Valore degli studi real-world nell’era degli studi clinici randomizzati

Roberto Bergamaschi
Centro di Ricerca Sclerosi Multipla
IRCCS Fondazione Mondino - Pavia
Gli studi clinici randomizzati (RCT) sono universalmente riconosciuti come gli studi ideali per studiare l’effetto di un farmaco.

Perché valutare l’efficacia e la safety di un farmaco nella RW?
FROM THE FIRST BOOK OF DANIEL (Old Testament 587-538 b.C.)

Please test your servants for ten days: Give us nothing but vegetables to eat and water to drink. Then compare our appearance with that of the young men who eat the royal food, and treat your servants in accordance with what you see.” So he agreed to this and tested them for ten days. At the end of the ten days they looked healthier and better nourished than any of the young men who ate the royal food. So the guard took away their choice food and the wine they were to drink and gave them vegetables instead.

Daniel's Training in Babylon

• It is suggested that the world’s first clinical trial was conducted by King Nebuchadnezzar.

• According to The Bible, the king, concerned about keeping his warriors in top physical condition, ordered his people to eat only meat and drink only wine.

• Yet several young men of royal blood, who liked to eat vegetables, objected.

• The king permitted the dissenters to follow the diet of veggies and water – but only for 10 days.

When the experiment ended, the vegetarians appeared better nourished than carnivores, so the king permitted the group to continue with their diet.

- GCP disregarded
- Do the participants signed informed consent?
- Ascertainment bias
- Allocation bias
- Outcome?
- Confounders (Divine intervention)?

The king certainly deserves credit for introducing two major components of a clinical trial:

(i) separate groups following different prescriptions
(ii) finite length of the trial, upon which the results are evaluated
PERINATAL LESSONS FROM THE PAST

James Lind (1716-94) of Edinburgh and the treatment of scurvy

Peter M Dunn

The Lind family moved to Edinburgh from Ayrshire in the 16th century. James Lind (senior) married Margaret Smellum in 1707 and they had a daughter, Joan, nine years before their son James was born on 4 October 1716.

James Lind received his schooling in Edinburgh before being apprenticed at the age of 15 in 1731 to George Langlands, a member of the Incorporation of Surgeons. After completing his training in 1739, he set off south and joined the Royal Navy as a surgeon’s mate. The next nine years were spent voyaging in the Mediterranean, off West Africa, and in the West Indies. In those days ships were cold, damp, and unwholesome, while the food consisted of putrid beef, rancid pork, mouldy biscuit and foul water. During these years, Lind carefully recorded all his observations, as his later writings show. By 1747 he had been promoted surgeon to HMS Salisbury, and it was during her cruise in the English Channel that year that there was a severe outbreak of scurvy and he was able to carry out his classic experiments on
STREPTOMYCIN TREATMENT OF PULMONARY TUBERCULOSIS

A MEDICAL RESEARCH COUNCIL INVESTIGATION

The following gives the short-term results of a controlled investigation into the effects of streptomycin on one type of pulmonary tuberculosis. The inquiry was planned and directed by the Streptomycin in Tuberculosis Trials Committee, composed of the following members: Dr. Geoffrey Marshall (chairman), Professor J. W. S. Blacklock, Professor C. Cameron, Professor N. B. Capon, Dr. R. Cruickshank, Professor J. H. Gaddum, Dr. F. R. G. Heaf, Professor A. Bradford Hill, Dr. L. E. Houghton, Dr. J. Clifford Hoyle, Professor H. Raistrick, Dr. J. G. Scadding, Professor W. H. Tytler, Professor G. S. Wilson, and Dr. P. D’Arcy Hart (secretary). The centres at which the work was carried out and the specialists in charge of patients and pathological work were as follows:

Brompton Hospital, London.—Clinician: Dr. J. W. Crofton, Streptomycin Registrar (working under the direction of the honorary staff of Brompton Hospital); Pathologists: Dr. J. W. Clegg, Dr. D. A. Mitchison.

Colindale Hospital (L.C.C.), London.—Clinicians: Dr. J. V. Hurford, Dr. B. J. Douglas Smith, Dr. W. E. Snell; Pathologists (Central Public Health Laboratory): Dr. G. B. Forbes, Dr. H. D. Holt.

Harefield Hospital (M.C.C.), Harefield, Middlesex.—Clinicians: Dr. R. H. Brent, Dr. L. E. Houghton; Pathologist: Dr. E. Nassau.

Bangour Hospital, Bangour, West Lothian.—Clinician: Dr. I. D. Ross; Pathologist: Dr. Isabella Purdie.

Killingbeck Hospital and Sanatorium, Leeds.—Clinicians: Dr. W. Santon Gilmour, Dr. A. M. Reeve; Pathologist: Professor J. W. McLeod.

Northern Hospital (L.C.C.), Winchmore Hill, London.—Clinicians: Dr. F. A. Nash, Dr. R. Shoulman; Pathologists: Dr. J. M. Alston, Dr. A. Mohun.

Sully Hospital, Sully, Glam.—Clinicians: Dr. D. M. E. Thomas, Dr. L. R. West; Pathologist: Professor W. H. Tytler.
THE CLINICAL TRIAL*

A. BRADFORD HILL, C.B.E., D.Sc., PH.D.†
The CONSORT Statement

- **CONS**olidated **S**tandards **O**f **R**eporting **T**rials
- Sviluppato da gruppi di editori e di ricercatori
- Inteso a migliorare il "reporting" di un CT, permettendo al lettore di capire la conduzione del trial e valutare la validità dei risultati
- Indica quali informazioni riportare sulla base dell’ evidenza empirica che la loro presenza è indispensabile per valutare l’affidabilità e la rilevanza dei risultati e che la loro assenza è associata a stime distorte
CHECKLIST PIU’ NOTE PER VALUTARE I RANDOMIZED CONTROLLED TRIALS (RCTs)

Reporting: descrizione di quello che è stato fatto - chiarezza della descrizione
- CONSORT 1996; aggiornamento nel 2010; 25 items
  [http://www.consort-statement.org](http://www.consort-statement.org)

