

FOCUS 1

Riproducibilità

Molti lavori pubblicati e citati sono difficili o impossibili da replicare

Anche se *Werner Heisenberg* ritiene che non sarà mai possibile, attraverso la “ragion pura”, arrivare a qualche verità assoluta, la ricerca della verità guida la ricerca scientifica. Il requisito imprescindibile di una ricerca è che deve essere riproducibile. Se sei l'unica persona che può osservare un risultato, quel risultato, molto probabilmente, non è valido. La scienza sostituisce i pregiudizi “pubblici e privati” con *evidenze verificabili*. La riproducibilità di un risultato è un segno di credibilità e lo definiamo convenzionalmente “*evidenza sperimentale*”. Spetta alla comunità scientifica verificarne la veridicità attraverso la replica dei risultati ottenuti. La *riproducibilità* non è soltanto un requisito scientifico fondamentale, ma diventa un *valore etico*, soprattutto quando le decisioni politiche (vedi DPCM) e le scelte terapeutiche si basano sui risultati della ricerca scientifica. C'è una crisi di riproducibilità nella scienza? La risposta è Sì! Anche se la Pandemia ha bruscamente interrotto questo processo le manifestazioni della crisi sono riaffiorate lo stesso, a volte conseguenze disastrose. In epoca pre-pandemica in un sondaggio della rivista *Nature*, ad esempio, circa il 90% dei ricercatori ha convenuto che esisteva una crisi "lieve" o "significativa", e tra il 40% e il 70% ha ipotizzato che la pressione esercitata sui ricercatori (pubblica o perisci) poteva contribuire alla irriproducibilità dei risultati. Il sondaggio del 2016 su 1.500 scienziati riferiva che il 70% non era riuscito a riprodurre almeno un altro esperimento di quelli pubblicati e di avere difficoltà oggettive nel riprodurre “esattamente” i propri. La [Food and Drug Administration statunitense](#) già periodo 1977-1990 aveva riscontrato “difetti significativi” nel 10-20% degli studi clinici. Nel 2009, il 2% degli scienziati ha ammesso di aver falsificato gli studi almeno una volta e il 14% di conoscere personalmente qualcuno che l'ha fatto. Su 49 lavori dal 1990 al 2003 con più di 1000 citazioni, 45 hanno affermavano che la terapia proposta era efficace. Di questi studi, il 16% è stato contraddetto da studi successivi, il 16% ha riscontrato una maggiore efficacia terapeutica, il 44% è stato replicato e il 24% è rimasto in gran parte confermato. In un articolo del 2012, **Glenn Begley**, un consulente biotecnologico che lavora presso AMGEN e **Lee Ellis**, dell'Università del Texas, hanno scoperto che solo l'11% di 53 studi preclinici sul cancro poteva essere replicato. Gli studi irriproducibili avevano una serie di caratteristiche in comune, tra cui il fatto che non erano stati eseguiti “in cieco,” mancanza di controlli positivi e negativi, un uso inappropriato di test statistici e indagini di laboratorio non sempre adeguatamente validate. Sebbene il 52% degli intervistati concordi sul fatto che vi sia una significativa 'crisi' di riproducibilità, meno del 31% pensa che la mancata riproduzione dei risultati pubblicati significhi che il risultato è probabilmente sbagliato, e la maggior parte afferma di fidarsi ancora della letteratura pubblicata. L'articolo di Ioannidis ha dato la consapevolezza della difficoltà a confermare evidenze scientifiche consolidate. Negli anni successivi, a “crisi di riproducibilità” emergeva come problema dominante in intere discipline con conseguenze finanziarie enormi. Uno studio del 2015 ha stimato che i ricercatori hanno speso circa 28 miliardi di dollari all'anno nei soli Stati Uniti per ricerche pre-cliniche risultate successivamente irriproducibili.

FOCUS 2

Le cause della irriproducibilità

I dieci paradigmi di Ioannides.

Nel 2005, John Ioannidis ha sostenuto, in modo scioccante e persuasivo, che la maggior parte dei risultati delle ricerche pubblicate nel campo della medicina erano “falsi” e che i fattori responsabili della loro irriproducibilità erano molteplici.

1) Alcuni ricercatori cercano nei loro risultati correlazioni logiche, molte volte torturano i numeri finché non trovano una relazione spuria "statisticamente significativa". Secondo gli statistici di professione i ricercatori biomedici hanno una scarsa comprensione della metodologia statistica e quindi impiegano inconsapevolmente le statistiche in modo improprio nella loro ricerca.

2) Le sperimentazioni sono condotte su campioni di ridotte dimensioni, e viene data eccessiva enfasi su risultati ottenuti da un numero esiguo di tests.

3) Si percepisce la tendenza a voler pubblicare studi che riportano effetti preliminari limitati e parziali sotto la pressione del “publish/perish”

4) C'è la tendenza diffusa a voler generare ad ogni costo nuove ipotesi o a ipotizzare improbabili correlazioni.

5) Flessibilità eccessiva nella sperimentazione in relazione al protocollo di partenza. I ricercatori possono consapevolmente o inconsciamente influenzare i loro dati per produrre i risultati desiderati, o combinare set di dati in modo tale da invalidare le loro conclusioni. Potendo scegliere tra più misure di una variabile spesso decidono di utilizzare quella che fornisce, probabilmente, un risultato statisticamente significativo. Procedure sperimentali li apparentemente legittime fin troppo facilmente scivolano sulla linea confusa delle manipolazioni illegittime intrinseche nelle tecniche di ricerca.

