

IL CRITICAL APPRAISAL SKILLS PROGRAMME (CASP): UNO STRUMENTO PER LA VALUTAZIONE CRITICA DELLA LETTERATURA SCIENTIFICA.

Chiara Baglioni*, Gemma Battagliese*, Elisabetta Berenci*, Valentina Cardi*, Silvia Carlucci*, Mariagrazia Iaciancio*, Olga Ines Luppino*, Alessandra Mancini*, Sabina Marianelli*, Danilo Russo*, Luigi Smimmo*, Carlo Buonanno°.

° Tutor project di ricerca I anno Scuola di Psicoterapia Cognitiva (SPC) di Roma

* Allievi specializzandi, project di ricerca Scuola di Psicoterapia Cognitiva (SPC) I anno Roma

Riassunto

L'articolo presentato ha lo scopo di descrivere il *Critical Appraisal Skills Programme (CASP)*, uno strumento valutativo che consente di stabilire in modo critico l'utilità e la validità degli studi pubblicati rispetto ad un argomento di particolare interesse scientifico, una procedura nata con l'obiettivo di colmare il gap esistente tra ricercatori e clinici.

Il programma nasce nel 1998 presso il *Centre of Evidence Based Medicine di Oxford* grazie alla collaborazione di professionisti di varia natura ed ha come finalità la valutazione critica della letteratura scientifica. L'analisi di validità, rilevanza clinica ed applicabilità degli studi analizzati sono gli obiettivi che è possibile raggiungere grazie ai sei questionari, costruiti in base alle differenti tipologie di studio esaminate: *randomized controlled trials (RCT)*, *review*, *studi qualitativi*, *studi case control*, *studi di coorte* e *studi sui test diagnostici*

Nell'ambito degli studi di natura clinica, l'obiettivo di tale diffusione è quello di promuovere l'utilizzo di uno strumento scientifico che sia di supporto al processo di decisione relativo alla scelta dei trattamenti da somministrare.

L'obiettivo di questo lavoro è quello di descrivere e discutere le principali caratteristiche di questo strumento, le potenzialità e i limiti, oltre che le possibilità di diffusione nel contesto italiano.

Parole chiave medicina evidence based, valutazione critica

THE CRITICAL APPRAISAL SKILLS PROGRAMME (CASP): AN INSTRUMENT OF CRITICAL EVALUATION OF SCIENTIFIC LITERATURE

Abstract

The present article has the aim to describe the *Critical Appraisal Skills Programme (CASP)* a method to critically assess the literature on a specific topic from the scientific literature.

It was born in 1998 at *Center of Evidence Based Medicine (Oxford)* from an international collaboration between clinicians and it received a great diffusion. *CASP* allows a scientific support to the decision making process and it seems an easy and flexible instrument to use.

This work will describe the main features, pros and cons and the potential Italian diffusion of *CASP*.

Keywords: evidence based medicine, critical appraisal

Introduzione

Gli studenti del primo anno della Scuola di Specializzazione in Psicoterapia Cognitiva di Roma (SPC) nell'ambito di un project loro proposto si sono interessati al CASP, un programma che promuove lo sviluppo di abilità di valutazione critica della letteratura relativa all'efficacia degli interventi clinici, il cui *background* teorico è quello della medicina basata su prove di efficacia (*evidence based medicine*).

La medicina basata su prove di efficacia è definita come “l'uso coscienzioso, esplicito e giudizioso delle prove empiriche disponibili, nel processo decisionale che riguarda la cura di ciascun paziente” (*Sackett et al.1996*).

Si articola in quattro fasi:

- 1) interrogarsi su specifiche questioni che riguardino il paziente o il problema presentato, l'intervento scelto, eventuali interventi alternativi e gli obiettivi che ci si propone di perseguire;
- 2) riferendosi alla letteratura scientifica esistente, rintracciare prove empiriche dell'efficacia di un particolare intervento;
- 3) valutare gli studi presenti in letteratura in maniera critica sulla base della loro validità, rilevanza clinica ed applicabilità;
- 4) scegliere l'intervento da utilizzare attraverso l'integrazione tra esperienza clinica e prove di efficacia, presupposto fondamentale della medicina *evidence based*.

