

Valutazione Multidimensionale dell'Innovatività: peculiarità interpretative

Giovanni Pappagallo,¹ Carmine Pinto,²

Michela Cinquini,³ Ivan Moschetti,³ Giovanni Ravasio⁴

¹ Epidemiologo Clinico, Gruppo Metodologico Linee Guida AIOM e AIRO

² Head Medical Oncology - Clinical Cancer Centre, IRCCS - AUSL di Reggio Emilia

³ Unità Revisioni Sistematiche e Produzione Linee Guida, Istituto Mario Negri, Milano

⁴ Direttore Economia Sanitaria, Monza

Gruppo di Lavoro multidisciplinare

Hanno contribuito all'analisi e al confronto sui temi nel corso di alcuni incontri, oltre agli Autori:

Patrizia Popoli (Dir. Centro Nazionale Ricerca e Valutazione Preclinica e Clinica dei Farmaci - ISS)

Anna Maria Marata (Coord. Commissione Regionale Farmaco, Regione Emilia-Romagna)

Lucia Magnano (Area Farmaco e Dispositivi Medici, Regione Emilia-Romagna)

Franco Locatelli (Dir. Dip. Oncoematologia e Terapia cellulare/genica - Osp. Bambino Gesù, Roma)

Antonio Cuneo (Dir. UOC Ematologia - Arcispedale Sant'Anna - AOU di Ferrara)

Il nuovo modello di valutazione multidimensionale proposto da AIFA per la classificazione dei farmaci innovativi e dei medicinali oncologici innovativi, basato su bisogno terapeutico, valore terapeutico aggiunto e qualità delle prove è oramai al terzo anno di applicazione.

Nello spirito di verifica e possibili aggiustamenti di questo processo decisionale, basato su criteri espliciti e trasparenti, come previsto dai documenti allegati alle delibere AIFA (519-2017 e 1535-2017) istitutive dei criteri e della relativa procedura di valutazione, un Gruppo di Lavoro multidisciplinare propone il documento "Valutazione Multidimensionale dell'Innovatività: peculiarità interpretative".

Come avvenuto nello scorso anno con lo "Schema per la preparazione del dossier di richiesta di innovatività dei farmaci"¹ da parte delle Aziende (avente lo scopo di rendere più rapidi e riproducibili i processi di valutazione di questi farmaci ex-ante alla loro introduzione nel SSN), il presente documento – risultato di un percorso di incontri con interlocutori pubblici e privati del nostro Sistema Sanitario – vuole favorire l'analisi e il confronto su questi temi tra Istituzioni, Società Scientifiche e Aziende.

1. INTRODUZIONE

Il nuovo modello di valutazione multidimensionale proposto da AIFA per la classificazione dei farmaci innovativi e dei medicinali oncologici innovativi, basato su *bisogno terapeutico, valore terapeutico aggiunto e qualità delle prove* è oramai al terzo anno di applicazione.

A mano a mano che aumentava il livello di confidenza degli Operatori coinvolti nelle procedure, sempre più si è andata a precisare la definizione dei tre punti sopra indicati:

- il bisogno terapeutico come *sintesi delle debolezze di efficacia e di tollerabilità delle proposte terapeutiche disponibili* per un determinato scenario clinico: "qualche cosa che manca e di cui si avverte il bisogno";
- il valore terapeutico aggiunto come *sintesi degli effetti desiderati e indesiderati* di interesse specifico in relazione al bisogno terapeutico: "la presenza di un guadagno netto e incontrovertibile rispetto all'esistente";

- la qualità delle prove come *sintesi del grado di confidenza degli effetti desiderati e indesiderati* di interesse specifico: “*un risultato netto affidabile, credibile e trasferibile alla popolazione oggetto del quesito*”.

Da qui la proposta del Gruppo di Lavoro di individuare nel sistema GRADE lo strumento di riferimento per l'intero processo di preparazione della richiesta di accesso al Fondo per l'Innovazione. Come noto il sistema GRADE (dalla formulazione del quesito clinico alla valutazione globale della qualità delle prove) consente di “superare” alcuni schematismi caratteristici dei processi registrativi dei farmaci (es. la rigida distinzione tra le diverse fasi della sperimentazione clinica; la distinzione tra *endpoint* primario ed *endpoint* secondari; il braccio di controllo dello studio registrativo come unico riferimento per la valutazione del farmaco stesso; il disegno dello studio come indicatore principe di qualità), avvicinando il processo di valutazione del farmaco al ragionamento clinico che sottende alla formulazione della proposta terapeutica.