6) Emergono pregiudizi intellettuali nella discussione dei risultati e il sospetto di conflitti di interesse.

7) Una certa tendenza a privilegiare una ricerca confermatrice nella certezza di ottenere risultati positivi, specialmente nelle aree di ricerca di tendenza.

8) Eccesso di libertà. La libertà incontrollata del ricercatore lo espone a sbagliare in tutti i modi sopra descritti. Minori sono i vincoli imposti dal protocollo sperimentale, maggiori sono le opportunità di modificare parametri in corso d'opera (work in progress). La mancanza di vincoli consente ai ricercatori di modificare i loro metodi a metà di uno studio mentre perseguono risultati pubblicabili e statisticamente significativi. I ricercatori spesso giustificano l'alterazione intermedia delle procedure di ricerca come "flessibilità", ma in pratica tale flessibilità spesso giustifica la riluttanza dei ricercatori ad accettare un risultato negativo (work in regress...)

Un articolo del 2011 ha stimato che fornire quattro "gradi di libertà del ricercatore" quattro possibilità di modificare il disegno sperimentale in corso, può portare a un tasso di falsi positivi del 61%..

9) Mancata condivisione dei risultati. I ricercatori condividono troppo raramente dati e metodologie una volta completati gli studi, dovrebbero effettuare un controllo reciproco che in realtà non avviene perché i ricercatori non sempre rendono disponibili i loro dati e i protocolli di studio. I repository online in cui i dati, i protocolli e i risultati possono essere archiviati e valutati dal pubblico cercano di migliorare l'integrità e la riproducibilità della ricerca.

10) L'appartenenza ad una “linea di pensiero” inibisce di fatto i tentativi di controllare i risultati, poiché gli studi di replicazione possono pregiudicare le convinzioni rassicuranti. Un'intera disciplina accademica può soccombere al “pensiero unico” e creare un consenso professionale

con una forte tendenza a respingere i risultati che ne mettono in discussione le basi. La schiacciante *omogeneità politica* degli accademici ha anche creato una cultura del “pensiero di gruppo” che distorce la ricerca accademica, dal momento che i ricercatori possono accettare prontamente risultati che confermano una visione del mondo liberale rifiutando al contempo conclusioni “conservatrici”. Il pensiero di gruppo politico colpisce particolarmente quei campi con ovvie implicazioni politiche, come la psicologia sociale e la scienza del clima.

FOCUS 3

John Ioannides

Un “influencers” per la qualità scientifica

Nato a New York City nel 1965, Ioannidis è cresciuto ad Atene dove si laureò in Medicina e Chirurgia con la menzione di miglior studente del *Athens College* e vincitore del *National Award of the Greek Mathematical Society* che lo ha definito come “un genio della statistica matematica medica”. Negli anni successivi presso la *Stanford University* che ha contribuito in maniera determinante alla definizione del EBM medicina basata sull'evidenza all'epidemiologia e alla ricerca clinica. Ioannidis è considerato il padre della “meta-ricerca” da lui definita come “lo studio degli studi”. L'articolo “*Why Most Published Research Findings Are False*” è il documento più scaricato nella *Public Library of Science*. Ioannidis afferma che la maggior parte delle ricerche pubblicate non soddisfa buoni standard scientifici di prova e che considera di fatto dei “falsi”.

Ha di fatto formalizzato la “crisi di replica” in diversi settori tra cui la genetica, gli studi clinici, neuroscienze, e la nutrizione. Il suo lavoro ha come oggetto ultimo di proporre soluzioni ai problemi nella ricerca e su come eseguire la ricerca in modo affidabile.

Nel articolo del 2016 “*Perché la maggior parte della ricerca clinica non è utile*” Ioannidis propone una riforma, caratterizzando alcuni punti affinché la ricerca medica possa essere realmente utile; in particolare auspica una medicina “centrata sul paziente” piuttosto che delle esigenze di medici, sponsor e degli stessi ricercatori. Insieme a Thomas Trikalinos ha coniato il fenomeno *Proteus* per descrivere la tendenza degli studi preliminari a trovare un effetto più ampio rispetto a quelli successivi.

The Atlantic ha dichiarato in un'edizione speciale su “Brave Thinkers” che Ioannidis “*potrebbe essere uno degli scienziati più influenti nel campo scientifico*”. Il *BMJ* lo ha definito come “l'inquisitore della scienza sciatta”. Ioannidis ha ricevuto numerosi premi e titoli onorifici. È membro delle Accademia Nazionale di Medicina degli Stati Uniti, della *Accademia Europea delle Scienze e delle Arti* ed insignito del *Einstein Fellow*. Nel 2019, è stato premiato con Robert S. Gordon, Jr. dal NIH come “Lecture in Epidemiology”.

