

***Statistica e Ricerca Medica per la Salute.***  
***Alcune note di metodo e chiose sulle applicazioni(\*)***

*Mario Bolzan & Fortunato Pesarin*  
*Dipartimento di Scienze Statistiche*  
*Università degli Studi di Padova*

**1. Il problema conoscitivo base della ricerca.**

Rispetto al passato i temi della salute dell'uomo si impongono oggi con sempre maggior rilevanza sia per le domande che provocano alla professionalità e coscienza del personale medico e sanitario e sia per le esigenze organizzative e di governo delle attività di cura e assistenza collegate.

Come in altre attività umane anche in medicina esistono delle condizioni di disagio di vario livello, che trovano la loro ragion d'essere nella percezione di evidenti consistenze di vuoti conoscitivi. La presenza di uno scarto tra ciò che si vorrebbe sapere e ciò che si conosce o che si ritiene necessario conoscere o capire per risolvere il disagio viene identificata come la causa, o una delle cause, del problema. Questa distanza rappresenta una dimensione del vuoto conoscitivo che sta alla base di molte situazioni del disagio percepito. In seguito, nelle successive fasi di soluzione del problema sarà necessario esplicitarlo in uno o più obiettivi e quesiti la cui espressione non sarà immediata né facile, ma dovrà essere formulata partendo dal linguaggio dei fondamenti delle varie discipline sostantive ma anche secondo un "linguaggio scientifico" universalmente condiviso. Il quesito è tipicamente una domanda, relativa ad uno o più fenomeni associati, di specifico interesse clinico, sanitario, sociale, ecc., che si caratterizza per il contenuto ma anche per il diverso livello conoscitivo (esplorativo, relazionale, sperimentale) ed a cui si intende dare risposta mediante una specifica azione di ricerca.

In tale dinamismo la sempre maggior disponibilità di accesso ad informazioni di vario livello—conoscenze e saperi già acquisiti, incluse le grandi basi di dati—permette di verificare se ciò che viene identificato come problema conoscitivo lo sia realmente o non sia piuttosto l'effetto di insufficiente utilizzo di quanto già realizzato e disponibile in rete od anche effetto di una sorta di incompletezza nelle informazioni evidenziate e/o evidenziabili tra quelle contenute nei fatti finora recepiti. Questo attento confronto in realtà potrà anche offrire elementi di riflessione per raffinare e migliorare il progetto ed il piano della ricerca programmata.

I dubbi, i quesiti e le ipotesi che nascono e maturano nell'uomo, sono il risultato di una riflessione intorno al reale empirico osservato e rilevato o, talvolta, anche solo congetturato; riflessione fatta con fantasia e desiderio di verità sulle cose che  
***(\*) da non diffondere senza l'autorizzazione degli autori***

circondano e riguardano l'uomo. Lo stupore sembra essere uno dei principali fattori aggreganti e stimolanti nella ricerca di soluzioni anche parziali. La formulazione del dubbio ed il conseguente avvio della ricerca si basa anche sull'intuizione. Tipicamente, ma non esclusivamente, tale dubbio nasce dalla riflessione sull'osservato ed esprime in maniera anche irrazionale un bisogno presente in ogni uomo di convergere verso la verità su di sé, sugli altri, sul mondo, sulla vita. Nello sviluppo e produzione delle conoscenze non bastano quindi solo segni oggettivi, comunque parziali, ma questi devono essere accompagnati da ipotesi e/o congetture anche elementari su quanto di sconosciuto e/o inadeguato ci sta attualmente provocando. L'attività di ricerca è la risposta al problema del bisogno conoscitivo dell'uomo, alla sua inquietudine: la fame di conoscenza è risolvibile dal possesso dei *saperi* che intendono rispondere al bisogno di *sapere* e *conoscenza* sempre presenti in ogni uomo.

*La meraviglia è la madre del bello e buono* (W. Goethe).

*Lo stupore è il seme della conoscenza* (F. Bacon).

*La vera intelligenza nella ricerca sgorga dalle sorgenti vive della meraviglia e dallo stupore* (P. Florenski).

## **2. La ricerca scientifica: una risposta razionale al bisogno di conoscenza.**

La risposta operativa ai quesiti conoscitivi si esprime con l'investigazione del reale –quando l'osservazione diretta non riesce a far emergere aspetti nascosti con le loro interazioni– talvolta anche provocata ad arte tramite appositi esperimenti (come è comune in fisica, in tecnologia e, al netto di eventuali connessi problemi etici, in biologia e medicina). Tale risposta si esprime con la ricerca che nello specifico di questa argomentazione può essere definita “scientifica”.

*La ricerca scientifica è un processo che affronta problemi conoscitivi principalmente sulla base delle evidenze empiriche prodotte dall'osservazione e cerca di dare risposte utili all'uomo attraverso l'approccio razionale ed i fondamenti delle scienze.* La ricerca scientifica è quindi produzione e investigazione di ipotesi sul modo d'essere della realtà. Questa non può offrire certezze assolute e definitive, ma offre risposte argomentate mediante il linguaggio scientifico dei *saperi* temporaneamente consolidati e accettati universalmente anche a seguito di analisi critica razionale delle sue affermazioni. Nella scienza la risposta in quanto generalizzata è universale, seppur sempre limitata e accompagnata dall'incertezza. Rappresenta pur sempre un'opportunità basata sull'evidenza empirica che esprime il mondo razionale (Cartwright and Runhardt, 2014).

L'approccio scientifico che caratterizza il procedere del ricercatore nella sua attività di produttore di nuove conoscenze deve essere inserito in una logica di processo dinamico e teso verso obiettivi raggiungibili coerenti con i bisogni percepiti. Il processo deve caratterizzarsi per la sistematicità del rispetto della

sequenzialità logica e coerente dei passaggi e delle fasi connesse al piano di azione. Si configura come un'azione di *investigazione*, non solo di tipo speculativo, sul modo di essere della realtà che si esprime *nell'evidenza empirica* oggettivamente raggiungibile dai sensi dell'uomo con gli strumenti (fisici od anche concettuali) attualmente disponibili. Basata quindi su riscontri oggettivi generalmente acquisiti tramite strumenti vagliati nel tempo e accettati razionalmente dalla comunità scientifica. Il processo deve caratterizzarsi per la *sistematicità* del rispetto della sequenzialità logica e coerente dei passaggi e delle fasi conseguenti al piano di azione a ciò preposto. Deve essere *critica*, cioè osservare e valutare senza anteporre giudizi precostituiti. Questo si realizza cercando, nel rispetto della corretta metodologia, di ridurre al minimo gli eventuali fattori di disturbo, noti o presunti tali, che possono agire in maniera asimmetrica e incontrollata interagendo con quanto viene realmente osservato. Questo approccio non è comunque immune da valutazioni e riflessioni critiche. Nella filosofia della scienza si è aperto un dibattito sul rapporto tra razionalità e irrazionalità nella pratica scientifica che procede dal *problema della verifica* attraverso la quale si cerca la conferma dei rapporti tra leggi e fatti secondo il tradizionale metodo sperimentale.

