

Approccio critico all'interpretazione dei trial cardiovascolari con risultati neutri o negativi

Alberto Aimo^{1,2}, Vincenzo Castiglione^{1,2}, Iacopo Fabiani², Paolo Morfino¹, Michele Emdin^{1,2}, Roberto Ferrari³, Luigi Tavazzi⁴, Aldo Pietro Maggioni⁵, Claudio Rapezzi^{3,4†}

¹Interdisciplinary Center for Health Sciences, Scuola Superiore Sant'Anna, Pisa

²UOC Cardiologia e Medicina Cardiovascolare, Fondazione Toscana Gabriele Monasterio, Pisa

³U.O. Cardiologia, Ospedale Universitario di Ferrara, Ferrara

⁴Maria Cecilia Hospital, GVM Care & Research, Cotignola (RA)

⁵Centro Studi ANMCO, Heart Care Foundation – Fondazione per il Tuo cuore, Firenze

Randomized controlled trials showing a significant benefit are met with enthusiasm because they may change the standard of care for patients who share the clinical and pathophysiologic characteristics of trial participants. Nonetheless, a well-designed and fully executed trial with neutral or negative findings also represents a critically important investigation deserving careful scientific scrutiny. In this paper we propose a 10-step approach to the interpretation of neutral or negative trials to exclude important methodological issues before concluding that the treatment really does not work. We will discuss this approach using the most classic trials of the past and some notable examples among superiority trials (mostly phase 3 trials) published over the last years.

Key words. Negative study; Neutral study; Randomized controlled trials; Study design.

G Ital Cardiol 2023;24(10):818-826

I trial clinici che mostrano un beneficio clinico significativo sono accolti con entusiasmo da ricercatori, medici e pazienti perché possono modificare lo standard di cura della condizione indagata. Tuttavia, uno studio ben progettato che dia risultati neutri (ovvero in cui non emergono differenze significative tra il trattamento sperimentale e la terapia standard) oppure negativi (cioè in cui il controllo sembra più efficace e/o l'intervento ha causato danni) merita un'attenta disamina per capire perché il suo endpoint primario non sia stato raggiunto. È fondamentale distinguere i trial che denotano una mancanza di efficacia dai trial in cui problemi nel disegno sperimentale impediscono di trarre conclusioni definitive circa l'efficacia del trattamento. Nel 2016, Pocock e Stone¹ hanno individuato alcune domande da porre quando un trial non raggiunge il suo endpoint primario. Un altro articolo nel quale alcuni punti attinenti al tema sono stati brevemente discussi è stato pubblicato recentemente². Nella rassegna attuale intendiamo affrontare nuovamente questo argomento e proporre un approccio più ampio in 10 passaggi all'interpretazione dei trial negativi/neutri (NN). Discuteremo i singoli passaggi facendo riferimento sia a trial "storici" che ad alcuni casi paradigmatici fra gli studi clinici, per lo più di fase 3, pubblicati

negli ultimi anni. Non prenderemo in considerazione studi di non inferiorità o di equivalenza, in quanto essi si basano su diverse assunzioni statistiche e richiederebbero una trattazione dedicata.

IPOTESI INTERPRETATIVE

1. Interruzione prematura

Un trial può essere interrotto prematuramente per vari motivi: evidenza di superiorità o chiara inefficacia del trattamento sperimentale in base a un'analisi preliminare, aumentato rischio di eventi avversi nei pazienti che assumono il trattamento, difficoltà di reclutamento, nuove informazioni sulla sicurezza e l'efficacia del trattamento sperimentale o decisione strategica dello sponsor qualora il trial sia sostenuto da finanziamenti privati e il contratto preveda la possibilità di interruzione dello studio in queste condizioni. Con l'eccezione dei casi in cui un'analisi preliminare mostra un'evidenza di superiorità, tutte le interruzioni premature comportano una riduzione della potenza statistica dello studio. Un esempio recente di interruzione prematura è rappresentato dallo studio di fase 4 LIFE (LCZ696 in Advanced Heart Failure)³, progettato per testare la superiorità di sacubitril/valsartan rispetto a valsartan in termini di tollerabilità, riduzione dei livelli biomarcatori cardiaci e benefici clinici in 400 pazienti con scompenso cardiaco e classe New York Heart Association (NYHA) IV. Come avvenuto per molti altri studi, il LIFE è stato interrotto prematuramente dopo il reclutamento di 335 pazienti a causa della pandemia COVID-19. In aggiunta, i pazienti arruolati hanno presentato una minore tolleranza al farmaco rispetto allo studio PARADIGM-HF (Prospective comparison of ARNI with ACEI to Determine Impact on Global Mortality and morbidity in Heart

© 2023 Il Pensiero Scientifico Editore

Ricevuto 22.06.2023; nuova stesura 18.07.2023; accettato 18.07.2023.

L.T. dichiara di aver ricevuto onorari da Servier in quanto membro di Trial Committee. A.P.M. dichiara di aver ricevuto finanziamenti da Bayer, AstraZeneca, Sanofi e Novartis. Gli altri autori dichiarano nessun conflitto di interessi.

[†]Deceduto.

Per la corrispondenza:

Dr. Alberto Aimo Scuola Superiore Sant'Anna e Fondazione Toscana Gabriele Monasterio, Piazza Martiri della Libertà 33, 56127 Pisa
e-mail: a.aimo@santannapisa.it, aimoalb@ftgm.it

Failure)^{3,4}. In effetti, 72 pazienti (18%) non hanno tollerato sacubitril/valsartan durante il breve periodo di run-in e 49 pazienti (29%) hanno interrotto il farmaco durante le 24 settimane di follow-up. A causa dell'interruzione prematura, 35 pazienti non hanno potuto concludere il periodo di follow-up. Nel complesso, il trial non ha mostrato differenze nel valore mediano del frammento N-terminale del propeptide natriuretico di tipo B (endpoint primario), né in diverse variabili cliniche considerate come endpoint secondari³. Chiaramente l'ottenimento di risultati neutri rappresenta un fallimento nella dimostrazione dell'ipotesi originale (in questo caso che sacubitril/valsartan sarebbe più efficace della terapia standard nei pazienti con scompenso cardiaco acuto avanzato), contrariamente agli studi di non inferiorità che vedono la loro realizzazione nel raggiungimento di risultati neutri, con intervalli di confidenza che non superino il margine di non inferiorità.

2. Potenza statistica insufficiente

Uno studio può avere una potenza statistica insufficiente perché il tasso di eventi osservati nel corso dello studio e/o l'entità dell'effetto sono risultati inferiori alle assunzioni iniziali. Un tasso di eventi inferiore al previsto può essere individuato durante il corso dello studio, anche senza bisogno di sapere quali pazienti assumono il trattamento sperimentale; gli effetti di un tasso di eventi minore del predetto possono essere limitati prolungando il periodo di reclutamento dei pazienti oppure la durata del follow-up. Questo è avvenuto ad esempio nello studio PARADISE-MI (Prospective ARNI versus ACE Inhibitor Trial to Determine Superiority in Reducing Heart Failure Events after Myocardial Infarction), che ha testato sacubitril/valsartan vs ramipril dopo un infarto miocardico in pazienti con evidenza di disfunzione sistolica e/o congestione polmonare senza storia di scompenso cardiaco. Il numero di pazienti è stato aumentato da 4650 a 5650 sulla base di una valutazione *ad interim* del tasso di eventi⁵. Al contrario, nello studio FLOWER-MI (Flow Evaluation to Guide Revascularization in Multivessel ST-Elevation Myocardial Infarction), l'assenza di un'analisi *ad interim* non ha permesso di rilevare la presenza di una netta discrepanza fra il tasso atteso dell'endpoint primario a 1 anno (10% con la strategia di rivascularizzazione coronarica guidata dalla riserva frazionale di flusso vs 15% con la strategia guidata dalla sola angiografia) rispetto al tasso osservato (6% vs 4%)⁶.