Il riferimento per la determinazione del valore terapeutico aggiunto è quindi rappresentato da quel trattamento attualmente considerato opzione terapeutica *standard* per la situazione specifica, indipendentemente dal fatto che esso sia stato o meno il braccio di controllo dello studio in esame. I parametri di valutazione da considerare primariamente sono quelli ritenuti dal consesso medico come essenziali² alla proposta terapeutica; devono quindi essere i migliori indicatori dei benefici attesi in relazione al *setting* di patologia considerato e degli eventuali danni ascrivibili al trattamento (reazioni avverse).

Il grado di confidenza (affidabilità) degli effetti rilevati su tali fattori rappresenta la base per il giudizio di qualità globale delle prove.

Valore terapeutico aggiunto (valutato con 5 livelli, *da massimo ad assente*) e qualità delle prove (valutata con 4 livelli da *alta* a *molto bassa*) concorrono entrambi alla valutazione globale dell'intervento in termini di innovatività. I giudizi sono indipendenti; per es. un valore aggiunto *massimo* rimane *massimo* (non viene riclassificato a un livello inferiore) anche in presenza di una qualità delle prove *moderata* o *bassa*. Sarà in sede di valutazione globale del dossier che si procederà alla sintesi delle due dimensioni: nel caso dell'esempio, la grande rilevanza dell'effetto osservato non potrà godere di una fiducia completa da parte del valutatore.

L'applicazione sistematica di questo approccio alla redazione delle richieste di accesso al Fondo per l'Innovazione ha inoltre consentito di individuare situazioni specifiche – che caratterizzano l'attuale, e sempre più la futura ricerca farmacologica – di problematica interpretazione da parte dei soggetti a diverso titolo coinvolti nei processi di sottomissione e valutazione delle richieste.

1. Parametri di valutazione

a. Determinazione del valore degli endpoint (cosiddetti) intermedi

In mancanza di regole definite utili al riconoscimento del valore clinico *intrinseco* di uno specifico endpoint intermedio (Disease-Free Survival - DFS; Progression-Free Survival - PFS; Metastasis-Free Survival - MFS, ecc.) per una specifica situazione clinica, l'uso di un tale parametro come *endpoint* essenziale (vedi sopra) dovrebbe essere adeguatamente documentato (documenti delle Agenzie Regolatorie, linee guida riconosciute, documenti di consenso formale prodotti da Società Scientifiche). Nel caso l'*endpoint* intermedio in questione venga invece proposto come *surrogato* di un *endpoint* di valenza clinica riconosciuta (Overall Survival - OS) dovrebbe essere disponibile un'adeguata dimostrazione di *surrogacy* (processo di validazione *at trial level*³ e definizione della soglia di *surrogacy*), che non abbia limitazione di applicabilità (*indirectness*) alla situazione in esame.

b. Patient Reported Outcomes (PROs) come endpoint essenziali

Il contributo dei PROs alla definizione del valore terapeutico aggiunto è fuori discussione.⁴

L’affidabilità di tali misure è tuttavia intrinsecamente legata a condizioni quali la corretta validazione (comprensiva della trasferibilità linguistica) e l’oggettiva rilevanza dell’effetto osservato (raggiungimento o meno della *Minimal Interesting Difference* - M.I.D.).⁵

Da non trascurare il fatto che la qualità dell’evidenza osservata risulta spesso bassa⁶ – *riscontro di risk of bias, imprecision, presenza di confronti multipli senza adeguata pianificazione statistica - multiplicity* – con intuibili conseguenze sulla qualità globale delle prove (*vedi più avanti nel paragrafo specifico*).

2. Analisi ad interim

Anche quando non chiaramente esplicitato, lo scopo delle analisi *ad interim* è l’interruzione anticipata della sperimentazione clinica, che può avvenire: (i) per la comparsa di effetti collaterali che vadano ad alterare sostanzialmente il rapporto tra benefici e danni attesi per la terapia; (ii) perché i dati fino a quel momento osservati sembrano concludere per una impossibilità di raggiungere l’obiettivo predefinito; (iii) perché l’efficacia del farmaco sembra emergere inequivocabilmente con i dati raccolti per l’analisi *ad interim* (evenienza di più frequente riscontro).

Nell’ultimo caso il rischio da evitare è di una sovrastima (*inflated estimate*) dell’effetto osservato nell’analisi *ad interim*, fenomeno osservato anche in presenza di adeguate strategie volte al contenimento della *multiplicity*.⁷ Vi è di conseguenza la necessità di individuare una soglia di *maturità* (in termini di percentuale di eventi osservati sul totale degli eventi predeterminati per l’analisi finale) che sia di sufficiente affidabilità.

Sulla base di quanto presente in letteratura,⁸ si suggerisce in questa sede di accettare come proponibile per quanto sopra il raggiungimento di almeno il 50% degli eventi predeterminati.

