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

Roberto Bergamaschi
Centro 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

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.**

Clinical research in ancient Babylon: methodologic insights from the book of Daniel. Grimes DA. Obstet Gynecol. 1995

Daniel's Training in Babylon

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

Clinical research in ancient Babylon: methodologic insights from the book of Daniel. Grimes DA. Obstet Gynecol. 1995

Archives of Disease in Childhood 1997;**76:**F64–F65

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

BRITISH MEDICAL JOURNAL

LONDON SATURDAY OCTOBER 30 1948

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. Reevie; 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 New England
Journal of Medicine

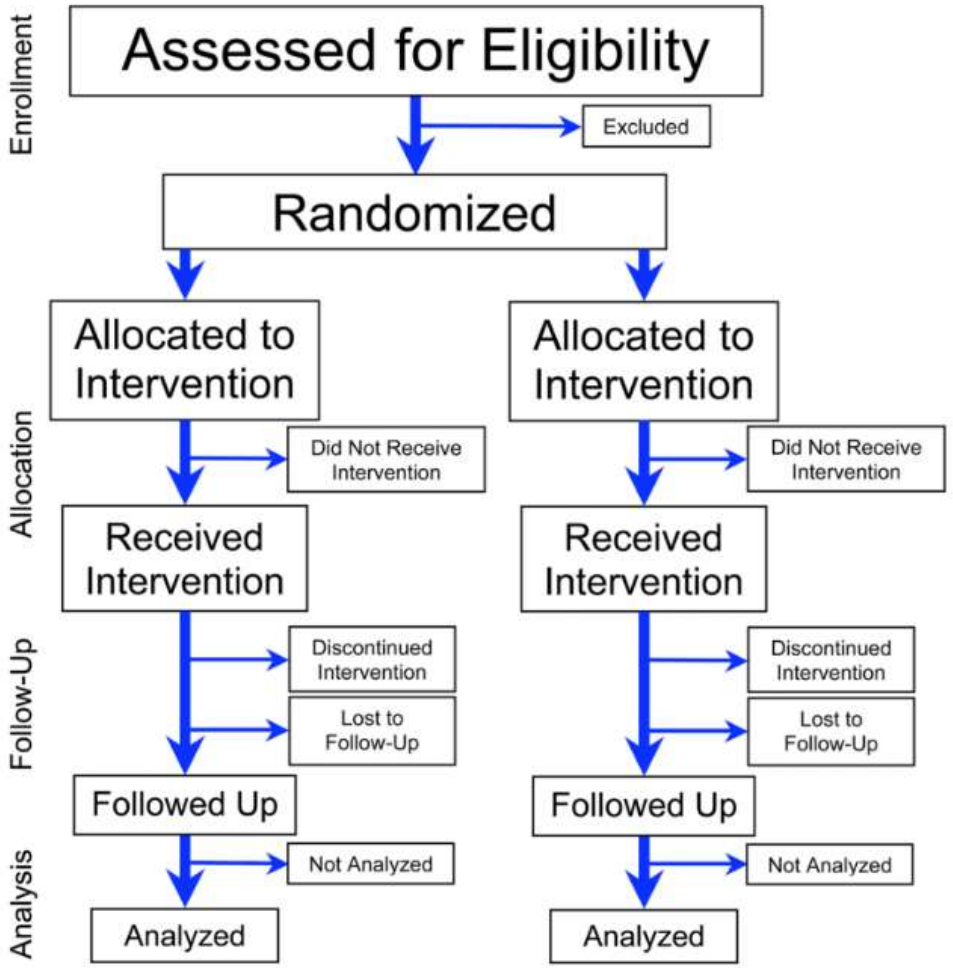
Volume 247 JULY 24, 1952 Number 4

THE CLINICAL TRIAL*

A. Bradford Hill, C.B.E., D.Sc., Ph.D.†

The CONSORT Statement

- **CON**solidated **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 la presenza è indispensabile per valutare l'affidabilità e la rilevanza dei risultati e che la loro assenza è associata a stime distorte



www.consort-statement.org

RCT

VALIDITA' INTERNA

⇒ evidenzia l'effetto del trattamento quando questo effettivamente esiste

⇒ minimizza i bias che possono produrre risultati "falsi"

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

RCT

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

RCT

Trattati e controlli devono avere la medesima tendenza a mostrare "naturalmente" l'esito di interesse

RANDOMIZZAZIONE
(assegnazione casuale a terapia/PL)

COMPARABILITA' DEI GRUPPI

- Favorisce la stessa (simile) distribuzione di variabili prognostiche note e non note tra i gruppi randomizzati

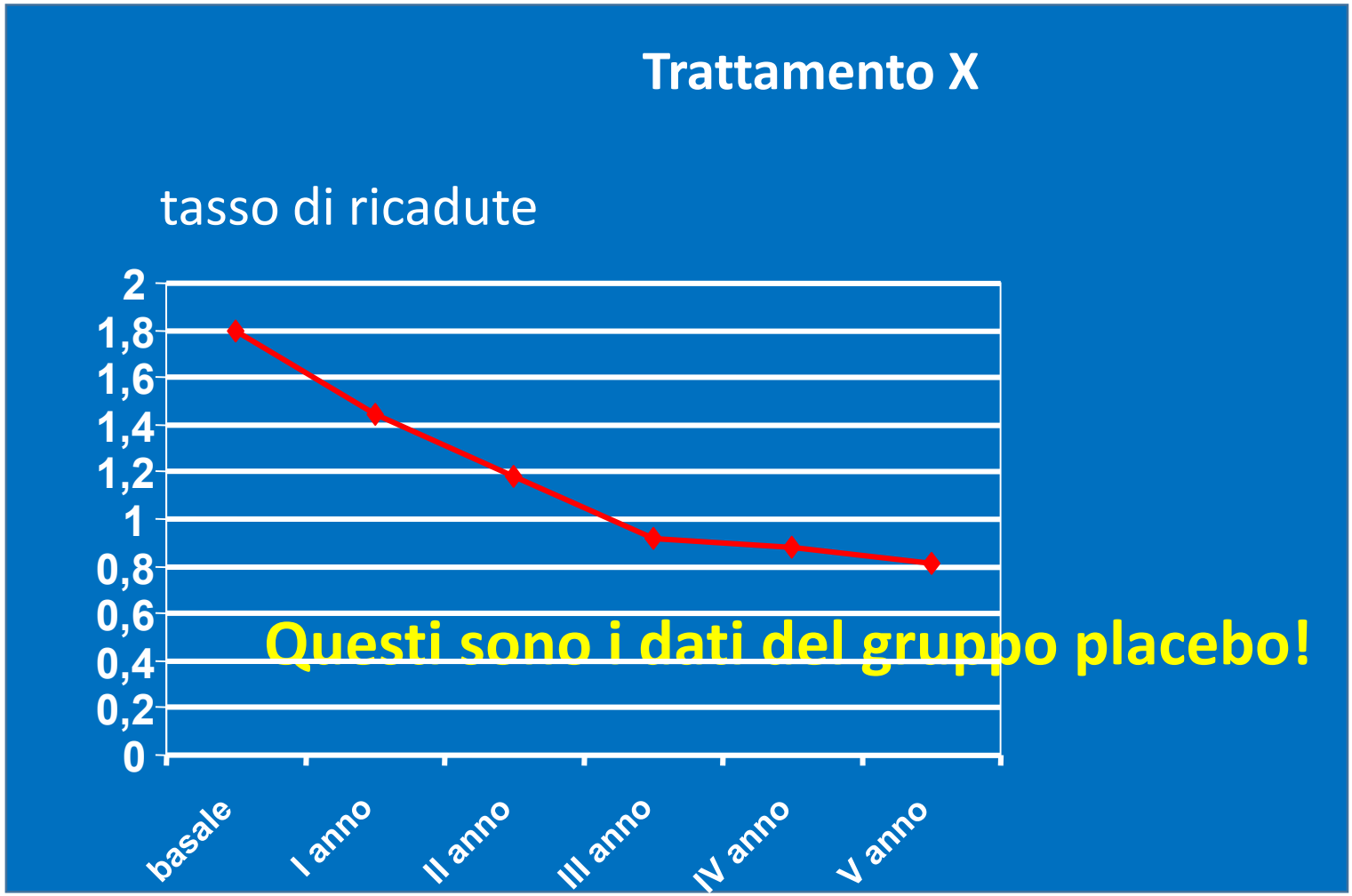
NOTA - non sempre la randomizzazione è efficace nel rendere comparabili i due gruppi per le variabili note e non note