Quando l'entità dell'effetto è inferiore al valore atteso, la potenza statistica calcolata stimando un effetto più consistente del trattamento finisce per essere insufficiente. Ad esempio, il trial DOREMI (Dobutamine Compared with Milrinone), che ha confrontato milrinone e dobutamina in pazienti con shock cardiogeno, è stato disegnato ipotizzando una notevole differenza di efficacia delle due terapie, con un tasso dell'endpoint primario del 20% inferiore nel gruppo trattato con milrinone, a fronte di una differenza effettiva del 5%⁷. In alcuni trial, in particolare negli studi di prevenzione, il periodo di follow-up potrebbe essere troppo breve per consentire di rilevare una differenza significativa fra i bracci di trattamento. Ad esempio, lo studio ACCELERATE (Assessment of Clinical Effects of Cholesteryl Ester Transfer Protein Inhibition with Evacetrapib in Patients at a High Risk for Vascular Outcomes) ha valutato evacetrapib, inibitore della proteina di trasferimento degli esteri del colesterolo (*cholesteryl ester transfer protein*, CETP), in soggetti con un rischio cardiovascolare aumentato. Questo studio è stato in-

terrotto dopo un periodo mediano di 26 mesi, rispetto ai 3 anni previsti, per la mancata evidenza di efficacia⁸. È interessante notare che lo studio REVEAL (Randomized Evaluation of the Effects of Anacetrapib through Lipid Modification), che ha valutato l'inibitore della CETP anacetrapib ed è stato condotto su una popolazione simile, è risultato positivo dopo un periodo mediano di 4.1 anni⁹.

L'adeguata potenza statistica di uno studio è teoricamente facile da verificare controllando il numero di pazienti arruolati e il tasso di eventi rispetto alle stime iniziali. In alcuni casi questa analisi può risultare complessa per vari motivi:

a) Perdita di potenza a causa del passaggio di pazienti fra i bracci sperimentali durante il corso dello studio (cross-over): un esempio è fornito dal trial STICH (Surgical Treatment for Ischemic Heart Failure), dove il 17% dei pazienti randomizzati alla terapia medica in ultima analisi ha subito un intervento di bypass aortocoronarico¹⁰. Un altro esempio è rappresentato dallo studio CABANA (Catheter Ablation vs Antiarrhythmic Drug Therapy for Atrial Fibrillation), dove molti pazienti nel braccio di terapia medica sono stati sottoposti ad ablazione durante gli 8 anni di follow-up¹¹. Come regola generale, quanto più è lungo lo studio, tanto maggiore è il tasso di cross-over, in particolare negli studi interventistici, in cui il rischio di cross-over è ineliminabile. Per lo più i pazienti (o i medici curanti di riferimento) che mutano il braccio di appartenenza del paziente lo fanno perché sono insoddisfatti del trattamento in atto, per cui il cross-over include nello studio una indebita selezione dei pazienti appartenenti a un braccio o all'altro scomponendone l'equilibrio e annullando il senso della randomizzazione. Tale presupposto si accompagna inevitabilmente ad una serie di ostacoli pratici, etici e logistici che non sempre possono essere facilmente risolti, rendendo tali studi meno rigorosi in termini scientifici. I trial con un alto tasso di cross-over spesso danno risultati neutri all'analisi "intention to treat" (ovvero l'analisi che si basa sugli intenti iniziali di trattamento, ignorando i cross-over avvenuti e quindi i trattamenti effettivamente somministrati), mentre possono dare risultati positivi all'analisi per protocollo ovvero l'analisi che raggruppa i pazienti in base alla terapia effettivamente erogata dal momento del cross-over. D'altro canto, uno studio interpretato con un'analisi per protocollo non è più assimilabile a un trial randomizzato. Una possibile strategia per ovviare a questo problema è il disegno degli studi, utilizzando un approccio bayesiano anziché frequentista. Secondo l'approccio bayesiano, l'inferenza è basata sul concetto di probabilità che può essere applicato sia alle osservazioni che, contrariamente all'approccio frequentista, alle ipotesi (la probabilità che può essere assegnata a un'ipotesi è un valore qualsiasi nell'intervallo tra 0 e 1). Secondo la metodica frequentista, invece, un'ipotesi è una proposizione vera o falsa (probabilità pari a 0 o 1). Avere a disposizione un range di probabilità potrebbe conferire dei benefici anche nell'arruolamento in uno studio clinico. Per evitare che un campione statistico sia troppo grande o troppo piccolo, una soluzione proposta è quella di un disegno tipo "Bayesian response adaptive randomization", anche se esistono ancora numerose criticità metodologiche non risolte¹² e la complessa necessità di predefinire, già nel protocollo iniziale, le metodologie adattative eventualmente utilizzate nello studio.

b) Numero di eventi incrementato da eventi minori o decisioni mediche, che sono spesso abbastanza uniformemente distribuiti tra i due gruppi di studio, diluendo così le differenze. L'endpoint primario dello studio PROactive (Prospective Pioglitazone Clinical Trial in Macrovascular Events) comprendeva sindrome coronarica acuta, rivascolarizzazione coronarica chirurgica o percutanea, rivascolarizzazione o amputazione degli arti inferiori. Il valore di p per questo endpoint è risultato non significativo ($p=0.095$). Viceversa, la combinazione di morte, infarto miocardico o ictus (un endpoint di maggiore rilevanza clinica perché comprende solo eventi spontanei e non decisionali come rivascolarizzazioni o amputazioni) si è verificato meno frequentemente nei pazienti in terapia con pioglitazone ($p=0.03$)¹³.

3. Problemi nella conduzione dello studio

In alcuni casi, problemi logistici legati alla popolazione reclutata o alla conduzione del trial possono avere un'importanza cruciale. L'affidabilità degli sperimentatori è fondamentale per avere dati attendibili. Al di là delle vere e proprie frodi scientifiche, l'arruolamento non corretto dei pazienti, l'attribuzione errata o la mancata segnalazione degli eventi, insieme all'inerzia complessiva dello sperimentatore, la scarsa aderenza del paziente, la non attenta vigilanza del centro di coordinamento o l'ingerenza eccessiva degli osservatori esterni possono portare a dati non attendibili o difficilmente interpretabili. Un esempio ormai classico è lo studio TOPCAT (Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist)¹⁴, che ha valutato la terapia con spironolattone in pazienti con scompenso cardiaco a frazione di eiezione preservata e globalmente non ha mostrato un miglioramento della prognosi nei pazienti in terapia con spironolattone. È stato osservato che i pazienti arruolati in Russia o Georgia hanno presentato un tasso di eventi notevolmente inferiore rispetto ai pazienti reclutati nelle Americhe¹⁵. Alla base ci sono probabilmente differenze nella selezione dei pazienti (con un più frequente arruolamento in Russia e Georgia di pazienti la cui diagnosi di scompenso cardiaco era più discutibile) e nel rispetto delle procedure sperimentali. Per quanto riguarda il secondo punto, sui campioni ematici dei pazienti provenienti da Russia e Georgia sono stati riscontrati bassi livelli di canrenone, metabolita dello spironolattone, il che suggerisce una scarsa aderenza al trattamento da parte dei pazienti arruolati¹⁶. Gravi errori nella conduzione di uno studio possono essere emendati solo con la ripetizione dello studio.