La verifica, o meglio il controllo, può avvenire attraverso due procedimenti:

i) **Induttivo**: dall'osservazione **certa**, o ritenuta tale, di fatti singoli si arriva, per successive generalizzazioni, all'enunciazione di un asserto di portata generale, ma incerto, **ossia verosimile, più che probabile**. In questo caso si verifica in realtà la connessione logica tra diversi enunciati e si considera una teoria tanto più rilevante quanti più enunciati la confermano (e/o non la smentiscono) e sono da essa spiegati.

ii) **Deduttivo o falsificazionista**: da una legge generale verosimile si passa a casi particolari empiricamente accertati. Dato che non è possibile stabilire la certezza di una legge che include un numero indefinito di casi al suo interno, si può e si deve cercare la sua **verosimile falsità** sottoponendola a seri e articolati tentativi miranti a metterla in difficoltà.

In sintesi una legge è più o meno forte a seconda della sua resistenza ai tentativi di falsificazione che avvengono sia tramite controlli razionali, anche logico-deduttivi, e sia tramite analisi empiriche, tipicamente mediante metodi statistici, su dati appositamente acquisiti a seguito di esperimenti espressamente progettati.

### 3. Statistica e Ricerca.

In questa circostanza sorge allora la domanda: a fronte di specifici problemi/obiettivi conoscitivi qual è il contributo che può dare la Statistica? La Statistica si configura come un approccio organizzato attorno a saperi condivisi dalla comunità internazionale e basato sull'evidenza empirica che intende dare risposte scientifiche di natura quantitativa allo studio di "universi" o

“popolazioni” (non solo umane), ossia di intere collettività di fatti. Non è solo un insieme di tecniche, come comunemente può essere intesa, ma una scienza che acquista la sua valenza piena nell’ambito della ricerca scientifica sull’insieme di unità omogenee, classificate come tali secondo noti, condivisi e protocollati criteri di inclusione ed esclusione, quali le popolazioni umane (soggetti sani, malati, a rischio, ecc.) oppure di entità anche complesse definite, come: ospedali, scuole, farmaci, configurazioni genetiche, produzioni industriali, ecc.. Intende offrire dapprima informazioni puntuali, anche se non univoche, sulle principali dimensioni/aspetti caratterizzanti ogni fenomeno empirico degno di interesse: la sua *consistenza* o proprietà di essere misurata secondo criteri oggettivi o almeno intersoggettivi condivisi e la sua attitudine ad assumere dimensioni diverse (non invariabili) rilevabili nel momento dell’osservazione, nota anche come *variabilità*. La Statistica non risponde a tutti i problemi di natura conoscitiva e scientifica, ma offre comunque utili contributi che si esplicano in ogni passaggio del processo di produzione dell’informazione. Per lo più si intende che un’affermazione è qualificabile come statistica se per essere conseguita, resa manifesta, comunicabile, archiviata e utilizzabile necessita di elaborare in modo essenziale informazioni contenute in dati fattuali rilevati e acquisiti (Casey et al., 2016). Il valore aggiunto della Statistica è rappresentato in particolare dalla capacità di dimensionare, spiegare e utilizzare quella caratteristica di ogni fenomeno empirico che è la variabilità e che costituisce l’elemento immanente e identificante l’incertezza presente in ogni affermazione puntuale, ma sintetica, compiuta sugli eventi empirici in studio. Il *dato* –messaggio, segnale elementare, percepibile ai sensi dell’uomo anche tramite strumenti fisici– diventa *statistico* nel momento in cui entra in un processo di produzione di informazioni statistiche finalizzate ad incrementare la conoscenza e a ridurre l’incertezza nel processo decisionale. Il dato statistico è la cosiddetta materia prima del processo di produzione delle informazioni statistiche, è un messaggio codificabile con il valore potenziale di diventare strumento decisionale –l’*informazione statistica*– nei processi di scelta fra ipotesi alternative. In buona sostanza si può affermare che tra i compiti principali della statistica vi è quello –non già di annullare l’incertezza nelle affermazioni conseguenti alla ricerca (proposito assolutamente irrealizzabile anche in linea di principio)– di progettare e applicare metodi di acquisizione e di analisi di dati empirici in grado di ridurre al minimo livello tale latente incertezza. Ossia di dotarsi e di usare metodi di analisi “massimamente efficienti nel conservare ed elaborare in forme intelligibili il contenuto informativo dei dati (Pesarin, 2018).

#### **4. La Ricerca sulla salute dell’uomo.**

Nello studio dell’insorgenza ed evoluzione di uno specifico evento associato alla salute dell’uomo, particolare attenzione viene assegnata alla *relazione di causalità*

o *causa-effetto* , cioè il nesso esistente fra l'evento in osservazione ed altri fattori agenti che ne possono essere in qualche misura causa o concausa. Sono (tipicamente malattie, comportamenti, successi, ecc.) legati all'uomo o ad una struttura del sistema o al servizio e che si ritiene che siano da ritenersi conseguenze di un fattore, o agente, generalmente esterno all'uomo/servizio (Feinstein, 1974). Nella ricerca applicata l'attenzione viene posta sulla circostanza del cosiddetto *cambiamento* di una determinata condizione dell'unità di osservazione, da una situazione precedente di assenza a quella di presenza dell'evento osservato (malattia, comportamento, gradimento, avvenuto successo, ecc.). La valutazione di tale cambiamento, osservato presso l'insieme di unità omogenee, oggetto d'indagine, si ha confrontando quanto attualmente osservato con ciò che è avvenuto in altri soggetti, il più possibilmente simili ai primi, i quali però non hanno conosciuto l'esperienza dell'*esposizione* e/o il contatto con il fattore in studio ritenuto essere, nell'ambito di quella ricerca, una possibile causa. La natura del processo è naturalmente caratterizzata, e perciò influenzata, da condizioni di incertezza. Le conclusioni, pertanto, non possono che essere affermazioni che sono tali solo se empiricamente argomentate e comunque e inevitabilmente associate ad un margine di affidabilità/probabilità/verosimiglianza.