FOCUS 4

John Ioannides nella tempesta pandemica

Durante la pandemia ricerche scientifiche viziate, condotte male e comunicate peggio

In un editoriale su *STAT* pubblicato il 17 marzo 2020, Ioannidis ha definito i lockdown “una reazione eccessiva rispetto a dati inaffidabili.” In piena pandemia Ioannidis ha ampiamente promosso uno studio di cui era stato coautore, “*Covid-19 Antibody Seroprevalence in Santa Clara County, California*”, pubblicato come preprint il 17 aprile 2020. Ha affermato che il numero di infezioni nella contea di Santa Clara era più di 50 volte superiore al conteggio ufficiale, portando il tasso di mortalità del virus dallo 0,1% allo 0,2%. Ioannidis ha concluso dallo studio che il coronavirus “*non è il problema apocalittico che pensavamo*”. Il messaggio ha trovato il favore dei

media di destra, ma ha scatenato una serie di polemiche accusando che i suoi test erano imprecisi e i suoi metodi erano sciatti. E che era il “terminale” di oscuri conflitti di interesse. Lo stesso **David H. Freedman**, suo amico e mentore affermava che lo studio di Santa Clara ha compromesso l'eccellente reputazione di Ioannidis, ipotizzando che le future generazioni di scienziati potrebbero ricordarlo come *“uno scienziato marginale che ha pompato un cattivo studio che ha supportato una folle teoria del complotto di destra nel mezzo di una massiccia crisi sanitaria”*. In recente articolo su *Scientific American*, uno studio legale indipendente ha stabilito l'assoluta assenza di conflitti finanziari. Questi avvenimenti fanno capire come l'editoria pandemica si sia costruita su scenari inquietanti manovrati da interessi occulti.

FOCUS 5

Il fattore P

La determinazione della significatività

Nel romanzo “Il segno dei quattro” Arthur Conan Doyle fa descrivere a Sherlock Holmes in maniera incisiva il significato di potenza statistica: *mentre il singolo individuo è un enigma irrisolvibile, quando è insieme agli altri diviene una certezza matematica*. È impossibile, per esempio, predire il modo in cui agirà un uomo, mentre è invece possibile dire con precisione cosa farà un certo numero di uomini messi insieme. L'individuo varia, ma le percentuali rimangono costanti. Così almeno dicono le statistiche. Possiamo ritenere uno studio “statisticamente potente” quando è in grado di descrivere sufficientemente dei risultati, anche se i suoi effetti sui dati sono minimi. In questa ottica gli studi che ripetono l'esperimento più volte ed usano un campione molto ampio sono i più potenti da un punto di vista statistico. Una potenza 0.8 significa che su dieci ipotesi corrette sottoposte a verifica, solo due saranno escluse perché i loro effetti non si riflettono sui dati, ed è questa, generalmente, la potenza considerata sufficiente nella maggior parte dei casi. Tuttavia questo obiettivo, nella sperimentazione clinica, non è sempre raggiungibile perché i grandi trials, quelli che prevedono “l'arruolamento” di molti soggetti, hanno il difetto, non trascurabile, di essere anche i più costosi.

Dalla ricerca di Ioannides, che bolla come poco attendibili la gran parte degli studi scientifici, è emerso ad esempio che negli studi delle neuroscienze la potenza statistica non supera lo 0.2, un valore molto distante dallo 0.8 teorico. Un aspetto da non trascurare è che di solito i ricercatori sperano di ottenere “risultati inattesi” e per questo cercano di verificare ipotesi spesso altamente improbabili.

Anche se è proprio cercando “l'impossibile” che l'uomo ha sempre realizzato e conosciuto il “possibile”, e coloro che si sono saggiamente limitati a ciò che sembrava possibile, non sono mai avanzati di un sol passo. Ioannides ritiene che nello specifico della sua disciplina, l'epidemiologia, ci si può aspettare che solo una ipotesi su dieci si riveli corretta. In ambiti di ricerca come la genomica e la proteomica, che analizzano grandi quantità di geni e proteine per dimostrare l'esistenza di correlazioni, anche una ipotesi su mille può dimostrarsi esatta.

Se la potenza statistica scende da 0.8 a 0.4 (il che è realistico in molti settori di ricerca) la conclusione è che più della metà dei risultati positivi sarebbe sbagliata e che, paradossalmente, i risultati negativi sono molto più affidabili. Il vantaggio statistico di uno studio con una potenza statistica pari a 0.8 è che ci saranno 875 risultati negativi, di cui solo 20 saranno falsi, garantendo così una accuratezza del 97 per cento. Purtroppo i ricercatori e le riviste di settore non sono molto interessati ai risultati negativi. Preferiscono enfatizzare quelli positivi, proprio quelli più soggetti ad errori. Un errore è tanto più pericoloso quanta più verità contiene. La causa più frequente di un errore è la ricerca della verità assoluta, e l'errore più grave per un ricercatore è avere la certezza di non commetterne. Questo forse spiega perché molte ricerche progettate male ed eseguite

peggio sfuggono ai revisori generalmente più propensi ad esporre ai redattori di riferimento l'importanza e la novità dell'articolo, piuttosto che i suoi punti deboli, ignorando sistematicamente di valutare la potenza statistica dello studio e che la forza di un ricercatore sta proprio nella potenza delle sue illusioni. **Victoria Stodden**, docente di statistica alla Stanford University, sostiene che la comprensione e l'utilizzo appropriato della statistica da parte dei ricercatori non è andata di pari passo con lo sviluppo di complesse tecniche matematiche per elaborare i dati. I ricercatori spesso usano tecniche inappropriate e largamente superate, solo perché sono quelle che conoscono meglio e gestiscono l'analisi dei dati con dei software che usano molto spesso in maniera automatica e senza conoscerli nel dettaglio. Una valutazione statistica può dimostrarci che siamo in errore, ma purtroppo ci dice poco o nulla su quale sia la verità.