4. Problemi nella scelta, definizione o interpretazione degli endpoint

Anche un endpoint apparentemente ben definito risente molto della soggettività dello sperimentatore e dell'oggettività dell'osservatore esterno. In effetti, l'attribuzione degli eventi può essere talora difficile, in particolare per osservatori indipendenti che valutano solo la documentazione clinica. Nello studio CHARM (Candesartan in Heart Failure: Assessment of Reduction in Mortality and Morbidity)-Preserved, che ha valutato la terapia con candesartan per il trattamento dello scompenso cardiaco a frazione di eiezione preservata, l'endpoint primario (morte cardiovascolare o ospedalizzazione per scompenso cardiaco) si è verificato in 324 e 372 pazienti nel gruppo randomizzato a candesartan e placebo, rispettivamente, secondo gli investigatori dello studio (hazard ratio [HR] 0.85, intervallo di confidenza [IC] 95% 0.73-0.98; $p=0.028$), ma in

333 vs 366 pazienti secondo gli osservatori esterni (HR 0.89, IC 95% 0.77-1.03; $p=0.051$)¹⁷, rendendo il trial neutro se interpretato in base al valore soglia di $p=0.05$.

L'appropriata definizione degli endpoint può avere un'importanza determinante. Nello studio OVERTURE (Omapatrilat Versus Enalapril Randomized Trial of Utility in Reducing Events), che ha confrontato omapatrilat vs enalapril in pazienti con scompenso cardiaco, i risultati hanno soddisfatto i criteri per la non inferiorità di omapatrilat vs enalapril, ma non per la superiorità relativamente all'endpoint primario (combinazione di morte per tutte le cause e ospedalizzazione per scompenso: HR 0.94, IC 95% 0.86-1.03; $p=0.187$)¹⁸. D'altro canto, un'analisi *post hoc* che ha utilizzato la stessa definizione di ospedalizzazione per scompenso cardiaco del trial SOLVD (Studies of Left Ventricular Dysfunction) ha mostrato che i pazienti trattati con omapatrilat hanno un rischio dell'11% inferiore ($p=0.012$)¹⁸.

Alcune definizioni degli endpoint si possono prestare a varie interpretazioni. Nello studio RELAX-AHF-2 (Relaxin in Acute Heart Failure 2), che ha valutato il vasodilatatore serelaxina nello scompenso cardiaco acuto, la definizione dell'endpoint "peggioramento di scompenso a 5 giorni" era piuttosto soggettiva (segni di peggioramento o sintomi di scompenso che richiedono un'intensificazione del trattamento), inoltre l'endpoint non è stato valutato da osservatori indipendenti¹⁹. Lo studio ATPCI (Efficacy and Safety of Trimetazidine in Patients with Angina Pectoris Having Been Treated by Percutaneous Coronary Intervention) ha fornito un altro esempio di endpoint composito con elementi di ambiguità, vale a dire "ricovero ospedaliero per eventi cardiaci"²⁰.

5. Selezione subottimale di endpoint compositi

Il numero di eventi determina la potenza dello studio. Raggiungere una potenza adeguata è più facile quando l'endpoint primario include molteplici elementi con una fisiopatologia in parte indipendente, più o meno sensibile agli effetti del trattamento. Un endpoint composito dovrebbe però avere rilevanza clinica e includere solamente elementi sovrapponibili che possano risentire del trattamento e la cui riduzione è parimenti importante sotto il profilo medico. Quest'ultimo requisito spesso non è soddisfatto. Nel trial COMMANDER-HF (A Study to Assess the Effectiveness and Safety of Rivaroxaban in Reducing the Risk of Death, Myocardial Infarction, or Stroke in Participants with Heart Failure and Coronary Artery Disease Following an Episode of Decompensated Heart Failure), il rationale per includere la morte non cardiovascolare nell'endpoint primario era abbastanza debole, anche se tale scelta poteva essere giustificabile nell'ottica di valutare le morti per causa emorragica o embolica²¹. Nel trial GALILEO (Global Study Comparing a Rivaroxaban-based Antithrombotic Strategy to an Antiplatelet-based Strategy after Transcatheter Aortic Valve Replacement to Optimize Clinical Outcomes), che ha analizzato la terapia con rivaroxaban dopo sostituzione valvolare aortica per via percutanea, non sono stati valutati la morte cardiovascolare e gli eventi tromboembolici arteriosi, piuttosto la morte per tutte le cause e gli eventi tromboembolici arteriosi o venosi²². Anche quando l'endpoint combinato viene scelto con accortezza, la sua composizione potrebbe avere una rilevanza cruciale nel determinare i risultati complessivi. Lo studio GISSI (Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto miocardico)-Prevenzione ha dato risultati

positivi, ma l'IC 95% per l'endpoint primario (combinazione di morte, infarto non fatale e ictus) era molto vicino alla neutralità (n-3 acidi grassi polinsaturi vs placebo: rischio relativo [RR] 0.90, IC 95% 0.81-0.99). È interessante notare che il trattamento ha ridotto maggiormente il rischio di morte per tutte le cause rispetto agli eventi non fatali (RR 0.86, IC 95% 0.76-0.97 e RR 0.99, IC 95% 0.84-1.16, rispettivamente) ed è risultato ancora più efficace sulla morte cardiaca improvvisa (RR 0.67, IC 95% 0.49-0.92; $p=0.010$)²³, suggerendo la probabile molteplicità degli effetti del farmaco in un contesto fisiopatologico complesso, ancora in parte oscuro. Lo studio CAPRICORN (Carvedilol Post-Infarct Survival Control in Left Ventricular Dysfunction) è risultato neutrale in relazione all'endpoint primario costituito da morte oppure ospedalizzazione per cause cardiovascolari (HR 0.92, IC 95% 0.80-1.07; $p=0.296$). Tuttavia, la terapia con carvedilolo ha determinato una riduzione significativa del rischio di morte per tutte le cause (HR 0.77, IC 95% 0.60-0.98; $p=0.031$)²⁴.

In alcuni studi è stato un elemento clinicamente meno rilevante a portare lo studio verso un risultato finale positivo. Un esempio è rappresentato dagli studi EMPEROR-Preserved (Empagliflozin Outcome Trial in Patients with Chronic Heart Failure and a Preserved Ejection Fraction) e EMPEROR-Reduced (Empagliflozin Outcome Trial in Patients with Chronic Heart Failure and a Reduced Ejection Fraction), nei quali l'endpoint primario è stato costituito da una combinazione di morte per cause cardiovascolari e ospedalizzazione per scompenso; gli effetti positivi di empagliflozin sull'ospedalizzazione per scompenso erano risultati più evidenti rispetto a quelli sulla morte per cause cardiovascolari^{25,26}. Per aggiungere un ulteriore grado di complessità, alcuni studi condotti in pazienti con una prognosi moderatamente grave, come CHARM-Preserved nello scompenso cardiaco a frazione di eiezione preservata, sarebbero stati positivi se gli eventi ricorrenti non fatali (ricoveri per scompenso) fossero stati inclusi nell'endpoint primario¹⁷.

