța și validitatea unei anumite scale (Hobart 2003; FDA 2009). Fiabilitatea, pe de altă parte, nu este preocupată de ceea ce reprezintă scorurile, ci de precizia scorurilor obținute. Ca atare, este legat de conceptul de incertitudine în măsurare care este utilizat în metrologie (Howarth & Redgrave 2008). Prin urmare, se poate gândi prin extensie, că scorurile observate sunt lipsite de erori. În consecință, cu cât fiabilitatea este mai bună, cu atât este mai ușor să detectați diferențele și schimbările. Fiabilitatea este exprimată ca un coeficient între 0 și 1 (unde 1 reprezintă fiabilitatea perfectă și nici o eroare); un coeficient de fiabilitate în jur de 0,80 sau mai mult este adesea considerat acceptabil. Uneori, capacitatea scorurilor de a reflecta schimbări (numită receptivitate) este considerată o a treia proprietate psihometrică. Proprietățile psihometrice ale scalelor de evaluare sunt de obicei testate în studiile efectuate cu acest scop specific (Hagell 2019; Hobart & Cano 2009; Hobart et al. 1996; Hobart 2003; Streiner & Norman 2014).

## Analiza numerelor

Statistica, ca și metrologia și psihometria, este o disciplină științifică proprie și una cu o istorie destul de lungă (Stiegler 1986). Ceea ce diferențiază astfel de discipline de majoritatea (dacă nu de toate) celelalte discipline este că ele sunt aplicabile indiferent de domeniul studiat, fie că este vorba de sănătate, agricultură, psihologie, paleontologie sau orice altceva. Scopul și funcția de bază ale statisticilor este de a rezuma, descrie și analiza date numerice, de obicei cu privire la asocieri de un anumit fel, cum ar fi diferențele și corelațiile. Numărul de teste statistice disponibile este vast și poate părea infinit. Cu toate acestea, pentru majoritatea situațiilor va fi suficient un număr relativ limitat de teste și metode. În plus, principiile importante de care trebuie să fii conștient în calitate de consumator de cercetare cantitativă sunt relativ puține. Aici nu vom intra în detalii cu privire la anumite teste statistice, ci vom





# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

trece în revistă câteva dintre principiile comune care sunt relevante pentru a ne da seama. Pentru mai multe detalii, vă rugăm să consultați oricare dintre numeroasele manuale disponibile (de exemplu, Norman & Streiner 2014, Swinscow & Campbell 1997 sau Munro 2005).

Analizele statistice pot fi clasificate ca univariate, bivariate și multivariate. Analizele univariate se referă la analiza unei variabile la un moment dat. În practică, aceasta implică în esență statistici descriptive care sunt utilizate pentru a rezuma datele. Analizele bivariate se referă la analiza a două variabile în același timp, de exemplu, corelația dintre două variabile sau comparațiile de grup în care o variabilă definește grupurile și cealaltă este cea care este comparată (de exemplu, compararea vârstelor între bărbați și femei). În analizele multivariate, sunt luate în considerare mai mult de două variabile. De exemplu, determinarea care dintre mai multe variabile sunt asociate cu un risc crescut de ceva. O variabilă poate fi definită ca o caracteristică care poate varia, de exemplu, sex, vârstă, tensiune arterială, anxietate, tratament, bunăstare. Cu alte cuvinte, variabilele sunt „lucrurile” pe care sunt colectate datele. În analizele datelor, se face adesea o distincție între variabilele dependente și cele independente. Variabila dependentă poate fi considerată ca rezultat sau efect, adică ceva care este influențat de (depinde) diferențe sau modificări ale altor variabile. Variabilele independente, pe de altă parte, sunt variabile care pot influența, explica sau cauza rezultatul sau efectul. De exemplu, în compararea vârstelor între bărbați și femei, vârsta este dependentă, iar sexul este variabila independentă. Într-un studiu clinic al unei terapii menite să îmbunătățească somnul, grupul de tratament (de exemplu, care primește terapia activă sau fără terapie) este variabila independentă, iar somnul este variabila dependentă (rezultatul). În cele din urmă, dacă studiem care dintre factorii vârstă, sexul și indicele de masă corporală (IMC) este cel mai puternic asociat cu bunăstarea percepută, primii trei (vârsta, sexul și IMC) sunt independenți, iar bunăstarea este variabila dependentă. Cu câteva excepții, orice variabilă poate fi fie dependentă, fie independentă, deoarece aceasta este definită de întrebarea de cercetare adresată. Dacă, de exemplu, ultimul exemplu de mai sus a fost reîncadrat ca, unul dintre care factori vârstă, sexul și bunăstarea percepută este cel mai puternic asociat cu IMC, variabila dependentă ar fi IMC, în loc de bunăstare (care ar fi un variabila independenta).

Pentru a complica lucrurile (?), termenii dependent și independent sunt folosiți și pentru a caracteriza tipul de observații (date) dintr-o analiză. Observațiile dependente înseamnă că observațiile sunt cumva legate și au o relație sistematică. Cel mai frecvent tip de observații dependente este atunci când datele sunt colectate de la aceleași persoane în momente diferite, de exemplu, înainte și după o intervenție. Observațiile independente, pe de altă parte, nu au această relație. Poate cel mai simplu mod de a distinge între observațiile dependente și cele independente este de a lua în considerare dacă datele provin de la aceleași persoane sau nu. Dacă o fac, observațiile sunt dependente, iar dacă nu, sunt independente. Uneori, observațiile dependente și independente sunt denumite și pereche și nepereche.

### *Statistici parametrice și neparametrice*



Co-funded by the  
Erasmus+ Programme  
of the European Union

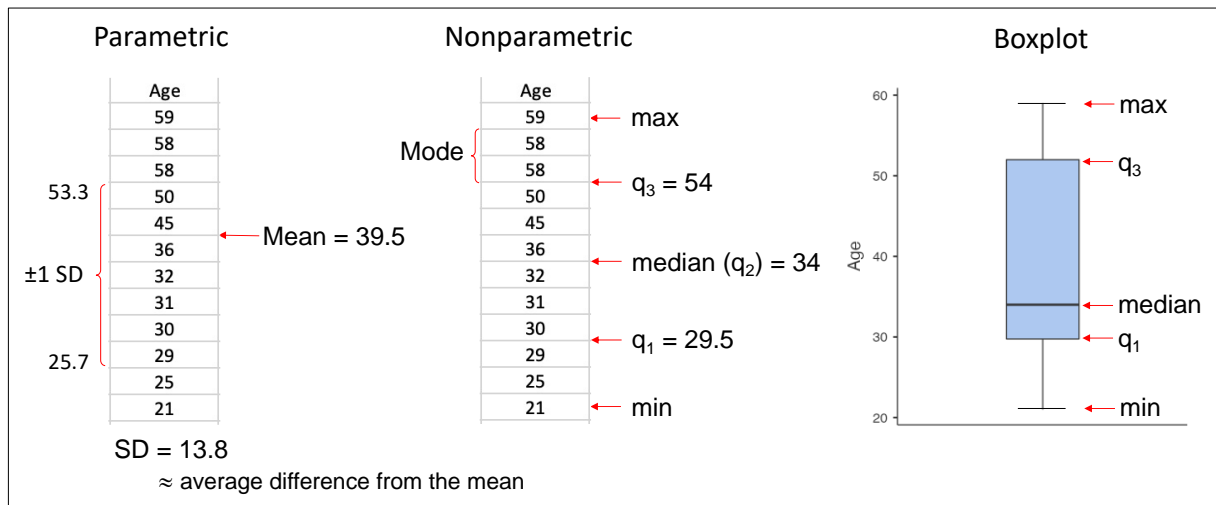
Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.

Pe lângă ce tip de analiză (descriptivă, de comparație, de corelare etc.), alegerea unui test statistic adecvat implică luarea în considerare a caracteristicilor datelor și a situației (Greenhalgh 1997a). Prin situație, înțelegem dacă observațiile sunt dependente sau independente și dacă există două sau mai mult de două grupuri implicate. Caracteristicile datelor care trebuie luate în considerare sunt dacă ele sunt nominale, ordinale sau interval/raport și distribuția lor (de obicei dacă sunt distribuite aproximativ normal sau simetric). Aceasta determină care dintre cele două tipuri principale de teste statistice este adecvată, cele parametrice sau neparametrice. Testele parametrice presupun că datele sunt cantitative (adică, interval sau raport) și provin dintr-o anumită distribuție (de obicei, distribuția normală), în timp ce testele neparametrice nu fac astfel de presupuneri, ci se bazează pe ordinea de șir, frecvențe sau proporții. Pentru a ilustra, vom lua în considerare statisticile descriptive parametrice și neparametrice.

Scopul principal al statisticii descriptive este de a rezuma un set mare de observații într-un mod care oferă o imagine clară și reprezentativă a datelor. Pentru a face acest lucru, există două tipuri principale de statistici descriptive, tendința centrală și dispersia (McCluskey & Lalkhen 2007). Tendința centrală exprimă valoarea „în medie” pentru o variabilă și depinde de caracteristicile datelor și este fie *obișnuită*, fie *mediană* sau *mijlocie*, putând fi cea mai bună alegere. Tendința centrală ar trebui să fie cuplată la măsura de dispersie, care descrie răspândirea datelor în jurul tendinței centrale. Măsurătorile comune ale dispersiei sunt min-max, interval, quartile și abaterea standard (SD). În plus, există și variația. În timp ce variația este folosită în mod obișnuit în calcularea diferitelor teste statistice, este rar (dacă vreodată) folosită pentru a descrie datele. Dintre statisticile descriptive menționate mai sus, media și SD (precum și varianța) sunt parametrice, iar celelalte sunt neparametrice.