Il documento "**Ridefinire la significatività statistica**", firmato da un gran numero di scienziati e matematici, propone che nei *"campi in cui la soglia per definire la significatività statistica per le nuove scoperte è $p < 0,05$, proponiamo una modifica a $p < 0,005$. Questo semplice passaggio migliorerebbe immediatamente la riproducibilità della ricerca scientifica in molti campi"*. Il loro fondamento logico è che *"una delle principali cause di non riproducibilità (è che gli) standard statistici di prova per affermare nuove scoperte in molti campi della scienza sono semplicemente troppo bassi. L'associazione di risultati 'statisticamente significativi' con $p < 0,05$ si traduce in un alto tasso di falsi positivi anche in assenza di altri problemi sperimentali, procedurali e di segnalazione"*.

Questo invito è stato successivamente criticato da un altro grande gruppo, il quale ha sostenuto che "ridefinire" la soglia non risolverà i problemi attuali, ne porterebbe di nuovi e che, alla fine, tutte le soglie dovevano essere giustificate caso per caso invece di seguendo le convenzioni generali. Sebbene gli statistici siano unanimi sul fatto che l'uso di $p < 0,05$ fornisce prove più deboli di quanto generalmente apprezzato, c'è una mancanza di unanimità su ciò che dovrebbe essere fatto al riguardo. Alcuni hanno sostenuto che i *metodi bayesiani* dovrebbero sostituire i valori p . Ciò non è avvenuto su vasta scala, in parte perché è complicato e in parte perché molti utenti diffidano delle specifiche delle distribuzioni precedenti in assenza di dati concreti.

I rischi derivanti dalla dipendenza dai valori p sono stati enfatizzati sottolineando che anche l'osservazione di $p = 0,001$ non era necessariamente una forte evidenza contro l'ipotesi nulla. Nonostante il fatto che il rapporto di verosimiglianza a favore dell'ipotesi alternativa sullo zero sia vicino a 100, se l'ipotesi fosse non plausibile, con una probabilità a priori di un effetto reale pari a 0,1, anche l'osservazione di $p = 0,001$ avrebbe un rischio falso positivo dell'8 per cento. Non raggiungerebbe nemmeno il livello del 5%.

È stato raccomandato di non utilizzare i termini "significativo" e "non significativo". I valori p e gli intervalli di confidenza dovrebbero essere ancora specificati, ma dovrebbero essere accompagnati da un'indicazione del rischio di falso positivo. È indispensabile incoraggiare campioni di dimensioni maggiori. Per migliorare la qualità delle repliche, sono spesso necessarie dimensioni del campione più grandi di quelle utilizzate nello studio originale. Sono necessarie dimensioni del campione più grandi perché le stime delle dimensioni degli effetti nel lavoro pubblicato sono spesso esagerate a causa del bias di pubblicazione e della grande variabilità del campionamento associata a piccole dimensioni del campione in uno studio originale. Inoltre, l'uso di soglie di significatività di solito porta a effetti gonfiati, perché in particolare con campioni di piccole dimensioni, solo gli effetti più grandi diventeranno significativi.

FOCUS 6

Un vizio di prospettive

Il ragionamento del medico e le aspettative dei pazienti

Marco Aurelio riteneva che *tutto ciò che vediamo è una prospettiva, non la verità*. Il gruppo di studio internazionale degli statistici ha riportato le diverse prospettive della medicina che possono condizionare il ragionamento medico e le aspettative dei pazienti, sommariamente:

Medicina: dal mito al mitocondrio

Abbiamo un'immagine mitica di ciò che la scienza e la medicina possono fare. Crediamo nella medicina come fornitore di conoscenza oggettiva, ma non riusciamo a riconoscere questo atteggiamento per quello che è: una credenza, che dovrebbe essere esaminata come tale nella sua relazione con un sistema di credenze più generale, cosa chiamiamo cultura.

Una promozione affrettata

La medicina è entrata solo di recente nei ranghi di ciò che chiamiamo "scienze dure". I numerosi modi in cui gli studi randomizzati non riescono a fornire risultati coerenti, o anche semplicemente applicabili, iniziano a chiederci se questa recente promozione non sia stata affrettata. Per secoli la medicina è stata concepita come rito magico o religioso, come tecnica o mestiere per i suoi aspetti pratici (chirurgia, ostetricia), o come disciplina letteraria che si basa sull'attenta lettura e glossatura di autori canonici dell'antichità (Galeno e Ippocrate) per la sua parte teorica (anatomia, farmacopea). Ora che la letteratura medica è secolarizzata e basata su fatti e osservazioni, si potrebbe presumere che la medicina applichi i metodi scientifici più rigorosi e fornisca una conoscenza basata sull'evidenza dei disturbi e delle condizioni per guidare i medici. Ma come mostrano il lavoro di Ioannidis una solida metodologia e criteri di pubblicazione selettivi non sono la norma. Né sono sufficienti di per sé per fornire il tipo di verità dura che i pazienti si aspettano.

Un malinteso dominante

La fede odierna nella medicina come scienza (e nella scienza come ricerca della verità assoluta e nella tecnologia come soluzione per tutto) è un malinteso dominante che non rende meno reale il bisogno di ricerca medica e l'enorme servizio che i medici rendono all'umanità. Tuttavia, questo malinteso è costoso finanziariamente, umanamente e intellettualmente: la fede cieca nei farmaci o nelle scoperte porta a trattamenti nella migliore delle ipotesi inefficaci, nel peggiore dei casi a quelli potenzialmente pericolosi. John Ioannidis e al suo team di studiosi per aver sfatato i miti dietro le "notizie" mediche: questo atto di interrogatorio sanitario ci costringe a pensare al corpo come a qualcosa di più complesso e prezioso di una macchina per essere risolto. La lezione da trarre da questa indagine non è certamente quella di tagliare i fondi per la ricerca medica, ma di riconoscere che le nostre aspettative tanto quanto le nostre metodologie devono essere ripensate.