RCT

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

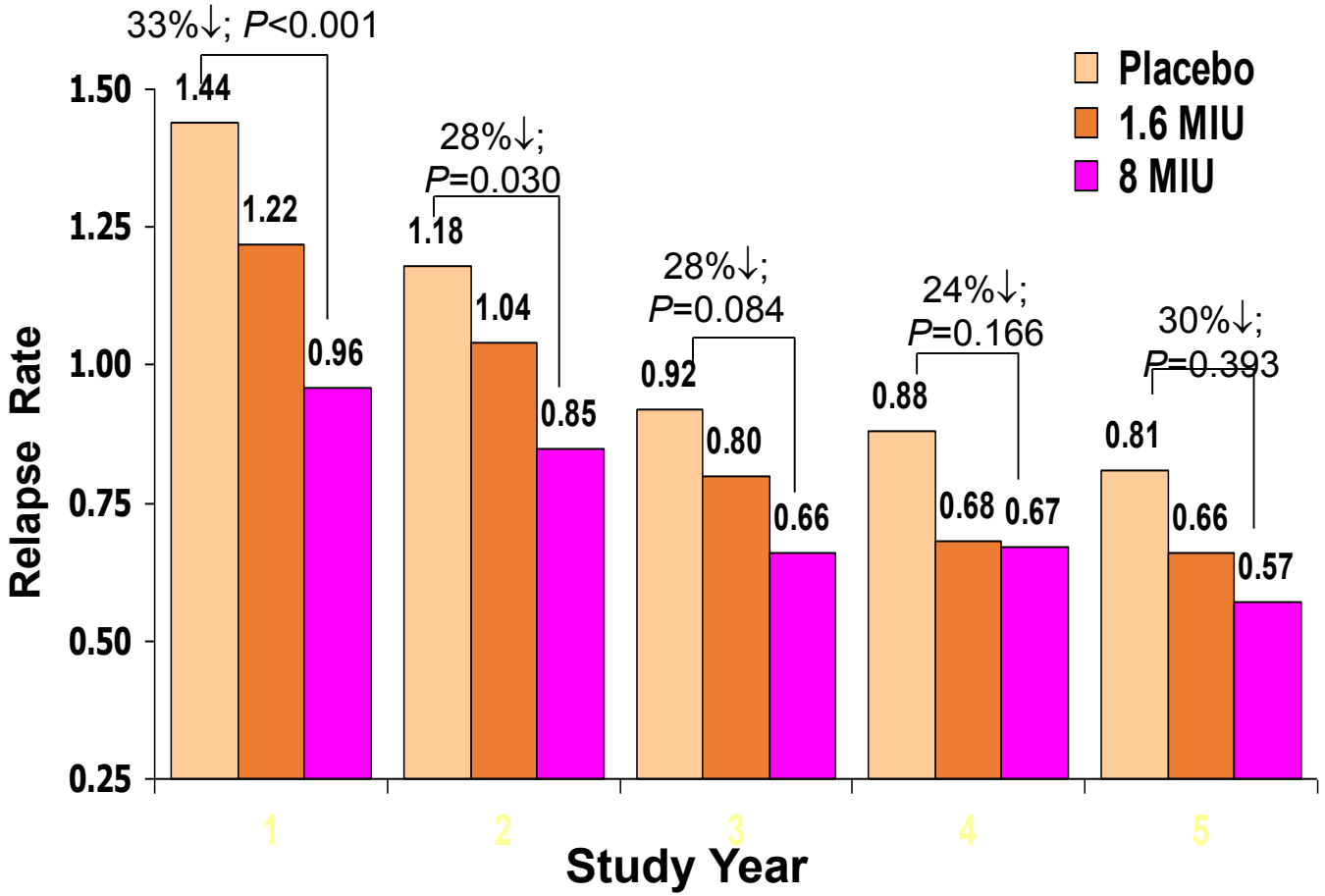
PLACEBO

COMPARABILITA' DEGLI EFFETTI



Durata del trattamento

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



The IFNB MS Study Group. Neurology. 1995;45:1277-1285.

RCT

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

CIECO

COMPARABILITA' DELLE OSSERVAZIONI

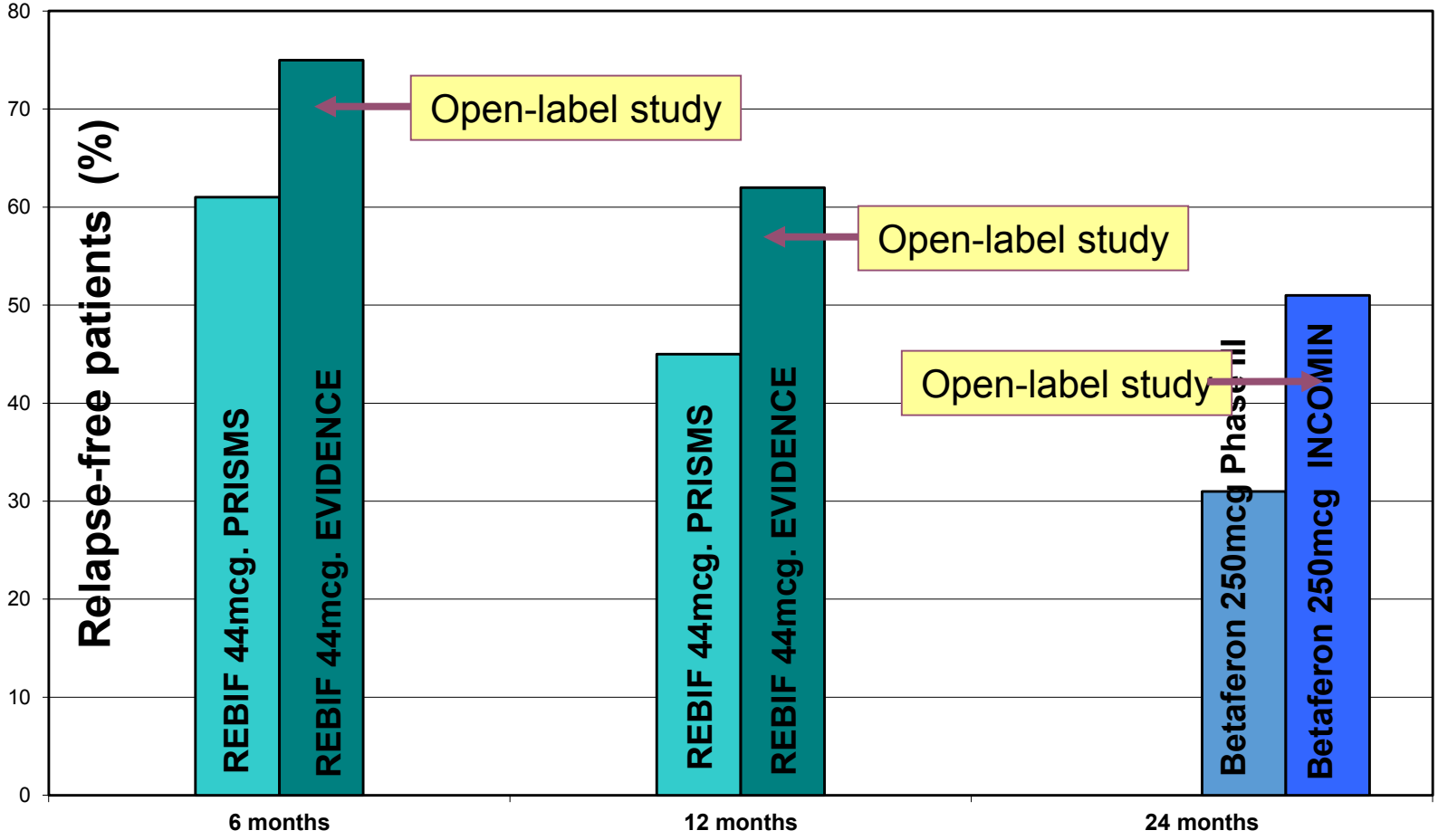
Protegge dall'effetto confondente di variabili che potrebbero presentarsi nel corso del follow-up

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 liberi da ricadute, …

Studi in “aperto” vs doppio-cieco
Le stime sono più ottimistiche



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

Dropouts nel corso dei RCT

Motivi

- trattamento scarsamente efficace
- effetti collaterali non tollerati

Effetti

- perdita di potenza dello studio
- minore affidabilità dei risultati

	Placebo	8 MIU
Annual exacerbation rates		
Completers	0.98	0.72
Dropouts	1.6	1.02
p Values within treatment group	0.006	0.152

Dropouts vs. Completers: Maggiore tasso di ricadute

IFNB Multiple Sclerosis Study Group, Neurology 1995

Table 1. Annual exacerbation rates by year of study

	Treatment group	N	Exac rate	Decrease (8 MIU vs placebo)	Significance (placebo vs 8 MIU)
Year 1	Placebo	123	1.44	33%	$p < 0.001$
	1.6 MIU	125	1.22		
	8 MIU	124	0.96		
Year 2	Placebo	110	1.18	28%	$p = 0.030$
	1.6 MIU	114	1.04		
	8 MIU	107	0.85		
Year 3	Placebo	96	0.92	28%	$p = 0.084$
	1.6 MIU	95	0.80		
	8 MIU	95	0.66		
Year 4	Placebo	82	0.88	24%	$p = 0.166$
	1.6 MIU	76	0.68		
	8 MIU	89	0.67		
Year 5	Placebo	56	0.81	30%	$p = 0.393$
	1.6 MIU	52	0.66		
	8 MIU	58	0.57		