Conduct: quello che viene fatto e come viene fatto
- Jadad 1996: ogni area; 5 items
- Pedro 2000: per valutare i trials inclusi nel database di fisioterapia PEDro; 11 items
- Chalmers 1981: terapia farmacologica; 32 items
- Reisch 1989: ogni area; 34 items
- Delphi List 1998: ogni area; 9 items
- Maastricht Amsterdam List (MAL) 1997: back pain; 4 items
- Cochrane Collaboration Risk of Bias Table 2008: 7 items
## CONSORT CHECKLIST (Reporting)

<table>
<thead>
<tr>
<th>Section</th>
<th>Item</th>
<th>Standard CONSORT Description</th>
<th>Extension for Nonpharmacologic Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title and abstract</td>
<td>1</td>
<td>How participants were allocated to interventions (e.g., “random allocation,” “randomized,” or “randomly assigned”)</td>
<td>In the abstract, description of the experimental treatment, comparator, care providers, centers, and blinding status</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background</td>
<td>2</td>
<td>Scientific background and explanation of rationale</td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>3</td>
<td>Eligibility criteria for participants and the settings and locations where the data were collected</td>
<td>When applicable, eligibility criteria for centers and those performing the interventions</td>
</tr>
<tr>
<td>Interventions</td>
<td>4</td>
<td>Precise details of the interventions intended for each group and how and when they were actually administered</td>
<td>Precise details of both the experimental treatment and comparator</td>
</tr>
<tr>
<td></td>
<td>4A</td>
<td>Description of the different components of the interventions and, when applicable, descriptions of the procedure for tailoring the interventions to individual participants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4B</td>
<td>Details of how the interventions were standardized</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4C</td>
<td>Details of how adherence of care providers with the protocol was assessed or enhanced</td>
<td></td>
</tr>
<tr>
<td>Objectives</td>
<td>5</td>
<td>Specific objectives and hypotheses</td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td>6</td>
<td>Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules</td>
<td>When applicable, details of whether and how the clustering by care providers or centers was addressed</td>
</tr>
<tr>
<td>Randomization-sequence generation</td>
<td>8</td>
<td>Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification)</td>
<td>When applicable, how care providers were allocated to each trial group</td>
</tr>
<tr>
<td>Allocation concealment</td>
<td>9</td>
<td>Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned</td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td>10</td>
<td>Who generated the allocation sequence, who enrolled participants,</td>
<td></td>
</tr>
</tbody>
</table>
## PEDro CHECKLIST (Conduct)

<table>
<thead>
<tr>
<th>PEDro scale</th>
<th>No</th>
<th>Yes</th>
<th>Where</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. eligibility criteria were specified</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2. subjects were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received)</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>3. allocation was concealed</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>4. the groups were similar at baseline regarding the most important prognostic indicators</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>5. there was blinding of all subjects</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>6. there was blinding of all therapists who administered the therapy</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>7. there was blinding of all assessors who measured at least one key outcome</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>8. measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>9. all subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analysed by “intention to treat”</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>10. the results of between-group statistical comparisons are reported for at least one key outcome</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>11. the study provides both point measures and measures of variability for at least one key outcome</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>
**VALIDITA’ INTERNA**

⇒ evidenzia l’effetto del trattamento quando questo effettivamente esiste

⇒ minimizza i bias che possono produrre risultati “falsi”

---

**RCT**

Consente di stabilire l’efficacia (capacità di modificare in senso favorevole la storia naturale) di un intervento (terapia) in una situazione controllata (ideale) minimizzando il rischio di errori sistematici e l’effetto dei confondenti
CONDIZIONI IDEALI

- Trattati e controlli devono avere la medesima tendenza a mostrare “naturalmente” l’esito di interesse

- Trattati e controlli devono avere la stessa tendenza a realizzare l’effetto atteso dal trattamento

- Le informazioni sull’occorrenza degli esiti di interesse devono essere raccolte con la stessa intensità nei trattati e nei controlli

BIAS

- RANDOMIZZAZIONE

- PLACEBO

- CIECO
Trattati e controlli devono avere la medesima tendenza a mostrare “naturalmente” l’esito di interesse

**NOTA** - non sempre la randomizzazione è efficace nel rendere **comparabili** i due gruppi per le variabili note e non note
Correzione della non comparabilità (sbilanciamento) tra gruppi

*a priori* (randomizzazione)

*a posteriori* (analisi)

stratificazione, appaiamento

analisi multivariata
Trattati e controlli devono avere la stessa tendenza a realizzare l’effetto atteso dal trattamento.

RCT

PLACEBO

COMPARABILITÀ' DEGLI EFFETTI
Questi sono i dati del gruppo placebo!

Durata del trattamento
IFN β-1b: Annual Relapse Rates Over 5 Years

Problemi legati ai criteri di selezione dei CT

• Inclusione di pazienti particolarmente “attivi”
  ➢ Criteri di inclusione
  ➢ Medico: tende a “favorire” i pazienti più “gravi”
  ➢ Paziente: maggiore adesione se più “grave”
Regressione verso la media

Lanciamo 60 dadi

1 $\leftarrow$ 10 dadi
2 $\leftarrow$ 10 dadi
3 $\leftarrow$ 10 dadi
4 $\leftarrow$ 10 dadi
5 $\leftarrow$ 10 dadi
6 $\leftarrow$ 10 dadi

Media = 3.5
Regressione verso la media

Lanciamo 60 dadi

1 \leftarrow 10 dadi
2 \leftarrow 10 dadi
3 \leftarrow 10 dadi
4 \leftarrow 10 dadi
5 \leftarrow 10 dadi
6 \leftarrow 10 dadi

\begin{center}
\textbf{Media} = 5.0
\end{center}
Rilanciamo 30 dadi

1 \(\leftarrow\) 5 dadi

2 \(\leftarrow\) 5 dadi

3 \(\leftarrow\) 5 dadi

4 \(\leftarrow\) 5 dadi

5 \(\leftarrow\) 5 dadi

6 \(\leftarrow\) 5 dadi

Media = 3.5

“Miglioramento” medio

= media iniziale – nuova media

= 5.0 - 3.5

= 1.5
Le informazioni sull’occorrenza degli esiti di interesse devono essere raccolte con la stessa intensità nei trattati e nei controlli.