Statisticile descriptive menționate mai sus sunt ilustrate în Figura 3 folosind un set de date cu vârstele (în ani) a unui eșantion mic de 12 persoane ( $n=12$ ). Pentru punere în evidență a ilustrării, vârstele au fost aranjate de la cel mai în vârstă la cel mai tânăr. Începând cu opțiunile parametrice, media este calculată ca fiind suma tuturor valorilor (în acest caz 474) împărțită la numărul de observații (adică, 12), ceea ce dă o vârstă medie de 39,5 ani. Mărimea dispersiei de utilizat cu media este SD, care poate fi interpretată ca răspândirea medie sub și peste medie. În acest caz, SD este de 13,8, deci, în medie, vârstele sunt repartizate cu 13,8 ani sub și peste 39,5 ani.



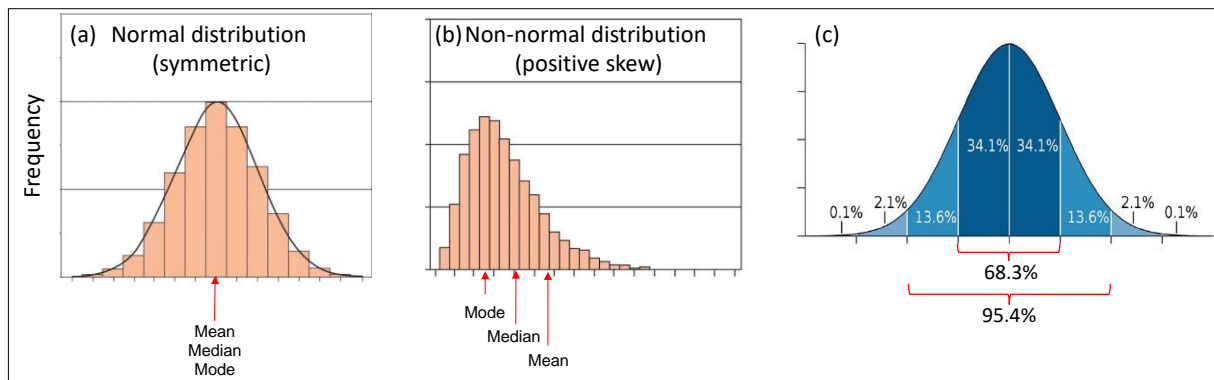
**Fig. 3** Statistici descriptive parametrice și neparametrice

Există două tendințe centrale neparametrice, obișnuită și mediană. *Valoarea obișnuită* este pur și simplu valoarea care apare cel mai frecvent, în acest caz 58. După cum se poate observa (Fig. 3), aceasta nu este o valoare deosebit de reprezentativă și, prin urmare, este rar folosită pentru a descrie tendința centrală a datelor cantitative. Poate fi folosită cu date ordinale, dar cea mai comună utilizare este cu date nominale. Mediana este valoarea din mijloc atunci când datele au fost ordonate. Când numărul de observații este par (ca în Fig. 3), mediana este media celor două puncte de date din mijloc (în acest caz 34 este media lui 32 și 36). Dispersia în jurul mediei este de obicei indicată de prima și a treia valoare a quartilei ( $q_1$  și, respectiv,  $q_3$ ). Acestea se bazează pe aceeași logică ca și mediana, care reprezintă a doua quartilă ( $q_2$ ). Astfel, cele trei quartile sunt valorile care împart observațiile în patru grupuri de dimensiuni egale. Prin urmare,  $q_1$  și  $q_3$  (care sunt folosite pentru a indica dispersia în jurul mediei) definesc intervalul în care se află jumătatea mijlocie a tuturor observațiilor. Acesta este uneori denumit intervalul inter-quartile (IQR), deși, conform definiției, IQR este pur și simplu diferența dintre  $q_3$  și  $q_1$  (în cazul nostru 24,5, adică 54 – 29,5). Uneori, cele trei quartile sunt denumite a 25-a ( $q_1$ ), a 50-a ( $q_2$ ) și a 75-a ( $q_3$ ) percentile. Un alt mod de exprimare a dispersiei este de a da cele mai mici și cele mai mari valori observate, adică min-max. Ca și în cazul quartilelor, acesta este adesea numit interval, deși, conform definiției, intervalul este pur și simplu diferența dintre cele mai mari și cele mai mici valori (în cazul nostru 38, adică 59 – 21). Pe cont propriu, min-max nu este foarte util, deoarece este foarte influențat, conform definiției, de valori aberante extreme și, prin urmare, este posibil să nu ofere informații foarte reprezentative. Cu toate acestea, poate fi util să se dea min-max în plus față de, de exemplu, SD sau  $q_1$ - $q_3$ , pentru a oferi o descriere mai completă a datelor. În extrema dreaptă în Figura 3 este o reprezentare grafică a ceea ce tocmai s-a discutat. Acest tip de grafic se numește boxplot sau box-and-whisker plot. Limitele inferioare și superioare ale casetei din mijlocul graficului reprezintă  $q_1$  și, respectiv,  $q_3$ . Linia orizontală din interiorul cutiei este mediana ( $q_2$ ), iar mustățile reprezintă cele mai mici și mai mari valori observate. Uneori, asteriscurile sunt folosite pentru a indica valori aberante extreme, dincolo de mustăți.

Mediana și quartilele sunt utilizate atunci când datele sunt fie ordinale, fie atunci când datele la nivel de interval sau raport nu sunt distribuite în mod normal. Acest lucru se datorează



faptului că aceste statistici neparametrice nu presupun date cantitative sau o distribuție specială a datelor. Pentru a vedea de ce apare acest caz, vom trece în revistă caracteristicile și logica distribuției normale (McCluskey & Lalkhen 2007). După cum este ilustrat în Figura 4 (panourile a și c), caracteristica importantă a distribuției normale este că este simetrică, adică frecvența (axa y) valorilor observate (axa x) crește constant și atinge vârfuri la mijlocul intervalului de valori, înainte ca acesta să scadă în aceeași rată cu care a crescut. Această curbă este uneori denumită o curbă în formă de clopot, deoarece seamănă cu un clopot vechi de biserică.



**Fig. 4** Distribuția normală (panourile a și c) și o distribuție non-normală cu oblic pozitiv (panoul b). O distribuție non-normală cu oblica negativă ar arăta același model, dar cu coada îndreptată spre stânga în loc de către dreapta.

După cum este ilustrat în panoul a, vârful distribuției normale coincide cu toate cele trei măsuri ale tendinței centrale. Valoarea de pe axa x care reprezintă vârful este cea mai comună valoare (adică obișnuită), împarte observațiile în două grupuri de dimensiuni egale (adică mediana) și este, de asemenea, media tuturor observațiilor. Totul se datorează simetriei distribuției. Panoul c din Figura 4 ilustrează, de asemenea, semnificația SD. Dacă mergem cu 1 SD peste și sub medie, acel interval va acoperi aproximativ 68% din toate observațiile (34,1% de ambele părți ale mediei) și dacă luăm 2 SDs peste și sub medie, intervalul va acoperi aproximativ 95% din toate observațiile (47,7% de ambele părți ale mediei). Cu toate acestea, atunci când distribuția nu este simetrică, aceste caracteristici se destramă, așa cum este ilustrat în panoul b din Figura 4. În primul rând, deoarece media este sensibilă la valori extreme, va fi influențată de asimetria distribuției, ceea ce o face mai puțin reprezentativă pentru date. Mediana, pe de altă parte, este insensibilă la distribuție și va reprezenta în continuare valoarea care împarte observațiile în două grupuri de dimensiuni egale. Prin urmare, mediana este măsura recomandată a tendinței centrale atunci când datele nu sunt conforme cu distribuția normală. În plus, deoarece distribuțiile non-normale sunt asimetrice, semnificația SD se destramă. Cu date sever denaturate, SD poate sugera, de exemplu, că diferența medie față de medie este o valoare care nu poate exista. Din nou, acest lucru va face ca SD să fie nerepresentativ pentru date și, în esență, neinterpretabil, în timp ce quartilele vor rămâne adecvate.

Examinarea de mai sus a statisticilor descriptive parametrice și neparametrice este, în termeni generali, reprezentativă și pentru alte statistici parametrice și neparametrice utilizate pentru,





de exemplu, comparații de grup, corelații și așa mai departe. Prin urmare, nu le vom detalia, dar vom concluziona că funcționează conform aceluiași principii ca și cele analizate mai sus. Cercetătorii preferă adesea metodele parametrice, deoarece acestea tind să ofere mai multă flexibilitate și informații mai utile. Cu toate acestea, cu puține excepții, există teste neparametrice disponibile ca alternativă la omologii lor parametrici, astfel încât nu este nevoie reală de o utilizare inadecvată a statisticilor parametrice, în special fără riscul de a obține rezultate false și nereprezentative, deoarece ipotezele nu sunt îndeplinite. Cu toate acestea, a judeca dacă a fost făcută o alegere corectă poate fi o sarcină dificilă, deoarece poate să nu fie evident dacă datele raportate într-un studiu clinic sunt compatibile cu distribuția normală sau nu. Dar există câteva semne care ar putea ajuta cititorul. În primul rând, dacă autorii, indiferent de motiv, prezintă atât media, cât și mediana, acest lucru poate oferi un indiciu, deoarece dacă aceste două valori diferă considerabil, acesta este un semn clar că datele nu urmează o distribuție normală. În al doilea rând, dacă valoarea SD este mare în raport cu medie (să zicem mai mare de jumătate din medie), acesta este, de asemenea, un semn că datele probabil nu sunt distribuite în mod normal (reamintim sensul SD așa cum este ilustrat în panoul c din Figura 4) . În cele din urmă, și indiferent de distribuție, statisticile parametrice precum media și SD nu sunt adecvate pentru datele calitative (adică, nominale și ordinale) (Greenhalgh 1997a; Svensson 2001).

Tabelul 2 enumeră diverse teste și metode statistice parametrice și neparametrice comune în funcție de tipul de analiză pe care o reprezintă, precum și de caracteristicile (nivelul și distribuția) datelor și situației.