La necessità dell'interpretazione

Più fondamentalmente, gli esperimenti e i risultati possono raramente, se non del tutto, essere classificati secondo il gold standard della "conoscenza certificata" che il pubblico richiede: anche lo studio più incontestabile deve essere interpretato prima che possa avere importanza intellettuale e terapeutica. Le "notizie" mediche diventano tali solo se presentate come potenzialmente rivoluzionarie o, più modestamente, utili. Estrapolare dall'ambiente clinico all'efficacia della vita reale richiede sempre un atto di fede, gradito quando allevia i dolori e salva vite, ma che non può mai essere fondato su una certezza scientifica assoluta. In altre parole, in medicina come in qualsiasi altro campo delle scienze umane, l'interpretazione è ciò che trasforma i dati in significato.

La stampa laica

Spesso, è il modo in cui la maggior parte delle notizie mediche "adatte alla stampa" viene dichiarata che è difettosa, piuttosto che i risultati stessi che nella maggior parte dei casi non dicono molto in termini di applicazione nella vita reale. I risultati statistici (validi o meno) non hanno valore prognostico. Le probabilità di ottenere un doppio sei con due dadi sono sempre una su trentasei, indipendentemente dall'età, dal sesso e dalla storia medica della persona che lancia i dadi. Non così con la probabilità di ottenere miglioramenti dal farmaco XYZ perché il 40% del gruppo di test lo ha fatto negli studi clinici.

Una analisi delle probabilità

Non siamo oggetti inanimati; non siamo intercambiabili: ogni corpo risponde in modo diverso e si presenta al lettino con la propria zuppa di geni, ormoni, emocromo, squilibrio dei liquidi, cellule invecchiate, vulnerabilità, temperamento e convinzioni e aspettative mediche (queste ultime hanno ripetutamente dimostrato di influenzare il risultato di qualsiasi trattamento). Quelle variabili non possono essere eliminate: anche la sperimentazione clinica meglio concepita può solo segnalare la probabilità, in determinate condizioni, di vedere questi e quei cambiamenti. Ma per ogni caso specifico, le probabilità possono essere analizzate solo individualmente, poiché è l'intero quadro medico e l'identità culturale del paziente che influenzeranno i risultati. Pochissimi farmaci sono farmaci mirati che affrontano la causa di una malattia o malattia: la maggior parte affronta i sintomi, a volte silenziosamente in modo obliquo (a chi non è stato prescritto un farmaco per la malattia X perché l'esperienza aveva dimostrato che potrebbe alleviare i sintomi di Z?).

Gli studi che puntano ai benefici medici di alcuni nutrienti o vitamine non sono lasciati ai guai. Può essere vero che uno, due o una dozzina di studi hanno dimostrato che le persone che ricevevano alte dosi di omega-3 mostravano rischi inferiori, ad esempio, di attacchi di cuore e cancro al colon rispetto a un gruppo di prova. Ciò non significa che se prendi alte dosi di olio di pesce, sarai protetto da tali malattie o addirittura ridurrai i rischi di contrarle.

FOCUS 7

La meta-ricerca

Le direttive della meta-ricerca

La meta-ricerca o anche " *ricerca sulla ricerca* ", utilizza il metodo scientifico per studiare come viene svolta la ricerca e dove è possibile apportare miglioramenti.

La meta-ricerca è stata descritta come "una visione a volo d'uccello della scienza".

Nelle parole di John Ioannidis , "La scienza è la cosa migliore che sia accaduta agli esseri umani ... ma possiamo farlo meglio".

Marcus R. Munafò e George Davey Smith sostengono, in un articolo pubblicato da *Nature* , che *la meta-ricerca dovrebbe promuovere la triangolazione, non solo la replica, in quanto la replica da sola ci porterà solo fino a un certo punto (e) potrebbe effettivamente peggiorare le cose ... Crediamo che una protezione essenziale contro le idee imperfette sia la triangolazione. Questo è l'uso strategico di più approcci per affrontare una domanda. Ogni approccio ha i propri presupposti, punti di forza e di debolezza non correlati. È meno probabile che i risultati che concordano tra le diverse metodologie siano artefatti Forse uno dei motivi per cui la replica ha catturato così tanto interesse è l'idea spesso ripetuta che la falsificazione sia al centro dell'impresa scientifica. Questa idea è stata resa popolare dalla massima degli anni '50 di Karl Popper secondo cui le teorie non possono mai essere dimostrate, ma solo falsificate . Tuttavia, un'enfasi eccessiva sulla ripetizione degli esperimenti potrebbe fornire un senso di certezza infondato sui risultati che si basano su un unico approccio. ... i filosofi della scienza sono andati avanti da Popper. Descrizioni*

migliori di come gli scienziati lavorano effettivamente includono ciò che l'epistemologo Peter Lipton ha chiamato nel 1991 "inferenza alla migliore spiegazione".