CIECO

COMPARABILITÀ DELLE OSSERVAZIONI

Protegge dall’effetto confondente di variabili che potrebbero presentarsi nel corso del follow-up
Cecità

• Doppio cieco non necessario
  • Hard end-point: parametro oggettivo non influenzato da errori o pregiudizi
    • Mortalità, variabili strumentali

• Doppio cieco necessario
  • Soft end-point: parametri che per loro natura hanno un’interpretazione non univoca e discutibile
    • Tasso di ricadute, % soggetti liber da ricadute, ...
Mantenimento del cieco nel corso del f-u

Questionario sulla condizione di cecità dello studio

• 143 PL, 158 bIFN-1a
• 99% degli examining non conosceva la terapia
• 32% dei pazienti hanno individuato correttamente il trattamento
  • 52% PL
  • 48% IFN

Jacobs, Ann Neurol 1996
RCT

- L’applicazione dei vari provvedimenti che idealmente devono assicurare la validità dello studio, non sempre ha pieno successo nella pratica

- Le criticità aumentano proporzionalmente con l’aumento delle variabili in studio
Analisi post-hoc

• Devono essere valutate molto criticamente
  • L’efficacia di una terapia può essere affermata solo sull’end-point principale

• Utili se presentate “onestamente”
  • basate su un razionale
  • possono fornire dati interessanti
  • suggerire futuri trial

• Problema del multiple testing
variabile X: gruppo A vs gruppo B  \( p < 0.05 \)

da probabilità di sbagliare nel ritenere significativa la differenza è 5%

variabile Y: gruppo A vs gruppo B  \( p < 0.05 \)

variabile Z: gruppo A vs gruppo B  \( p < 0.05 \)

variabile K: gruppo A vs gruppo B  \( p < 0.05 \)

la probabilità di sbagliare nel ritenere significativa una differenza è 14,26%
Acqua gassata vs. Acqua naturale nella prevenzione delle ricadute nella SM

• Tre variabili:
  • Sesso (maschi vs femmine)
  • Età (< 40 anni vs ≥ 40)
  • Disabilità (EDSS < 3.5 vs ≥ 3.5)

• Probabilita’ di trovare un risultato SIGNIFICATIVO (p < 0.05) per puro caso:
  • 5% con la sola analisi principale
  • 18,5% con 4 analisi (principale+ 3 sottogruppi)
  • 40% con le 10 analisi possibili (come sopra piu’ tutte le 6 combinazioni)
Motivi
• trattamento scarsamente efficace
• effetti collaterali non tollerati

Effetti
• perdita di potenza dello studio
• minore affidabilità dei risultati

Dropouts nel corso dei RCT
<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>8 MIU</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annual exacerbation rates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completers</td>
<td>0.98</td>
<td>0.72</td>
</tr>
<tr>
<td>Dropouts</td>
<td>1.6</td>
<td>1.02</td>
</tr>
<tr>
<td>p Values within treatment group</td>
<td>0.006</td>
<td>0.152</td>
</tr>
</tbody>
</table>

Dropouts vs. Completers: Maggiore tasso di ricadute

IFNB Multiple Sclerosis Study Group, Neurology 1995
<table>
<thead>
<tr>
<th>Year</th>
<th>Treatment Group</th>
<th>N</th>
<th>Exacerbation Rate</th>
<th>Decrease (8 MIU vs placebo)</th>
<th>Significance (placebo vs 8 MIU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Placebo</td>
<td>123</td>
<td>1.44</td>
<td>33%</td>
<td>(p &lt; 0.001)</td>
</tr>
<tr>
<td></td>
<td>1.6 MIU</td>
<td>125</td>
<td>1.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 MIU</td>
<td>124</td>
<td>0.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Placebo</td>
<td>110</td>
<td>1.18</td>
<td>28%</td>
<td>(p = 0.030)</td>
</tr>
<tr>
<td></td>
<td>1.6 MIU</td>
<td>114</td>
<td>1.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 MIU</td>
<td>107</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Placebo</td>
<td>96</td>
<td>0.92</td>
<td>28%</td>
<td>(p = 0.084)</td>
</tr>
<tr>
<td></td>
<td>1.6 MIU</td>
<td>95</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 MIU</td>
<td>95</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Placebo</td>
<td>82</td>
<td>0.88</td>
<td>24%</td>
<td>(p = 0.166)</td>
</tr>
<tr>
<td></td>
<td>1.6 MIU</td>
<td>76</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 MIU</td>
<td>89</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Placebo</td>
<td>56</td>
<td>0.81</td>
<td>30%</td>
<td>(p = 0.393)</td>
</tr>
<tr>
<td></td>
<td>1.6 MIU</td>
<td>52</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 MIU</td>
<td>58</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Are clinical trials of therapeutic agents for MS long enough?

The 1990s has seen an unprecedented growth in therapeutic trials of potential treatments for multiple sclerosis (MS), which has culminated in the licensing of three preparations of interferon beta for this disease. Physicians are being bombarded with material emphasising the therapeutic excellence of these drugs. But how secure are the data that form the basis of the claims?

Rudge, The Lancet, 1999
Sulfasalazine MS Study

Noseworthy et al, Neurology 1998
Sulfasalazine MS Study

Cumulative probability of progressing

Progressive Patients

Placebo

Sulfasalazine

p = 0.098

Noseworthy et al, Neurology 1998
La durata di un clinical-trial RC è di breve-medio termine

L’occorrenza di un evento rilevante è spesso a medio-lungo termine
Studi di Estensione dei RCT

VANTAGGI

➢ Lunga durata dell’osservazione (conferma dei risultati a breve termine)

ASPETTI CRITICI

✓ Coorti pre-selezionate
✓ Perdita della cecità
✓ Perdita di popolazione di confronto (estrapolazioni, coorti storiche)
✓ Progressivo incremento dei drop-out (autoselezione)
A prospective open-label study of glatiramer acetate: over a decade of continuous use in multiple sclerosis patients

CC Ford¹, KP Johnson², RP Lisak³, HS Panitch⁴, G Shifroni⁵, JS Wolinsky⁶ and The Copaxone® Study Group

Estensione dello studio: Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial.