6. Trappola nosografica

La "trappola nosografica" è la situazione in cui la popolazione studiata è eterogenea, anche perché situazioni cliniche diverse vengono accomunate in definizioni diagnostiche approssimative, e il trattamento non ha la stessa efficacia in tutta la popolazione. Ad esempio, farmaci anti-fibrotici o farmaci che agiscono sulla via di segnalazione del guanosin-monofosfato ciclico potrebbero rivelarsi efficaci in alcuni pazienti con scompenso a frazione di eiezione preservata ma non in altri^{27,28}. La valutazione di strategie mirate in una popolazione ampia ed eterogenea (ad esempio rivaroxaban in pazienti con scompenso cardiaco²¹ o gli acidi grassi n-3 in pazienti ad alto rischio cardiovascolare)²⁹ potrebbe fornire ulteriori esempi di trappole nosografiche. In questi casi il rischio competitivo di eventi non associati agli obiettivi prioritari dello studio è molto elevato.

Lo scompenso a frazione di eiezione preservata rappresenta una condizione sistemica e altamente eterogenea, che include un'ampia gamma di quadri fisiopatologici e clinici. La possibilità di identificare selettivamente sottogruppi con caratteristiche prognostiche e terapeutiche peculiari permetterebbe l'implementazione di terapie specifiche e maggiormente efficaci per ciascun gruppo di pazienti. L'alta variabilità clinica dell'insufficienza cardiaca a frazione di eiezione preservata rende ardua la progettazione di trial clinici così come predirne l'efficacia, suggerendo che un approccio "one size fits all" non sia metodologicamente adeguato. Lo sviluppo

di tecnologie computazionali che consentano di analizzare grandi quantità di dati (es. provenienti dall'imaging) potrebbe rivelarsi un'utile risorsa per identificare forme della malattia che seguono dei pattern comuni³⁰.

Negli studi "adattativi", emersi in particolare durante la recente pandemia, alcune trappole possono essere previste ed entro certi limiti prevenute o contenute. Nel caso del COVID-19 la trappola era strutturale: una diagnosi omnicomprensiva, con eziologia agevolmente verificabile (infezione da coronavirus) che però poteva manifestarsi con pressoché tutti i sintomi possibili, anche in rapporto alla logistica preferenziale del virus nell'individuo infetto e/o alla risposta individuale al virus. Lo vediamo con chiarezza meglio ora, che si sta cercando di fenotipizzare il "long COVID", a seguito della recente pubblicazione dell'Organizzazione Mondiale della Sanità³¹. I sintomi possibili includono manifestazioni cliniche così varie e diverse da rendere praticamente inutile utilizzarla: non ne esce un quadro clinico e fisiopatologico interpretabile e quindi un programma razionale di ricerca, non orientato unicamente sui sintomi. Durante la pandemia COVID-19 una trappola è stata l'imprevedibile variabilità nel tempo, anche breve, del tasso di inflazione locale dell'epidemia (e quindi dell'arruolamento negli studi in corso) o delle sue caratteristiche prevalenti che potevano inficiare obiettivi, numerosità e tempi previsti degli studi. Da queste difficoltà si è rinviato, fino a dominare lo scenario della ricerca clinica, il modello adattativo, che ha consentito di condurre molti studi in parallelo da centri esperti organizzati in vaste reti, per lo più nazionali (alcuni allargati poi a livello internazionale), con una metodologia uniforme, mantenuta attiva cambiando gli obiettivi, un farmaco dopo l'altro, che se efficaci venivano direttamente inclusi nella pratica clinica anche senza attendere mesi o anni di verifica. Spesso la pratica esplorativa iniziale era strutturata in due tempi: il primo con pochi centri molto qualificati, anche tecnologicamente, per identificare in breve tempo quali farmaci fossero promettenti e quali no, avendo la possibilità di aumentare la numerosità del campione per sostanze sperimentate che potevano diventare farmaci. Un secondo tempo allargato alla rete connessa, per testare la reale rilevanza terapeutica del farmaco testato nel mondo reale. Questa, *in nuce*, è stata la "fabbrica" delle terapie per il COVID-19, con una caratteristica specifica dei protocolli: larga disponibilità a verifiche degli effetti clinici durante gli studi (*ad interim*) da parte di Comitati Etici che verificavano in particolare le solidità delle assunzioni preliminari, suggerendo poi ai Comitati esecutivi (che ignoravano i dati) se continuare, modificare o interrompere gli studi, secondo i principi dei trial adattativi o "a piattaforma"³² coerenti con l'urgenza imposta dal tentativo di limitare i morti da COVID-19. In piena pandemia era necessario vigilare per non perdere tempo. Nonostante il dispiego delle forze in campo, nella lotta del mondo al COVID-19 la farmacologia clinica non ha avuto un ruolo primario, e il "mondo scientifico" che si è volentiersamente buttato sulla ricerca clinica, tanto meno. Tuttavia, in qualche caso, ha mostrato agilità organizzativa, rapidità di collegamenti e intese, contenibile competitività tra reti e paesi e capacità di individuare rapidamente strumenti metodologici adatti all'emergenza: i trial adattativi, appunto, e subito dopo i trial pragmatici. Ma soprattutto si è imposta la digitalizzazione dei sistemi sanitari mettendo in evidenza la ricchezza e la rapida disponibilità dei dati consentite dai sistemi elettronici nel monitoraggio quotidiano dello stato di salute del pianeta.

7. Assenza di prerequisiti fisiopatologici per l'efficacia del trattamento

I risultati neutri o negativi di alcuni studi potrebbero essere attribuiti, almeno in parte, alla scelta del campione, rappresentato da popolazioni in cui l'efficacia del trattamento è improbabile. Lo studio MITRA-FR (Percutaneous Repair with the MitraClip Device for Severe Functional/Secondary Mitral Regurgitation) ha arruolato soggetti con un grado di insufficienza mitralica "proporzionale" alla gravità della disfunzione del ventricolo sinistro³³. Questi pazienti avevano una minore probabilità di rispondere alla riparazione percutanea della valvola mitrale con dispositivo Mitraclip rispetto a quelli in cui il grado di insufficienza era "sproporzionato" rispetto all'entità della disfunzione ventricolare sinistra, manifestando una componente disfunzionale propria della valvola mitralica, come nello studio COAPT (Cardiovascular Outcomes Assessment of the MitraClip Percutaneous Therapy for Heart Failure Patients with Functional Mitral Regurgitation)^{34,35}. Nello studio COAPT, in cui i criteri di inclusione erano qualitativi (entità del rigurgito di grado 3+ o 4+) e differenti da quelli dello studio MITRA-FR (area effettiva dell'orifizio rigurgitante >20 mm²), il valore medio dell'area rigurgitante era di 45 mm² e di 31 mm² nel secondo. Nell'insufficienza mitralica secondaria non degenerativa, l'efficacia dell'intervento probabilmente è proporzionale all'entità dell'orifizio rigurgitante valutato al top della terapia medica ottimale per lo scompenso. Infatti, una riduzione totale dell'endpoint composto di morte per tutte le cause e scompenso cardiaco nel gruppo trattato con riparazione percutanea rispetto alla sola terapia medica (HR 0.53, IC 95% 0.44-0.64) rende solidi i dati dello studio COAPT esteso³⁶. Gli studi in proposito si susseguono e le informazioni emergenti orientano verso i criteri COAPT come riferimenti per l'indicazione alla procedura interventistica sulla mitrale nei pazienti con scompenso cardiaco, ventricolo sinistro dilatato e frazione di eiezione ridotta³⁷. È quindi plausibile che alcuni trial si siano rivelati neutri o negativi perché hanno arruolato pazienti per i quali era improbabile che il trattamento fosse efficace (oppure in cui la numerosità relativa dei pazienti appropriati fosse insufficiente).