Il CASP si propone come strumento in grado di aiutare il clinico ad affrontare in maniera guidata e più consapevole la terza fase del processo appena descritto.

Il presente studio ha la finalità di approfondire ed applicare il CASP per permetterne una sua valutazione.

Cos'è il *Critical Appraisal*?

Il *Critical Appraisal* è uno strumento utilizzato per l'analisi sistematica degli studi pubblicati, al fine di valutarne validità e rilevanza. Attraverso lo studio critico dell'assetto metodologico, esso consente di stabilire la spendibilità nella pratica clinica degli interventi presi in esame. In altre parole, il CASP interviene a sostegno del clinico nella fase di valutazione critica della letteratura e gli fornisce un utile strumento per valutare la solidità delle prove d'efficacia cui è stato sottoposto un trattamento. Lo scopo generale è colmare il *gap* esistente tra ricerca e pratica clinica. Attraverso una serie di domande di ordine metodologico sulla raccolta dei dati, sui metodi di analisi e su come vengono presentati i risultati, il CASP facilita la comprensione delle conclusioni di un articolo ed aiuta a determinare in che modo esse possano influenzare il processo decisionale nella pratica clinica: durante le consultazioni con il paziente, il clinico tiene conto delle migliori evidenze disponibili nella ricerca clinica per decidere quale sia il modo migliore per impostare il trattamento.

“Per determinare quali siano le migliori evidenze esistenti in letteratura, occorrono le abilità di valutazione critica necessarie per la comprensione dei risultati della ricerca e per valutarne la validità e quindi se è utile nel processo decisionale” (Hill & Spittlehouse 2001).

CASP: nascita e diffusione

Nel 1998 nasce ad Oxford il CASP *International Network* (CASPi), una rete di collabora-

zione tra coloro che svolgono la professione clinica. Il CASPi si propone di promuovere la collaborazione internazionale, avvalendosi del lavoro di organizzazioni indipendenti locali che organizzino seminari volti a sviluppare abilità di ricerca e di valutazione critica, favorendo lo sviluppo e la diffusione di nuovi materiali e metodi di apprendimento. I programmi di training si basano su *workshop* della durata di almeno tre ore, in cui ci si avvale di lezioni frontali introduttive e di piccoli gruppi di lavoro in cui si realizzano esperienze pratiche guidate supervisionate. Nello specifico, nelle fasi iniziali, il conduttore introduce un set di domande strutturate per la valutazione critica ed organizza successivamente piccoli gruppi, ciascuno con il compito di esaminare un articolo. Alla fine, il conduttore discute l'esito del lavoro dei gruppi e fornisce *feedback* relativi alla valutazione, con l'obiettivo di rinforzare e condividere l'apprendimento. I *training* per acquisire le abilità di valutazione critica si sono diffusi in molti paesi. Prima è stata la Spagna nel 1997, seguita da Norvegia, Argentina, Romania ed Ungheria. Programmi simili sono stati sviluppati, a partire dal 2000, in Francia, Canada, Messico, Colombia e Brasile.

La valutazione critica: strumenti guida

Non tutti i contributi scientifici presenti in letteratura risultano essere automaticamente di buona qualità ed è sulla base di questa consapevolezza che la consultazione e l'analisi del materiale scientifico dovrebbe basarsi su un processo sistematico di valutazione critica che permetta a chiunque, ed al clinico in particolare, di decidere se una ricerca è apprezzabile sul piano della validità, rilevanza clinica ed applicabilità.