10 years
108/232 (46%)

Johnson KP et al. Neurology 1995
Limiti dei RCT

• Condotti su popolazioni selezionate in setting protetti
• Condotti in Centri ultraspecializzati o, viceversa, poco specializzati (aree dove il sistema sanitario non offre disponibilità)
• Effetto trial
  ➢ Per la sola partecipazione ad uno studio, un soggetto riceve benefici, indipendentemente dal braccio di trattamento assegnato, incluso placebo
• Dimensioni ridotte
  ➢ Difficile evidenziare differenze di efficacia nei sottogruppi
• Durata del follow-up relativamente breve
  ➢ Difficile far emergere effetti avversi rari o che si verificano a distanza di tempo
• Molto costosi
Storia naturale

Disegno di studio RCT

Condizione innaturale

Farmaci individuati come efficaci, sono indicati solo per sottocategorie di pazienti

Pazienti selezionati

Breve osservazione
VALIDITA’ DEGLI STUDI CLINICI

VALIDITA’ INTERNA

⇒ evidenzia l’effetto del trattamento quando questo effettivamente esiste

⇒ minimizza i bias che possono produrre risultati “falsi”

VALIDITA’ ESTERNA

⇒ consente di generalizzare le conclusioni dello studio alle popolazioni reali al di fuori delle condizioni ideali (artificiali) del RCT
• Elevata validità INTERNA, ma scarsa validità ESTERNA
  ➢ Dubbia generalizzabilità dei risultati su popolazioni con caratteristiche diverse
  ➢ Pochi dati su comorbidità, terapie concomitanti, compliance reale alle terapie

Sempre più difficili da condurre per questioni etiche
Quesiti non risolti dai RCT

• Efficacia e sicurezza del trattamento in studio
  • a lungo termine
  • in pazienti «particolari» (in età pediatrica, in gravidanza)
  • in pazienti con comorbidità

• Comparazione del trattamento in studio con altre terapie

• Combinazione e sequenza del trattamento in studio con altre terapie (posizionamento)
Clinical research is not relevant to practice

- Traditional RCTs study the effectiveness of treatments delivered to carefully selected populations under ideal conditions.
- This makes it difficult to translate results to the real world.
- Even when we do implement a tested intervention into everyday clinical practice, we often see a "voltage drop"—a dramatic decrease in effectiveness.

“If we want more evidence-based practice, we need more practice-based evidence.”

Pazienti trattati secondo la pratica clinica

Tutti i pazienti sono trattati, compresi quelli con comorbidità

Pazienti trattati secondo il protocollo

Criteri di inclusione e di esclusione
To determine whether an intervention produces the expected result under ideal circumstances.

To study Efficacy

To measure the degree of beneficial effect under “real world” clinical settings.

To study Effectiveness

To study Efficacy

To study Effectiveness

RCT

RW studies
Studi osservazionali in RW

VANTAGGI
- Lunga durata
- End point rilevanti per la vita del paziente (disabilità, progressione)
- Comportamento del trattamento nella pratica quotidiana
- Minori costi

Buona VALIDITÀ ESTERNA

ASPETTI CRITICI
- Mancanza di randomizzazione
- Mancanza di cecità
- Popolazioni di confronto

Scarsa VALIDITÀ INTERNA
COME MIGLIORARE LA VALIDITÀ' INTERNA DEGLI STUDI OSSERVAZIONALI

• Disporre di fonti di dati affidabili
• Ottimizzare la qualità dello studio
• Utilizzare approcci statistici adeguati per:
  • minimizzare i bias
  • rendere le popolazioni confrontabili
COME MIGLIORARE LA VALIDITÀ' INTERNA DEGLI STUDI OSSERVAZIONALI

• Disporre di fonti di dati affidabili

• **Ottimizzare la qualità dello studio**

• Utilizzare approcci statistici adeguati per:
  • minimizzare i bias
  • rendere le popolazioni confrontabili
Griglie per la valutazione critica della qualità degli studi

- Ne esistono per ogni tipo di studio
- Consentono una lettura guidata degli articoli
- Pongono domande su modalità specifiche di conduzione dello studio che devono essere rispettate affinché i risultati siano affidabili
- Aiutano ad individuare eventuali distorsioni sistematiche
CHECKLIST PIU’ NOTE PER VALUTARE GLI STUDI NON RANDOMIZZATI
(NRs: SPERIMENTALI - OSSERVAZIONALI)

Reporting: descrizione di quello che è stato fatto - chiarezza della descrizione

✔ **Strobe** (von Elm 2007): cohort, case control cross sectional; 22 items


Conduct: quello che viene fatto e come viene fatto

✔ **New Castle Ottawa Scale**: cohort studies: case control; 8 items

✔ **Downs and Black instrument**: cohort studies

✔ **Criteria of the Cochrane EPOC Group**: controlled before after studies; interrupted time series analysis
STROBE CHECKLIST (Reporting: descrizione di quello che è stato fatto-chiarezza della descrizione)