Lo studio CIRT (Cardiovascular Inflammation Reduction Trial) è stato interrotto prematuramente perché il farmaco antinfiammatorio metotrexato non ha dimostrato alcun segno di efficacia in pazienti con aterosclerosi³⁸. Va sottolineato che i criteri di inclusione dello studio CIRT non prevedevano la documentazione di attivazione infiammatoria³⁸, contrariamente allo studio positivo CANTOS (Canakinumab Antiinflammatory Thrombosis Outcome Study), in cui terapie antinfiammatorie sono state somministrate a pazienti con elevati livelli di proteina C-reattiva ad alta sensibilità³⁹. Nello studio EVEREST (Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study With Tolvaptan)⁴⁰, il tolvaptan, antagonista orale del recettore V₂ della vasopressina, si è rivelato in grado di aumentare il sodio sierico nei pazienti con iponatremia⁴¹ ed è ragionevolmente più efficace nei pazienti con scompenso acuto e iponatremia. Tuttavia, i pazienti con livelli di sodio sierico basale ≤137 mEq/l rappresentavano appena il 27% dell'intera coorte arruolata. Nel trial PROTECT (Effects of Rolofylline, a New Adenosine A₁ Receptor Antagonist on Symptoms, Renal Function, and Outcomes in Patients With Acute Heart Failure), solo il 14% dei pazienti aveva una clearance della creatinina <30 ml/min⁴² sebbene ci fossero prerequisiti teorici per prevedere un maggiore beneficio della rolofillina (un antagonista del recettore A₁ dell'adenosina) nei pazienti con severa disfunzione renale^{43,44}.

8. Cancellazione algebrica

In pratica è un esempio estremo della categoria precedente, caratterizzata dall'eterogeneità della popolazione studiata. La cancellazione algebrica si verifica quando i risultati di uno studio sono positivi in alcuni pazienti e negativi in altri, oltre i confini della probabilità casuale. Un esempio "storico" è rappresentato dallo studio DIG (Digitalis Investigation Group), che è stato globalmente neutro, con un RR di mortalità pari a 0.99 (IC 95% 0.91-1.07; p=0.80)⁴⁵. I pazienti con frazione di eiezione <45% e in ritmo sinusale con concentrazioni sieriche di digossina da 0.5 a 0.8 ng/ml mostravano un tasso di mortalità del 6.3% inferiore rispetto al placebo (IC 95% 2.1-10.5%), mentre i pazienti con livelli di digossina da 0.9 a 1.1 ng/ml e in particolare quelli con livelli ≥1.2 ng/ml esibivano una mortalità più elevata (aumento del 2.6%, IC 95% 3.0-8.3% e 11.8%, IC 95% 5.7-18.0%, rispettivamente)⁴⁶.

Nonostante lo studio DIG sia un caso limite, le analisi di sottogruppo mostrano spesso effetti ampiamente diversi tra sottogruppi differenti, ognuno dei quali richiederebbe numerosità molto diverse per testare ipotesi di superiorità (o inferiorità). Va ricordato infatti che la potenza statistica di un trial viene tipicamente calcolata solo per l'endpoint primario, rendendo tutte le altre analisi secondarie e le analisi di sottogruppo non confermate, al massimo generatrici di ipotesi.

Un caso particolare di cancellazione algebrica è rappresentato dallo studio ASCEND (A Study of Cardiovascular Events in Diabetes), che ha arruolato 15480 pazienti diabetici, senza malattie cardiovascolari note, a ricevere 100 mg di aspirina o placebo per una mediana di 7.4 anni. L'endpoint primario (composito di morte vascolare, infarto miocardico, ictus o attacco ischemico transitorio) si è verificato nell'8.5% dei soggetti del gruppo con aspirina rispetto al 9.6% del gruppo col placebo (p=0.01). La riduzione dei principali eventi cardiovascolari è stata dell'1.1%, mentre l'aumento del rischio assoluto di sanguinamento maggiore è stato dello 0.9%⁴⁷. Questo risultato suggerisce effetti diversi (e potenzialmente anche avversi) in pazienti con caratteristiche diverse.

9. Trappola della p

In alcuni casi è possibile che il valore di p per l'endpoint primario si avvicini molto alla soglia di significatività statistica, rendendo difficile l'interpretazione del trial. Due esempi degni di nota sono gli studi CHARM-Preserved (p=0.051)¹⁷ e PARAGON-HF (Prospective Comparison of ARNI with ARB Global Outcomes in Heart Failure with Preserved Ejection Fraction) (p=0.06)⁴⁸, che hanno testato rispettivamente candesartan e sacubitril/valsartan nello scompenso cardiaco a frazione di eiezione preservata.

A fronte di risultati di questo tipo, vari autori hanno proposto di superare la dicotomia fra risultati positivi o negativi in base al valore di p relativo all'endpoint primario⁴⁹⁻⁵². È stato proposto di valutare l'entità dell'effetto e le relative stime di errore piuttosto che valori puntuali di p⁵³. Riviste prestigiose, tra cui il *New England Journal of Medicine*, hanno limitato l'utilizzo dei valori puntuali di p, in particolare per le analisi secondarie⁵⁴. L'approccio pragmatico seguito dalla Food and Drug Administration (FDA) nella valutazione dei risultati dello studio PARAGON-HF è consistito nel valutare tutte le evidenze disponibili, inclusi gli eventi degli endpoint primari riferiti dagli investigatori, e un'analisi di sottogruppo

che suggeriva una maggiore efficacia di sacubitril/valsartan per valori di frazione di eiezione <57%. Poiché la totalità delle informazioni era positiva, la FDA ha concesso un'indicazione estesa all'utilizzo di sacubitril/valsartan che includeva pazienti con scompenso e frazione di eiezione "al di sotto della soglia di normalità"^{51,55}.

Un approccio bayesiano al disegno degli studi sostituirebbe la dicotomia fra risultati positivi e negativi con uno spettro di probabilità e quindi tenderebbe a risolvere alla radice il problema dell'interpretazione di risultati borderline.

10. Il trattamento non funziona davvero

Quando un trial con una potenza statistica adeguata e senza significativi problemi metodologici presenta un valore di p relativo all'endpoint primario lontano dalla significatività statistica, la conclusione ragionevole è che il trattamento non è più efficace del suo controllo in quello specifico contesto clinico. Uno studio che dimostri che un determinato trattamento non funziona è importante tanto quanto un trial con risultati positivi, poiché chiarisce che quella linea di ricerca non è conveniente in termini di tempo e risorse e perché mostra quali terapie è opportuno non mettere in atto.

CHE COSA CI DOBBIAMO CHIEDERE QUANDO SIAMO DI FRONTE A UNO STUDIO NEUTRO/NEGATIVO?

Lo studio è realmente neutro o negativo?

Come discusso in precedenza e riassunto nella Figura 1, la metodologia di un trial deve essere attentamente esaminata prima di trarre conclusioni definitive circa i risultati.