In tale ricerca, le **condizioni necessarie** per poter affermare di essere in presenza di una relazione di causalità fra un agente di rischio e un evento osservato (modello elementare di causa-effetto) sono : a) *Consistenza*; b) *Intensità*; c) *Specificità*; d) *Sequenzialità temporale degli eventi*; e) *Logicità*. Laddove non si verifichi almeno una di queste condizioni, l'associazione fra i due eventi è da considerarsi spuria, cioè verosimilmente imputabile a cause diverse dalla relazione di causalità. L'esistenza dei presupposti a), b) e c) è valutata in base a metodi e tecniche statistiche; i criteri d) ed e) afferiscono alla conoscenza sostantiva (Feinstein 1974; Armitage and Berry, 1998).

Come già affermato e descritto l'attività di ricerca sulla salute si basa su *evidenze*, alcune delle quali anche soltanto potenziali e che, all'interno di uno specifico problema conoscitivo, potrebbero diventare *dati*. Evidenze che possono non avere –come in ogni ambito di ricerca– le medesime capacità informative. Queste devono essere note al ricercatore e all'utilizzatore al fine di saperle valorizzare adeguatamente.

Lo Schema 1 che segue descrive i diversi e crescenti livelli di informazione offerti dalle differenti tipologie di approcci e disegni di indagine e studio.

Le *opinioni*, tipicamente affermazioni non adeguatamente argomentate, offrono un livello primitivo e generalmente poco informato e specifico sull'evento in oggetto, possono costituire orientamenti generali, comunque da verificare, per un'utile base per riflessioni successive che potranno trovare nei disegni di indagine organizzati una risposta adeguata ai bisogni conoscitivi. Lo *studio di casi*, necessario in presenza di popolazioni rare o con bassa diffusione, è un approccio che deve integrarsi con analisi e simulazioni provenienti dagli studi

osservazionali descrittivi più articolati e su popolazioni più ampie. Gli *studi esplorativi* condotti mediante indagini o *survey trasversali*, fotografano la situazione di una popolazione in una particolare circostanza storica e offrono un primo quadro della diffusione delle condizioni di interesse specifico (malattia, malessere, e fattori di rischio, ecc.) nella popolazione obiettivo –le frequenze marginali delle modalità di esposizione e di salute sono stime campionarie puntuali della loro consistenza nella popolazione generale– senza con ciò permettere di fornire stime attendibili delle relazioni di causalità.

Schema 1. Livello di potenzialità informativa dell'evidenza empirica prodotta dalle differenti impostazioni e disegni di indagine.

ALTA EVIDENZA	INFORMAZIONI	
↑	1. Revisione sistematica di studi sperimentali (farmacosorveglianza, metanalisi, ecc.) .	Efficacia basata sulla evidenza.
	2. Singoli Studi Sperimentali, Randomizzati.	Efficacia sperimentale.
	3. Studi osservazionali: Longitudinali di Coorte, Caso-Controllo.	Incidenza, Prevalenza, Analisi di causalità.
	4. Studi esplorativi, Trasversali, Cross-over.	Stime della diffusione dei fattori di rischio e delle condizioni di salute nella popolazione.
	5. Studio di casi.	Simulazione di ipotesi di causalità.
	6. Opinioni.	Orientamenti generali, introduttivi.
BASSA EVIDENZA		

Gli *studi osservazionali* più attrezzati offerti dai disegni *longitudinali di coorte e caso-controllo* offrono –con diversa attendibilità– stime non distorte delle misure della relazione di causalità fra fattori di rischio ed effetto (rischio assoluto, relativo, attribuibile, ecc.).

Decisamente più articolato ma più informativo è lo *studio sperimentale* ove, nella sua formulazione più tipica, viene *programmato un apposito esperimento* puntualmente descritto in un *protocollo* al fine di *valutare l'efficacia* di un trattamento o di un intervento, mediante il confronto degli eventi osservati in un gruppo di unità che ricevono il trattamento sotto esame con quelli osservati in un altro gruppo, comparabile al primo, di pazienti sottoposti al cosiddetto trattamento di controllo. L'informazione prodotta si individua nella cosiddetta *efficacia sperimentale*. Il criterio caratterizzante di questi disegni è la *randomizzazione* (tecnicamente: l'attribuzione “a caso” dei trattamenti ai soggetti elicitati, ossia assicurando a questi la medesima probabilità di ricevere il trattamento attivo) non solo delle unità ma anche ai singoli gruppi ai quali saranno assegnati casualmente

i diversi livelli sperimentali secondo le modalità (cieco, doppio cieco, ecc.) e declinazioni ormai note e collaudate nella ricerca clinica (Armitage and Berry, 1996; Deaton and Cartwright, 2016)

I soggetti dei diversi gruppi devono essere reclutati, trattati e seguiti durante lo stesso periodo di tempo o comunque secondo modalità condivise e protocollate in termini scientifici così da ridurre massimamente effetti indesiderati associati a cause al di fuori del controllo dello sperimentatore. Nel disegno sperimentale più elementare con un agente (variabile indipendente) ed un effetto atteso (variabile dipendente), il ricercatore manovra con ampia autonomia la variabile indipendente, comunque conservando precisa memoria delle scelte effettuate e allineate con quanto indicato nel protocollo, ed è interessato a verificare alcune ipotesi sul corrispondente comportamento della variabile dipendente.

Questo tipo di modello consente di verificare integralmente ipotesi predittive sulla relazione *causa-effetto* tra le variabili in gioco. Una specifica applicazione in ambito medico è rappresentato dal “**trial clinico**”, studio *farmacologico, biomedico* o *salute-correlato* sull'uomo, che segue dei protocolli predefiniti. Lo scopo ultimo è quello di verificare che una nuova terapia sia **sicura, efficace e possibilmente migliore** di quella normalmente impiegata e correttamente somministrata. Tali studi vengono preceduti dagli studi pre-clinici i quali analizzano l'efficacia dei farmaci o di altro trattamento sanitario, richiedono l'attivazione ed esecuzione di procedure **scientificamente adeguate**. La fase preliminare è un'estesa **ricerca pre-clinica**, per ottenere risultati che dimostrino il potenziale “distruttivo” ed efficace della cura nei confronti della malattia. La ricerca pre-clinica deve determinare le potenzialità e caratteristiche farmacologiche o applicative delle nuove tecniche (quali le molecole nelle terapie “*genetiche*”), sia con ricerche sugli animali, sia con studi sperimentali su cellule, anche dei malati sotto cura (i cosiddetti studi “*in vitro*”).