IFNB Multiple Sclerosis Study Group, Neurology 1995

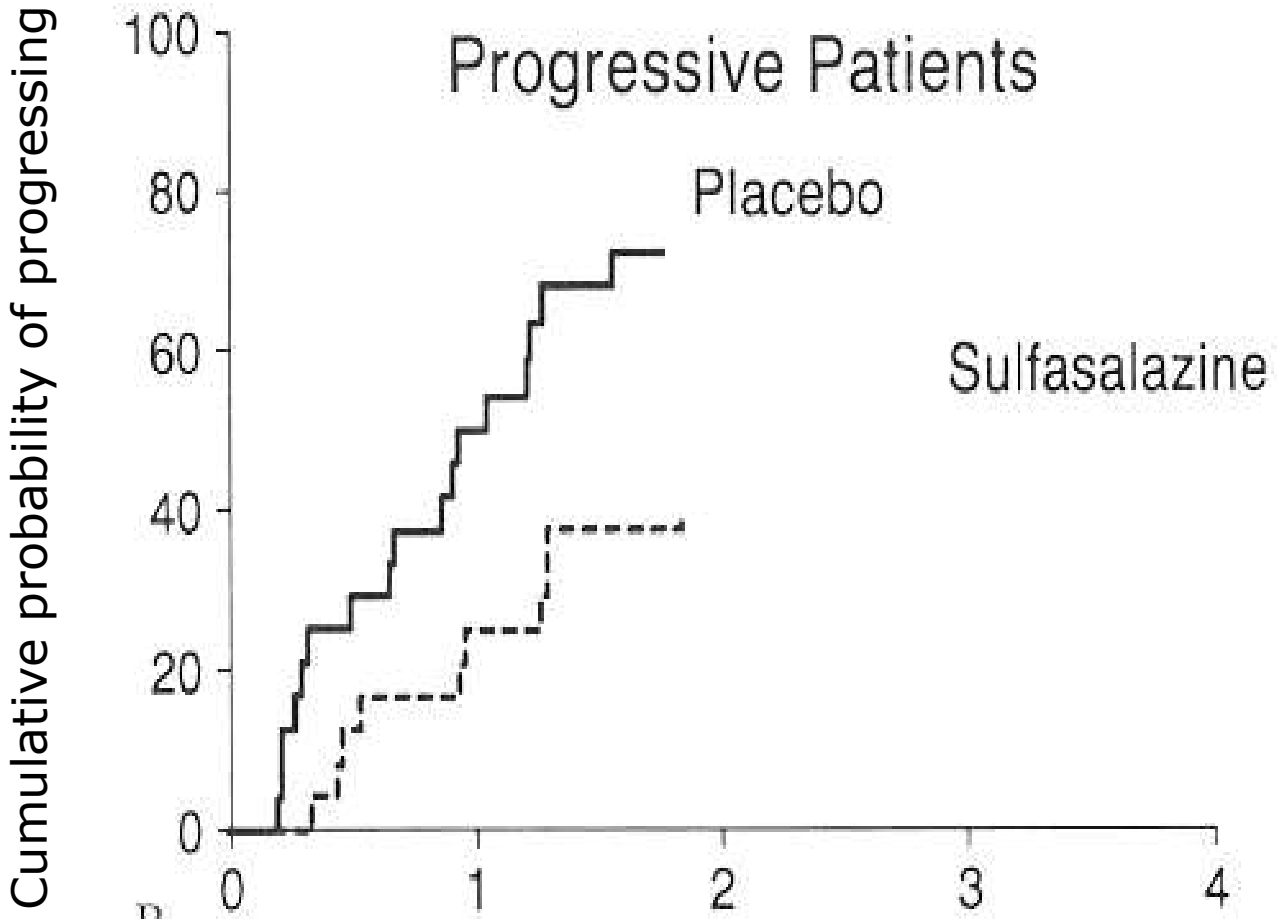
Durata del Follow-up

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?

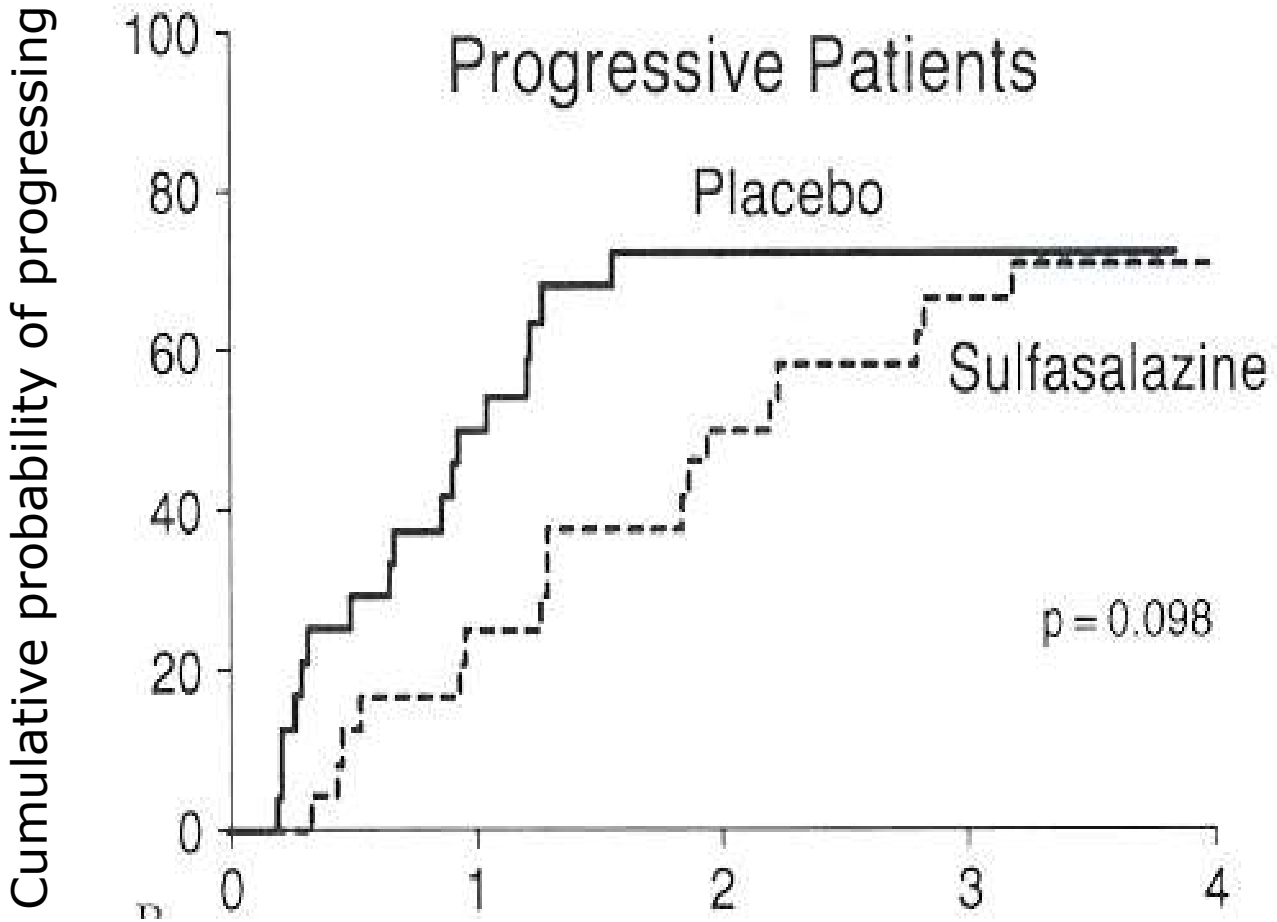
Rudge, The Lancet,1999

Sulfasalazine MS Study



Noseworthy et al, Neurology 1998

Sulfasalazine MS Study



Noseworthy et al, Neurology 1998

La durata di un clinical-trial RC è di breve-medio termine

2-5 anni

L'occorrenza di un evento rilevante (EDSS 6, SP) è a medio-lungo termine

> 10 anni

Studi di Estensione dei RCT

VANTAGGI

- Lunga durata dell'osservazione (conferma dei risultati a breve termine)

ASPETTI CRITICI

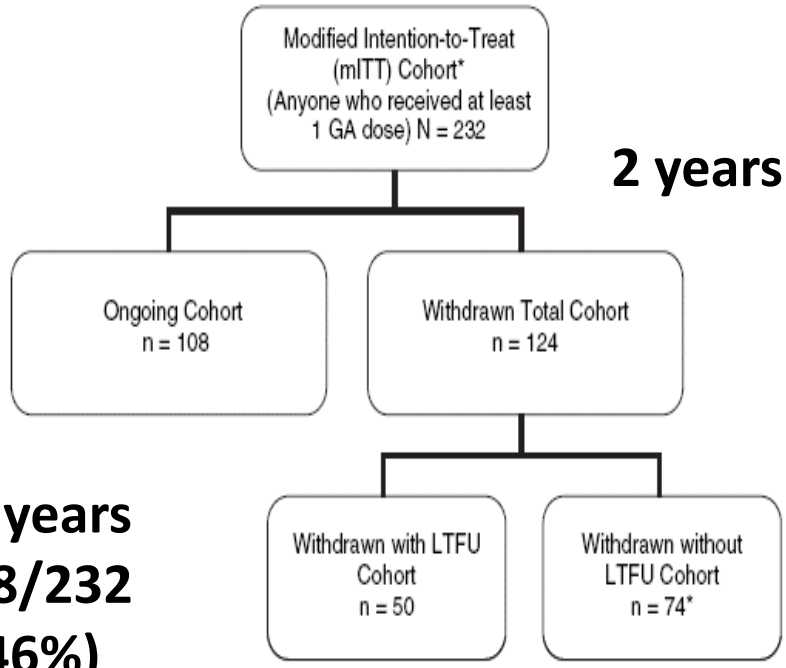
- ✓Coorti pre-selezionate
- ✓Perdita della cecità
- ✓Perdita di popolazione di confronto (estrapolazioni, coorti storiche)
- ✓Progressivo incremento dei drop-out (autoselezione)

ARTICLE Multiple Sclerosis 2006; 12: 309–320

A prospective open-label study of glatiramer acetate: over a decade of continuous use in multiple sclerosis patients

CC Ford[1], KP Johnson[2], RP Lisak[3], HS Panitch[4], G Shifroni[5], JS Wolinsky[6] 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.



2 years

10 years
108/232
(46%)

Johnson KP et al. Neurology 1995

A prospective open-label study of glatiramer acetate: over a decade of continuous use in multiple sclerosis patients

CC Ford[1], KP Johnson[2], RP Lisak[3], HS Panitch[4], G Shifroni[5], JS Wolinsky[6] and The Copaxone® Study Group

Table 5 EDSS data at 10 years/LTFU

EDSS measure	Ongoing ($n=108$)	Withdrawn with LTFU[a] ($n=50$)	*P* value
EDSS score			
Mean ± SD	3.06 ± 1.78	5.22 ± 2.21	<0.0001[c]
Median	2.50	6.00	
Range	0.0–8.0	1.0–9.0	
EDSS change from GA start			
Mean ± SD	0.50 ± 1.65	2.24 ± 1.86	<0.0001[c]
Median	0.50	2.25	
Range	−3.5 to 5.5	3.0–5.5	
Categorical analysis			
Clinically stable/improved	67/108 (62%)	14/50 (28%)	<0.0001[d]
Patients reaching EDSS 4, 6, or 8[b]			
EDSS 4 *n/N* (%)	20/84 (24%)	25/37 (68%)	<0.0001[d]
EDSS 6 *n/N* (%)	8/106 (8%)	23/46 (50%)	<0.0001[d]
EDSS 8 *n/N* (%)	1/108 (1%)	5/50 (10%)	<0.0125[e]

Clinical research is slow

- To most people, randomized controlled trials (RCTs) are the mainstay of clinical research.
- But traditional RCTs are slow and expensive—and rarely produce findings that are easily put into practice.