Quando vengono rilevati problemi metodologici, lo studio dovrebbe essere considerato non conclusivo anziché neutro o negativo. Questo è ad esempio il caso dei trial con una potenza statistica insufficiente. Un altro esempio di evidenza non conclusiva è rappresentato dagli studi in cui l'endpoint primario è una combinazione di eventi diversi e solo alcuni di essi sono associati all'effetto del trattamento, mentre gli altri, distribuiti in entrambi i gruppi, determinano semplicemente un rumore di fondo che diluisce l'effetto reale del trattamento. Anche gli studi in cui la popolazione interessata è molto eterogenea e il trattamento sembra avere un effetto solo su alcuni sottogruppi potrebbero essere etichettati come non conclusivi. In tali situazioni, gli endpoint secondari e le analisi esplorative potrebbero fornire un'interpretazione alternativa dei risultati della sperimentazione. In alcuni casi, l'evidenza delle analisi

INTERPRETAZIONE CRITICA DEGLI STUDI CLINICI NEUTRI O NEGATIVI

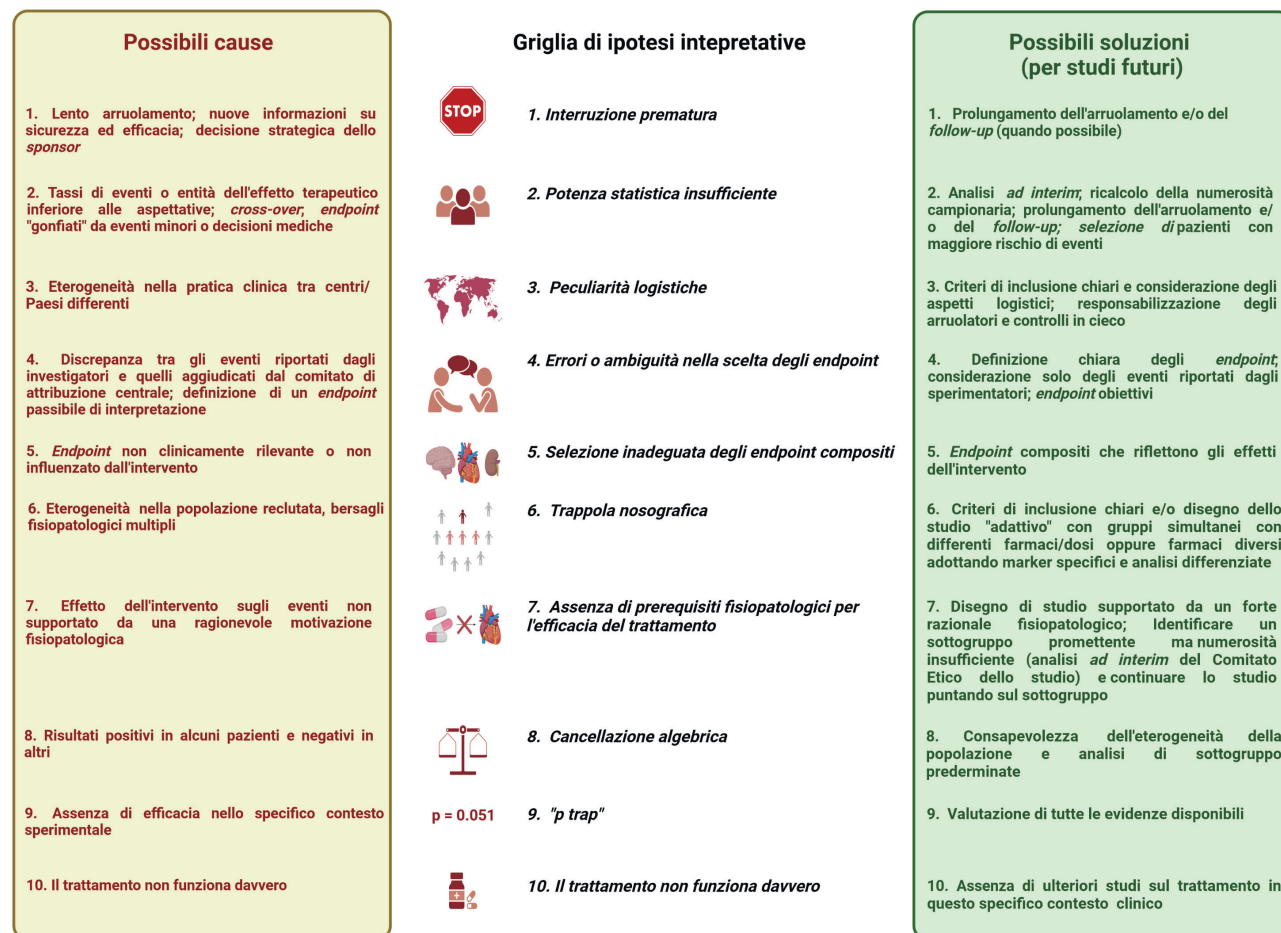


Figura 1. Possibile schema per l'interpretazione di trial neutri o negativi.

si esplorative, al di là del risultato primario, potrebbe essere considerata così forte che gli autori potrebbero affermare che uno studio neutro o negativo secondo i criteri convenzionali, possa, almeno in attesa di verifiche ulteriori meglio orientate, essere considerato positivo. Questo può avere un profondo impatto sulla ricezione dello studio da parte del pubblico e, in casi eccezionali, anche influenzare le linee guida e le agenzie di regolamentazione. Ad esempio, anche se candesartan, spironolattone e sacubitril/valsartan non hanno raggiunto gli endpoint primari nei rispettivi studi di valutazione di efficacia rispetto a placebo in pazienti con scompenso cardiaco a frazione di eiezione preservata (CHARM-Preserved¹⁷, TOPCAT¹⁴ e PARAGON-HF⁴⁸), i risultati dell'analisi nei sottogruppi di pazienti con frazione di eiezione vicina al limite inferiore della normalità sono stati considerati abbastanza robusti dall'American Guideline Committee da indurre a raccomandare questi farmaci nei pazienti con scompenso cardiaco e frazione di eiezione "nella parte inferiore dello spettro di normalità"⁵⁶. Un esempio ancora più notevole è lo studio CAPRICORN, che ha testato carvedilolo nello scompenso cardiaco con frazione di eiezione ridotta²⁴. Questo studio non ha dimostrato una differenza significativa tra i due bracci di trattamento in termini di endpoint primario (morte per ogni causa oppure ospedalizzazione per cause cardiovascolari), tuttavia ha evidenziato un beneficio del carvedilolo sulla mortalità per ogni causa, tanto da convincere la FDA ad approvare il farmaco per il trattamento dello scompenso cardiaco. Questi casi sono molto rari e queste decisioni sono di solito supportate da una forte motivazione fisiopatologica sottostante, nonché da prove di efficacia e sicurezza del farmaco in altri studi.

Cosa possiamo imparare da uno studio "realmente" neutro o negativo?