Questo è necessario perché quando le ipotesi scientifiche che hanno portato alla messa a punto di un nuovo trattamento (relative ad una nuova terapia o cura) sono state simulate in laboratorio, saggiate e provate in circostanze controllate, il trattamento potrà essere somministrato ai pazienti su larga scala. Il trial clinico è il fondamento stesso della *medicina basata sull'evidenza*, offrendo modalità organizzate e scientifiche per le migliori prove e sperimentazioni possibili e ottenendo risultati sostanziali sui vantaggi e gli svantaggi dei diversi trattamenti. Senza i trial clinici, il progresso nella lotta contro le malattie sarebbe bloccato, poiché nessuna sperimentazione è uguale ad un'altra sebbene condotte su uno stesso agente o popolazione. Nei fatti, le sperimentazioni non si configurano come delle mere simulazioni artificiali sul campo.

Negli studi sperimentali la *randomizzazione* è vincolo ricorrente il quale non si risolve nella definizione della numerosità campionaria come che esistessero “numeri magici statisticamente rilevanti”. Il compito principale dell'attribuzione “a caso” dei soggetti ai trattamenti è quello di “neutralizzare” il più possibile

l'intervento di propensioni distorcenti anche indotte dall'inconscio desiderio dello sperimentatore di far emergere oltre il lecito effetti desiderati, con ciò implicitamente violando il principio di neutralità dello sperimentatore nei riguardi dell'esperimento.

Una dimensione assunta troppo spesso a vincolo assoluto è rappresentata dalla dimensione campionaria dell'esperimento. Le domande: "Quanti casi?", "Quanti pazienti da trattare?" assumono spesso toni estremi e paralizzanti. La numerosità campionaria è una (importante) dimensione dell'esperimento, è strettamente associata a: i) il livello di *potenza* desiderata dall'esperimento, ossia al tasso di credito da attribuire ai conseguenti risultati; ii) la possibile differenza tra i diversi e alternativi trattamenti; iii) la variabilità del fenomeno in studio. Sicuramente la numerosità campionaria influenza la valutazione dei costi della sperimentazione prevista nel protocollo di studio, ma è anche influenzata –dal punto di vista etico– dalla stima dell'efficacia del trattamento. Vi è da osservare, a questo proposito, che solitamente uno studio clinico randomizzato è condotto su un campione di soggetti che per essere individuati hanno dovuto subire, sia per ragioni etiche e sia organizzative, tutto un insieme di momenti di selezione da parte: a) del responsabile sanitario, b) del team preposto allo studio; c) dal finanziatore dello studio; d) dal soggetto medesimo, che può o meno aderire allo studio. Per tali ragioni l'insieme dei soggetti che entrano nello studio non può essere visto –in senso stretto– come "un campione casuale" estratto dalla popolazione cui il trattamento e/o farmaco è diretto. Si tratta in questa fase di considerare la presenza dei cosiddetti *selection-biased samples*. Ovviamente, questo tipo di studio è sicuramente idoneo per la valutazione di efficacia del trattamento e/o del farmaco. Però, non potendo essere basato su "un campione casuale dalla popolazione target", non consente una immediata stima quantitativa dell'eventuale effetto a livello della popolazione di interesse (Altman, 2018). Effetto che, pertanto, dovrà essere quantificato con le dovute cautele, ricorrendo anche alle conoscenze acquisite tramite studi pre-clinici, bio-chimici, ecc. Dovrà inoltre essere verificato dai sempre più attenti contributi –quali ad esempio la *farmacosorveglianza* e la *meta-analisi*– di *medicina basata sull'evidenza* (Cowen, 2019).

## **5. Le aree di applicazione della ricerca sulla salute**

Gli ambiti di ricerca sulla salute dell'uomo spaziano fra molteplici finalità, applicazioni e potenzialità informative (v. Schema 2) ed offrono risposte sempre più specifiche, ma fra loro altamente correlate e funzionali al miglioramento dell'organizzazione dei servizi e delle fasi di cura e assistenza della salute umana. Fra questi assumono particolare importanza applicazioni in settori di ricerca marginalizzati nel passato quali gli studi in *Sanità Pubblica* ove la disponibilità di dati amministrativi potenziati con l'appaiamento dei dati sanitari del singolo



paziente possono fornire informazioni a completo e diretto beneficio dell'organizzazione dei servizi sanitari (Diehr et al., 1999; Brogini et al., 2004; Kindig, 2007; Busetto et al., 2013; Hayat et al., 2017). La strumentazione della statistica multivariata offre notevoli potenzialità di analisi in tale area di ricerca potendo disporre di basi dati consistenti ed omogenee a livello territoriale almeno regionale (Cartwright and Runhardt, 2014; Hayat et al., 2017; Buja et al., 2018).

Schema 2. Principali tipologie di ambito di ricerca sulla salute dell'uomo.

AMBITO	STRUMENTO	OBIETTIVO	POPOLAZIONE TARGET	PRODOTTI CONOSCITIVI
RICERCA DI BASE	Studi Pre-Clinici	Attività di assistenza	Gruppi di pazienti/ assistiti volontari	Prevalenze, relazioni statistiche
RICERCA CLINICA SPERIMENTALE	Studi Clinici Controllati	Efficacia Sperimentale	Campione statistico di pazienti	Incidenze, significatività cliniche e statistiche
RICERCA CLINICA POPULATION BASED	Studi Osservazionali	Efficacia osservata	Tutti i pazienti/assistiti	Misure del Rischio
RICERCA DI SANITA' PUBBLICA	Studi di Costi e Benefici	Efficienza	Popolazione generale assistita	Costi standard, unitari, indicatori produttività e rendimento del processo

## 6. Il contributo della statistica nella sperimentazione clinica: l'Inferenza Statistica.

Esaminando il risultato dell'analisi dei dati forniti dal campione al fine di trarne delle conclusioni da poter estendere alla popolazione target, occorre formulare tre domande fondamentali:

- i) Le conclusioni sono corrette per gli individui che compongono il campione?
- ii) Il campione “rappresenta bene la popolazione (selezionata)” da cui è stato effettivamente estratto?
- iii) Il campione “rappresenta bene la popolazione target”?

La risposta a queste tre domande deriva dai concetti di *validità interna* e di *validità esterna* di uno studio statistico (Sackett, 1979; Deaton and Cartwright, 2016).