**Tabelul 2.** Metode statistice și utilizarea lor în funcție de nivelurile de date, distribuție și situație <sup>a</sup>

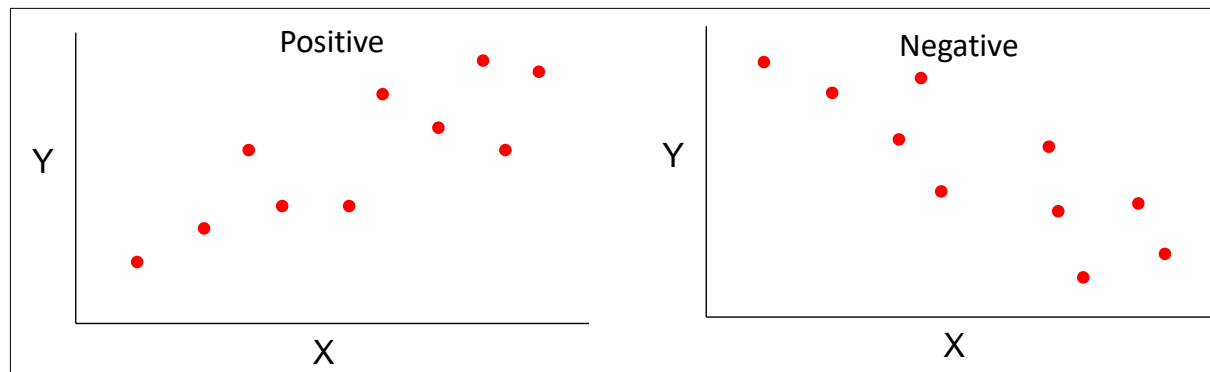
Data level / Distribution	Description	Correlation	Group comparison			
			Independent observations		Dependent observations	
			2 groups	>2 groups	2 time points	>2 time points
<ul style="list-style-type: none"> <li>Nominal</li> </ul>	<ul style="list-style-type: none"> <li>Frequencies, %</li> <li>Mode</li> </ul>	<ul style="list-style-type: none"> <li>Phi coefficient</li> <li>Cramér's V</li> </ul>	<ul style="list-style-type: none"> <li>Chi<sup>2</sup> test</li> <li>Fisher's exact test</li> </ul>		<ul style="list-style-type: none"> <li>McNemar's test</li> </ul>	
<ul style="list-style-type: none"> <li>Ordinal</li> <li>Interval / non-normal</li> <li>Ratio / non-normal</li> </ul>	<ul style="list-style-type: none"> <li>Median (md)</li> <li>Quartiles (q1-q3)</li> <li>Min – Max</li> <li>Range</li> </ul>	<ul style="list-style-type: none"> <li>Spearman's rho (<math>r_s</math>)</li> <li>Kendall's tau</li> <li>Kendall's W</li> </ul>	<ul style="list-style-type: none"> <li>Mann-Whitney U-test</li> </ul>	<ul style="list-style-type: none"> <li>Kruskal-Wallis' H-test</li> </ul>	<ul style="list-style-type: none"> <li>Wilcoxon's signed-rank test</li> </ul>	<ul style="list-style-type: none"> <li>Friedman's test</li> <li>Skilling-Mack test</li> </ul>
<ul style="list-style-type: none"> <li>Interval / normal</li> <li>Ratio / normal</li> </ul>	<ul style="list-style-type: none"> <li>Mean</li> <li>Standard deviation (SD)</li> </ul>	<ul style="list-style-type: none"> <li>Pearson's <math>r</math></li> <li>Intra-class correlation (ICC)</li> </ul>	<ul style="list-style-type: none"> <li>Independent samples t-test</li> <li>Analysis of covariance (ANCOVA)</li> </ul>	<ul style="list-style-type: none"> <li>Analysis of variance (ANOVA)</li> <li>ANCOVA</li> </ul>	<ul style="list-style-type: none"> <li>Paired t-test</li> <li>ANCOVA</li> </ul>	<ul style="list-style-type: none"> <li>Repeated-measures ANOVA</li> <li>ANCOVA</li> </ul>

<sup>a</sup> Zona umbrită reprezintă statistici parametrice

## Corelație și regresie

Împreună cu comparațiile de grup, analizele privind asocierea dintre variabile sunt printre cele mai frecvente în științele sănătății. Deși comparațiile de grup se adresează și asocierilor (adică între grupuri de oameni și variabila dependentă), tipurile mai specifice de analize pentru a studia asociațiile sunt corelația și regresia. În timp ce corelația și regresia au asemănări, ele și diferă una de cealaltă. Corelația se referă la relația liniară dintre două variabile. Acest lucru poate fi abordat grafic prin intermediul diagramelor de dispersie, unde valorile fiecărei persoane pe două variabile (să le numim X și Y) sunt reprezentate grafic și

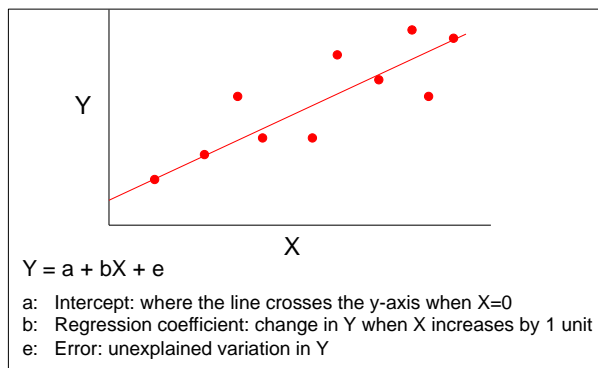
forma diagraamelor este inspectată în ceea ce privește cât de aproape sunt diagraamele de o linie dreaptă imaginară și dacă asocierea tinde să fie pozitivă (valorile pe Y tind să crească pe măsură ce valorile pe X cresc) sau negativă (Figura 5).



**Fig. 5** Graficul de dispersie ipotetic al relațiilor liniare pozitive și negative dintre variabilele X și Y.

Corelația este un mijloc de cuantificare a tipului de relații ilustrate în Figura 5. Această cuantificare se face printr-un coeficient de corelație, care poate lua valori între -1 și 1 și exprimă măsura în care punctele urmează o linie dreaptă. Valorile negative reprezintă o relație negativă, iar valorile pozitive reprezintă o relație pozitivă; cu cât coeficientul este mai aproape de -1 sau 1, cu atât relația liniară este mai puternică. Este important să rețineți că coeficientul exprimă liniaritatea (adică proximitatea punctelor de o linie dreaptă imaginară) și nimic altceva. Aceasta înseamnă că anumite modele de asocieri, cum ar fi relațiile în formă de U, nu sunt reflectate de coeficientul de corelație. Prin urmare, este înțelept să nu ne bazăm doar pe corelație, ci și să reprezentăm datele, deoarece acest lucru poate dezvălui modele importante care nu sunt reflectate de coeficient. Cel mai comun coeficient de corelație este coeficientul de corelație produs-moment al lui Pearson ( $r$ ). Cu toate acestea, aceasta este o metodă parametrică care, printre altele, presupune că ambele variabile sunt cantitative (interval/raport) și distribuite normal. Atunci când nu este acesta cazul, trebuie utilizată în schimb corelația neparametrică, cum ar fi corelația de rang a lui Spearman ( $\rho$ ).

După cum s-a văzut mai devreme, corelația nu spune dacă asocierea este cauzală. Odată cu regresia, suntem într-o poziție oarecum mai bună pentru a aborda cauzalitatea, deși măsura în care pot fi făcute inferențe cauzale se bazează și pe aspecte legate de proiect și alte aspecte (Greenhalgh 1997b; Guyatt et al. 1995). Motivul este că regresia nu se referă doar la relația dintre variabile, ci permite ca variația unei variabile (variabila dependentă) să fie explicată sau prezisă de la alta (variabila independentă). Există multe tipuri de analize de regresie disponibile și alegerea depinde, în mare măsură, de tipul de variabilă dependentă a cărei valoare se dorește să prezică. Exemplele includ regresia logistică (variabila dependentă este dihotomică, de exemplu, bolnavă sau sănătoasă), regresia ordinală (variabila dependentă este ordinală), regresia multinomială (variabila dependentă este nominală cu mai mult de două categorii) și *regresia poisson* (variabila dependentă este un număr de evenimente). Cu toate acestea, logica de bază este aceeași indiferent de tipul de regresie și va fi ilustrată aici prin intermediul regresiei liniare, adică atunci când variabila dependentă este cantitativă. Acest lucru este ilustrat în Figura 6.



**Fig. 6** Regresia liniară pentru a prezice dependentă variabila  $Y$  din variabila independentă  $X$ .

Regresia identifică linia dreaptă care urmează observațiile (punctele) cât mai aproape posibil, care este exprimată ca o ecuație care spune valoarea estimată a variabilei dependente  $Y$  pe baza variabilei independente  $X$ :  $Y = a + bX$ . În această ecuație, **a** reprezintă interceptul (adică valoarea în care linia dreaptă traversează axa  $y$ , cu alte cuvinte valoarea lui  $Y$  când valoarea lui  $X$  este 0), iar **b** este coeficientul de regresie (sau panta). Deoarece  $b$  este înmulțit cu  $X$ , aceasta înseamnă că  $b$  reprezintă modificarea lui  $Y$  atunci când  $X$  crește cu o unitate. Deoarece variabila independentă nu explică niciodată toată variația în  $Y$ , există și un termen de eroare ( $e$ ) în ecuație. Pe lângă informațiile menționate mai sus, o analiză de regresie va oferi și informații despre proporția de variație în  $Y$  care poate fi explicată prin  $X$ . Această valoare se numește valoarea  $R^2$ .