<table>
<thead>
<tr>
<th>Item No</th>
<th>Recommendation</th>
</tr>
</thead>
</table>
| **Title and abstract** | 1. Indicate the study’s design with a commonly used item in the title or the abstract  
(b) Provide in the abstract an informative and balanced summary of what was done and what was found |
| **Introduction** | 2. Explain the scientific background and rationale for the investigation being reported |
| **Objectives** | 3. State specific objectives, including any prespecified hypotheses |
| **Methods** | 4. Present key elements of study design early in the paper |
| **Study design** | 5. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection |
| **Participants** | 6. (a) Cohort study—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up  
Case-control study—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls  
Cross-sectional study—Give the eligibility criteria, and the sources and methods of selection of participants  
(b) Cohort study—For matched studies, give matching criteria and number of exposed and unexposed  
Case-control study—For matched studies, give matching criteria and the number of controls per case |
| **Variables** | 7. Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable |
| **Data sources/measurement** | 8. For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group |
| **Bias** | 9. Describe any efforts to address potential sources of bias |
| **Study size** | 10. Explain how the study size was arrived at |
| **Qualitative variables** | 11. Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why |
| **Statistical methods** | 12. (a) Describe all statistical methods, including those used to control for confounding  
(b) Describe any methods used to examine subgroups and interactions  
(c) Explain how missing data were addressed |
NEW CASTLE CHECKLIST (Conduct: quello che viene fatto e come viene fatto)

Selection
1) Representativeness of the exposed cohort
   a) truly representative of the average _______________ (describe) in the community ✿
   b) somewhat representative of the average ______________ in the community ✿
   c) selected group of users eg nurses, volunteers
   d) no description of the derivation of the cohort
2) Selection of the non exposed cohort
   a) drawn from the same community as the exposed cohort ✿
   b) drawn from a different source
   c) no description of the derivation of the non exposed cohort
3) Ascertainment of exposure
   a) secure record (eg surgical records) ✿
   b) structured interview ✿
   c) written self report
   d) no description
4) Demonstration that outcome of interest was not present at start of study
   a) yes ✿
   b) no

Comparability
1) Comparability of cohorts on the basis of the design or analysis
   a) study controls for ______________ (select the most important factor) ✿
   b) study controls for any additional factor ✿ (This criteria could be modified to indicate specific control for a second important factor.)

Outcome
1) Assessment of outcome
   a) independent blind assessment ✿
   b) record linkage ✿
   c) self report
   d) no description
ACCURATEZZA DIAGNOSTICA

**Reporting:** descrizione di quello che è stato fatto - chiarezza della descrizione

✓ *Stard* (Bossuyt 2003); 25 items

**Conduct:** quello che viene fatto e come viene fatto

✓ *Quadas* (Whiting 2003); 14 items
STARD CHECKLIST (Reporting: descrizione di quello che è stato fatto-chiarezza della descrizione)

<table>
<thead>
<tr>
<th>Section and topic</th>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title, abstract, and keywords</td>
<td>1</td>
<td>Identify the article as a study of diagnostic accuracy (recommend MeSH heading “sensitivity and specificity”)</td>
</tr>
<tr>
<td>Introduction</td>
<td>2</td>
<td>State the research questions or aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups</td>
</tr>
<tr>
<td>Methods:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>3</td>
<td>Describe the study population: the inclusion and exclusion criteria and the settings and locations where the data were collected</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Describe participant recruitment: was this based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Describe participant sampling: was this a consecutive series of participants defined by selection criteria in items 3 and 4? If not, specify how participants were further selected</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Describe data collection: was data collection planned before the index tests and reference standard were performed (prospective study) or after (retrospective study)?</td>
</tr>
<tr>
<td>Test methods</td>
<td>7</td>
<td>Describe the reference standard and its rationale</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Describe technical specifications of material and methods involved, including how and when measurements were taken, or cite references for index tests or reference standard, or both</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Describe definition and rationale for the units, cut-off points, or categories of the results of the index tests and the reference standard</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Describe the number, training, and expertise of the persons executing and reading the index tests and the reference standard</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>Were the readers of the index tests and the reference standard blind (masked) to the results of the other test? Describe any other clinical information available to the readers.</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>12</td>
<td>Describe methods for calculating or comparing measures of diagnostic accuracy and the statistical methods used to quantify uncertainty (e.g., 95% confidence intervals)</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Describe methods for calculating test reproducibility, if done</td>
</tr>
<tr>
<td>Results:</td>
<td>14</td>
<td>Report when study was done, including beginning and ending dates of recruitment</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>Report clinical and demographic characteristics (e.g., age, sex, spectrum of presenting symptoms, comorbidity, current treatments, and recruitment centre)</td>
</tr>
</tbody>
</table>
Table 2: The QUADAS tool

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes</th>
<th>No</th>
<th>Unclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the spectrum of patients representative of the patients who will receive the test in practice?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>2. Were selection criteria clearly described?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>3. Is the reference standard likely to correctly classify the target condition?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>6. Did patients receive the same reference standard regardless of the index test result?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>8. Was the execution of the index test described in sufficient detail to permit replication of the test?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>9. Was the execution of the reference standard described in sufficient detail to permit its replication?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>10. Were the index test results interpreted without knowledge of the results of the reference standard?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>11. Were the reference standard results interpreted without knowledge of the results of the index test?</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
</tbody>
</table>
RIASSUMENDO

Individuare il disegno di studio: scegliere le scale appropriate

Ho necessità di pubblicare o valutare un articolo per la pubblicazione: Checklist sul Reporting

Ho necessità di leggere e capire se le conclusioni sono valide: Checklist sul Conduct
The GRACE Checklist: A Validated Assessment Tool for High Quality Observational Studies of Comparative Effectiveness