I trial NN rappresentano, negli ultimi anni, una gran parte degli studi clinici in ambito cardiovascolare e sono comunque importanti per il progresso della scienza e per la cura del paziente tanto quanto gli studi positivi⁵⁷. Alcuni studi hanno rivoluzionato la nostra comprensione dei meccanismi fisiopatologici alla base di alcune malattie. Ad esempio, il fallimento dei maggiori trial che hanno studiato i vasodilatatori nello scompenso cardiaco acuto, come la prostaciclina⁵⁸ e la nesiritide⁵⁹, ha portato alla comprensione che il modello emodinamico è una semplificazione eccessiva della complessità fisiopatologica dello scompenso cardiaco acuto. Inoltre, un'analisi approfondita dei risultati di uno studio NN può fornire strumenti essenziali per gli investigatori e gli sponsor per migliorare la progettazione di studi futuri. Uno studio negativo può offrire informazioni sulla sicurezza e/o l'efficacia di un trattamento, che devono essere considerate nello sviluppo

di agenti terapeutici simili. In aggiunta, uno studio NN può servire come base per un calcolo più realistico della potenza statistica per gli studi futuri, basandosi su eventi osservati in un contesto clinico specifico. Inoltre, l'individuazione di una potenza insufficiente in una precedente sperimentazione NN potrebbe indurre l'inclusione di misure prestabilite volte a prevenire tale limitazione in studi futuri che esaminino la stessa patologia o un trattamento simile, come, ad esempio, una valutazione in cieco dell'incidenza degli eventi durante l'esecuzione dello studio e una stima della potenza basata sull'incidenza "effettiva" dell'evento, seguita da un prolungamento del periodo di arruolamento e/o di follow-up (quando ragionevole). In terzo luogo, il mancato raggiungimento della significatività statistica per un endpoint composito primario potrebbe supportare l'adozione di combinazioni di eventi che riflettono più da vicino la prognosi di una determinata malattia negli studi futuri. In quarto luogo, documentare l'eterogeneità nella popolazione arruolata in uno studio potrebbe spingere all'adozione di criteri di inclusione più stringenti e potrebbe supportare la progettazione di uno studio dedicato per una popolazione selezionata sulla base di una forte motivazione fisiopatologica. Infine, gli studi negativi hanno l'importante compito di bloccare ulteriori indagini su terapie dannose o inefficaci.

In sintesi, nonostante i grandi progressi nella progettazione degli studi clinici, si possono ancora incontrare diversi problemi metodologici quando si analizzano i trial di superiorità pubblicati anche su riviste di assoluto prestigio. La capacità di riconoscere possibili problematiche metodologiche è importante per interpretare correttamente gli studi NN e per decidere se e come proseguire la ricerca su specifiche strategie di trattamento.

RIASSUNTO

Gli studi clinici controllati randomizzati che mostrano un beneficio significativo sono accolti con entusiasmo perché possono cambiare lo standard di cura per i pazienti che condividono le caratteristiche cliniche e fisiopatologiche dei partecipanti alla sperimentazione. Tuttavia, anche uno studio ben progettato e portato a termine con risultati neutri o negativi costituisce un'indagine importante e meritevole di un'attenta valutazione scientifica. In questo documento proponiamo un approccio in 10 passaggi per l'interpretazione di studi neutri o negativi al fine di escludere importanti questioni metodologiche prima di concludere che il trattamento non funzioni realmente. Discuteremo questo approccio facendo riferimento sia a studi importanti del passato che ad alcuni esempi significativi fra gli studi clinici, per lo più di fase 3, pubblicati negli ultimi anni.

Parole chiave. Disegno di studio; Studio controllato randomizzato; Studio negativo; Studio neutro.

BIBLIOGRAFIA

1. Pocock SJ, Stone GW. The primary outcome fails – what next? *N Engl J Med* 2016;375:861-70.
2. Rapezzi C, Aimo A, Fabiani I, et al. Critical reading of cardiovascular trials with neutral or negative results. *Eur Heart J* 2023. doi: 10.1093/eurheartj/ehad353 [Epub ahead of print].
3. Mann DL, Givertz MM, Vader JM, et al.; LIFE Investigators. Effect of treatment

- with sacubitril/valsartan in patients with advanced heart failure and reduced ejection fraction: a randomized clinical trial. *JAMA Cardiol* 2022;7:17-25.
4. McMurray JJ, Packer M, Desai AS, et al.; PARADIGM-HF Investigators and Committees. Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N Engl J Med* 2014;371:993-1004.
5. Pfeffer MA, Claggett B, Lewis EF, et al.; PARADISE-MI Investigators and Com-

- mittees. Angiotensin receptor-neprilysin inhibition in acute myocardial infarction. *N Engl J Med* 2021;385:1845-55.
6. Puymirat E, Cayla G, Simon T, et al.; FLOWER-MI Study Investigators. Multivessel PCI guided by FFR or angiography for myocardial infarction. *N Engl J Med* 2021;385:297-308.
7. Mathew R, Di Santo P, Jung RG, et al. B. Milrinone as compared with dobu-

tamine in the treatment of cardiogenic shock. *N Engl J Med* 2021;385:516-25.

8. Lincoff AM, Nicholls SJ, Riesmeyer JS, et al.; ACCELERATE Investigators. Evacetrapib and cardiovascular outcomes in high-risk vascular disease. *N Engl J Med* 2017;376:1933-42.

9. HPS3/TIMI 55-REVEAL Collaborative Group; Bowman L, Hopewell JC, Chen F, et al. Effects of anacetrapib in patients with atherosclerotic vascular disease. *N Engl J Med* 2017;377:1217-27.

10. Velazquez EJ, Lee KL, Deja MA, et al.; STICH Investigators. Coronary-artery bypass surgery in patients with left ventricular dysfunction. *N Engl J Med* 2011;364:1607-16.

11. Packer DL, Mark DB, Robb RA, et al.; CABANA Investigators. Effect of catheter ablation vs antiarrhythmic drug therapy on mortality, stroke, bleeding, and cardiac arrest among patients with atrial fibrillation: the CABANA randomized clinical trial. *JAMA* 2019;321:1261-74.

12. Ventz S, Alexander BM, Trippa L. Bayesian adaptive randomization in dose-finding trials. *JAMA Netw Open* 2018;1:e186075.

13. Dormandy JA, Charbonnel B, Eckland DJ, et al.; PROactive Investigators. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled trial. *Lancet* 2005;366:1279-89.

14. Pitt B, Pfeffer MA, Assmann SF, et al.; TOPCAT Investigators. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med* 2014;370:1383-92.

15. Pfeffer MA, Claggett B, Assmann SF, et al. Regional variation in patients and outcomes in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist (TOPCAT) trial. *Circulation* 2015;131:34-42.

16. de Denus S, O'Meara E, Desai AS, et al. Spironolactone metabolites in TOPCAT – new insights into regional variation. *N Engl J Med* 2017;376:1690-2.

17. Yusuf S, Pfeffer MA, Swedberg K, et al.; CHARM Investigators and Committees. Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial. *Lancet* 2003;362:777-81.

18. Packer M, Califf RM, Konstam MA, et al. Comparison of omapatrilat and enalapril in patients with chronic heart failure: the Omapatrilat Versus Enalapril Randomized Trial of Utility in Reducing Events (OVERTURE). *Circulation* 2002;106:920-6.

19. Metra M, Teerlink JR, Cotter G, et al.; RELAX-AHF-2 Committees Investigators. Effects of serelaxin in patients with acute heart failure. *N Engl J Med* 2019;381:716-26.

20. Ferrari R, Ford I, Fox K, et al.; ATP-CI Investigators. Efficacy and safety of trimetazidine after percutaneous coronary intervention (ATPCI): a randomised, double-blind, placebo-controlled trial. *Lancet* 2020;396:830-8.

21. Zannad F, Anker SD, Byra WM, et al.; COMMANDER HF Investigators. Rivaroxaban in patients with heart failure, sinus rhythm, and coronary disease. *N Engl J Med* 2018;379:1332-42.

22. Dangas GD, Tijssen JGP, Wöhrle J, et al.; GALILEO Investigators. A controlled trial of rivaroxaban after transcatheter aortic-valve replacement. *N Engl J Med* 2020;382:120-9.

23. Dietary supplementation with n-3 polyunsaturated fatty acids and vitamin E after myocardial infarction: results of the GISSI-Prevenzione Trial. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto miocardico. *Lancet* 1999;354:447-55.