La validità interna esprime e misura quanto i risultati di uno studio sono corretti per il campione di individui che sono stati effettivamente studiati. E' “*interna*” in quanto si applica alle condizioni del particolare gruppo di individui studiati, e non

necessariamente agli altri. Per esempio, negli studi clinici, la validità interna dipende dalla correttezza di impostazione dello studio stesso, da una corretta randomizzazione ai trattamenti dei soggetti elicitati, dalla scelta di buone tecniche diagnostiche e da un loro corretto utilizzo, da una metodologicamente corretta elaborazione dei dati, ecc. La validità interna rappresenta comunque *una condizione necessaria ma non sufficiente* perché uno studio sia utile, ossia informativo e credibile.

La validità *esterna* rappresenta il grado di generalizzabilità delle conclusioni tratte da uno studio. Per esempio, nel caso di uno studio epidemiologico clinico, essa risponde alla domanda “*Supponendo che i risultati di un determinato studio siano corretti e attendibili, essi si possono applicare anche ai miei pazienti?*”? E, in aggiunta alla precedente, “*A quali altre tipologie di pazienti possono essere applicati?*” (Reiss, 2017).

In questo processo, occorre ovviamente distinguere tra stime in ottica predittiva, concernente l'applicazione anche a soggetti diversi da quelli effettivamente esaminati, e verifica d'ipotesi sulla presenza attiva di effetti terapeutici. Ad esempio, va qui ancora ricordato che usualmente le sperimentazioni cliniche vengono di fatto condotte su soggetti consenzienti compresenti in una o più strutture ospedaliere e che pertanto rientrano tra i cosiddetti campioni da una popolazione selezionata: ossia costituiscono campioni selection-biased. Le risultanze delle analisi di questi campioni non sono quindi direttamente generalizzabili alla cosiddetta popolazione target. Generalizzabilità che dovrà quindi essere condotta con estrema cautela e che non sarà sempre possibile effettuare a meno di non adottare delle metodologie adeguate. A questo proposito va sottolineato che qualora nelle elaborazioni statistiche con dati selection-biased vengano adottati per problemi di verifica d'ipotesi dei metodi cosiddetti non-parametrici, essenzialmente dipendenti dai soli dati osservati, allora vi può essere una maggiore e meno rischiosa generalizzabilità (Pesarin, 1990; Pesarin and Salmaso, 2010). Si consideri ancora la seguente domanda: “*Questo farmaco ha dato risposte significative sui soggetti effettivamente osservati?*”, questa potrà essere tranquillamente identificata nella seguente: “*Il farmaco dà risposte significative sulla popolazione selezionata, quella di cui il campione effettivo è stato estratto?*” L'identificazione o sovrapponibilità delle due domande, e quindi delle risposte della prima sulla seconda è operazione molto delicata e complessa e può richiedere il ricorso anche a nozioni e conoscenze esterne ai dati effettivamente raccolti nel trial. In altre parole, la validità esterna misura il grado di idoneità (verità) dell'assunto secondo cui gli individui studiati con il campione sono “uguali o almeno somiglianti” ad altri pazienti non osservati, ma potenzialmente osservabili, affetti dalla stessa condizione.

I contributi della statistica nella ricerca medica sono stati –in linea generale– e come già ribadito notevoli. In particolare, ma non esclusivamente, nella sperimentazione clinica l'approccio inferenziale della metodologia statistica è

rilevante e qualificato. Permette di acquisire informazioni statistiche (media, frequenza, distribuzione di una caratteristica, relazione tra caratteristiche, ecc.) su di una popolazione usando le osservazioni di un campione da essa estratto. Gli strumenti principali dell'Inferenza Statistica sono l'informatore campionario (o statistica campionaria), ottenuto mediante una o più operazioni, non esclusivamente aritmetiche, su dati campionari. Questo strumento ha le seguenti caratteristiche:

- i) Per ogni campione si ottiene un solo risultato.
- ii) Il risultato è un numero realmente conseguito o conseguibile nell'analisi.
- iii) Variando il campione il risultato può variare, ossia il valore conseguito non è sempre invariante, pertanto si dice che il processo genera una distribuzione campionaria.

I prodotti principali dell'Inferenza Statistica sono fondamentalmente due: a) la *Stima campionaria*; b) la *Verifica d'Ipotesi*.

La verifica d'ipotesi è una procedura statistica che porta ad accettare (temporaneamente) come vera, o respingere come falsa, una proposizione che esplicita con il linguaggio statistico un problema conoscitivo proprio dello specifico ambito di applicazione. Quest'ultimo dovrà essere tradotto in linguaggio statistico ad esempio come: "La media aritmetica del livello di un parametro clinico nei pazienti trattati con X è maggiore (o semplicemente diverso o minore) rispetto ai pazienti trattati con Y o non trattati con X".

Ogni ipotesi statistica è una affermazione relativa a condizioni reali/empiriche (cliniche, sanitarie, ecc.) espresse mediante il linguaggio e la simbologia della statistica. Si procede da una *ipotesi* base o *nulla*, indicata con  $H_0$ , che viene posta in *alternativa* da una indicata con  $H_1$ . Le operazioni di verifica della verosimiglianza dell'ipotesi avvengono sempre e comunque in condizioni di incertezza –salvo lo studio su di un fenomeno immutabile nel tempo e nello spazio (più ipotetico che reale)– poiché si basano sempre su osservazioni parziali (campione) della realtà empirica, per definizione incontenibile e indefinibile in uno spazio di ipotesi circoscritto. Lo strumento operativo statistico che utilizza le informazioni parziali del campione è la *funzione test* (test:  $t$ ,  $Z$ ,  $F$ , chi-quadrato, ecc.) che si esprime come un informatore standardizzato applicato ai dati campionari e di cui si conoscono le proprietà sotto certe condizioni. Le conclusioni sulla verosimiglianza delle ipotesi hanno un margine di affidabilità che la scienza statistica riesce a dimensionare, prevedere e controllare. Questa operazione di dimensionamento dell'affidabilità dei tests si esprime eminentemente con il dimensionamento di due tipologie di errori: *errore di I tipo  $\alpha$*  e *errore di II tipo  $\beta$* . Questi cosiddetti errori, che non sono e non vanno confusi con dei veri e propri "sbagli", per definizione sono delle probabilità che corrispondono alle misure del "grado di accadimento" di conseguire delle conclusioni sbagliate.

### **6.1 Errore di I tipo $\alpha$ .**

E' l'errore che si commette quando si rifiuta  $H_0$  quando questa è vera (probabilità di un risultato significativo spurio). E' abituale fissarla a 0,05, vale a dire una probabilità del 5%. È la misura della significatività statistica del test. Il dimensionamento di questa probabilità naturalmente condiziona anche la numerosità campionaria: quanto più piccola si desidera sia  $\alpha$ , tanto maggiore deve essere la dimensione campionaria ove si voglia mantenere la medesima potenzialità dell'indagine. Questa probabilità, comunemente anche detta del falso positivo, è la probabilità di classificare un soggetto (affermazione) come positivo quando in realtà è negativo.