Nancy A. Dreyer, PhD, MPH; Allison Bryant, MPH; and Priscilla Velentgas, PhD

2016  *JMCP*  Journal of Managed Care & Specialty Pharmacy  1107
<table>
<thead>
<tr>
<th>Component Item</th>
<th>Scoring as Fit for Purpose: Sufficient (+), Insufficient (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td></td>
</tr>
</tbody>
</table>
| D1. Were treatment and/or important details of treatment exposure adequately recorded for the study purpose in the data source(s)? Note: not all details of treatment are required for all research questions. | (+) Yes, reasonably necessary information to determine treatment or intervention was adequately recorded for study purposes (e.g., for drugs, sufficient detail on dose, days supplied, route, or other important data. For vaccines, consider the importance of batch, dose, route, and site of administration, etc. For devices, consider type of device, placement, surgical procedure used, serial number, etc.).  
(-) No, data source clearly deficient, or not enough information in article.  |
| D2. Were the primary outcomes adequately recorded for the study purpose (e.g., available in sufficient detail through data sources)?    | (+) Yes, information to ascertain outcomes were adequately recorded in the data source (e.g., if clinical outcomes were ascertained using ICD-9-CM diagnosis codes in an administrative database, the level of sensitivity and specificity captured by the codes were sufficient for assessing the outcome of interest).  
(-) No, data source clearly deficient (e.g., the codes captured a range of conditions that was too broad or narrow, and supplementary information such as that from medical charts was not available), or not enough information in article. |
| D3. Was the primary clinical outcome(s) measured objectively rather than subject to clinical judgment (e.g., opinion about whether the patient's condition has improved)? | (+) Yes, clinical outcomes were measured objectively (e.g., hospitalization, mortality).  
(-) Not applicable; primary outcome not clinical (e.g., PROs).  
(-) No (e.g., clinical opinion about whether patient's condition improved) or not enough information in article. |
| D4. Were primary outcomes validated, adjudicated, or otherwise known to be valid in a similar population?                          | (+) Yes, outcomes were validated, adjudicated, or based on medical chart abstractions with clear definitions (e.g., a validated instrument was used to assess patient-reported outcomes [e.g., SF-12 Health Survey]; a clinical diagnosis via ICD-9-CM code was used, with formal medical record adjudication by committee to confirm diagnosis or other procedures to achieve reasonable sensitivity and specificity; and billing data were used to assess health resource utilization).  
(-) No, or not enough information in article. |
| D5. Was the primary outcome(s) measured or identified in an equivalent manner between the treatment/intervention group and the comparison group? | (+) Yes.  
(-) No, or not enough information in article. |
| D6. Were important covariates that may be known confounders or effect modifiers available and recorded? Important covariates depend on the treatment and/or outcome of interest (e.g., body mass index should be available and recorded for studies of diabetes; race should be available and recorded for studies of hypertension and glaucoma). | (+) Yes, most if not all important known confounders and effect modifiers available and recorded (e.g., measures of medication dose and duration).  
(-) No, at least 1 probable known confounder or effect modifier not available and recorded (as noted by authors or as determined by user's clinical knowledge), or not enough information in article. |

*Dreyer et al, JMCP 2016*
<table>
<thead>
<tr>
<th>Component Item</th>
<th>Scoring as Fit for Purpose: Sufficient (+), Insufficient (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methods</strong></td>
<td></td>
</tr>
</tbody>
</table>
| M1. Was the study (or analysis) population restricted to new initiators of treatment or those starting a new course of treatment? Efforts to include only new initiators may include restricting the cohort to those who had a washout period (specified period of medication nonuse) before the beginning of study follow-up. | (+) Yes, only new initiators of the treatment of interest were included in the cohort, or for surgical procedures and devices, including only patients who never had the treatment before the start of study follow-up.  
(-) No, or not enough information in article. |
| M2. If 1 or more comparison groups were used, were they concurrent comparators? If not, did the authors justify the use of historical comparison groups? | (+) Yes, data were collected during the same time period as the treatment group (concurrent), or historical comparators were used with reasonable justification (e.g., when it is impossible for researchers to identify current users of older treatments or when a concurrent comparison group is not valid, as when uptake of new product is so rapid that concurrent comparators differ greatly on factors related to the outcome).  
(-) No, historical comparators used without being scientifically justifiable, or not enough information in article. |
| M3. Were important confounding and effect-modifying variables taken into account in the design and/or analysis? Appropriate methods to take these variables into account may include restriction, stratification, interaction terms, multivariate analysis, propensity score matching, instrumental variables, or other approaches. | (+) Yes, most if not all important covariates that would be likely to change the effect estimate substantially were accounted for (e.g., measures of medication dose and duration).  
(-) No, some important covariates were available for analysis but not analyzed appropriately, or at least 1 important covariate was not measured, or not enough information in article. |
| M4. Is the classification of exposed and unexposed person-time free of "immortal time bias," i.e., "immortal time" in epidemiology refers to a period of cohort follow-up time during which death (or an outcome that determines end of follow-up) cannot occur. | (+) Yes.  
(-) No, or not enough information in the article. |
| M5. Were any meaningful analyses conducted to test key assumptions on which primary results are based (e.g., were some analyses reported to evaluate the potential for a biased assessment of exposure or outcome, such as analyses where the impact of varying exposure and/or outcome definitions was tested to examine the impact on results)? | (+) Yes, and primary results did not substantially change.  
(-) Yes, and primary results changed substantially.  
(-) None reported, or not enough information in article. |


GRACE = Good Research for Comparative Effectiveness; ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification; PRO = patient-reported outcomes.
Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making

1. *A priori*, determine and declare that a study is a Hypothesis Evaluation Treatment Effectiveness (HETE) study or an Exploratory study based on conditions outlined below.

2. Post a HETE study protocol and analysis plan on a public study registration site prior to conducting the study analysis.

3. Publish HETE study results with attestation to conformance and/or deviation from the study protocol and original analysis plan. Possible publication sites include a medical journal, or a publicly available web-site.

4. Enable opportunities to replicate HETE studies (i.e., for other researchers to be able to reproduce the same findings using the same data set and analytic approach). The ISPE companion paper lists information that should be reported in order to make the operational and design decisions behind a RWD study transparent enough for other researchers to reproduce the conduct of the study.

5. Perform HETE studies on a different data source and population than the one used to generate the hypotheses to be tested unless it is not feasible (e.g., another data set is not available).

6. Authors of the original study should work to publicly address methodological criticisms of their study once it is published.