- In fact, after an **average of 17 years** only 14% of research findings will have led to widespread changes in care.[1]

The evidence paradox

- More than 18,000 RCTs are published each year—in addition to tens of thousands of other clinical studies.
- Yet systematic reviews consistently find that **we don't have enough evidence to effectively inform the clinical decisions** providers and patients must make.

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."

Green, LW. *American Journal of Public Health*, 2006.

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 CT

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

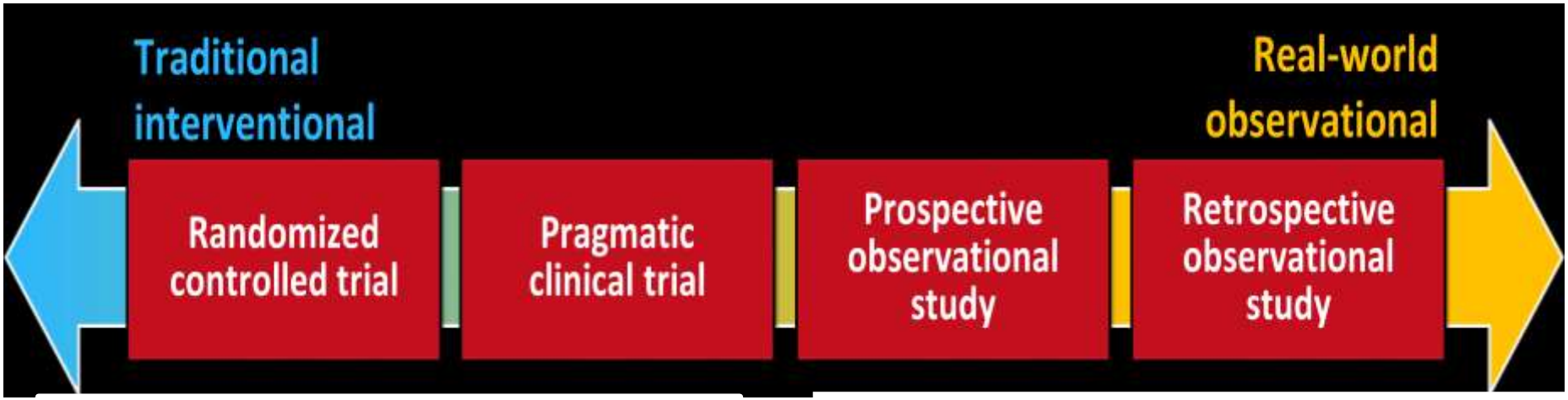
Limiti dei RCT

- **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

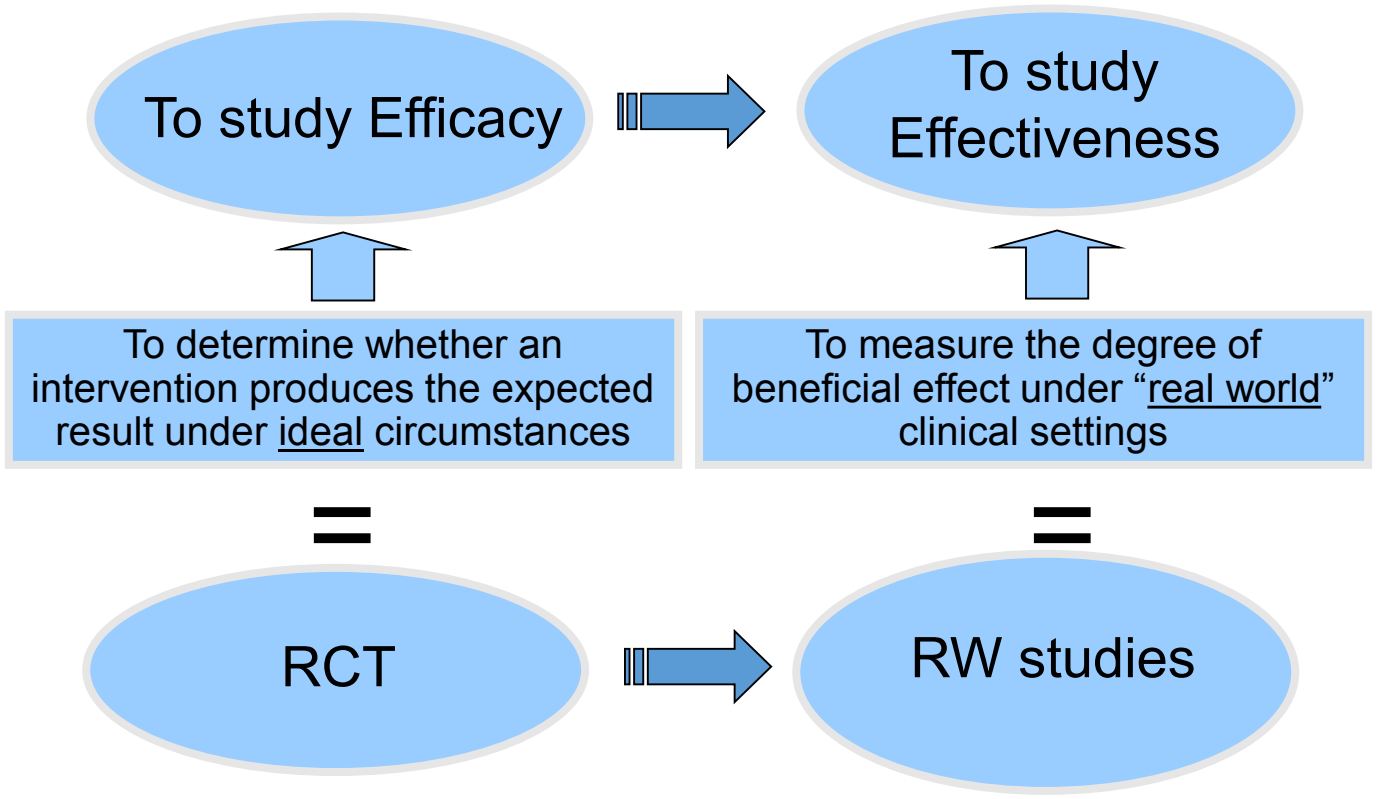
Quesiti non risolti dai RCT

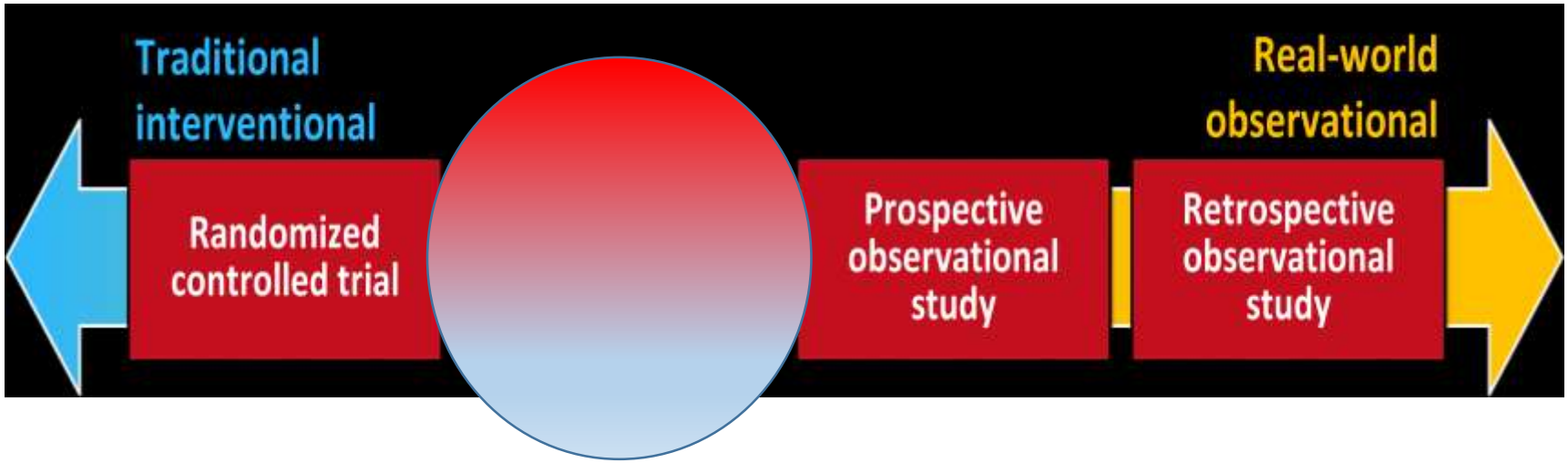
- Efficacia e sicurezza del trattamento in studio
  - a lungo termine
  - in pazienti «particolari» (in età pediatrica, in gravidanza)
  - in pazienti con comorbidità
- Comparazione del trattamento in studio con altre terapie
- Combinazione e sequenza del trattamento in studio con altre terapie (posizionamento)



Pazienti trattati secondo il protocollo ⇔ **Pazienti trattati secondo la pratica clinica**

Criteri di inclusione e di esclusione ⇔ **Tutti i pazienti sono trattati, compresi quelli con comorbidità**


To study Efficacy
To study Effectiveness
To determine whether an intervention produces the expected result under ideal circumstances
To measure the degree of beneficial effect under "real world" clinical settings
=
=
RCT
RW studies


Traditional interventional
Real-world observational
Randomized controlled trial
Prospective observational study
Retrospective observational study

A **pragmatic clinical trial** (**PCT**), sometimes called a **practical clinical trial** (**PCT**), is a clinical trial that focuses on correlation between treatments and outcomes in real-world health system practice rather than focusing on proving causative explanations for outcomes