Nelle usuali applicazioni tipici valori di  $\alpha$  sono: 0,10, 0,05, 0,01.

### **6.2 Errore di II Tipo $\beta$ .**

Errore che consiste nell'accettare  $H_0$ , essendo questa falsa. O meglio: la probabilità di rifiutare  $H_1$  (accettando  $H_0$ ) quando  $H_1$  è vera. Il suo complemento a 1 è la *potenza* del test:  $1 - \beta$ . *La potenza è la probabilità di ottenere una significatività statistica quando realmente la sostanza provata è efficace.* La potenza quindi è la probabilità di rifiutare  $H_0$  quando è falsa.

Tipici valori di  $\beta$  sono: 0.20, 0.15, 0.10, 0.05. Quanto più piccola si desidera  $\beta$  (e quindi più alta la potenza), tanto maggiore deve essere la dimensione del campione ovviamente a parità di errore di I tipo  $\alpha$  e a parità di effetto del trattamento provato. Questa probabilità, comunemente anche detta del falso negativo, è la probabilità di classificare un soggetto (affermazione) come negativo quando in realtà è positivo.

Vi sono anche altri errori, poiché altre sono le circostanze in cui si fanno scelte sempre in condizioni di incertezza.

### **6.3 Errore di III tipo.**

Errore che consiste nel considerare superiore un trattamento che in realtà è inferiore. Ad esempio, una siffatta tipologia di errore può ricorrere quando con un cosiddetto test a due code, in cui nell'analisi l'effetto sia risultato significativo, si voglia poi inferire sulla effettiva direzione, positiva o negativa, dell'effetto testato.

### **6.4 Errore $\alpha$ globale.**

È l'errore  $\alpha$  che si commette nel fare confronti multipli (ad es: a seguito di livelli significativi dell'ANOVA fra più di due campioni indipendenti, si voglia poi inferire su un eventuale ordinamento degli effetti campionari effettivamente riscontrati rispetto ai livelli di trattamento).

### **6.5 Il livello di Significatività $p$ .**

Per meglio comprendere gli errori sopra descritti è opportuno precisare il significato compiuto del *livello o misura della probabilità*  $p$ . Questa rappresenta la probabilità che, subordinatamente all'ipotesi nulla, il test adottato “superi il”, ossia sia “più estremo del”, valore effettivamente osservato nell'analisi assumendo che valori grandi del test siano evidenza contro l'ipotesi nulla. Tanto più il valore  $p$  è “piccolo”, tanto più il valore osservato del test è “grande”, ossia tanto meno verosimile risulta l'ipotesi nulla (quella assunta a priori temporaneamente come vera espressione di quanto già appurato e consolidato come vero e perciò in genere da smentire). Essa pertanto sarà rifiutata, in favore dell'ipotesi alternativa  $H_1$ , se questa  $p$  è inferiore al livello  $\alpha$  convenuto. Ciò non significa avere ottenuto un successo, come viceversa affermare che c'è stato insuccesso se fosse  $p > \alpha$ .

In sintesi, il valore di  $p$  si può definire come la probabilità di osservare un valore della *funzione test* uguale a o “più lontano di” quello osservato nell'ipotesi  $H_0$  di nessuna efficacia. Di fatto il valore  $\alpha$  per un test  $T$  non è nient'altro che il valore critico del test statistico  $p$  associato a  $T$ . Si può interpretare come una prima misura dell'evidenza contro l'ipotesi nulla. Va sempre interpretata alla luce della numerosità campionaria. Un suo valore, ad es.  $p = 0.01$ , a parità di altre condizioni esprime un'evidenza negativa. A parità di effetto, con funzioni test ben costruite il valore è tanto più piccolo in probabilità quanto più alta è la numerosità campionaria. Per cui, a parità di numerosità campionaria, l'effetto del trattamento deve essere tanto più forte affinché l'evidenza del risultato dell'inferenza conservi la potenza desiderata. Per questa ragione si consiglia sempre di formulare anche una stima intervallare dell'effetto, da discutere attentamente con gli esperti e operatori (medici, epidemiologi, sanitari, ecc.). L'ampiezza dell'intervallo è in relazione inversa della numerosità campionaria e della variabilità campionaria. Con grandi campioni, piccole differenze senza alcuna importanza clinica possono risultare statisticamente significative ma non essere –appunto– clinicamente rilevanti. Se si volesse estendere l'analisi anche alla “*significatività clinica*” occorrerebbe riformulare le ipotesi da sottoporre a verifica statistica: invece di puntare l'attenzione su un unico valore ad esempio “*della media*”, definire invece  $H_0$  come “*l'effetto medio del trattamento è interno ad un intervallo clinicamente stabilito*” e procedere poi con le analisi statistiche appropriate. Le analisi che conducono ad affermare che un risultato è “*statisticamente non significativo*” non permettono altri tipi di interpretazioni quali quelle di essere in presenza di una “*prova di equivalenza*”. Per la verifica di questa tipologia di ipotesi si devono avviare analisi con presupposti differenti rispetto a quelli usuali. Infatti, tali analisi si propongono di provare che la differenza media fra gli effetti di due agenti/azioni/trattamenti (farmaci, terapie, ecc.) è interna ad un determinato intervallo indicato come clinicamente irrilevante per giustificare l'asserzione di essere in presenza di un effetto empirico di natura clinica, terapeutica, ecc., del trattamento. Sono studi impegnativi per i quali viene richiesta una numerosità

campionaria in generale abbastanza più consistente degli altri studi sperimentali. Ad ogni modo, i risultati della ricerca medica vanno sempre giudicati secondo la loro importanza declinata dalla prospettiva clinica, epidemiologica, sanitaria, ecc., e non solamente secondo la loro “significatività” statistica. Il valore di  $p$  non basta da solo, ma deve essere accompagnato dalle altre evidenze dell’esperimento. Fra tutti si ricorda il trattamento dei *persi di vista* frequente in molti trial. Significatività statistica non è sempre sinonimo di significatività o rilevanza clinica (Schor and Karten, 1966).

## 7. Statistica ed Etica .

In ambito medico la ricerca scientifica deve caratterizzarsi per la sua capacità di servire il bene comune nella salute, nella società e per l’uomo, cercando di offrire conoscenze e strumenti che tutelino la dignità delle persone in particolare di coloro che si trovano in condizioni di difficoltà.