In effetti, alcune previsioni di una crisi imminente nel meccanismo di controllo della qualità della scienza e della necessità di una meta-ricerca possono essere fatte risalire a diversi decenni, soprattutto

Il fisico **John Derek de Solla Price** il padre della scientometria predisse che la scienza avrebbe potuto raggiungere la "senilità" come risultato della propria crescita esponenziale.

Una certa letteratura odierna sembra confermare questa profezia "traboccante", lamentando il decadimento sia dell'attenzione che della qualità

Il filosofo e storico della scienza **Jerome R. Ravetz** ha predetto nel suo libro del 1971 *Scientific Knowledge and Its Social Problems* che la scienza - nella sua progressione dalla "piccola" scienza composta da comunità isolate di ricercatori, alla "grande" scienza o "tecno-scienza" - subirebbe grossi problemi nel suo sistema interno di controllo della qualità. Ravetz ha riconosciuto che la struttura degli incentivi per gli scienziati moderni potrebbe diventare disfunzionale, ora nota come l'attuale sfida "pubblica o muori", creando incentivi perversi a pubblicare qualsiasi risultato, per quanto dubbio. Secondo Ravetz, la qualità nella scienza è mantenuta solo quando c'è una comunità di studiosi legati da un insieme di norme e standard condivisi, i quali sono tutti disposti e in grado di ritenersi reciprocamente responsabili.

Lo storico **Philip Mirowski** ha offerto una diagnosi simile nel suo libro del 2011 *Science Mart* (2011). *(nel titolo, la parola "Mart" si riferisce al gigante della vendita al dettaglio "Walmart", usata da Mirowski come metafora per la mercificazione della scienza)*.

Nell'analisi di Mirowski, la qualità della scienza crolla quando diventa una merce negoziata in un mercato. Mirowski fa risalire il decadimento della scienza alla decisione delle grandi società di chiudere i loro laboratori interni e di esternalizzare l'attività di ricerca nel tentativo di ridurre i costi e aumentare i profitti.

Le società successivamente hanno spostato la loro ricerca dalle università a un'opzione ancora più economica: le organizzazioni di ricerca a contratto (CRO). La crisi del sistema di controllo della qualità della scienza sta influenzando l'uso della scienza per la politica. L'economista **Noah Smith** suggerisce che un fattore della crisi è stata la sopravvalutazione della ricerca nel mondo accademico e la sottovalutazione della capacità di insegnamento, specialmente in campi con poche importanti scoperte recenti.

La teoria del sistema sociale, secondo il sociologo tedesco **Niklas Luhmann** offre un'altra lettura della crisi. Secondo questa teoria, ciascuno dei sistemi come 'economia', 'scienza', religione, media e così via comunica utilizzando il proprio codice, vero / falso per la scienza, profitto / perdita per l'economia, nuove / non-notizie per i media; secondo alcuni sociologi la mediatizzazione della scienza, sua mercificazione e la sua politicizzazione - come risultato dell'accoppiamento strutturale tra i sistemi, hanno portato a una confusione dei codici del sistema originale. Se il codice vero / falso della scienza viene sostituito da quelli degli altri sistemi, come profitti / perdite, notizie / non notizie, l'operazione della scienza entra in una crisi interna.

FOCUS 8

Come costruire un report riproducibile

Le direttive della meta-ricerca

La *preregistrazione degli studi* è una recente innovazione nell'editoria scientifica per affrontare la crisi della replica attraverso l'uso di rapporti registrati. Il formato del rapporto registrato richiede che gli autori presentino una descrizione dei metodi di studio e delle analisi prima della raccolta dei dati. Una volta che il metodo e il piano di analisi sono stati vagliati attraverso la revisione tra pari, la pubblicazione dei risultati è provvisoriamente garantita, a seconda che gli autori seguano il protocollo proposto. Uno degli obiettivi dei rapporti registrati è aggirare il pregiudizio di pubblicazione verso risultati significativi che possono portare all'implementazione di pratiche di ricerca discutibili e incoraggiare la pubblicazione di studi con metodi rigorosi.

I metodi per affrontare la crisi includono la pre-registrazione di studi scientifici e sperimentazioni cliniche, nonché la fondazione di organizzazioni come CONSORT e EQUATOR Network che emettono linee guida per la metodologia e il reporting.

Sono in corso sforzi per riformare il sistema di incentivi accademici, per migliorare il processo di revisione tra pari, per ridurre l'uso improprio delle statistiche, i pregiudizi nella letteratura scientifica e per aumentare la qualità e l'efficienza complessive del processo scientifico.

Nel luglio 2016 l'organizzazione olandese per la ricerca scientifica ha messo a disposizione 3 milioni di euro per gli studi di replicazione. Il finanziamento è finalizzato alla replica basata dei dati esistenti ed alla raccolta di nuovi dati.

Plos One e Science Exchange hanno proposto un servizio che mette in contatto ricercatori e laboratori chiamato "Reproducibility initiative" attraverso cui i ricercatori, possono chiedere dietro pagamento, la convalida del loro lavoro ad un laboratorio indipendente, o una organizzazione può richiedere di riprodurre un lavoro a cui è particolarmente interessato.

Alcuni anni fa Laura e John Arnold due miliardari statunitensi hanno messo a disposizione oltre un milione di dollari per riesaminare i cinquanta articoli di oncologia più influenti pubblicati tra il 2010 e 2012.