PCTs: Fewer exclusions allow for a broader subset of participants

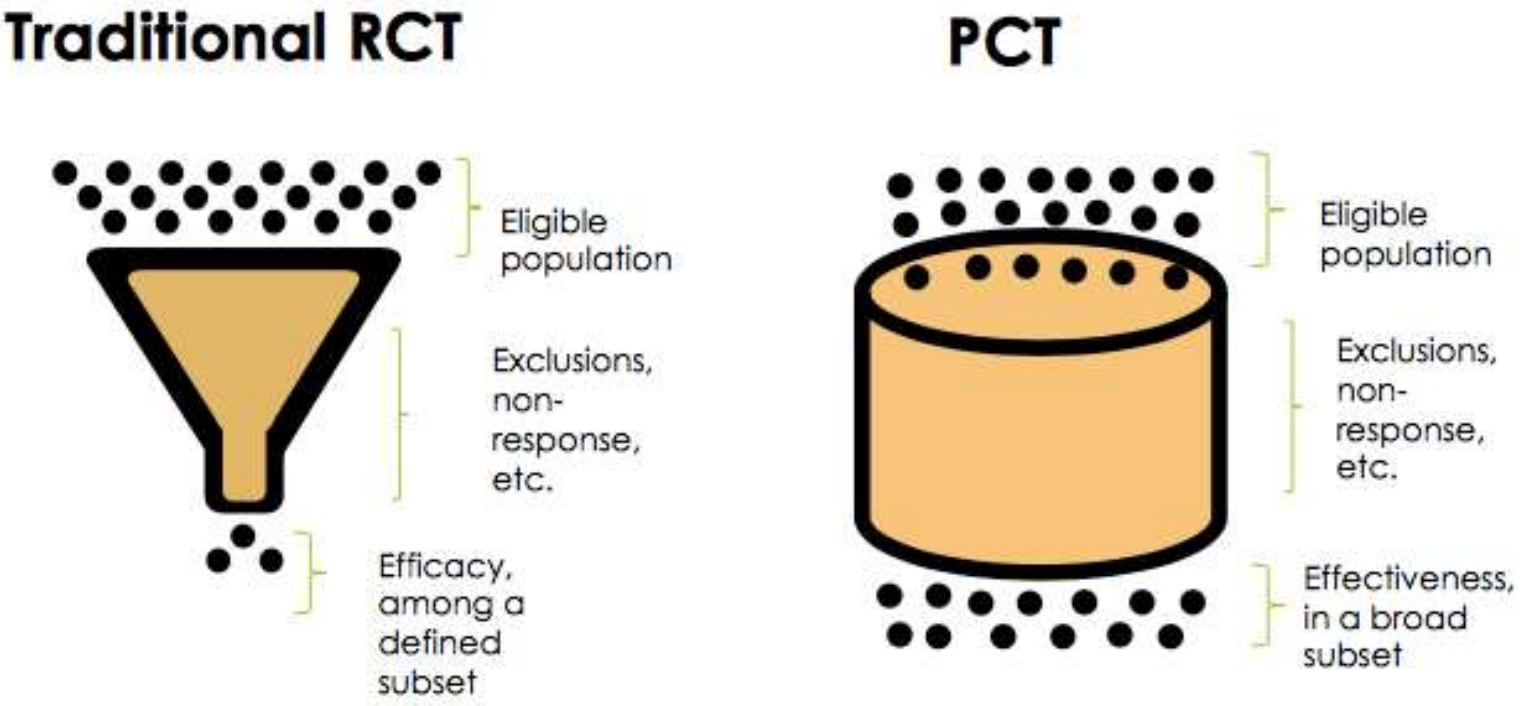


Figure provided by Gloria Coronado, PhD, Kaiser Permanente Center for Health Research

RCT vs Pragmatic trials

Table 1 | Key differences between trials with explanatory and pragmatic attitudes, adapted from a table presented at the 2008 Society for Clinical Trials meeting by Marion Campbell, University of Aberdeen

Question	Efficacy—can the intervention work?	Effectiveness—does the intervention work when used in normal practice?
Setting	Well resourced, "ideal" setting	Normal practice
Participants	Highly selected. Poorly adherent participants and those with conditions which might dilute the effect are often excluded	Little or no selection beyond the clinical indication of interest
Intervention	Strictly enforced and adherence is monitored closely	Applied flexibly as it would be in normal practice
Outcomes	Often short term surrogates or process measures	Directly relevant to participants, funders, communities, and healthcare practitioners
Relevance to practice	Indirect—little effort made to match design of trial to decision making needs of those in usual setting in which intervention will be implemented	Direct—trial is designed to meet needs of those making decisions about treatment options in setting in which intervention will be implemented

Zwarenstein, et al. BMJ 2008;337:a2390

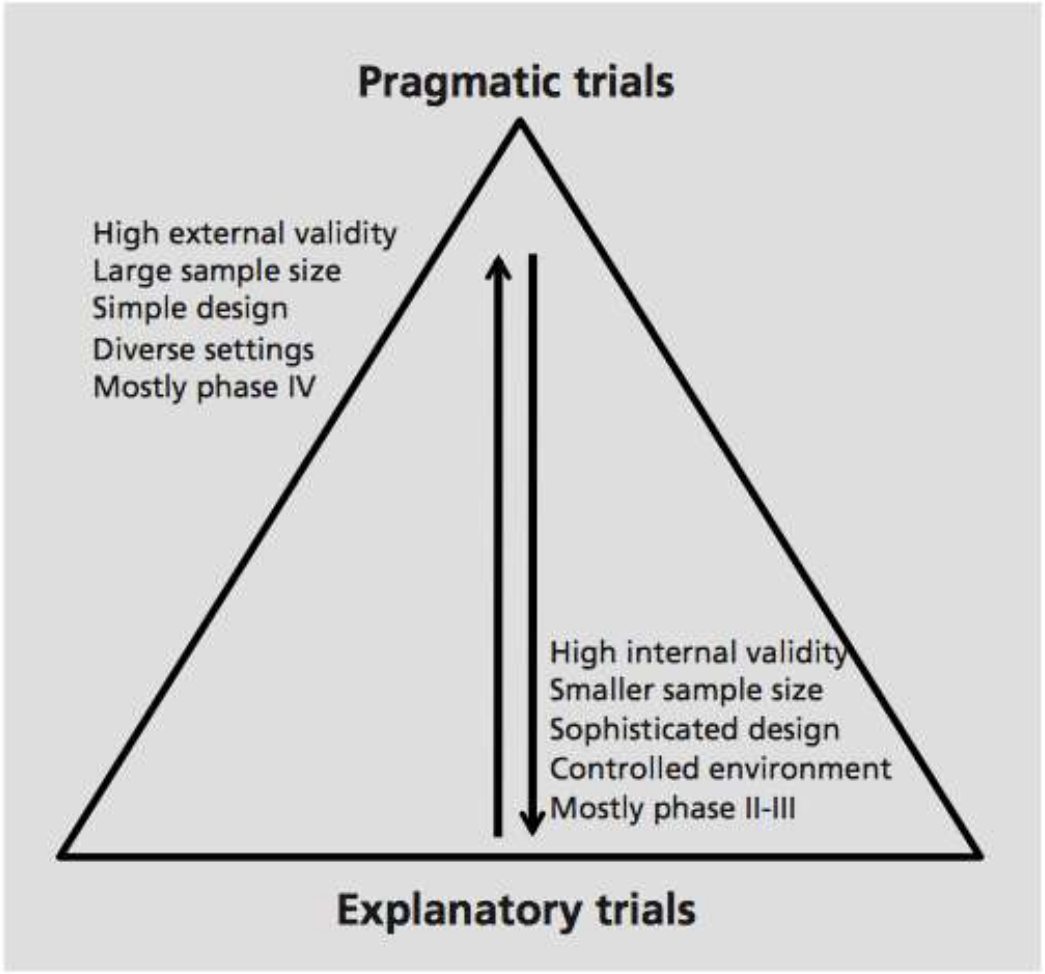
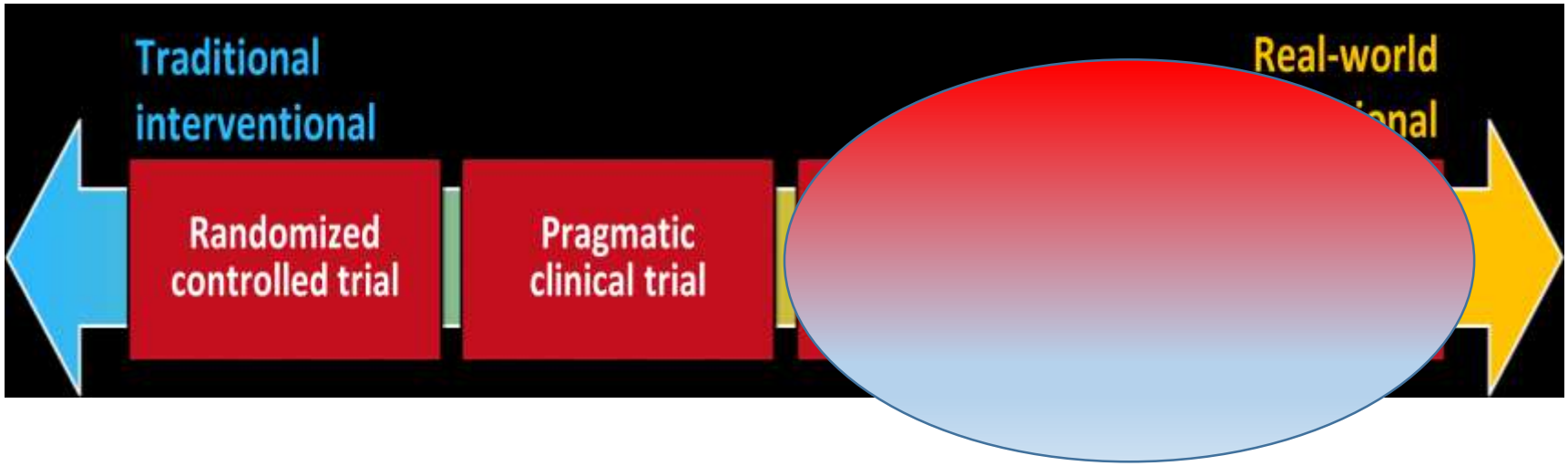


Figure 1. Schematic of the relationship between explanatory and pragmatic trials. The wide base of the pyramid depicts the relatively higher proportion of explanatory trials.