24. Dargie HJ. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet* 2001;357:1385-90.

25. Packer M, Anker SD, Butler J, et al.; EMPEROR-Reduced Trial Investigators. Cardiovascular and renal outcomes with empagliflozin in heart failure. *N Engl J Med* 2020;383:1413-24.

26. Anker SD, Butler J, Filippatos G, et al.; EMPEROR-Preserved Trial Investigators. Empagliflozin in heart failure with a preserved ejection fraction. *N Engl J Med* 2021;385:1451-61.

27. Aimo A, Senni M, Barison A, et al. Management of heart failure with preserved ejection fraction: from neurohormonal antagonists to empagliflozin. *Heart Fail Rev* 2023;28:179-91.

28. Senni M, Paulus WJ, Gavazzi A, et al. New strategies for heart failure with preserved ejection fraction: the importance of targeted therapies for heart failure phenotypes. *Eur Heart J* 2014;35:2797-815.

29. Nicholls SJ, Lincoff AM, Garcia M, et al. Effect of high-dose omega-3 fatty acids vs corn oil on major adverse cardiovascular events in patients at high cardiovascular risk: the STRENGTH randomized clinical trial. *JAMA* 2020;324:2268-80.

30. Shah SJ, Borlaug BA, Kitzman DW, et al. Research priorities for heart failure with preserved ejection fraction: National Heart, Lung, and Blood Institute Working Group Summary. *Circulation* 2020;141:1001-26.

31. WHO Solidarity Trial Consortium. Repurposed antiviral drugs for COVID-19 – interim WHO solidarity trial results. *N Engl J Med* 2021;384:497-511.

32. Park JH, Detry MA, Murthy S, et al. How to use and interpret the results of a platform trial users' guide to the medical literature. *JAMA* 2022;327:67-74.

33. Obadia JF, Messika-Zeitoun D,

Leurent G, et al.; MITRA-FR Investigators. Percutaneous repair or medical treatment for secondary mitral regurgitation. *N Engl J Med* 2018;379:2297-306.

34. Grayburn PA, Sannino A, Packer M. Proportionate and disproportionate functional mitral regurgitation: a new conceptual framework that reconciles the results of the MITRA-FR and COAPT trials. *JACC Cardiovasc Imaging* 2019;12:353-62.

35. Stone GW, Lindenfeld J, Abraham WT, et al.; COAPT Investigators. Transcatheter mitral-valve repair in patients with heart failure. *N Engl J Med* 2018;379:2307-18.

36. Stone GW, Abraham WT, Lindenfeld J, et al.; COAPT Investigators. Five-year follow-up after transcatheter repair of secondary mitral regurgitation. *N Engl J Med* 2023;388:2037-48.

37. McMurray JJV, Metra M. Transcatheter repair of secondary mitral regurgitation. *N Engl J Med* 2023;388:2097-8.

38. Ridker PM, Everett BM, Pradhan A, et al.; CIRT Investigators. Low-dose methotrexate for the prevention of atherosclerotic events. *N Engl J Med* 2019;380:752-62.

39. Ridker PM, Everett BM, Thuren T, et al.; CANTOS Trial Group. Antiinflammatory therapy with canakinumab for atherosclerotic disease. *N Engl J Med* 2017;377:1119-31.

40. Konstam MA, Gheorghiadu M, Burnett JC Jr, et al.; Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study With Tolvaptan (EVEREST) Investigators. Effects of oral tolvaptan in patients hospitalized for worsening heart failure: the EVEREST Outcome Trial. *JAMA* 2007;297:1319-31.

41. Schrier RW, Gross P, Gheorghiadu M, et al.; SALT Investigators. Tolvaptan, a selective oral vasopressin V2-receptor antagonist, for hyponatremia. *N Engl J Med* 2006;355:2099-112.

42. Voors AA, Dittrich HC, Massie BM, et al. Effects of the adenosine A1 receptor antagonist rolofylline on renal function in patients with acute heart failure and renal dysfunction: results from PROTECT (Placebo-Controlled Randomized Study of the Selective Adenosine A1 Receptor Antagonist Rolofoylline for Patients Hospitalized with Acute Decompensated Heart Failure and Volume Overload to Assess Treatment Effect on Congestion and Renal Function). *J Am Coll Cardiol* 2011;57:1899-907.

43. Dittrich HC, Gupta DK, Hack TC, et al. The effect of KW-3902, an adenosine A1 receptor antagonist, on renal function and renal plasma flow in ambulatory patients with heart failure and renal impairment. *J Card Fail* 2007;13:609-17.

44. Givertz MM, Massie BM, Fields TK, Pearson LL, Dittrich HC; CKI-201 and CKI-202 Investigators. The effects of KW-3902, an adenosine A1-receptor antagonist, on

diuresis and renal function in patients with acute decompensated heart failure and renal impairment or diuretic resistance. *J Am Coll Cardiol* 2007;50:1551-60.

45. Digitalis Investigation Group. The effect of digoxin on mortality and morbidity in patients with heart failure. *N Engl J Med* 1997;336:525-33.

46. Rathore SS, Curtis JP, Wang Y, Bristow MR, Krumholz HM. Association of serum digoxin concentration and outcomes in patients with heart failure. *JAMA* 2003;289:871-8.

47. ASCEND Study Collaborative Group; Bowman L, Mafham M, Wallendszus K, et al. Effects of aspirin for primary prevention in persons with diabetes mellitus. *N Engl J Med* 2018;379:1529-39.

48. Solomon SD, McMurray JJV, Anand IS, et al.; PARAGON-HF Investigators and Committees. Angiotensin-neprilysin inhibition in heart failure with preserved ejection fraction. *N Engl J Med* 2019;381:1609-20.

49. Matthews RAJ. Moving towards the post $p < 0.05$ era via the analysis of credibility. *Am Stat* 2019;73:202-12.

50. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature* 2019;567:305-7.

51. O'Connor CM. Meet me in the middle: lessons from the cardiorenal advisory committee for sacubitril/valsartan in HFpEF. *JACC Heart Fail* 2021;9:161-3.

52. Tavazzi L. La significatività statistica nella ricerca clinica. *G Ital Cardiol* 2020;21:1-5

53. Chavalarias D, Wallach JD, Li AH, Ioannidis JP. Evolution of reporting p values in the biomedical literature, 1990-2015. *JAMA* 2016;315:1141-8.

54. Harrington D, D'Agostino RB Sr, Gattsonis C, et al. New guidelines for statistical reporting in the Journal. *N Engl J Med* 2019;381:285-6.

55. US Food & Drug Administration. Prescribing Information Entresto [https://www.](https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/207620s018lbl.pdf)

[accessdata.fda.gov/drugsatfda_docs/label/2021/207620s018lbl.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/207620s018lbl.pdf) [ultimo accesso 18 luglio 2023].

56. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA Guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *J Am Coll Cardiol* 2022;79:e263-421.

57. O'Connor CM. Why negative trials are positive for heart failure patients. *JACC Heart Fail* 2016;4:329-30.

58. Califf RM, Adams KF, McKenna WJ, et al. A randomized controlled trial of eplerenone therapy for severe congestive heart failure: The Flolan International Randomized Survival Trial (FIRST). *Am Heart J* 1997;134:44-54.

59. O'Connor CM, Starling RC, Hernandez AF, et al. Effect of nesiritide in patients with acute decompensated heart failure. *N Engl J Med* 2011;365:32-43.