3. Analisi per sottogruppi

Il campo di applicazione della richiesta di accesso al Fondo per l’Innovazione è definito dal quesito iniziale (in forma di P.I.C.O. Popolazione, Intervento, Confronto, Outcome).⁹

Non infrequentemente la “P.” obiettivo del quesito corrisponde a una sottopopolazione delle evidenze di riferimento, la quale può essere stata o meno oggetto di stratificazione, e per la quale siano state previste o meno misure di tipo statistico al fine di contenere la *multiplicity*. Vengono così a definirsi tre evenienze contraddistinte da diverso grado di affidabilità:

- a. **sottogruppo determinato a posteriori (*post hoc*)**, l’analisi condotta sul quale sarà inevitabilmente viziata da *risk of bias, multiplicity* e molto probabilmente da *imprecision*, con conseguente *downgrade* di (almeno) due livelli. Qualsiasi conclusione tratta da questo tipo di analisi non potrà che essere *suggestiva*.
- b. **sottogruppo prespecificato (*prespecified*)** e oggetto di stratificazione, a carico del quale è comunque presente *multiplicity* e probabilmente *imprecision*, con conseguente *downgrade* di (almeno) un livello. Una conclusione tratta da questo tipo di analisi dovrebbe essere comunque considerata almeno *dubbia*.
- c. **sottogruppo prepianificato (*preplanned*)** al fine di contenere la *multiplicity*, a carico del quale potrebbe essere comunque presente *imprecision*, con conseguente *downgrade* di un livello. Solo per questa situazione una conclusione clinica può essere considerata come *dimostrativa*.

4. Confronti indiretti

Non infrequentemente il trattamento attualmente considerato quale opzione terapeutica

standard per la situazione specifica ("C." di P.I.C.O.) non corrisponde al braccio di controllo dello studio in esame. Altrettanto spesso tale *standard* terapeutico è stato a sua volta braccio sperimentale a confronto con il medesimo braccio di controllo dello studio in esame. Da qui la necessità di operare una valutazione comparativa indiretta dei due trattamenti di interesse fondata sulla disponibilità di un braccio di controllo comune.

Per qualsiasi tipologia di confronto indiretto è cruciale poter provare la sovrapponibilità (*similarity*) dei bracci di controllo (dei due trattamenti a confronto) costituenti il braccio di controllo comune; questo in termini sia di caratteristiche prognostiche basali sia di fattori che possono influenzare l'esito dei trattamenti (*effect modifiers*). Qualora siano presenti più studi suscettibili di creare una rete di evidenze comprendente confronti diretti e indiretti, sarà allora possibile stimare il confronto di interesse attraverso una *network meta-analysis* (NMA), per la quale ulteriori requisiti da verificare saranno l'omogeneità degli effetti relativi ai confronti diretti disponibili e la consistenza (sovrapponibilità degli effetti) tra stime dirette e indirette.

Conseguentemente, quando ci si trovi a effettuare un confronto indiretto si renderà frequentemente necessario operare numerosi *downgrade* qualitativi; in particolare:

- a. **in caso di confronto indiretto tradizionale (*Indirect Adjusted Comparison*)** sarà quasi inevitabile operare un *downgrade* per *indirectness* di almeno due livelli (per insufficiente dimostrazione di similarità del braccio di controllo comune); il *downgrade* potrà essere attenuato in caso di applicazione di metodiche di aggiustamento (*population-adjusted estimates of treatment effects*) dei parametri sopra definiti;¹⁰
- b. **nel caso sia disponibile una NMA**, potranno essere applicati i recenti criteri GRADE¹¹ (non si potrà prescindere in questo caso da un supporto metodologico avanzato), in base ai quali eventuali *downgrade* qualitativi saranno dipendenti da:
 - i. qualità dei confronti diretti disponibili (in termini convenzionali di *risk of bias, imprecision, inconsistency*);
 - ii. qualità dei confronti indiretti e della rete in senso globale (in termini di *intransitivity, incoherence, imprecision*).

5. Qualità globale delle prove

Originariamente il sistema GRADE aveva proposto di considerare come giudizio riassuntivo della qualità globale delle evidenze considerate il più basso livello di qualità assegnato a un *outcome* essenziale.¹² Nel nostro Paese – sia in sede di Regione Emilia-Romagna¹³ sia più recentemente in sede di Istituto Superiore di Sanità¹⁴ – è stato suggerito, in presenza di evidenze di beneficio e di danno (*outcome* essenziali) entrambe a favore del trattamento oggetto di valutazione, di considerare come riferimento il livello di qualità dell'*outcome* essenziale che si ritenga (nel caso presente il proponente la richiesta) più possa influire sulla formulazione della proposta terapeutica.