Patsopoulos et al. Dialogues in Clinical Neuroscience - Vol 13 . No. 2 . 2011

Studi Pragmatici

Richiedono **sforzi organizzativi e finanziari** confrontabili, se non superiori, agli RCT (la dimensione campionaria deve essere molto ampia per tenere conto della eterogeneità del campione)


Traditional
interventional
Real-world
Randomized
controlled trial
Pragmatic
clinical trial

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 VALIDITA' ESTERNA

ASPETTI CRITICI

- ✓Mancanza di randomizzazione
- ✓Mancanza di cecità
- ✓Popolazioni di confronto

Scarsa VALIDITA' INTERNA

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

Neurological Sciences
https://doi.org/10.1007/s10072-018-3610-0

ORIGINAL ARTICLE

The Italian multiple sclerosis register

Maria Trojano[1] · Roberto Bergamaschi[2] · Maria Pia Amato[3] · Giancarlo Comi[4] · Angelo Ghezzi[5] · Vito Lepore[6,7] · Maria Giovanna Marrosu[8] · Paola Mosconi[7] · Francesco Patti[9] · Michela Ponzio[10] · Paola Zaratin[10] · Mario Alberto Battaglia[10,11] · on behalf of the Italian Multiple Sclerosis Register Centers Group

Received: 19 June 2018 / Accepted: 16 October 2018

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

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 1 GRACE Checklist: Components and Response Guide

Component Item	Scoring as Fit for Purpose: Sufficient (+), Insufficient (-)
Data	
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 1 GRACE Checklist: Components and Response Guide

Component Item	Scoring as Fit for Purpose: Sufficient (+), Insufficient (-)
Methods	
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.

Source: Dreyer NA, Velentgas P, Westrich K, Dubois R. The GRACE checklist for rating the quality of observational studies of comparative effectiveness: a tale of hope and caution.[12]

GRACE = Good Research for Comparative Effectiveness; ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification; PRO = patient-reported outcomes.

Dreyer et al, JMCP 2016

Received: 21 July 2017 | Revised: 26 July 2017 | Accepted: 28 July 2017

DOI: 10.1002/pds.4297

WILEY

ORIGINAL REPORT

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

Marc L. Berger[1] | Harold Sox[2] | Richard J. Willke[3] | Diana L. Brixner[4] | Hans-Georg Eichler[5] | Wim Goettsch[6] | David Madigan[7] | Amr Makady[6] | Sebastian Schneeweiss[8] | Rosanna Tarricone[9] | Shirley V. Wang[8] | John Watkins[10] | C. Daniel Mullins[11]

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.

Recommendations for good procedural practices for Hypothesis Evaluating Treatment Effectiveness Studies

Berger et al, 2017

Treatment decisions in multiple sclerosis — insights from real-world observational studies

Maria Trojano[1], Mar Tintoré[2], Xavier Montalban[3], Jan Hillert[5], Tomas Kalincik[6], Pietro Iaffaldano[1], Tim Spelman[6], Maria Pia Sormani[8] and Helmut Butzkueven[6]

Abstract | The complexity of multiple sclerosis (MS) treatment means that doctors and decision-makers need the best available evidence to make the best decisions for patient care. Randomized controlled trials (RCTs) are accepted as the gold standard for assessing the efficacy and safety of any new drug, but conclusions of these trials do not always aid in daily decision-making processes. Indeed, RCTs are usually conducted in ideal conditions, so can measure efficacy only in restricted and unrepresentative populations. In the past decade, a growing number of MS databases and registries have started to produce long-term outcome data from large cohorts of patients with MS treated with disease-modifying therapies in real-world settings. Such observational studies are addressing issues that are otherwise difficult or impossible to study. In this Review, we focus on the most recently published observational studies designed to identify predictors of poor outcome and treatment response or failure, and to evaluate the relative and long-term effectiveness of currently used MS treatments. We also outline the statistical approaches that are most commonly used to reduce bias and limitations in these studies, and the challenges associated with the use of 'big MS data' to facilitate the implementation of personalized medicine in MS.

Trojano et al., Nature Reviews, Neurology, 2017

Box 1 | **Assessing the quality of real-world observational studies**

Several factors can affect the quality and reliability of real-world observational studies. For that reason, several parameters should be considered when assessing their quality. We used the following criteria to identify most of the studies included in this Review:

- Treatment details and primary outcomes are adequately recorded
- Primary outcomes are appropriate and objectively measured
- Confounders of treatment effect are adequately recorded and taken into account in the analysis
- The statistical methods for reducing bias are properly used
- Sensitivity analyses are used to explore residual confounding
- Study limitations are openly acknowledged and discussed

Trojano et al., Nature Reviews, Neurology, 2017

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

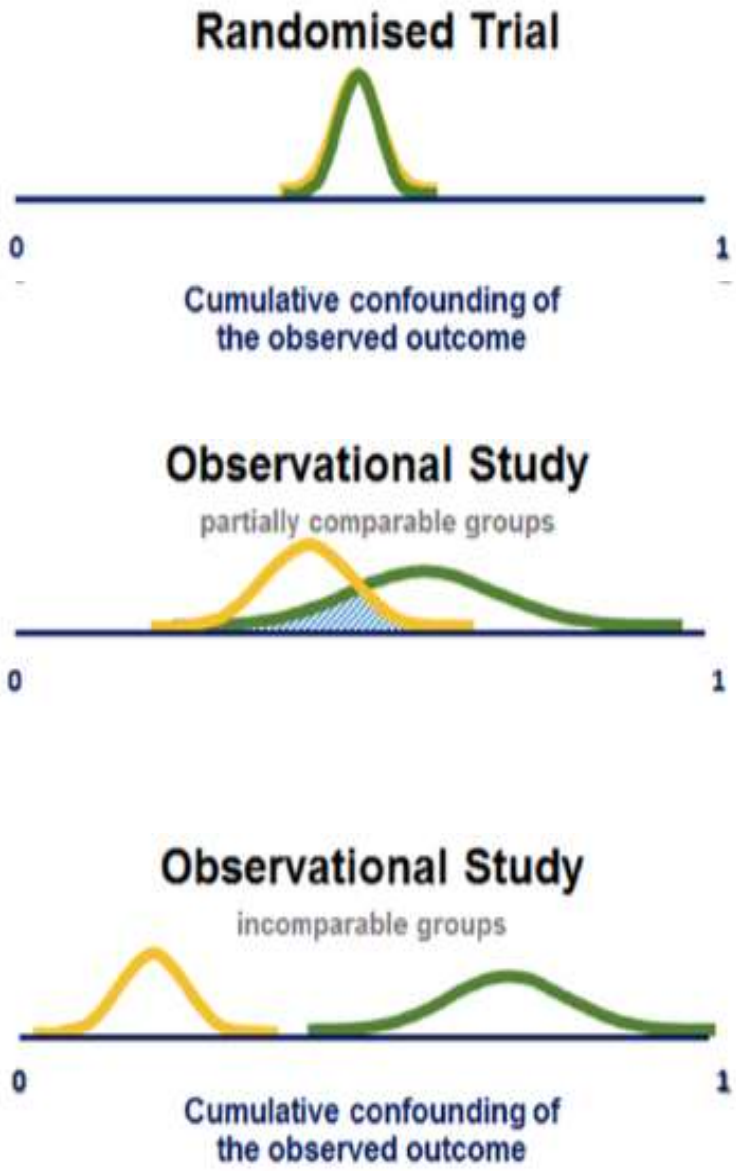
- Disporre di fonti di dati affidabili
- Ottimizzare la qualità dello studio
- **Utilizzare approcci statistici adeguati per:**
  - **minimizzare i bias**
  - rendere le popolazioni confrontabili

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

MRI: magnetic resonance imaging.

Kalincik and Butzkueven, Multiple Sclerosis Journal, 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 assignation 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

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, MD[3]

Ann Neurol, 2007

Modelli di regressione di Cox aggiustati per il

Propensity Score per stabilire l'omogeneità tra gruppi (trattati vs. non trattati)

Covariate incluse nel modello: età d'esordio, sesso, durata di malattia, numero di ricadute nell'ultimo anno, EDSS