Il CASP è costituito da sei strumenti di valutazione che si differenziano sulla base delle più comuni metodologie di ricerca utilizzate nella letteratura clinica¹ :

- *Randomized controlled trials (RCT)*
- *Reviews*
- *Qualitative Studies*
- *Case Control Studies*
- *Cohort Studies*
- *Diagnostic Tests*

Questi strumenti si articolano in differenti domande, pur condividendo la medesima struttura di base. Quest'ultima è composta da tre sezioni:

1) Sezione A: validità dello studio (identificazione del problema di ricerca, pianificazione del disegno sperimentale, fase delle osservazioni);

2) Sezione B: adeguatezza dei risultati (analisi dei dati);

3) Sezione C: rilevanza clinica ed applicabilità/generalizzabilità dello studio (interpretazione dati, comunicazione risultati).

Ogni questionario è costituito da dieci o dodici domande. In ogni caso, per procedere all'analisi dell'intero articolo, è necessario che le prime due domande di *screening* ricevano una risposta affermativa. Le domande di *screening* forniscono indicazioni su alcuni punti chiave:

- centralità, chiarezza della domanda di ricerca posta e sua utilità ("La domanda su cui si basa lo studio è chiara e specifica?" oppure "Gli scopi della ricerca sono enunciati con chiarezza?");

¹ Nei paragrafi successivi sarà fornita una descrizione più dettagliata delle diverse categorie di studio e dei relativi questionari ad esse applicabili.

- adeguatezza del metodo di ricerca impiegato nello studio (“Il metodo di ricerca utilizzato per lo studio è corretto?”; “Il metodo utilizzato è il più adatto?”). Idealmente queste risposte dovrebbero essere fornite con la sola lettura dell’*abstract* o della prima pagina dell’articolo.

I Questionari

Nei paragrafi successivi, proveremo a fornire una descrizione dei differenti questionari.

Il questionario di valutazione delle rassegne sistematiche

Le rassegne condotte per valutare l’efficacia di specifici trattamenti psicologici hanno come obiettivi riassumere e discutere una serie di articoli scientifici che hanno come oggetto una determinata strategia di cura. Una rassegna sistematica cerca di seguire lo stesso livello di rigorosità scientifica che si applica alla costruzione di un esperimento scientifico. Secondo Davies e Crombie (1998), le rassegne sistematiche prendono in considerazione un’ampia quantità di articoli scientifici su un determinato argomento, sintetizzano i risultati di ciascuno studio imparzialmente e forniscono un riassunto generale delle conclusioni, includendo considerazioni ed implicazioni. Il questionario relativo alla valutazione delle rassegne sistematiche è stato sviluppato sulla base delle linee guida pubblicate da Oxman et al. (1994) ed è articolato in dieci domande, due di *screening* ed otto di valutazione. Le domande di *screening* servono a valutare se la rassegna risponde ad una domanda appropriata e specifica rispetto ad uno o più aspetti di un trattamento e se considera gli studi adeguati per raggiungere tale fine. Questa seconda domanda viene approfondita con linee guida più dettagliate nella domanda tre, con la quale si identificano i *databases* utilizzati per la ricerca bibliografica e si stabilisce se gli autori abbiano incluso articoli non pubblicati (*abstract* di presentazioni a convegni o articoli acquisiti tramite contatto con gli esperti sull’argomento), oltre che articoli non in lingua inglese. La quarta domanda è finalizzata a stabilire se la selezione degli articoli è stata condotta da più giudici indipendentemente o se è stato seguito un valido sistema di *scoring*. Le domande dalla cinque alla sette si riferiscono al metodo con cui i risultati sono presentati nella rassegna. Due sono i metodi che possono essere seguiti: il metodo qualitativo, con il quale gli articoli vengono raggruppati sulla base dell’argomento, oppure il metodo quantitativo, anche detto meta-analisi, una tecnica statistica che serve a valutare i risultati di più ricerche indipendenti.

Infine, le domande dalla otto alla dieci si riferiscono all’organizzazione e alla valutazione della discussione dei risultati e dunque quanto i risultati degli studi considerati siano generalizzabili all’intera popolazione e quanto siano rilevanti per definire l’efficacia di un trattamento.