6. Studi orientati dal bersaglio molecolare / medicina di precisione

Si riscontra per queste tipologie di studio la necessità di una precisa collocazione delle evidenze disponibili: dalla fase di sviluppo del biomarcatore all'impiego di routine (*ESMO Scale for Clinical Actionability of Molecular Targets*).¹⁵

Punti cruciali per la valutazione delle prove in questo campo sono la correttezza del processo di validazione del biomarcatore (*validazione analitica vs. validazione clinica vs. utilità clinica*)¹⁶ e la disponibilità del test specifico anche in sede periferica (a evitare possibili *downgrade* per *indirectness*). Per le particolari tipologie di studio (*basket* e *umbrella trials*) e i disegni sperimentali (*adaptive design*) specificamente orientati a un aumento dell'efficienza della sperimentazione:¹⁷

- a. spesso non è prevista la randomizzazione, con conseguente elevato rischio di *bias*;
- b. sono previsti molti confronti statistici, con conseguente rischio di *multiplicity*;
- c. le sottopopolazioni individuate dal biomarcatore sono spesso di piccole dimensioni campionarie, con conseguente rischio di risultati falsamente negativi;
- d. sono previste interruzioni precoci (programmate), con rischio di sopravvalutazione degli effetti (*inflated estimate*, v. #2);

In ogni caso, qualora l'aumentata efficienza venga a esitare in un effetto clinico particolarmente rilevante e in assenza di *imprecision*, vi è la possibilità di aumentare (*upgrade*) il livello di confidenza derivante dal disegno dello studio.¹⁸

Da ultimo riveste particolare interesse la possibilità di considerare le sottopopolazioni molecolari alla stregua di “malattie rare” (v. criteri di classificazione Innovatività 18.09.2017).

BIBLIOGRAFIA

1. Pinto C, et al. *Economia & Politica del Farmaco*, 2018 2: 4–15 e *Sanità24* (21-06-2018).
2. Guyatt GH, et al. *J Clin Epidemiol* 2011;64:395–400.
3. Buyse M, et al. *Biom J*. 2016 Jan;58(1):104–32. doi: 10.1002/bimj.201400049. Epub 2015 Feb 12. Review.
4. Basch E. The missing voice of patients in drug safety reporting. *N Engl J Med*. 2010;362(10):865–869.
5. Jayadevappa R, et al. *J Clin Epidemiol*. 2017 Sep;89:188–198. doi: 10.1016/j.jclinepi.2017.06.009. Epub 2017 Jul 1.
6. Calvert M, et al. CONSORT PRO Group. *JAMA*. 2013 Feb 27;309(8):814–22.
7. Bassler D, et al. *JAMA*. 2010;303(12):1180–1187.
8. Freidlin B & Korn EL. *Clinical Trials* 2009; 6: 119–125.
9. PICO Framework - PubMed Health <https://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0029906/>.
10. Phillipppo DM, et al. NICE DSU Technical Support Document 18. December 2016. www.nicedsu.org.uk.
11. Brignardello-Petersen R, et al. *Journal of Clinical Epidemiology* 93 (2018) 36–44.
12. Brozek et al, *Allergy* (2009) 64:1109–1116.
13. De Palma R, et al. Dossier n. 172/2009 - La produzione di raccomandazioni cliniche con il metodo GRADE. L'esperienza sui farmaci oncologici. http://asr.regione.emilia-romagna.it/wcm/asr/collana_dossier/doss172.htm.
14. Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure. Manuale metodologico per la produzione di linee guida di pratica clinica. v.1.3.2 aprile 2019.
15. <https://www.esmo.org/Policy/ESCAT>.
16. Perez-Gracia JL, et al. *Cancer Treatment Reviews* 53 (2017) 79–97.
17. Janniaud P, et al. *Cancer Treatment Reviews* 73 (2019) 20–30.
18. Guyatt GH, et al. *J Clin Epidemiol* 64 (2011) 1311–1316.

Disclosure: Il presente lavoro è il risultato degli incontri del Gruppo di Lavoro multidisciplinare, con il coordinamento di *Economia Sanitaria srl*, presentato in un evento pubblico

Provenienza: non commissionato, non sottoposto a peer-review.

Pubblicato: 2 marzo 2020

Copyright: © 2020 *Economia Sanitaria srl*, Monza (Italy)

E.mail: g.ravasio@economiasanitaria.it

Testata: “*Economia & Politica del Farmaco e delle Tecnologie Sanitarie*”

Registrazione: Tribunale di Milano n° 201 del 05/04/1997

Direttore Responsabile: Giovanni Ravasio

Editore: *Economia Sanitaria srl* - Via Medici, 39 - 20900 Monza (MB)