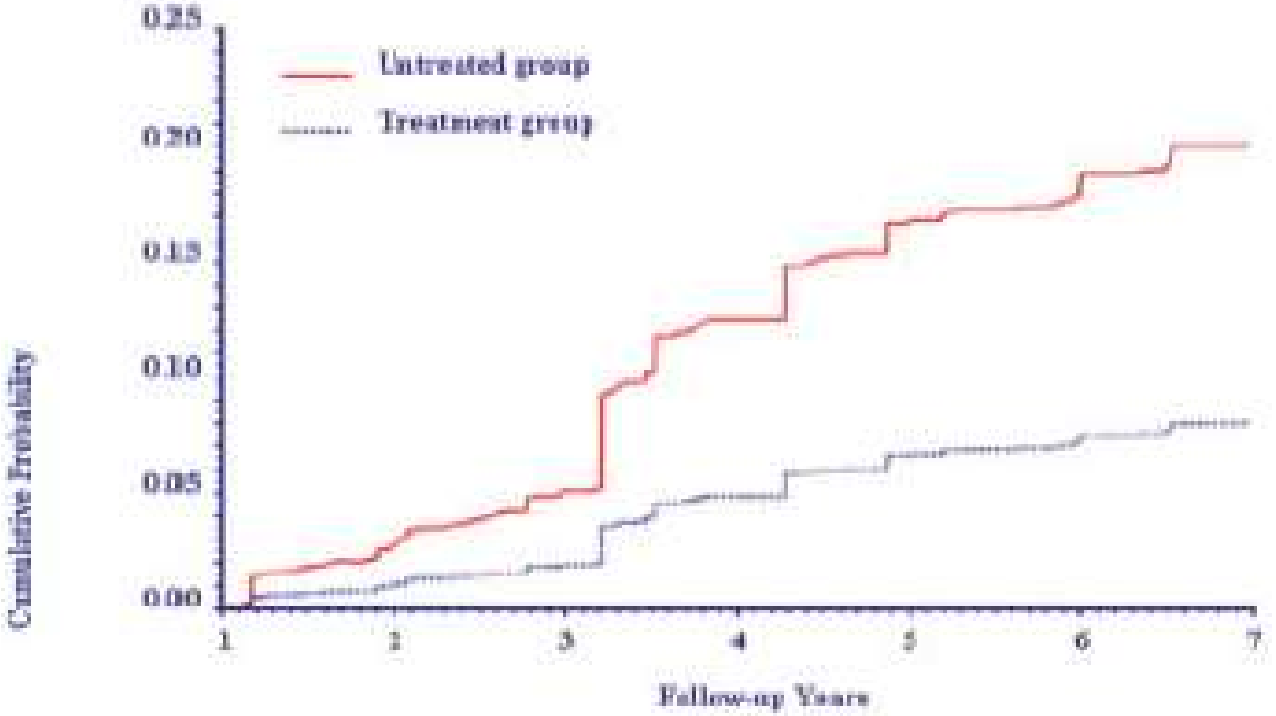


Fig 1. Propensity score–adjusted survival curves for end point: time from first visit to secondary progression. Cumulative probability represents the estimated proportion of patients reaching the end point. Solid line *indicates untreated group;* dashed line *indicates treatment group.*

Trojano et al, Ann Neurol 2007

COME MIGLIORARE LA 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
    - **Bayesian score**

ELSEVIER

Journal of the Neurological Sciences 189 (2001) 13–21

Journal of the Neurological Sciences

www.elsevier.com/locate/jns

Predicting secondary progression in relapsing–remitting multiple sclerosis: a Bayesian analysis

Roberto Bergamaschi [a,*], Carlo Berzuini [b], Alfredo Romani [a], Vittorio Cosi [a]

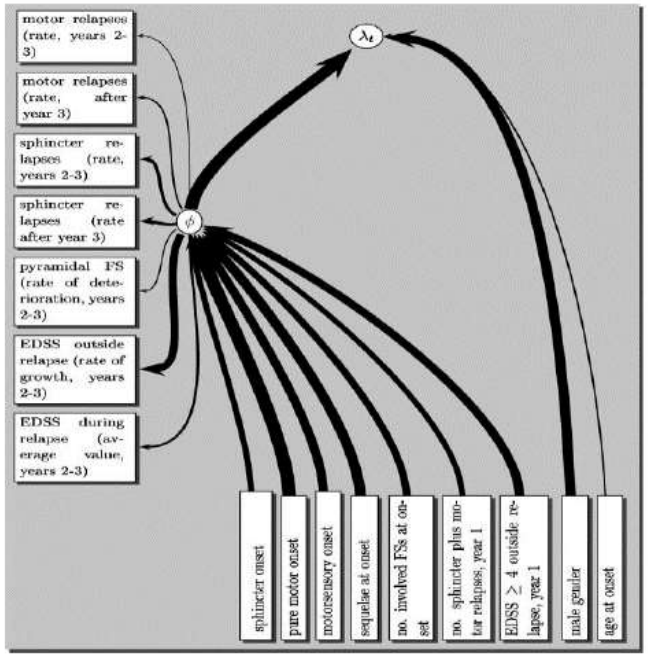


Table 1 Estimates of the Bayesian risk associated with early clinical predictors observed within 1 year of disease onset

	Mean LRR	Mean log LRR	95% CI
Age at onset (in decades)	1.05	0.05	1.02 to 1.09
Female sex	0.39	−1.07	0.17 to 0.78
Sphincter onset	2.98	0.93	1.10 to 6.10
Pure motor onset	2.11	0.62	0.90 to 4.20
Motor–sensory onset	2.40	0.81	1.15 to 4.41
Sequel after onset	1.76	0.52	1.04 to 2.88
Functional systems involved at onset	1.39	0.32	1.16 to 1.64
Sphincter plus motor relapses	2.10	0.71	1.56 to 2.89
EDSS ⩾4 outside relapse	2.28	0.44	0.40 to 6.50

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 Cosi

J Neurol Neurosurg Psychiatry 2007;**78**:757–759. doi: 10.1136/jnnp.2006.107052

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

Immunomodulatory therapies delay disease progression in multiple sclerosis

Multiple Sclerosis Journal, 2012

Roberto Bergamaschi[1], Silvana Quaglini[2], Eleonora Tavazzi[1], Maria Pia Amato[3], Damiano Paolicelli[4], Valentina Zipoli[3], Alfredo Romani[1], Carla Tortorella[4], Emilio Portaccio[3], Mariangela D'Onghia[4], Francesca Garberi[2], Valeria Bargiggia[1] and Maria Trojano[4]

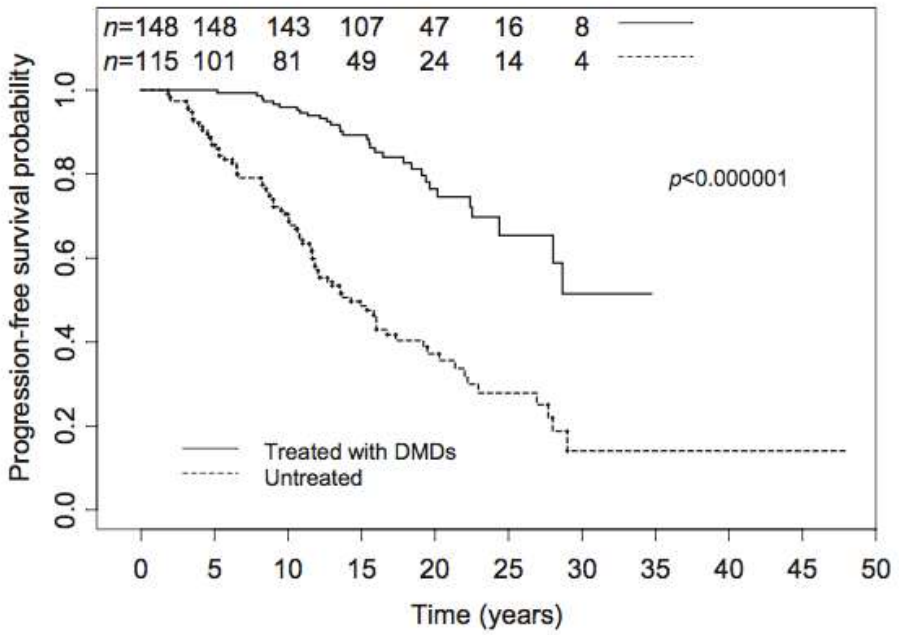


Figure 2. Secondary progression-free survival for high risk patients (fourth BREMS quartile) divided into two groups of patients: untreated and treated with DMDs.

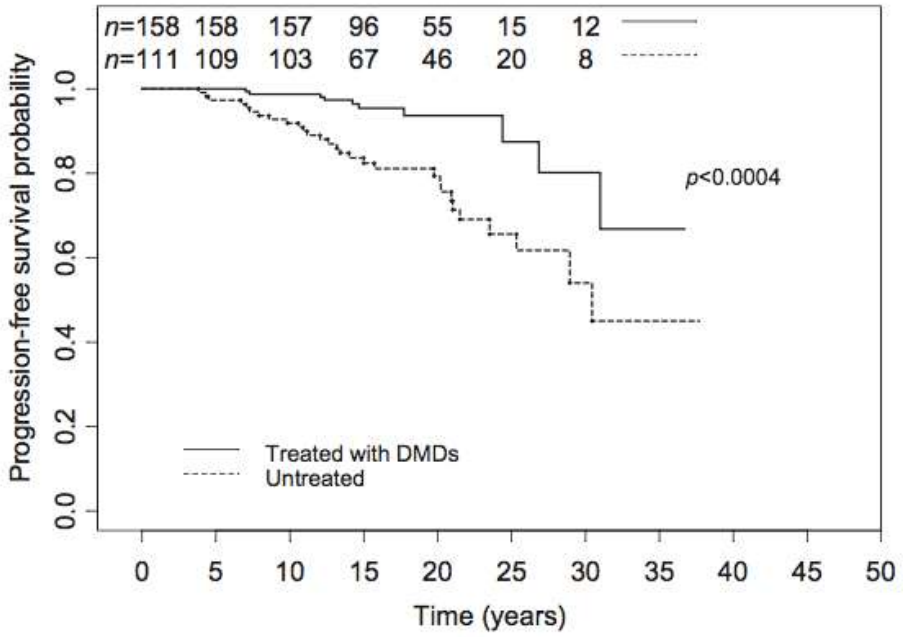


Figure 3. Secondary progression-free survival for low risk patients (first BREMS quartile) divided into two groups of patients: untreated and treated with DMDs.