7. Include key stakeholders (patients, caregivers, clinicians, clinical administrators, HTA/payers, regulators, manufacturers) in designing, conducting, and disseminating HETE studies.
WEBSITE EBM

Centro Cochrane Italiano
http://www.cochrane.it/it

Gruppo Italiano per la Medicina Basata sulle Evidenze - GIMBE
http://www.gimbe.org
COME MIGLIORARE LA VALIDITA’ INTERNA DEGLI STUDI OSSERVAZIONALI

• Disporre di fonti di dati affidabili
• Ottimizzare la qualità dello studio
• Utilizzare approcci statistici adeguati per:
  • minimizzare i bias
  • rendere le popolazioni confrontabili
COME MIGLIORARE LA VALIDITÀ' INTERNA DEGLI STUDI OSSERVAZIONALI

- Disporre di fonti di dati affidabili
- Ottimizzare la qualità dello studio
- Utilizzare approcci statistici adeguati per:
  - minimizzare i bias
  - rendere le popolazioni confrontabili
Errore sistematico e validità interna di uno studio

I risultati di uno studio sono tanto più validi (probabilmente veri) quanto meno esso è affetto da errori sistematici. Valutare la qualità metodologica di uno studio (validità interna) significa verificare se e in che misura lo studio è affetto da errori sistematici.
Table 1. Classes of bias in observational data.

<table>
<thead>
<tr>
<th>Bias</th>
<th>Origin</th>
<th>Mitigation strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indication bias</td>
<td>Non-random treatment exposure</td>
<td>Multivariable adjusted models, matching</td>
</tr>
<tr>
<td>Attrition bias</td>
<td>Between-group difference in follow-up duration</td>
<td>Pairwise censoring</td>
</tr>
<tr>
<td>Detection bias</td>
<td>Differences in follow-up protocols</td>
<td>Models adjusted for follow-up density (e.g. frequency of visits or MRI scans)</td>
</tr>
<tr>
<td>Immortal time bias</td>
<td>Systematic differences in the definitions of study entry</td>
<td>Modelling time-dependent covariates</td>
</tr>
<tr>
<td>Will Rogers phenomenon</td>
<td>Changing diagnostic criteria</td>
<td>Sensitivity analyses excluding historical cohorts</td>
</tr>
<tr>
<td>Recall bias</td>
<td>Systematic differences in the proportion of retrospective data</td>
<td>Sensitivity analyses using only prospectively recorded data</td>
</tr>
<tr>
<td>Selection bias</td>
<td>Preferential inclusion of subpopulations in registries</td>
<td>Sensitivity analysis using only population-based cohorts</td>
</tr>
<tr>
<td>Unidentified bias</td>
<td>Missing information for confounders of disease outcomes</td>
<td>Estimation of the robustness to hidden bias (e.g. Rosenbaum bounds)</td>
</tr>
</tbody>
</table>

MRI: magnetic resonance imaging.

T Kalincik and H Butzkueven, MSJ 2016
Indication bias

In a randomised trial, confoundings of study outcomes are balanced between the treatment groups as a result of randomisation.

In observational studies, treatment assignment is a function of multiple factors, many of which are associated with disease outcomes and therefore act as their confounders. Where sufficient overlap in the confounders between the compared groups exists (dashed area), well-balanced cohorts can be extracted from the existing data using the appropriate analytical methodology (matching or weighting).

Where no such overlap exists, balanced comparison is not possible and the risk of erroneous inference is significant should this remain undetected.

Kalincik and Butzkueven, Multiple Sclerosis Journal, 2016
Attrition bias originates from **systematic difference in the follow-up duration** between the compared cohorts, which results in a difference in the probability of reaching study endpoints. The optimal mitigation strategy is pairwise censoring, when the matched patient pair is censored when one of the patients is censored.

*Kalincik and Butzkueven, Multiple Sclerosis Journal, 2016*
COME MIGLIORARE LA VALIDITA’ INTERNA DEGLI STUDI OSSERVAZIONALI

• Disporre di fonti di dati affidabili
• Ottimizzare la qualità dello studio
• Utilizzare approcci statistici adeguati per:
  • minimizzare i bias
  • rendere le popolazioni confrontabili
COME MIGLIORARE LA VALIDITA’ INTERNA DEGLI STUDI OSSERVAZIONALI

• Disporre di fonti di dati affidabili
• Ottimizzare la qualità dello studio
• Utilizzare approcci statistici adeguati per:
  • minimizzare i bias
  • rendere le popolazioni confrontabili
    • Propensity score
Rendere le popolazioni confrontabili

New Natural History of Interferon-β–Treated Relapsing Multiple Sclerosis

Maria Trojano, MD,1 Fabio Pellegrini, MscStat,2 Aurora Fuiani, MD,1 Damiano Paolicelli, MD,1 Valentina Zipoli, MD,3 Giovanni B. Zimatore, MD,1 Elisabetta Di Monte, MD,1 Emilio Portaccio, MD,3 Vito Lepore, MD,1 Paolo Livrea, MD,1 and Maria Pia Amato, MD3 Ann Neurol, 2007

• 1504 SMRR (1103 treated with IFN, 401 not treated)
• Cox’s regression models adjusted by Propensity Score (PS) in order to obtain homogeneity between groups (treated vs. not treated)
• Covariates included in the model: age at onset, gender, disease duration, number of relapses during the last year, EDSS
• Patients not overlapping in propensity score were excluded
• Propensity score curves adjusted by Cox’s models
Fig 1. Propensity score–adjusted survival curves for end point: time from first visit to secondary progression. Cumulative probability represents the estimated proportion of patients reaching the end point. Solid line indicates untreated group; dashed line indicates treatment group.