Quanto affermato assume il pieno significato di fronte alle seguenti due domande –non alternative ma distinte– avanzate dall’evidenza scientifica clinica. La prima domanda proviene dal medico curante, che ha in cura il sig. Gervasio e coerentemente con le finalità del suo status professionale, può venire così espressa: “*Qual è il miglior trattamento per la condizione (clinica) del paziente sig. Gervasio?*”.

La comunità medica avanza la seguente seconda domanda per affrontare la cura di una malattia  $M$  diffusa fra la popolazione generale dei propri pazienti: “*Quale trattamento (terapia) può essere considerato migliore per la malattia  $M$ ?*”

Quest’ultima domanda viene affrontata dai ricercatori clinici professionisti della ricerca scientifica mediante la sperimentazione clinica che si pongono l’obiettivo di produrre evidenze scientifiche forti e utili ai pazienti di oggi e di domani.

Entrambe le domande interpellano dimensioni diverse ma complementari dell’etica, quella individuale dell’operatore e quella collettiva della comunità scientifica. L’etica individuale si occupa di rispondere alla domanda: “*Il paziente riceve il trattamento e la cura considerate maggiormente indicate per la sua (attuale) condizione?*”. L’etica collettiva intende rispondere, o meglio esige la risposta, alla domanda: “*Viene garantito il raggiungimento del progresso scientifico nel modo più efficiente/efficace possibile al benessere collettivo (dei pazienti) presente e futuro?*” (Pegoraro et al., 2016).

Il dibattito è aperto da tempo e fin dagli anni ’70 veniva affermato: “... uno dei problemi più seri della ricerca clinica è sottoporre i pazienti al rischio di lesioni, sofferenze od altri inconvenienti in sperimentazioni: i) con troppo pochi individui per raggiungere conclusioni valide; ii) con più soggetti di quanto richiesto dall’obiettivo proposto; iii) o con modalità di randomizzazione o “doppio cieco” inadeguate” (May, 1975).

Gli studi, in particolare quelli basati sulla sperimentazione, devono essere scientificamente ed eticamente fondati (Ambroz et al., 1978). Gli elementi di criticità etica negli studi sperimentali sono fondamentalmente i seguenti:

- i) (In) Certezza fra diverse soluzioni terapeutiche equivalenti (revisione scientifica).
- ii) (In) Sufficiente conoscenza dei farmaci o terapie.
- iii) Qualità dello studio.
- iv) Adeguata documentazione sulle proprietà cliniche della terapia/profilassi, sotto sperimentazione.

La statistica non è d'altronde esente da rischi analoghi. Errori nell'uso della statistica possono verificarsi a tutti gli stadi di un'indagine e uno di essi può essere sufficiente a vanificare l'intero lavoro. Uno studio può essere concepito o eseguito perfettamente, ma se viene analizzato in maniera scorretta le sue conseguenze possono essere non meno gravi di quelle di uno studio sbagliato in partenza. Infine non è etico pubblicare risultati scorretti o fuorvianti o omissivi (McClimans, 2010; Altman, 2018).

Le implicazioni etiche di una ricerca carente dal punto di vista statistico possono venire così riassunte:

- i) l'inadeguato coinvolgimento dei pazienti, esposti a rischi ed inconvenienti ingiustificati;
- ii) il cattivo uso delle risorse, compreso il tempo del ricercatore impiegabile in attività più utili;
- iii) le conseguenze della pubblicazione di risultati erronei, implicanti spesso la necessità di intraprendere ulteriore lavoro, altrimenti non necessario (May, 1975).

Queste fonti di distorsione trovano oggi una maggiore attenzione nella pratica della *medicina basata sull'evidenza* (Best Evidence Medicine) che si configura come un processo di apprendimento, in cui l'assistenza dei propri pazienti crea la necessità di informazioni clinicamente importanti e di affidabilità nota su diagnosi, prognosi, terapia e altri problemi clinici e sanitari. Apprendimento e pratica che permettono di valutare criticamente le informazioni per la loro validità (studio metodologicamente corretto, vicinanza alla verità) e utilità (esportabilità e applicabilità clinica efficace), per la capacità di integrarle con la propria esperienza clinica e applicarla nella pratica ed infine valutare le proprie prestazioni in termini di risultato finale oltre che di prodotto e prestazione.

La Statistica Medica fornisce alla sperimentazione clinica il cosiddetto "gold standard" per stabilire inferenze tipo causa-effetto ed offre allo sperimentatore clinico gli strumenti per operare sulla base di una attenta combinazione di teoria, pratica ed etica (Sackett, 1979).

Nello Schema 3 vengono sinteticamente messe a confronto le diverse finalità della Statistica e dell'Etica al fine di riassumere quanto già descritto e di valorizzarne i diversi contributi.

Schema n.3. Finalità della Statistica e dell'Etica nella ricerca.

<b>STATISTICA vs ETICA</b>	
<b>Statistica</b>	<b>Etica</b>
Persegue ciò che è <i>vero</i> .	Persegue ciò che è <i>giusto</i> .
Si basa sulla verosimiglianza	Si basa sulla moralità, sul fare il bene allo/dell'uomo.
Riguarda ciò che <i>potrebbe</i> accadere	Riguarda ciò che <i>dovrebbe</i> accadere
Opera su eventi incerti	Opera su eventi incerti

Anche nella **Carta Europea dei Ricercatori** (2005) viene dichiarato che tra i “*principi generali e requisiti applicabili ai ricercatori*” sono espressamente contemplati: “*la libertà di ricerca, i principi etici, la responsabilità professionale e la diffusione e valorizzazione dei risultati*”. I ricercatori d'altronde sono esplicitamente invitati “*ad aderire alle pratiche etiche riconosciute e ai principi etici fondamentali applicabili nella o nelle loro discipline nonché alle norme etiche stabilite dai vari codici nazionali, settoriali o istituzionali*”.

Queste linee potranno essere tanto più autorevoli quanto più sapranno offrire alla ricerca scientifica una credibilità superiore ai timori associati ai problemi che intende risolvere e di dimostrare la capacità di ristabilire la centralità alla persona mediante i cambiamenti che propone.

*Lo scienziato non è colui che da le giuste risposte ma colui che sa porre le giuste domande* (C. Levy Strauss).

*Pensare bene è principio della morale* (Blaise Pascal ).