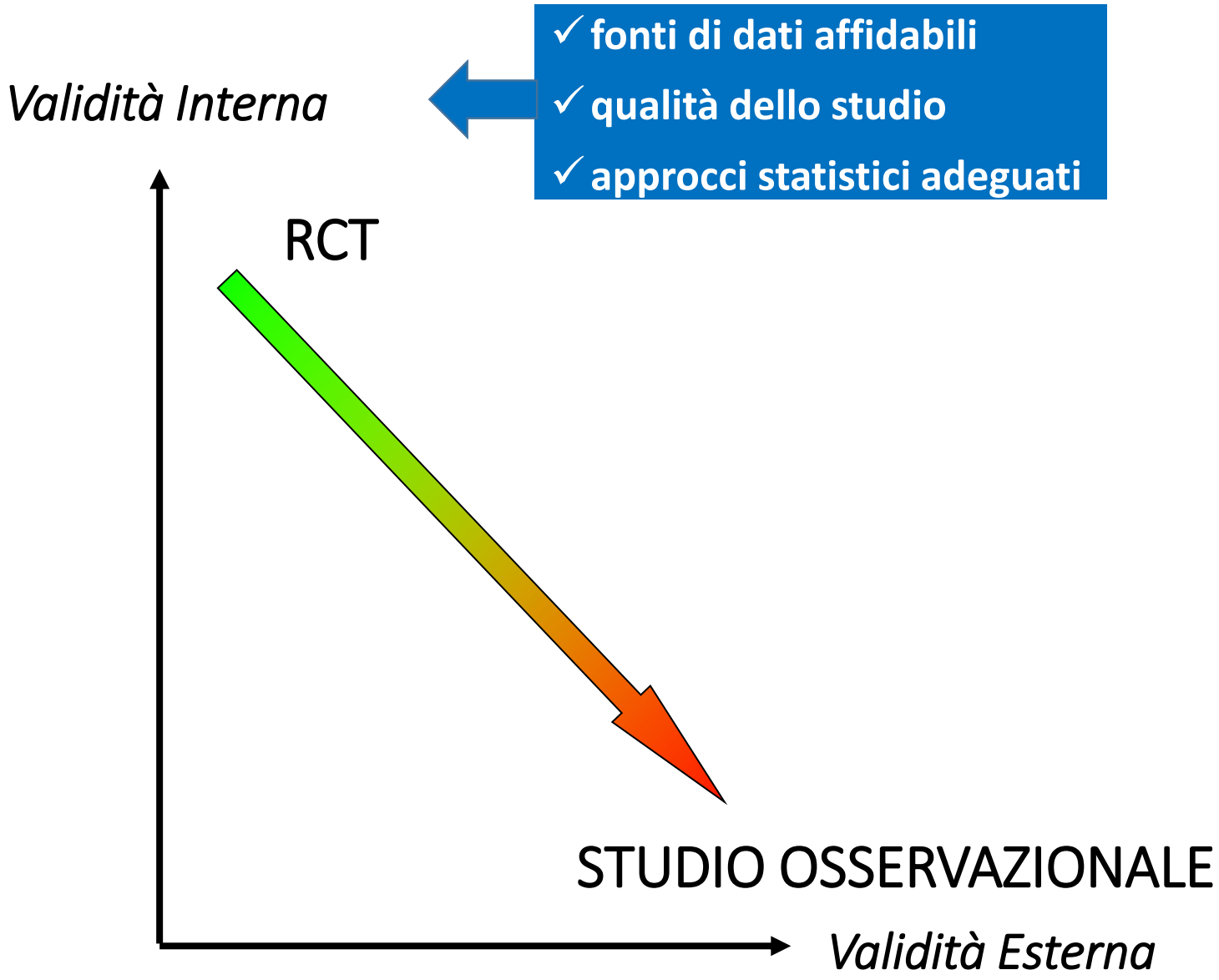
L’utilizzo di approcci statistici appropriati può consentire di migliorare il

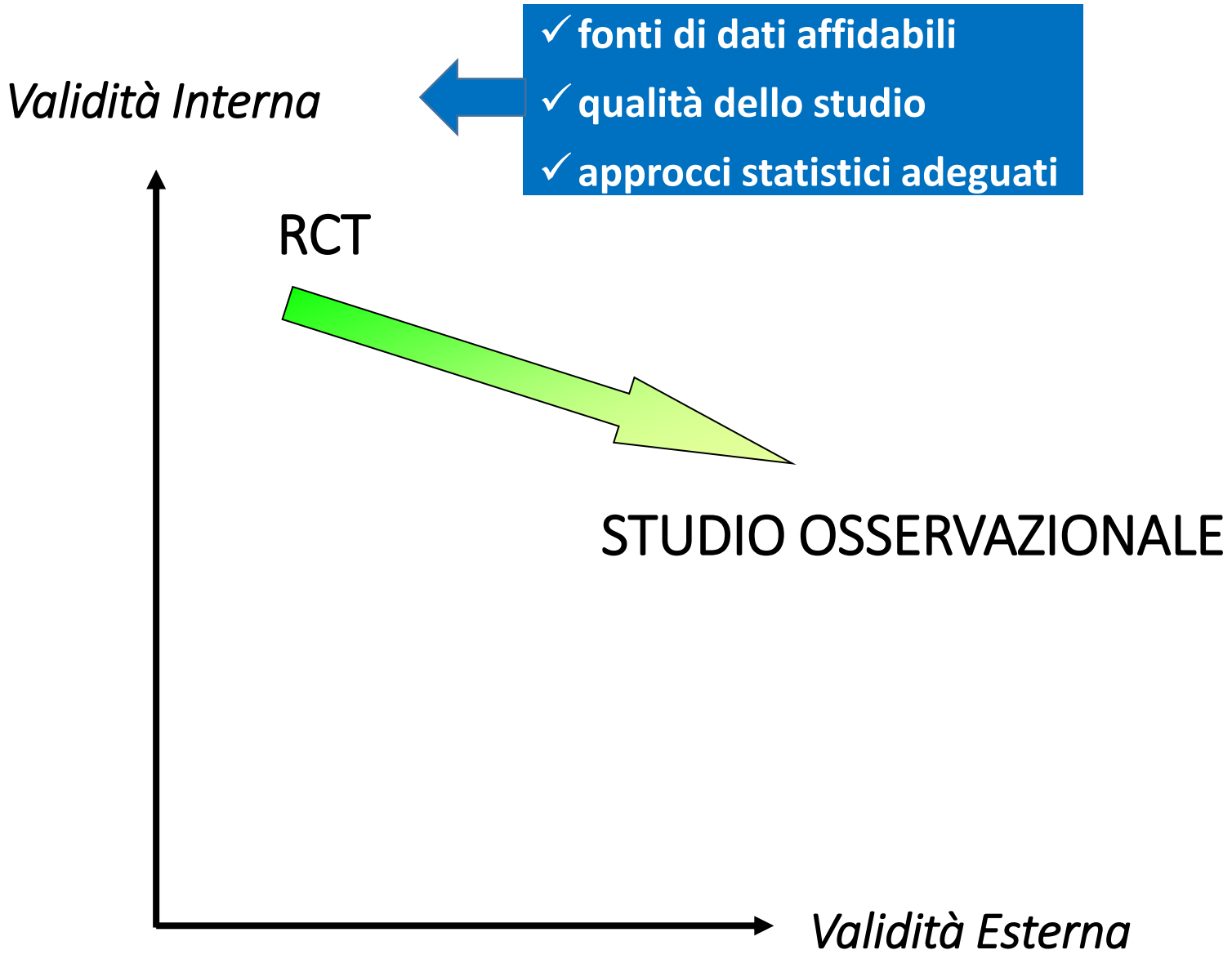
disegno degli studi osservazionali

“randomizzazione a posteriori”

Analisi di pazienti confrontabili per:

- propensione ad essere trattati (Propensity score)
- rischio di evoluzione sfavorevole (Bayesian score)


✓ fonti di dati affidabili
✓ qualità dello studio
✓ approcci statistici adeguati
Validità Interna
RCT
STUDIO OSSERVAZIONALE
Validità Esterna


✓ fonti di dati affidabili
✓ qualità dello studio
✓ approcci statistici adeguati
Validità Interna
RCT
STUDIO OSSERVAZIONALE
Validità Esterna

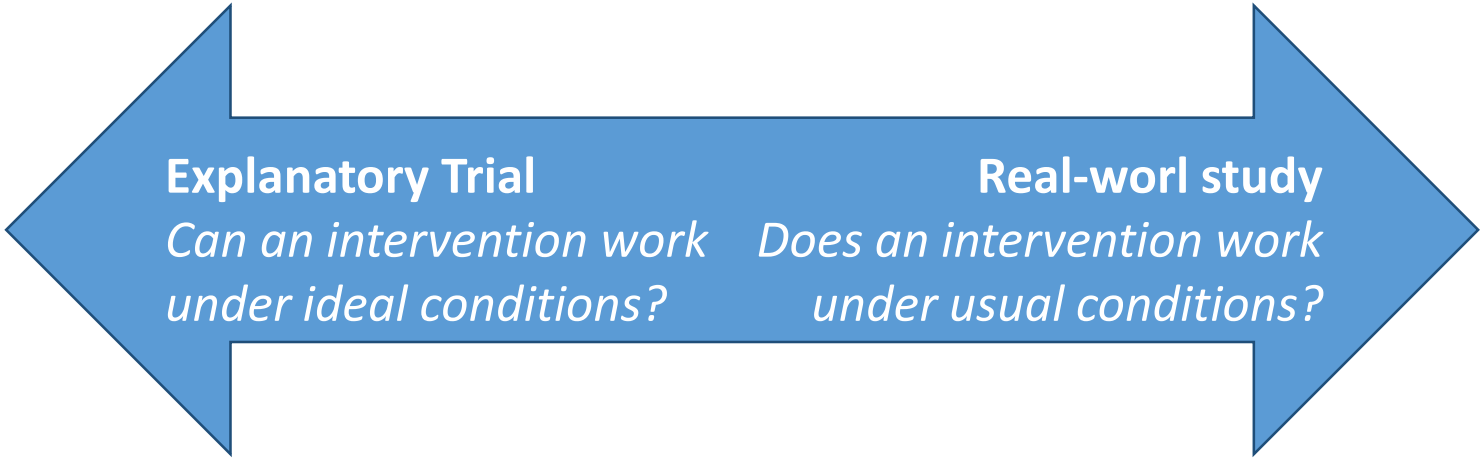
Circumstances under which observational studies are valuable and can be supportive of randomised clinical trials

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

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**



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. "

Archibald Cochrane *Effectiveness and Efficiency: Random Reflection on Health Services, Nuffield Provincial Trust, London, 1972*