Cel mai puternic lucru al regresiei constă în faptul că nu ne limităm să luăm în considerare o singură variabilă independentă, ci suntem capabili să studiem simultan influența mai multor variabile asupra variației lui  $Y$ . Aceasta se numește regresie multiplă și reprezintă unul dintre cele mai comune tipuri a metodelor statistice multivariate. În regresia multiplă, vom obține același tip de informații ca și cea analizată mai sus, dar pentru fiecare dintre variabilele independente (explicative) ținând cont și de influența celorlalte variabile independente în analiză. De exemplu, dacă suntem interesați să explicăm variația greutatei corporale din activitatea fizică, putem include și, de exemplu, vârsta și sexul în analiza noastră, deoarece există motive să credem că acestea sunt, de asemenea, asociate cu greutatea. Rezultatele ne vor spune apoi în ce măsură putem explica variațiile de greutate din activitatea fizică, controlând în același timp influența vârstei și sexului.

Alte metode statistice multivariate comune în științele sănătății includ (dar nu se limitează la) analiza factorială (identificarea variabilelor care au ceva în comun), analiza traseului (în esență o variație a regresiei în care este posibil să se țină seama și de asocieri indirecte), și analiza covarianței (ANCOVA; comparație de grup care controlează sau ia în considerare variabile suplimentare; de exemplu, compararea efectului unei intervenții între grupurile de studiu, luând în considerare diferențele dintre grupuri la începutul studiului).



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

## Testarea ipotezei și valoarea P

Majoritatea analizelor statistice se referă la testarea ipotezelor. De exemplu, grupurile diferă? Există o asociere? Există un risc crescut? Și așa mai departe. În mod formal, ceea ce este testat este ceva numit ipoteza nulă, de obicei notă  $H_0$ . Dacă luăm în considerare o analiză a unei diferențe de grup potențial, ipoteza nulă spune că nu există nici un efect, adică nici o diferență între grupuri. Teoretic, se consideră că ipoteza nulă este valabilă, cu excepția cazului în care există dovezi suficient de puternice împotriva acesteia. În acest caz, ipoteza nulă este respinsă în favoarea ipotezei alternative (numită  $H_1$ ). Ipoteza alternativă este opusul ipotezei nule și înseamnă că există un efect, în acest caz o diferență între grupuri. Această logică nu se aplică doar comparațiilor de grup, ci și practic la toate tipurile de analize statistice. De exemplu, în corelație ipoteza nulă spune că nu există asociere între variabile, în regresie ipoteza nulă spune că Y nu poate fi explicat prin X, în evaluarea riscului ipoteza nulă spune că nu există risc crescut și așa mai departe. Termenul general folosit indiferent de situație este „efect”, care astfel se poate referi la diferențe, asocieri, putere explicativă, risc sau orice altceva este studiat.

Deci, cum determinăm când există dovezi suficient de puternice împotriva ipotezei nule? Acest lucru se realizează prin evaluarea probabilității ca efectul observat (sau un efect mai mare) să fi avut loc dacă ipoteza nulă este adevărată. Această probabilitate este denumită valoarea P. Ca toate probabilitățile, valoarea P poate varia în principiu între 0 și 1, deși în practică nu este niciodată 0 sau 1, deoarece există întotdeauna o mică șansă ca aceasta să fie greșită. Regula generală de la aproximativ 100 de ani este că ipoteza nulă este respinsă dacă valoarea P este  $<0,05$ , adică dacă există mai puțin de 5% probabilitate de a vedea un efect (de exemplu, diferență) de dimensiunea observată, sau mai mare, dacă ipoteza nulă este adevărată. Când valorile P sunt  $<0,05$ , acesta este de obicei denumit un rezultat semnificativ statistic (Greenhalgh 1997 b). În analizele cu mize mari, cum ar fi intervențiile în care rezultatul se referă dacă oamenii vor supraviețui sau nu, acest prag este de obicei stabilit la un nivel mai scăzut (de exemplu,  $P<0,001$ ) pentru a fi cât mai sigur posibil. De asemenea, în studiile în care sunt efectuate multe teste de ipoteză, este obișnuit (și recomandabil) să se ajusteze valorile P rezultate în jos, deoarece influența șanselor crește pentru fiecare test care este efectuat. O astfel de ajustare comună este ceva numit corecție Bonferroni (Bland & Altman 1995; Greenhalgh 1997 b). Cu toate acestea, se recomandă ca valorile P exacte (nu numai dacă sunt deasupra sau sub, de exemplu, 0,05) să fie raportate și interpretarea lor să nu fie una binară. De exemplu, diferența practică dintre valorile P de 0,052 și 0,049 este neglijabilă (Whitley & Ball 2002a; Cohen 1994).

Este deosebit de important să fim conștienți de faptul că valoarea P, oricât de mică, se referă doar la semnificația statistică. Cu alte cuvinte, nu spune nimic despre dimensiunea efectului observat sau dacă acesta are vreo semnificație clinică sau nu (Matthews & Altman 1996; Altman & Bland 1995). Acest lucru se datorează faptului că valoarea P depinde nu numai de efect (de exemplu, dimensiunea unei diferențe), ci și de dimensiunea eșantionului, precum și de dispersie. Pe măsură ce dimensiunea eșantionului crește și dispersia scade, valoarea P va scădea. De exemplu, cu 9 persoane în fiecare dintre cele două grupuri cu un SD de 15, o





# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

diferență medie de aproximativ 10 va fi semnificativă ( $P < 0,05$ ), în timp ce același nivel de semnificație statistică va fi observat cu o diferență medie de aproximativ 6 dacă sunt 25 de persoane în fiecare grup și cu o diferență medie de aproximativ 1 dacă există 900 de persoane în fiecare grup (Norman & Streiner 2014). Cu alte cuvinte, eșantioanele foarte mari sunt mai mult sau mai puțin legate de a da rezultate semnificative statistice, chiar dacă în practică pot fi banale. Pentru a obține o mai bună înțelegere a semnificației rezultatelor, este obișnuit să se prezinte nu numai valorile  $P$  și statisticile descriptive asociate, ci și ceva numit intervale de încredere. Un interval de încredere ne spune în ce interval va fi probabil efectul adevărat. Cel mai obișnuit este raportarea intervalelor de încredere de 95%, care reprezintă intervalul de valori în care putem avea încredere 95% că efectul este adevărat (de exemplu, diferența de grup) (Greenhalgh 1997b; Sheldon 2000). Dacă acest interval include sau este aproape de ceea ce poate fi considerat un efect banal sau inexistent, probabil că există motive întemeiate de a pune sub semnul întrebării relevanța practică a constatărilor. Trebuie remarcat faptul că, la fel ca valorile  $P$ , intervalele de încredere sunt, de asemenea, influențate de dimensiunea și dispersia eșantionului.

## *Câți oameni?*

Problema dimensiunii eșantionului a fost abordată mai devreme în acest raport și va fi acum explorată puțin mai detaliat. După cum sa menționat în secțiunea anterioară, un aspect important în planificarea unui studiu care implică comparații de grup (de exemplu, un studiu pentru a evalua eficacitatea unei intervenții) este acela de a vă asigura că grupurile de studiu sunt suficient de mari pentru a detecta un astfel de efect statistic. Pe de altă parte, dacă grupurile sunt prea mari, analizele pot indica efecte semnificative din punct de vedere statistic, care de fapt sunt triviale. Prin urmare, este de dorit să existe mostre suficient de mari, dar nu prea mari. Pe lângă motivele tehnice statistice, un alt motiv important pentru a avea o dimensiune a eșantionului bine echilibrată este unul etic. Ar fi de o legitimitate etică îndoielnică să se efectueze un studiu cu un eșantion atât de mic încât este foarte improbabil să se detecteze un efect chiar dacă acesta există. Și pe de altă parte, ar fi, de asemenea, discutabil dacă este o bună practică etică să expunem mult mai mulți oameni decât este necesar la povara suplimentară care înseamnă participarea la un studiu clinic.

Modul în care cercetătorii abordează problema obținerii unor dimensiuni bine echilibrate ale eșantionului este prin realizarea unei analize de putere, care este utilizată pentru a determina câți participanți sunt necesari pentru a detecta un anumit efect (Whitley & Ball 2002b; Chan 2003). Puterea se referă la probabilitatea de a detecta statistic un efect. Investigatorii tind de obicei să obțină cel puțin 80% și adesea 90% putere. Deși procedura exactă diferă în funcție de tipul de analiză planificat, ingredientele de bază necesare sunt mărimea efectului de interes (de exemplu, diferența care este relevantă clinic), la ce valoare  $P$  urmează să fie detectat efectul (de exemplu,  $< 0,05$ ), dispersia așteptată a măsurării rezultatului (de exemplu, SD) și puterea dorită (de exemplu, 80%). Odată ce aceste valori au fost stabilite, se poate calcula numărul de persoane necesare în fiecare grup. Partea complicată a procedurii apare când se decide efectul sau diferența minimă care este de importanță clinică și dispersia



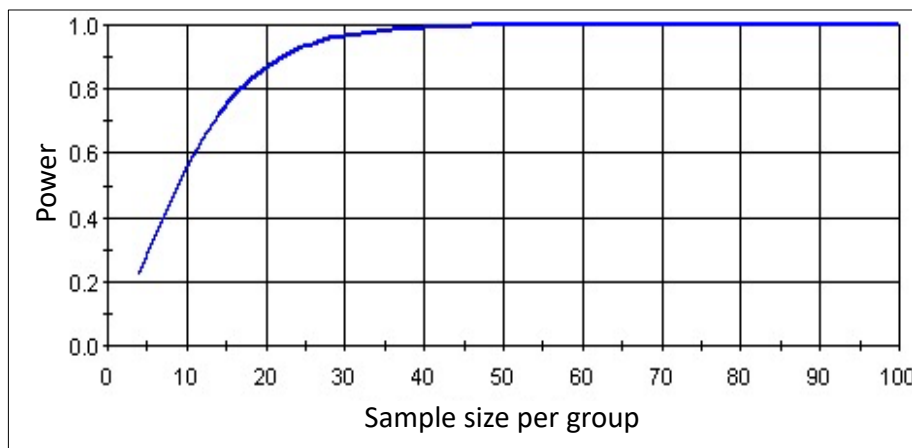
Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



așteptată. Cu unele variabile de rezultat bine stabilite, cum ar fi tensiunea arterială, poate fi relativ ușor să decideți dimensiunea minimă a efectului, în timp ce poate fi mai dificil cu măsurători de rezultat bazate pe scară de evaluare. Pentru unele astfel de scale, studiile în care s-a estimat o diferență minim importantă în scoruri pot fi disponibile, dar adesea va fi o chestiune de judecată bazată pe experiența clinică. Dispersia așteptată a măsurării rezultatului este, de asemenea, dificil de știut înainte de a fi colectate date și, de obicei, aceasta se bazează pe presupuneri învățate din studiile anterioare. Cu toate acestea, orice considerații și ipoteze care au fost făcute la efectuarea analizei de putere ar trebui rezumate atunci când rezultatele studiului sunt în cele din urmă raportate. În general, există un efect de platou în analizele de putere. Adică, se câștigă multă putere atunci când crește o dimensiune relativ mică a eșantionului, în timp ce se câștigă mai puțin odată ce un anumit număr de oameni sunt în studiu. Acest lucru este ilustrat în Figura 7.