## **8. Conclusione.**

La ricerca è una attività di équipe non di élite, vede coinvolte figure professionali, sensibilità e saperi differenti. E' fondamentale trovare un'unità nella diversità, ovvero trovare una base comune e un linguaggio che rendano comparabili e confrontabili i diversi contributi tecnici e scientifici per saper apprezzare anche i risultati non univoci. La statistica nella sua forte identità culturale si propone di favorire maggiormente la diffusione della cultura della metodologia della ricerca quantitativa nello studio di popolazioni anche attraverso una capacità di ascolto e di comunicazione e socializzazione di propri saperi per far crescere la cultura ed il pensiero scientifico. La scienza e la pratica medica devono saper sviluppare una maggiore sensibilità alla valorizzazione e utilizzazione dell'evidenza empirica attraverso l'approccio scientifico quantitativo nella consapevolezza che questo atteggiamento rappresenta la soluzione razionale più avanzata in condizioni di



limitate risorse (Cowen, 2019). L'incontro fra le diverse culture e figure professionali coinvolte nel processo devono riconoscere la diversità, la particolarità di ogni ambito di ricerca e di ogni ricerca, valutando l'importanza del contesto storico, culturale e operativo della quotidianità e dell'individualità degli attori e produrre una costruttiva sinergia finalizzata al bene dell'uomo, non solo delle scienze e nemmeno quello –talvolta– corporativo delle professioni.

## **Bibliografia**

Altman D., (2018): Un uso scorretto della statistica non è etico - *Il Pensiero Scientifico* <https://pensiero.it/in-primo-piano/.../un-uso-scorretto-della-statistica-non-è-etico> di Douglas Altman 6 giu 2018.

Ambroz A, Chalmers TC, Smith H, Schroeder B, Freiman Ja, Shareck EP., (1978). Deficiencies of randomized control trials. *Clin Res*; 26: 280 A.

Armitage P., Berry G., (1996): "STATISTICAL METHODS in MEDICAL RESEARCH"; (Ed. Italiana a cura di M. Bolzan) McGraw-Hill Libri Italia s.r.l.

Brogini A., Bolzan M. and Slanzi D. (2004). Identifying a Bayesian Network for the problem: Hospital and families: the analysis of patient satisfaction with their stay in hospital" In :*APPLIED BAYESIAN STATISTICAL STUDIES IN BIOLOGY AND MEDECINE M. Di Bacco, G. D'Amore, F. Scalfari, Eds. Kluwer Academic Publishers, 41-72.*

Buja A., Sartor G., Scioni M., Vecchiato A., Bolzan M., Rebba V., Sileni V., Chiarion P., Angelo C., Montesco M., Del Fiore P., Baldo V., Rossi C. R., (2018). Estimation of Direct Melanoma-related Costs by Disease Stage and by Phase of Diagnosis and Treatment According to Clinical Guidelines. *Acta Derm Venereol*; 98: 218–224.

Busetto M., Fusco V., Corbella F., Bolzan M., (2013). Predictive factors for oropharyngeal mycosis during radiochemotherapy for head and neck carcinoma and consequences on treatment duration. Results of mycosis in radiotherapy (MIR): A prospective longitudinal study. *RADIOTHERAPY AND ONCOLOGY*, ISSN: 0167-8140, doi: 10.1016/j.radonc.2013.05.040.

Cartwright, N.L. and R. Runhardt, (2014). "Measurement", in N.L. Cartwright and E. Montuschi (eds.), *Philosophy of Social Science: A New Introduction*, Oxford: Oxford University Press, pp. 265–287.

Casey J.A., Schwartz B. S., Stewart W. F., and Adler N.E. (2016). Using Electronic Health Records for Population Health Research: A Review of Methods and Applications . *Annu. Rev. Public Health* 2016. 37:61–81.

Cowen N.(2019), Street-level Theories of Change: Adapting the Medical Model of Evidence-based Practice for Policing. New York University School of Law and Nancy Cartwright, Durham University and UCSD CHES *Working Paper* No. 2019-02 [Produced as part of the Knowledge for Use (K4U) Research Project] Durham University February 2019.

Deaton A., and Cartwright N.,(2016). Understanding and misunderstanding randomized controlled trials. CHES *Working Paper* No. 2016-05 ISSN 2053-2660 The K4U project has received funding from the European Research Council (ERC) under the European Union's.

Diehr P., Yanez D., Ash A., Hornbrook M., Lin D. Y., (1999). METHODS FOR ANALYZING HEALTH CARE UTILIZATION AND COSTS . *Annu. Rev. Public Health*. 1999. 20:125–44

Feinstein A.R. (1974). A survey of the statistical procedures in general medical journals. *Clin Pharmacol Ther*;15:97-107.

Hayat M. J., Powell A., Johnson Tessa J., Cadwell B.L.. (2017). Statistical methods used in the public health literature and implications for training of public health professionals. *PLOS ONE* | <https://doi.org/10.1371/journal.pone.0179032> June 7, 2017

Kindig David A. (2007). Understanding Population Health Terminology. *The Milbank Quarterly*, Vol. 85, No. 1, 2007 (pp. 139–161) c 2007 Milbank Memorial Fund. Published by Blackwell Publishing.

May W.W. (1975). Etica e ricerca - Research Italy. <https://www.researchitaly.it/etica-e-ricerca/>. The composition and function of ethical committees. *J Med Ethics* 1975; 1: 23-9.

McClimans L., (2010), "A theoretical framework for patient-reported outcome measures", *Theoretical Medicine and Bioethics*, 31: 225–240.

Pegoraro R., Picozzi M., Spagnolo A. G. (2016) : LA CONSULENZA DI ETICA CLINICA IN ITALIA. LINEAMENTI E PROSPETTIVE. Piccin Ed., pp 218. Padova.

Pesarin F. (1990): On a nonparametric combination methods for dependent permutation test with applications. *Psychotherapy and Psychosomatics*; 54 (2-3): 172-179.

Pesarin F. (2018) Multi-one-sided tests for multivariate ordered categorical data with medical applications. *BIOSTAT at 25 Invited Essays in Theoretical, Biomedical and Social Statistics*. M. Di Bacco , F. Scalfari Editors. Edizioni ETS, Pisa, Italy.

Pesarin F. Salmaso L. (2010). *Permutation test for complex data: Theory, Application and Software*; Wiley, Chichester, UK.

Reiss J. (2017). Against External Validity. Durham University *CHESS Working Paper* No. 2017-03 Durham University November 2017.

Sackett D.L. (1979). Bias in analytic research. *J Chronic Dis*; 32:51-63.

Schor S., Karten I. (1966). Statistical evaluation of medical journal manuscripts. *JAMA*; 195: 1123-8.