Laura e John Arnold sono due miliardari statunitensi **fondatori** della LJAF una fondazione che fa parte del Giving Pledge, un club di miliardari che si sono impegnati a donare una grande frazione del loro reddito per cause filantropiche durante la loro vita.

Alcuni anni fa la LJAF ha messo a disposizione oltre un milione di dollari per riesaminare i cinquanta articoli di oncologia più influenti pubblicati tra il 2010 e 2012. Tra gli obiettivi compariva anche quello di identificare (smascherare) quei lavori con dati non proprio attendibili anche alla luce degli innumerevoli falsi prodotti dalla reazione al "publish or perish". Arthur Bloch, causticamente ricordandoci la *legge di Fett* che *non bisogna mai replicare un esperimento riuscito. Se la riproducibilità è un problema, conduci l'esperimento una volta sola.*

Seconda legge di Finale, qualunque sia il risultato di un esperimento, ci sarà sempre qualcuno pronto fraintenderlo o a falsificarlo, o ad accettarlo perché perfettamente in accordo con il suo punto vista. L'European Research Council ha di fatto snobbato la proposta della "Reproducibility initiative" dichiarando che *"gli studi che cercano di riprodurre risultati già ottenuti non sono nelle nostre priorità e non prevedono l'attribuzione di fondi"* Se è vero che la diffidenza verso gli altri nasce anche dalla sfiducia in noi stessi c'è poco da stare allegri.

FOCUS 9

I criteri di selezione ed accettazione di un report

Il ruolo dei revisori è fondamentale

Alcuni editori hanno creato veri e propri monopoli capaci addirittura di gestire e orientare le carriere scientifiche di ricercatori e in particolare gli interessi delle Company che, direttamente o indirettamente, gravitano intorno alla salute e alla malattia. Nel 1997 l'anglo-olandese Elsevier vantava un utile di 633.6 miliardi di lire con un margine di guadagno del 35%. E poiché il costo dell'abbonamento alle riviste negli ultimi anni è aumentato in maniera vertiginosa, gli utili continuano ad aumentare. Attualmente sono circa 2000 testate a raccogliere il 95% degli articoli. Queste testate diventano sempre più irraggiungibili e ambite perché assicurano al ricercatore la notorietà, che a sua volta apre le porte ai finanziamenti.

È innegabile l'insieme degli interessi che orbitano intorno al non rinunciare e l'editoria scientifica finisce tristemente per assomigliare sempre di più ad un circolo privato elitario che riporta risultati, simile ad un mostro che si nutre dell'interazione tra forze di mercato interne ed esterne alla scienza e incapace di soddisfare le esigenze di una comunità scientifica moderna (Ben Goldacre). Quello che colpisce è la tracotanza dell'editoria scientifica che non si ritiene responsabile di aver pubblicato ricerche che si sono rivelate fraudolente o prive di senso.

Non è solo per quello che facciamo che siamo ritenuti responsabili, ma anche per quello che non facciamo.

Anche quando si tratta di frodi vere e proprie spesso impiegano molto tempo per denunciarle e si rifiutano di indagare su comportamenti scorretti diffusi. Questo è scoraggiante e irritante. Suggestisce come la *crisi della riproducibilità* non sia una rivoluzione metodologica specifica, ma un sintomo di un sistema scientifico che deve essere rivisto a più livelli.

Anche le frodi vere e proprie spesso impiegano molto tempo per essere ripudiate, con alcune università e riviste che trascinano i piedi e rifiutano di indagare su comportamenti scorretti diffusi. Questo è scoraggiante e irritante. Suggestisce che la crisi della replicazione non sia una rivalutazione metodologica specifica, ma un sintomo di un sistema scientifico che necessita di ripensamento a molti livelli.

Non possiamo semplicemente insegnare agli scienziati come scrivere articoli migliori. La pressione per pubblicare molti documenti favorisce coloro che possono metterli insieme rapidamente - e un modo per essere veloci è essere disposti a tagliare gli angoli.

"Le riviste ora ritirano circa 1.500 articoli all'anno - un aumento di quasi 40 volte rispetto al 2000, e un cambiamento drammatico anche se si tiene conto del raddoppio o del triplo approssimativo degli articoli pubblicati all'anno", sostiene Ivan Oransky del Retraction Watch. "I giornali sono migliorati", riportando maggiori dettagli sui documenti ritirati e migliorando il loro processo di ritiro.

Anche altri cambiamenti nelle pratiche scientifiche comuni sembrano aiutare. Ad esempio, le preregistrazioni - che annunciano come condurrà la tua analisi prima di fare lo studio - portano alla pubblicazione di risultati più nulli. Dopo tonnellate di risorse spese per dimostrare la portata del problema, lottare per più ritrattazioni, insegnare metodi statistici migliori e tentare di trascinare la frode allo scoperto, i documenti continuano a non replicarsi quanto i ricercatori sperano, e le carte cattive sono ancora ampiamente citate - suggerendo che gran parte del problema non è ancora stata toccata. Il fattore di impatto o "impact factor" è un indicatore sintetico, che misura il numero medio di citazioni ricevute in un particolare anno da articoli pubblicati in una rivista scientifica nei due anni precedenti.

È "prodotto" dalla Thomson Reuters Corporation, divisione Healthcare & Science, una delle più grandi agenzie d'informazione scientifica al mondo. La sua affidabilità è assicurata dalla banca dati

che li genera, che non è una fondazione senza fini di lucro, ma una azienda privata che persegue i suoi più che leciti profitti.