**Fig. 7** Exemplu de diagramă care ilustrează relația dintre putere și dimensiunea eșantionului necesară per grup pentru a detecta o diferență medie de 10 la  $P < 0.05$  cu un SD de 10 în fiecare grup.

Odată ce a fost efectuată o analiză de putere, trebuie făcută o estimare a ratei așteptate de abandon, deoarece de obicei nu se poate aștepta ca toți participanții recrutați să finalizeze studiul. Dacă, de exemplu, s-a estimat că sunt necesare 25 de persoane în fiecare grup pentru a avea o putere de 90% (Figura 7) și se dovedește că 10 persoane din orice motiv nu finalizează studiul, puterea reală la final va fi de aproximativ 70%. Pentru a compensa acest lucru, majoritatea studiilor recrutează mai mulți participanți decât s-a estimat în analiza puterii.

Estimarea cerințelor privind dimensiunea eșantionului prin intermediul analizei puterii se face în primul rând în studii intervenționale și alte studii care se preocupă în primul rând de comparațiile de grup. În studiile mai explorative, folosind, de exemplu, regresia sau alte metode statistice multivariate, principala problemă nu este de obicei atât de preocupată de a avea suficienți participanți pentru a detecta un anumit efect, ci mai degrabă de a avea un eșantion suficient de mare pentru a obține rezultate interpretabile și stabile. În timp ce această cerință este complexă și diferă între metode, ca regulă generală, statisticienii tind să recomande un minim de 50-100 + cel puțin 10 sau în jur de, per variabilă sau parametru estimat (Norman & Streiner 2014). De exemplu, dacă intenționăm să facem analize de



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

regresie multiplă cu 10 variabile independente, va trebui să avem date complete de la cel puțin 50-100 + 100 (10 variabile x 10 observații), adică 150-200 de persoane.

## Punând totul împreună

Principiile revizuite mai sus vin cu o serie de avertismente pentru practicantul HCP. De exemplu, nu este suficient ca un studiu să fie conceput ca un RCT, dar trebuie luată în considerare și validitatea sa internă și externă. Aceasta include, dar nu se limitează la întrebări precum

- Eșantionul de studiu a fost reprezentativ pentru pacienții întâlniți în practica clinică zilnică?
- Este variabila primară de rezultat al studiului relevantă din punct de vedere clinic?
- Rezultatul studiului primar a fost evaluat sau măsurat folosind un instrument de încredere și valid?
- Metodele utilizate pentru analiza datelor au fost adecvate pentru a îndeplini obiectivele studiului și pentru datele disponibile în studiu?
- Sunt rezultatele studiului semnificative din punct de vedere clinic și interpretabile?

În timp ce astfel de probleme ar fi trebuit abordate în timpul procesului de evaluare intercollegială în cazul în care se renunță la publicarea lucrărilor de cercetare, acesta nefiind neapărat cazul și, chiar dacă este, pot rămâne publicate atunci când se critică un studiu în funcție de propria practică clinică și de context. Prin urmare, este important ca HCP și alți utilizatori finali ai studiilor clinice să fie conștienți de aceste provocări. Aceasta nu înseamnă că fiecare trebuie să fie un expert în metodologia cercetării; de multe ori este suficient să fii conștient de probleme și să aplici această conștientizare cu bun simț și experiență. În secțiunea următoare vom exemplifica problemele menționate mai sus, analizând un exemplu din lumea reală.

### *Un exemplu din lumea reală*

Pentru a ilustra punctele subliniate mai sus, vor fi analizate aspectele cheie ale unui RCT dublu orb, controlat cu placebo. RCT a implicat persoane cu boala Parkinson (PD)<sup>1</sup> care suferă de wear-off, adică episoade cu eficacitate redusă a levodopei și simptome crescute de PD înainte de apariția următoarei doze. RCT a durat 13 săptămâni și s-a referit la un medicament care să fie utilizat împreună cu levodopa pentru a reduce problemele legate de uzură (wear-off). Scopul studiului a fost de a determina efectul terapiei combinate asupra activităților din viața

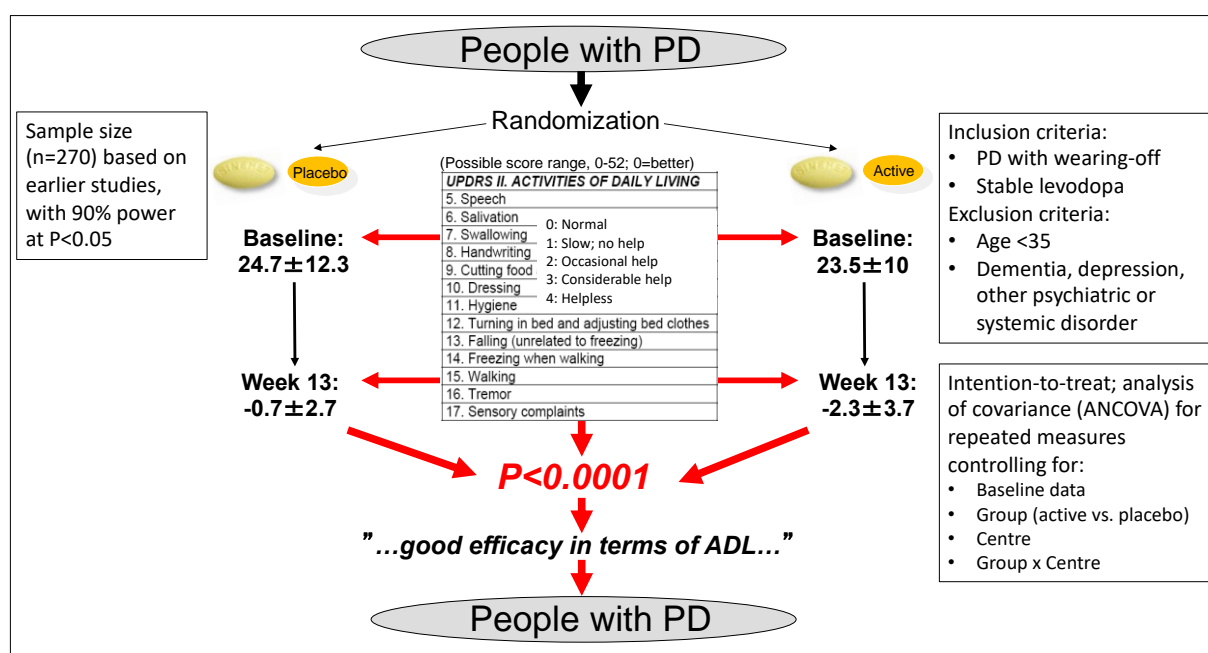
---

<sup>1</sup> PD este o tulburare neurodegenerativă progresivă frecventă asociată cu simptome motorii de bradikinezie (încetinirea mișcării), rigiditate musculară, tremor și tulburări posturale, precum și caracteristici non-motorii cum ar fi depresia, anxietatea, tulburările de somn, oboseala, tulburările cognitive, disautonomia și durerea. Se crede că principalele simptome motorii se datorează pierderii de dopamină din creier, iar tratamentul medicamentos simptomatic urmărește restabilirea nivelurilor de dopamină din creier, unde levodopa (un precursor al dopaminei) este compusul principal.



de zi cu zi (ADL), care a fost astfel rezultatul principal al studiului<sup>2</sup>. Măsura principală a rezultatului din studiu a fost partea a II-a a Unified PD Rating Scale (UPDRS; Fahn et al. 1987), care este o scară de aur de evaluare clinică standard în PD și constă din patru părți principale: Proces de gândire, comportament și dispoziție (partea I), ADL (partea II), Simptome motorii (partea III) și Complicațiile terapiei (partea IV).

Studiul și principalele sale caracteristici de proiectare și rezultatele sunt ilustrate în Figura 8. Pe baza unui calcul de putere pentru a detecta o diferență între tratamente cu o putere de 90% la  $P < 0,05$ , pe baza rezultatelor din studiile anterioare care utilizează partea II a UPDRS (UPDRS II) studiul a recrutat 270 de participanți, dintre care 174 au fost randomizați pentru a primi medicamentul de studiu activ și 96 au fost randomizați pentru a primi pastile placebo inactive (grupul de control). Eșantionul a fost evaluat la momentul inițial și 13 săptămâni mai târziu, iar modificarea din cadrul fiecărui grup a fost comparată la încheierea studiului conform principiului intenției de tratare<sup>3</sup>. Datele au fost analizate utilizând analiza covarianței (ANCOVA), controlând datele inițiale, grupul (activ vs. placebo), centrul de studiu și interacțiunea grup x centru de studiu. Diferența de scoruri UPDRS II între loturile de studiu a fost semnificativă ( $P < 0,0001$ ) în favoarea medicamentului activ de studiu și s-a ajuns la concluzia că terapia adjuvantă are o eficacitate bună în ceea ce privește ADL, ceea ce este astfel concluzia care trebuie generalizată la marea majoritate a populației de persoane cu PD din afara studiului.