Trojano et al, Ann Neurol 2007
COME MIGLIORARE LA VALIDITÀ' INTERNA DEGLI STUDI OSSERVAZIONALI

1. Disporre di fonti di dati affidabili
2. Ottimizzare la qualità dello studio
3. Utilizzare approcci statistici adeguati per:
   - minimizzare i bias
   - rendere le popolazioni confrontabili
     - Propensity score
     - Bayesian score
Predicting secondary progression in relapsing–remitting multiple sclerosis: a Bayesian analysis

Roberto Bergamaschi, Carlo Berzuini, Alfredo Romani, Vittorio Cosi

Table 1  Estimates of the Bayesian risk associated with early clinical predictors observed within 1 year of disease onset

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Mean LRR</th>
<th>Mean log LRR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at onset (in decades)</td>
<td>1.05</td>
<td>0.05</td>
<td>1.02 to 1.09</td>
</tr>
<tr>
<td>Female sex</td>
<td>0.39</td>
<td>-1.07</td>
<td>0.17 to 0.78</td>
</tr>
<tr>
<td>Sphincter onset</td>
<td>2.98</td>
<td>0.93</td>
<td>1.10 to 6.10</td>
</tr>
<tr>
<td>Pure motor onset</td>
<td>2.11</td>
<td>0.62</td>
<td>0.90 to 4.20</td>
</tr>
<tr>
<td>Motor–sensory onset</td>
<td>2.40</td>
<td>0.81</td>
<td>1.15 to 4.41</td>
</tr>
<tr>
<td>Sequel after onset</td>
<td>1.76</td>
<td>0.52</td>
<td>1.04 to 2.88</td>
</tr>
<tr>
<td>Functional systems involved at onset</td>
<td>1.39</td>
<td>0.32</td>
<td>1.16 to 1.64</td>
</tr>
<tr>
<td>Sphincter plus motor relapses</td>
<td>2.10</td>
<td>0.71</td>
<td>1.56 to 2.89</td>
</tr>
<tr>
<td>EDSS ≥4 outside relapse</td>
<td>2.28</td>
<td>0.44</td>
<td>0.40 to 6.50</td>
</tr>
</tbody>
</table>
Early prediction of the long term evolution of multiple sclerosis: the Bayesian Risk Estimate for Multiple Sclerosis (BREMS) score

Roberto Bergamaschi, Silvana Quaglini, Maria Trojano, Maria Pia Amato, Eleonora Tavazzi, Damiano Paolicelli, Valentina Zipoli, Alfredo Romani, Aurora Fuiani, Emilio Portaccio, Carlo Berzuini, Cristina Montomoli, Stefano Bastianello, Vittorio Casi


0.05 x age (in decades)
+ (-1.07) (if female gender)
+ 0.93 (if sphincter onset)
+ 0.62 (if pure motor onset)
+ 0.81 (if motor-sensory onset)
+ 0.32 x number of neurological functional systems involved at onset
+ 0.52 (if sequel after onset)
+ 0.71 x number of sphincter and motor relapses
+ 0.44 (if EDSS 4.0 within the first year of disease)

= BREMS score
Bayesian Risk Estimate for Multiple Sclerosis
Figure 2. Secondary progression-free survival for high risk patients (fourth BREMS quartile) divided into two groups of patients: untreated and treated with DMDs.

Figure 3. Secondary progression-free survival for low risk patients (first BREMS quartile) divided into two groups of patients: untreated and treated with DMDs.
Bayesian Risk Estimate for Multiple Sclerosis at Onset

$0.05 \times \text{age (in decades)} + (-1.07) \text{ (if female gender)} + 0.93 \text{ (if sphincter onset)} + 0.62 \text{ (if pure motor onset)} + 0.81 \text{ (if motor-sensory onset)} + 0.32 \times \text{number of neurological functional systems involved at onset} + 0.52 \text{ (if sequel after onset)}$

= BREMSO score

Bayesian Risk Estimate for Multiple Sclerosis at Onset
BREMSO: a simple score to predict early the natural course of multiple sclerosis


**Kaplan-Meier survival estimates**

Log-rank test P<0.001

Risk of becoming secondary progressive in MS patients. MS patients with a high BREMSO score (> 3rd quartile) are compared with MS patients with a low BREMSO score (< 1st quartile)
L’utilizzo di approcci statistici appropriati può consentire di migliorare il disegno degli studi osservazionali.

“randomizzazione a posteriori”

Analisi di pazienti confrontabili per:
- propensione ad essere trattati (Propensity score)
- rischio di evoluzione sfavorevole (Bayesian score)
Validità Interna

RCT

Validità Esterna

STUDIO OSSERVAZIONALE

✔ fonti di dati affidabili
✔ qualità dello studio
✔ approcci statistici adeguati
Validità Interna

RCT

Validità Esterna

STUDIO OSSERVAZIONALE
Circumstances under which observational studies are valuable and can be **supportive** of randomised clinical trials

<table>
<thead>
<tr>
<th>Circumstance</th>
<th>Impact on clinical practice integrated into guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>When large studies are needed to ascertain an outcome (e.g., to assess infrequent or long-term effects)</td>
<td>Provides important information about benefit-risk evaluation for the health practitioner</td>
</tr>
<tr>
<td>When treatment adherence might have an impact on outcome</td>
<td>Obtaining information about the behaviour of the target group in relation to the outcome</td>
</tr>
<tr>
<td>When a timely result is needed</td>
<td>In situations in which it would be politically or ethically unacceptable to deny access to an intervention</td>
</tr>
<tr>
<td>When multiple treatment solutions are available</td>
<td>Impact on guidelines</td>
</tr>
<tr>
<td>When wanting to explore population subsets</td>
<td>Providing associations and hypotheses, which have to be explored further</td>
</tr>
</tbody>
</table>

Adapted from Dreyer et al. Health Affairs. 2010 and from Heikinheimo et al. 2017
RCTs - RW studies continuum

• RW studies are not abandonment of the scientific methods
• RW studies don’t take away from basic science or diminish the importance of traditional RCTs – Just a balance is needed
• No clinical trial is completely explanatory or RW
• RCTs and RW studies exists on a continuum

Explanatory Trial
Can an intervention work under ideal conditions?

Real-world study
Does an intervention work under usual conditions?
Conclusioni

- Gli studi RW sono complementari agli RCT
- La validità interna degli studi RW è migliorabile
- La disponibilità sempre maggiore di grandi database rappresenta una opportunità per la conduzione di studi RW
- I risultati degli studi RW permettono di colmare lacune nel panorama scientifico, purché condotti con rigore metodologico, accurato controllo della qualità dei dati, impiego di approcci statistici adeguati

“between measurements based on randomized controlled trials and benefit ... there is a gulf which has been much underestimated. ”