Dal 1990 ad oggi sono sempre più numerosi i ricercatori che, pur non mettendo in discussione la validità dell'impact factor, criticano l'uso che ne viene fatto non tanto per la superficialità del giudizio legato all'impact che vede *"i ricercatori valutati un tanto al kilo"* (Rémi Chauvin), quanto piuttosto che incoraggi ed alimenti la "pubblicomania", la frenesia di pubblicare ad ogni costo (con il rischio potenziale di falsi) sulle riviste più prestigiose. Nei paesi dove i ricercatori sono in gran parte precari, e pertanto soggetti ad una mobilità "fisiopatologica", l'impact factor è di fatto la prassi principale di valutazione. E quando non c'è una posizione lavorativa stabile tutto viene rimesso in discussione ogni due o tre anni. Di volta in volta il ricercatore deve trovare non solo un laboratorio universitario o una istituzione che lo assuma, ma anche i finanziamenti pubblici o privati per le ricerche e spesso per il suo stesso stipendio.

Tutto ruota sulla sua reputazione scientifica, e questa è legata al numero e al prestigio delle sue pubblicazioni misurate attraverso l'impact factor. Alcune università definiscono già al momento della stipula dell'ingaggio il numero delle citazioni che i futuri articoli prodotti da un ricercatore o da un gruppo dovranno aver collezionato nel corso dell'anno. L'imperativo è: publish or perish.

C'era da aspettarselo che questa corsa frenetica alla pubblicazione provocasse distorsioni metodologiche e il ricorso a strategie compensative e/o difensive.

Alcuni ricercatori ricorrono quindi alla prassi, discutibile, che consiste nel pubblicare due o tre volte lo stesso articolo in riviste diverse, apportando qualche lieve modifica o piccole integrazioni.

Altri a ricorrere all'espedito dell'autocitazione. Altri ancora citano in via prioritaria i loro amici, che a loro volta si impegnano a citarli, in modo da aumentare reciprocamente il loro "impact" e la loro visibilità nella comunità scientifica. Questo determina anche la proliferazione delle "firme di favore" inserendo tra gli autori persone che non hanno partecipato alla ricerca, né direttamente né indirettamente. Presenza molte volte imposta dal rais di turno che può condizionare la vita stessa del ricercatore all'interno del laboratorio o del dipartimento in cui opera.

Cosa ancora più grave quando questa accanita competizione spinge a falsificare i dati sperimentali, e questo per un vero ricercatore è come barare al solitario.

Friedhelm Herrmann, brillante e stimato clinico universitario tedesco è passato alle cronache alcuni anni fa in quanto autore di clamorosi falsi scientifici.

A seguito di accuse e sospetti una commissione, istituita dal CNR tedesco per verificare l'attendibilità delle sue ricerche, arrivò alla conclusione che dei 347 articoli pubblicati, almeno 94 contenevano dei dati falsificati e più di un centinaio sollevavano seri dubbi.

Modificare dei dati clinici e pubblicarli su riviste scientifiche qualificate rischia di incoraggiare e diffondere, attraverso le linee guida, terapie che possono mettere in pericolo la vita stessa dei pazienti. Il caso Herrmann non è stato purtroppo un caso isolato e fa riflettere sulle particolari "caratteristiche antropologiche" di alcuni ricercatori e sull'uso strumentale che viene fatto della "conoscenza". Un ricercatore che truca i dati è un fatto di per sé inimmaginabile, incomprensibile ed irrazionale: è come barare al solitario.

Tuttavia il "falso" pone l'interrogativo inquietante su cosa possa spingere uno scienziato a comportarsi da falsario ed ancora, come sia possibile che ricercatori disonesti abbiano ricevuto un credito tale da far loro raggiungere posizioni apicali nelle strutture della ricerca e dell'Università unitamente a funzioni di educatore. E quando lo scandalo è diventato straripante ed inarrestabile è arrivata l'indignazione accademica convenzionale e scontata, quanto il rituale dell'elogio.

Non esiste delitto, inganno, trucco, imbroglio e vizio che non viva della sua segretezza. Ma i segreti, caro prof. Friedhelm, hanno il difetto di essere svelati. Ne era fermamente convinto **Benjamin Franklin** quando affermava che *tre persone possono conservare saldamente un segreto se due di loro sono morte*. Una nuova sostanza chimica, isolata e purificata da un lichene, ha

mostrato di possedere una sorprendente efficacia in vitro su diverse linee di cellule tumorali. Questi risultati, ottenuti nei laboratori di una università del nord America, sono stati riportati in un lavoro inviato per la pubblicazione a 304 riviste che hanno provveduto a sottoporlo ai loro referees. Circa la metà (157) delle riviste hanno accettato di pubblicare il lavoro senza alcun particolare commento. Questa notizia si perderebbe nel normale "traffico editoriale" se non nascondesse dei retroscena quanto meno inquietanti. La ricerca in questione, e i risultati riportati, erano completamente inventati, l'autore aveva firmato il lavoro servendosi di uno pseudonimo, l'università presso cui erano stati condotti gli esperimenti addirittura non era mai esistita. Inoltre il disegno sperimentale era zeppo di errori ed i risultati erano discussi in maniera a dir poco imbarazzante. **John Bohannon**, biologo giornalista, autore di questa "deliziosa beffa", voleva in questo modo mettere a nudo la fragilità dei criteri adottati dai referees di alcune riviste scientifiche per l'accettazione di un lavoro.