**Fig. 8** Un RCT dublu-orb, controlat cu placebo, după punerea pe piață (fază IV) din Europa de Nord, a unui medicament adjuvant (adăugarea levodopa) pentru boala Parkinson (PD) cu privire la efectele acestuia asupra activităților vieții zilnice.

<sup>2</sup> Nu sunt furnizate referiri la studiu sau denumirea medicamentului testat, deoarece acest lucru este irelevant pentru scopul acestei ilustrații.

<sup>3</sup> O strategie de analiză a datelor în studiile intervenționale, în care participanții sunt analizați în funcție de grupul cărora le-a fost repartizat la începutul studiului, indiferent dacă au continuat sau nu acel tratament.





# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

## *A fost eșantionul de studiu reprezentativ pentru pacienții întâlniți în practica clinică zilnică?*

Studiul a fost realizat în Europa de Nord. La identificarea participanților la studiu, au fost excluși cei care aveau vârsta sub 35 de ani, sufereau de demență, depresie sau alte tulburări psihiatrice, sistemice sau metabolice. Aceasta înseamnă că populația studiată (adică tipul de persoane la care se intenționează să se generalizeze rezultatele) nu include persoane cu, de exemplu, comorbidități mentale, care sunt frecvente în PD. Deși nu a fost raportat, se poate presupune că majoritatea participanților la studiu erau caucazieni de moștenire nord-europeană. O implicație a acestui fapt este că este incert în ce măsură rezultatele sunt reprezentative și aplicabile oamenilor din alte zone geografice și culturale. Acest lucru poate fi relevant în ceea ce privește, de exemplu, modul în care a fost evaluat rezultatul primar (ADL) în studiu (adică relevanța și semnificația diferitelor sarcini ADL pot diferi în funcție de mediul cultural), precum și din perspectivă biomedicală, deoarece farmacologia poate diferi în funcție de etnie (Johnson 1997; Shas 2015).

## *Este variabila principală a rezultatului studiului relevantă clinic?*

Se poate spune că pare rezonabil să se considere performanța activității în viața de zi cu zi (ADL) ca fiind un rezultat relevant din punct de vedere clinic în studierea valorii unei intervenții pentru persoanele care trăiesc cu o tulburare progresivă pe termen lung, cum ar fi PD. Într-adevăr, se poate argumenta că acest rezultat este mai relevant dintr-o perspectivă clinică și centrată pe persoană decât, de exemplu, efectele asupra simptomelor motorii, deoarece simptomele reduse pot ușura performanța ADL, dar aceasta din urmă este influențată și de alți factori (Organizația Mondială a Sănătății 2001).

## *Rezultatul primar al studiului a fost evaluat sau măsurat folosind un instrument de încredere și valid?*

Variabila de rezultat primară a fost ADL și aceasta a fost evaluată utilizând UPDRS II, care pretinde să reprezinte ADL. Este alcătuit din 13 itemi care sunt evaluați de HCP pe baza informațiilor dintr-un interviu clinic cu pacientul. Fiecare item este punctat de la 0 („normal” sau absența problemelor) la 4 („severă” sau incapacitatea de a îndeplini sarcina), iar scorurile itemilor sunt însumate pentru a oferi un scor total care reprezintă nivelul de limitare ADL al persoanei, de la 0 (fără limitări) la 52 (limitări severe). Deși UPDRS II este utilizat pe scară largă și a fost considerat „recomandat” pentru evaluarea dizabilității în PD (Shulman 2016), acesta a fost, de asemenea, criticat. În timp ce fiabilitatea UPDRS II, în general, a fost considerată acceptabilă, validitatea sa a fost pusă sub semnul întrebării în mod repetat (Forjaz & Martinez-Martin 2006; Hagell 2019; Hariz et al. 2003; Ramaker et al. 2002; van Hilten et al. 1994). Problema principală este că doar șase din cei 13 itemi ai scalei reprezintă performanța activității, în timp ce elementele rămase se referă la deprecieri. Aceasta înseamnă că nu este clar ce reprezintă de fapt scorurile UPDRS II.

Mai mult, la citirea publicării integrale a RCT luate în considerare aici, se dovedește că, în timp ce UPDRS II a fost măsura principală a rezultatului, anchetatorii au folosit și două scale de evaluare suplimentare pentru a evalua performanța ADL. Cu toate acestea, spre deosebire de





# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

măsura de rezultat primară, UPDRS II, niciuna dintre acestea nu a arătat diferențe semnificative între grupurile de studiu.

*Metodele utilizate pentru analiza datelor au fost adecvate pentru a îndeplini obiectivele studiului și pentru datele disponibile în studiu?*

Datele au fost analizate conform principiului intenției de tratare, care este în general recomandat, deoarece seamănă mai bine cu practica clinică și evită potențialul risc de a pretinde un efect de tratament care de fapt nu există din cauza, de exemplu, a efectelor adverse care îi determină pe oameni să oprească tratamentul.

O analiză de putere a fost efectuată la începutul studiului pentru a determina dimensiunea necesară a eșantionului pentru a detecta o diferență de rezultat cu o putere de 90% la  $P < 0,05$ . Deoarece rezultatul studiului a fost foarte semnificativ statistic, cu o valoare  $P$  foarte mică ( $< 0,0001$ ), este clar că studiul nu a fost sub putere. Cu toate acestea, este mai puțin clar dacă a implicat de fapt mai mulți oameni decât era necesar. Acest lucru se datorează valorii  $P$  foarte mici în combinație cu faptul că nu știm ce dimensiune a unui efect (adică, cât de mare este o diferență între grupuri) care a fost utilizată ca diferență relevantă clinic în analiza puterii (doar că s-a bazat pe studii anterioare folosind UPDRS II). Vom reveni la acest punct în scurt timp.

Variabila rezultatului a fost evaluată folosind UPDRS II, care este o scală de evaluare și, prin urmare, reprezintă date calitative (ordinale). Cu toate acestea, datele au fost prezentate folosind media și SD și analizate folosind ANCOVA, care toate sunt statistici parametrice care presupun date cantitative. Deși măsura în care acest lucru ar fi influențat (părtinitoare) rezultatele este în esență necunoscută. Cu toate acestea, deoarece datele scalei de evaluare nu sunt statistici parametrice adecvate (Greenhalgh 1997a; Svensson 2001), acest lucru poate pune în pericol validitatea rezultatelor.

*Sunt rezultatele studiului semnificative din punct de vedere clinic și interpretabile?*

La sfârșitul studiului a fost observată o diferență medie de 1,6 pe UPDRS II (care reprezintă aproximativ 3% din intervalul posibil de scor 0-52). O întrebare relevantă din punct de vedere clinic este, prin urmare, ce înseamnă aceasta de fapt și dacă reprezintă o diferență semnificativă din punct de vedere clinic. Studiile anterioare care abordează diferența minimă importantă a scorurilor UPDRS II au sugerat că îmbunătățirile oricăror dintre aproximativ 2 și 5 pot fi semnificative, în funcție de stadiul bolii (Hagell 2019). Diferența observată de 1,6 este sub acest interval și nu se știe cât de reprezentativă este această diferență medie raportată, deoarece mediana poate fi mai adecvată. În plus, nu se știe ce înseamnă această diferență în termeni practici. Adică, cum au beneficiat de fapt oamenii cu această îmbunătățire. Faptul că UPDRS II reprezintă atât performanță în activitate, cât și deficiențe complică și mai mult această chestiune.

În cele din urmă, este important să fim conștienți de faptul că scorurile și rezultatele medii din studiile clinice care se bazează pe comparații de grup nu spun nimic despre rezultatele persoanelor individuale. Acest lucru se datorează faptului că sunt date de grup. În principiu, eșecul de a demonstra un efect se poate datora faptului că unii participanți s-au îmbunătățit,



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.

iar unii s-au deteriorat, rezultând nicio schimbare sau diferență medie, când de fapt un subgrup de oameni poate experimenta beneficii clare. Aceasta este o preocupare generală care este dificil de ocolit fără analize de subgrup efectuate cu atenție sau luând în considerare eroarea de măsurare la nivel individual de persoană (Hobart et al. 2010).





# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

## Referințe

Alrubaiy L, Hutchings HA, Williams JG. Assessing patient reported outcome measures: A practical guide for gastroenterologists. *United European Gastroenterology Journal* 2014; 2(6): 463–470. <https://doi.org/10.1177/2050640614558345>

Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311: 485. <https://doi.org/10.1136/bmj.311.7003.485>

Andrich D, Marais I. *A course in Rasch measurement theory: Measuring in the educational, social and health sciences*. Singapore: Springer Nature Singapore Pte Ltd, 2019.

Bland JM, Altman DG. Multiple significance tests: the Bonferroni method; *BMJ* 1995; 310(6973): 170. <https://doi.org/10.1136/bmj.310.6973.170>

Campbell NR. *Physics: The elements*. London: Cambridge University Press, 1920. <https://archive.org/details/physicstheelemen029733mbp/page/n11/mode/2up>

Cano SJ, Hobart JC. The problem with health measurement. *Patient Preference and Adherence* 2011; 5: 279-290. <https://doi.org/10.2147/PPA.S14399>

Chan YH. Randomised controlled trials (RCTs) - Sample size: The magic number? *Singapore Medical Journal* 2003; 44(4):172-174. [www.sma.org.sg/smj/4404/4404bs1.pdf](http://www.sma.org.sg/smj/4404/4404bs1.pdf)

Cohen J. The earth is round ( $p < .05$ ). *American Psychologist* 1994; 49(12): 997–1003. <https://doi.org/10.1037/0003-066X.49.12.997>

Elfil M, Negida A. Sampling methods in clinical research; An educational review. *Emergency* 2017; 5(1): e52. <https://doi.org/10.22037/emergency.v5i1.15215>

Fahn S, Elton RL, members of the UPDRS development committee. Unified Parkinson's Disease Rating Scale. In: Fahn S, Marsden CD, Calne DB, Goldstein M (Eds.) *Recent Developments in Parkinson's Disease*, Vol. 2. Florham Park: MacMillan Healthcare Information, 1987, pp. 153-163.

FDA. *Patient-reported outcome measures: use in medicinal product development to support labelling claims*. Washington DC: Food and Drug Administration, 2009. [www.fda.gov/media/77832/download](http://www.fda.gov/media/77832/download)

Forjaz MJ, Martinez-Martin P. Metric attributes of the Unified Parkinson's Disease Rating Scale 3.0 battery: Part II, construct and content validity. *Movement Disorders* 2006; 21(11): 1892–1898. <https://doi.org/10.1002/mds.21071>





# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

Greenhalgh T. Statistics for the non-statistician. I: Different types of data need different statistical tests. *BMJ* 1997a; 315(7104): 364-366. [www.ncbi.nlm.nih.gov/pmc/articles/PMC2127256/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2127256/)

Greenhalgh T. Statistics for the non-statistician. II: "Significant" relations and their pitfalls. *BMJ* 1997b; 315(7105): 422-425. [www.ncbi.nlm.nih.gov/pmc/articles/PMC2127270/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2127270/)

Grimes DA, Schulz KF. An overview of clinical research: The lay of the land. *Lancet* 2002; 359(9300): 57-61. [https://doi.org/10.1016/S0140-6736\(02\)07283-5](https://doi.org/10.1016/S0140-6736(02)07283-5)

Guyatt G, Walter S, Shannon H, Cook D, Jaeschke R, Heddle N. Basic statistics for clinicians: 4. Correlation and regression. *Canadian Medical Association Journal* 1995; 152(4): 497-504. [www.ncbi.nlm.nih.gov/pmc/articles/PMC1337703/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1337703/)

Hagell P. Measuring activities of daily living in Parkinson's disease: On a road to nowhere and back again? *Measurement* 2019; 132: 109-124. <https://doi.org/10.1016/j.measurement.2018.09.050>

Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. *International Journal of Nursing Studies* 2009; 46(3): 380-393. <https://doi.org/10.1016/j.ijnurstu.2008.10.007>

Haidich AB. Meta-analysis in medical research. *Hippokratia* 2010; 14(Suppl 1): 29-37. [www.ncbi.nlm.nih.gov/pmc/articles/PMC3049418/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049418/)

Hariz GM, Lindberg M, Hariz MI, Bergenheim AT. Does the ADL part of the Unified Parkinson's Disease Rating Scale measure ADL? An evaluation in patients after pallidotomy and thalamic deep brain stimulation. *Movement Disorders* 2003; 18(4): 373-381. <https://doi.org/10.1002/mds.10386>

Hobart J. Rating scales for neurologists. *Journal of Neurology Neurosurgery & Psychiatry* 2003; 74 (Suppl 4): iv22-iv26. <https://doi.org/jnnp.74.suppl.4.iv22>

Hobart JC, Lamping DL, Thompson AJ. Evaluating neurological outcome measures: the bare essentials. *Journal of Neurology Neurosurgery & Psychiatry* 1996; 60(2): 127-130. <https://doi.org/10.1136/jnnp.60.2.127>

Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technology Assessment* 2009; 13(12). <https://doi.org/10.3310/hta13120>

Hobart JC, Cano SJ, Warner TT, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? *Journal of Neurology Neurosurgery and Psychiatry* 2010; 81: 1044-1048. <http://dx.doi.org/10.1136/jnnp.2009.201392>



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

Howarth P, Redgrave F. *Metrology – In short, 3rd edition*. Euramet: Albertslund, 2008. [www.dfm.dk/wp-content/uploads/2020/12/Metrology\\_in\\_short\\_3rd\\_ed.pdf](http://www.dfm.dk/wp-content/uploads/2020/12/Metrology_in_short_3rd_ed.pdf)

Impellizzeri FM, Bizzini M. Systematic review and meta-analysis: A primer. *International Journal of Sports Physical Therapy* 2012; 7(5): 493-503. [www.ncbi.nlm.nih.gov/pmc/articles/PMC3474302/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3474302/)

Johnson JA. Influence of race or ethnicity on pharmacokinetics of drugs. *Journal of Pharmaceutical Sciences* 1997; 86(12): 1328-1333. <https://doi.org/10.1021/js9702168>

Joint Committee for Guides in Metrology (JCGM). *International vocabulary of metrology - Basic and general concepts and associated terms (VIM), 3rd edition*. JCGM, 2012. [www.bipm.org/en/committees/jc/jcgm/publications](http://www.bipm.org/en/committees/jc/jcgm/publications)

Kang M, Ragan BG, Park JH. Issues in outcomes research: an overview of randomization techniques for clinical trials. *Journal of Athletic Training* 2008; 43(2): 215-221. <https://doi.org/10.4085/1062-6050-43.2.215>

Kuhn TS. The function of measurement in modern physical science. *Isis* 1961; 52(2): 161-193. <https://doi.org/10.1086/349468>

Matthews JNS, Altman DG. Interaction 2: compare effect sizes not P values. *BMJ* 1996; 313: 808. <https://doi.org/10.1136/bmj.313.7060.808>

McCluskey A, Lalkhen AG. Statistics II: Central tendency and spread of data. *Continuing Education in Anaesthesia Critical Care & Pain* 2007; 7(4): 127-130. <https://doi.org/10.1093/bjaceaccp/mkm020>

Miller CJ, Smith SN, Pugatch M. Experimental and quasi-experimental designs in implementation research. *Psychiatry Research* 2020; 283: 112452. <https://doi.org/10.1016/j.psychres.2019.06.027>

Munro BH. *Statistical methods for health care research*. Philadelphia: Lippincott Williams & Wilkins, 2005.

Noordzij M, Dekker FW, Zoccali C, Jager KJ. Study designs in clinical research. *Nephron Clinical Practice* 2009; 113(3): c218-221. <https://doi.org/10.1159/000235610>

Norman GR, Steiner DL. *Biostatistics: The Bare Essentials*. Shelton: PMPH-USA Limited, 2014.

Pendrill L. *Quality assured measurement: Unification across social and physical sciences*. Cham: Springer Nature Switzerland AG, 2019.



Program Committee of the Institute of Objective Measurement. Definition of objective measurement. Institute of Objective Measurement, 2000. [www.rasch.org/define.htm](http://www.rasch.org/define.htm)

Ramaker C, Marinus J, Stiggelbout AM, van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. *Movement Disorders* 2002; 17(5): 867–876. <https://doi.org/10.1002/mds.10248>

Rasch G. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut, 1960.

Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312: 71-72. <https://doi.org/10.1136/bmj.312.7023.71>

Shas RR. Inter-ethnic differences in drug response: Implications for drug development and complying with drug regulation. *Clinical Research and Regulatory Affairs* 2015; 32(3): 90-100. <https://doi.org/10.3109/10601333.2015.1064131>

Sheldon TA. Estimating treatment effects: Real or the result of chance? *Evidence-Based Nursing* 2000; 3(2): 36-39. <http://dx.doi.org/10.1136/ebn.3.2.36>

Shorten A, Moorley C. Selecting the sample. *Evidence Based Nursing* 2014; 17(2): 32-33. <http://dx.doi.org/10.1136/eb-2014-101747>

Shulman LM, Armstrong M, Ellis T, Gruber-Baldini A, Horak F, Nieuwboer A, Parashos S, Post B, Rogers M, Siderowf A. Disability rating scales in Parkinson's disease: critique and recommendations. *Movement Disorders* 2016; 31: 1455-1465. <https://doi.org/10.1002/mds.26649>

Stevens SS. On the theory of scales of measurement. *Science* 1946; 103(2684): 677-680. <https://doi.org/10.1126/science.103.2684.677>

Stigler SM. *The History of Statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Belknap Press of Harvard University Press, 1986.

Streiner DL, Norman GR. *Health Measurement Scales. A Practical Guide to Their Development and Use (5<sup>th</sup> edition)*. Oxford: Oxford University Press, 2014.

Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaires. *Journal of Rehabilitation Medicine* 2001; 33(1): 47-48. <https://doi.org/10.1080/165019701300006542>

Swinscow TDV, Campbell MJ. *Statistics at square one (9th edition)*. London: BMJ Publishing Group, 1997. [www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one](http://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one)



Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. *Value in Health* 2004; 7(Suppl 1): S22-26. <https://doi.org/10.1111/j.1524-4733.2004.7s106.x>

van Hilten JJ, van der Zwan AD, Zwinderman AH, Roos RAC. Rating impairment and disability in Parkinson's disease: Evaluation of the Unified Parkinson's Disease Rating Scale. *Movement Disorders* 1994; 9(1): 84-88. <https://doi.org/10.1002/mds.870090113>

Whitley E, Ball J. Statistics review 3: Hypothesis testing and P values. *Critical Care* 2002a; 6(3): 222-225. <https://doi.org/10.1186/cc1493>

Whitley E, Ball J. Statistics review 4: Sample size calculations. *Critical Care* 2002b; 6(4): 335-341. <https://doi.org/10.1186/cc1521>

World Health Organization. *International classification of functioning, disability and health: ICF*. Geneva: World Health Organization, 2001. <https://apps.who.int/iris/handle/10665/42407>

World Medical Association. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. World Medical Association, 2018. [www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/](http://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/)

Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation* 1989; 70(12): 857-860. [www.rasch.org/memo44.htm](http://www.rasch.org/memo44.htm)

Wright BD, Masters GN: Rating scale analysis. Chicago: Mesa Press, 1982. [www.rasch.org](http://www.rasch.org)

Wright BD, Stone MH: Best test design. Chicago: Mesa Press, 1979. [www.rasch.org](http://www.rasch.org)







# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

## Anexa 1: Scurte explicații ale termenilor comuni în cercetarea clinică

Termen	Explicație
Risc absolut	Riscul de a experimenta un anumit eveniment (de exemplu, o boală sau o complicație) într-o anumită perioadă de timp. Exprimat ca procent de persoane care experimentează evenimentul și de cei care ar putea să-l experimenteze. De exemplu, dacă 26 din 100 de persoane vor avea demență în timpul vieții, riscul absolut este de 26/100 sau 26%.
Evaluare	Aprecierea structurată sau nestructurată a ceva. Evaluările nestructurate se fac în mod obișnuit, de exemplu „Mă simt stresat” și „această persoană pare să experimenteze durere”. Evaluările structurate în cercetarea clinică implică în mod obișnuit prezența sau absența a ceva (de exemplu, un diagnostic) pe baza unor criterii, precum și utilizarea diferitelor scale de evaluare raportate de către sine sau de observator cu privire la prezența/absența, frecvența sau amploarea/severitatea a ceva (de exemplu, simptome, sprijin, încredere, bunăstare). Valorile numerice rezultate sunt fie de natură pur calitativă (de exemplu, 0 = bolnav, 1 = sănătos; date nominale), fie ordonate (de exemplu, 0 = niciunul, 1 = ușoară, 2 = moderată, 3 = severă; date ordinale) care nu acceptă calcularea validă, de exemplu, a mărimii sau procentelor diferențelor sau modificărilor.
Interferență/ influență	Ceva care distorsionează rezultatele unui studiu. De exemplu, dacă un anumit subgrup de persoane este supra-reprezentat într-un studiu, acest lucru poate influența rezultatele.
Importanța clinică	Gradul în care un studiu abordează o problemă de importanță pentru practica clinică. Alternativ, gradul în care rezultatul unui anumit studiu (de exemplu, diferența de rezultat între persoanele care primesc tratamente diferite) este important în practica clinică.
Focus grup	Un grup de indivizi reunit pentru a discuta un subiect sau a răspunde la anumite întrebări referitoare la un anumit subiect. Folosit de obicei în studii calitative și cu metode mixte (interviuri de grup țintă/focus).
Teorie solidă	O abordare de cercetare calitativă care urmărește să dezvolte teorii bazate pe observații din lumea reală. Datele sunt de obicei analizate continuu în timpul colectării lor și utilizate pentru a informa interviurile viitoare.
Consimțământ informat	Un principiu etic care cere cercetătorilor să obțină participarea voluntară a subiecților, după ce i-au informat despre studiu și despre posibilele riscuri și beneficii ale acestuia și despre ce înseamnă să participe.



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

Intenția de a trata (ITT)	Un principiu de analiză a datelor în studiile intervenționale, în care datele sunt analizate în funcție de grupul căruia i-a fost repartizat fiecare participant la începutul studiului, indiferent dacă a continuat sau nu acel tratament, a abandonat prematur sau chiar a făcut un studiu greșit sau tratament. Adică, fiecare pacient randomizat în studiul clinic ar trebui inclus în analiza primară în cadrul grupului de tratament respectiv, căruia i-au fost repartizați la randomizare. Opusul este denumit analiză per-protocol (PP)
Interviu	Metodă de colectare a datelor în care un intervievator pune întrebări unui respondent, fie față în față, fie prin alte mijloace (de exemplu, telefon, apel video). Poate fi structurat (respectarea strictă a unui chestionar standardizat), semi-structurat (o listă de subiecte care urmează să fie tratate) sau nestructurat (fără plan predeterminat cu privire la conținutul sau fluxul de informații).
Măsurare	Cuantificare liniară a ceva care este exprimat folosind o unitate definită care își menține dimensiunea indiferent de instrumentul folosit și indiferent cine sau ce persoană sau lucru relevant este măsurat. Valorile numerice rezultate sunt liniare, fie fără zero absolut (de exemplu, temperatură conform Celcius sau Fahrenheit; date de interval), fie cu zero absolut (de exemplu, greutate în kg).
Meta-analiza	Integrarea cantitativă a rezultatelor mai multor studii cantitative care abordează o anumită întrebare de cercetare. De exemplu, mai multe studii clinice privind efectul unui anumit tratament pot fi rezumate într-o meta-analiză pentru a oferi rezultate noi și integrate asupra eficacității acestuia. Omologul său calitativ se numește meta-sinteză.
Cercetare cu metode mixte	Cercetare care colectează și analizează atât date calitative, cât și cantitative.
Ipoteza nulă	O ipoteză care spune că nu există niciun efect, de exemplu, diferență între grupuri.
Raportul de cote	Cotele sunt probabilitatea ca un eveniment să se producă împărțită la probabilitatea ca acel eveniment să nu se producă. Un raport de cote este compararea cotelor pentru două rezultate care descriu asocierea lor.
Observația participantă	O metodă de colectare a datelor în care cercetătorul participă și observă un anumit grup sau cultură.
Valoarea P	Probabilitatea ca un rezumat statistic al datelor (de exemplu, diferența de medie a eșantionului între două



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

	grupuri) să fie egală sau mai mare decât valoarea sa observată în ipoteza nulă.
Cercetare calitativa	Investigarea fenomenelor, de obicei într-o manieră aprofundată și holistică, prin colectarea și analiza materialelor narative, de obicei transcripții textuale ale interviurilor (dar și alte texte). Un obiectiv comun este de a obține o înțelegere mai bună și aprofundată a ceva, de exemplu impactul trăirii cu o anumită tulburare.
Cercetare cantitativă	Studii care descriu și investighează asocierile dintre fenomene folosind date numerice, de obicei bazate pe evaluări și/sau măsurători. Are ca scop realizarea de predicții și generalizări.
Studiu controlat randomizat (RCT)	Un design cu adevărat experimental, de obicei referitor la efectul unei intervenții sau al unui tratament. În afară de tratamentul studiat, un RCT implică o condiție de control (un grup de participanți care primesc un tratament alternativ sau fără tratament) și randomizare, adică lăsarea hazardului să decidă dacă participanții vor primi tratamentul în studiu sau alternativ/niciun tratament.
Risc relativ	Uneori numită și rație de risc. O estimare a riscului pentru ceva dintr-un grup în comparație cu (relativ cu) altul, calculată prin împărțirea riscului absolut pentru un grup, la riscul absolut pentru altul. De exemplu, dacă persoanele cu antecedente familiale ale unei anumite afecțiuni au un risc de 12% de a dezvolta ei înșiși afecțiunea, iar cei fără antecedente familiale au un risc de 3%, riscul relativ de a dezvolta afecțiunea este de 4 ori (12/3) mai mare în rândul persoanelor cu antecedente familiale.
Fiabilitate	Definită ca raportul dintre valoarea adevărată și valoarea observată (adevărat + eroare), fiabilitatea este un indice al absenței erorii în cuantificarea sau categorizarea a ceva. Indicele poate varia între 0 și 1, unde valorile mai mari reprezintă o mai bună fiabilitate și mai puțină eroare. Ca atare, fiabilitatea este legată de precizie și acord. Folosit în principal cu scale de evaluare și alte instrumente de evaluare structurate și legat de conceptul de incertitudine în știința măsurării (metrologie). Fiabilitatea poate fi estimată în diverse moduri și din diferite perspective, de exemplu, intern la o scală de evaluare (coerență internă), între evaluatori (inter-evaluați) sau între puncte de timp (test-retestare).
Dreptul de retragere	Un principiu etic care înseamnă că orice participant la un studiu de cercetare are dreptul de a se retrage din (părăsi)



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



# EUROTEQ

Clinical measurement literacy for  
EUROpean Transparency and EQuality in health

	studiu în orice moment, fără a fi nevoit să explice de ce și fără implicații pentru îngrijirea sa clinică.
Saturare	Un concept folosit de obicei în colectarea datelor calitative, indicând un punct în care datele suplimentare (de obicei interviurile) nu oferă nicio informație nouă.
Semnificație statistică	Definit în mod tradițional ca o valoare a probabilității (valoarea P) $<0,05$ (5%). Valoarea P este definită ca probabilitatea ca, de exemplu, diferența medie reală între două grupuri să fie egală sau mai extremă decât valoarea sa observată.
Revizuire sistematică	Rezumat riguros al rezultatelor cercetării cu privire la o anumită întrebare sau domeniu de cercetare, bazat pe o căutare sistematică a bazelor de date din literatura de specialitate, eșantionarea sistematică a surselor și procedurile sistematice de extragere a datelor. Poate fi însoțită de o meta-analiză (cantitativă) sau meta-sinteză (calitativă).
Analiză tematică	O metodă de analiză a datelor calitative (texte, de obicei transcrierea interviurilor) care implică identificarea temelor descriptive și dezvoltarea unor teme analitice generale care dezvoltă noi concepte și înțelegerea fenomenului investigat. Ca atare, de obicei implică o abordare inductivă. Strâns legat de analiza calitativă a conținutului.
Valabilitate	(a) Un aspect central al calității studiilor de cercetare cantitativă. Două aspecte centrale sunt valabilitatea internă și cea externă. Valabilitatea internă se referă la măsura în care rezultatele unui studiu pot fi de încredere (de exemplu, că tratamentul studiat este responsabil pentru efecte) și include aspecte precum metodele statistice și eșantionul utilizat în analize. Validitatea externă se referă la măsura în care rezultatele vor rezista asupra variațiilor în persoane, în cadru, timp sau în măsurarea rezultatelor. Valabilitatea externă este denumită și generalizare. (b) Un aspect central al calității și proprietăților psihometrice ale scalelor de evaluare și ale altor metode de colectare a datelor cantitative, în care se referă la gradul în care datele generate de un instrument reprezintă variabila pe care se intenționează să o reprezinte.



Co-funded by the  
Erasmus+ Programme  
of the European Union

Reference number: 2020-1-SE01-KA202-077806

This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.