Alcuni punti ritenuti importanti dal questionario sulle rassegne sistematiche vengono applicati raramente nella pratica, ma non sempre per mancanza di rigorosità scientifica. Per esempio, l’inclusione di articoli non pubblicati viene consigliata dal questionario CASP al fine di bilanciare i risultati positivi con quelli negativi, dato che i risultati positivi vengono pubblicati con più facilità. Tuttavia, va sottolineato che questo punto non considera che gli articoli pubblicati hanno ricevuto un processo di revisione che potrebbe renderli non paragonabili e di qualità superiore agli articoli non pubblicati. Allo stesso modo, dato che le riviste ad *impact factor* più alto e note a livello internazionale sono attualmente per la maggior parte in lingua inglese, gli studi pubblicati su riviste nazionali potrebbero essere di minore qualità.

Il questionario di valutazione degli studi caso-controllo

Il caso-controllo (*case-control*) è un tipo di disegno sperimentale epidemiologico usato per identificare i fattori di rischio che possono contribuire ad una specifica patologia mediante la comparazione tra soggetti che presentano tale patologia (“casi”) e soggetti che non la presentano, ma che sono simili ai casi per altre caratteristiche (“controlli”). Il questionario relativo alla valutazione degli studi caso-controllo si compone di 11 domande, suddivise in tre sezioni. La sezione A (validità), composta dalle due domande di screening e da quattro domande dettagliate. Le domande dalla tre alla sei valutano:

- l’adeguatezza dei metodi adottati per il campionamento dei casi e dei controlli, considerando se le differenze tra i gruppi sulle variabili non di interesse siano state minimizzate;
- la valutazione dell’esposizione a fattori di rischio o protettivi per entrambi i gruppi;
- la valutazione delle strategie utilizzate per minimizzare eventuali fattori confondenti (genetici, ambientali, socioeconomici).

La sezione B (rilevanza clinica) è composta da tre domande che consentono di valutare la forza dell’associazione tra l’esposizione ai fattori di rischio e la presenza della patologia, e la credibilità dei risultati. In particolare, rispetto a quest’ultimo punto il questionario suggerisce di considerare i criteri di *Bradford Hills* (relazione temporale, forza, plausibilità biologica, ecc.). La sezione C (generalizzabilità) è composta da due domande che tengono conto della generalizzabilità dei risultati e della loro coerenza con quelli provenienti da altri tipi di studi (*RCT*, *reviews* sistematiche, studi di coorte).

Il questionario di valutazione degli studi di coorte

Uno studio di coorte è uno studio osservazionale che segue nel tempo l’evoluzione di un gruppo di soggetti identificati chiaramente in base a determinate caratteristiche.

La selezione dei gruppi di confronto è effettuata in funzione dell’esposizione (esposti/non esposti) dei soggetti della coorte a interventi o fattori specifici. Uno studio di coorte può essere prospettico se definisce il gruppo di partecipanti e lo segue nel tempo o retrospettivo se identifica i fattori di rischio attraverso la ricostruzione anamnestica dei soggetti presi in esame. Il questionario relativo agli studi di coorte si compone di dodici domande. La sezione A comprende cinque domande, oltre alle due di *screening*. Le prime sono più dettagliate e valutano la validità dello studio indagando in che modo sia stata reclutata la coorte, oltre che l’eventuale intervento di *bias* di selezione che potrebbero compromettere la generalizzabilità dei risultati. Nella quarta e quinta domanda si valuta se le misure utilizzate siano oggettive o soggettive, se l’esposizione e l’*outcome* siano stati accuratamente misurati in modo tale da minimizzare gli errori, se il metodo di misura sia simile nei due gruppi e infine se i soggetti e/o lo sperimentatore siano *blinded* rispetto all’esposizione. La sesta domanda valuta se gli autori hanno identificato importanti fattori concorrenti e se hanno tenuto conto di questi fattori nel disegno e/o nelle analisi. La settima riguarda informazioni rispetto alla completezza e alla durata del *follow-up*.

Le domande della sezione B si riferiscono alla validità dei risultati dello studio rispetto alla stima del rischio e alla credibilità dei risultati (criteri di *Bradford Hills*). La sezione C, infine, riguarda la generalizzabilità dei risultati.

Gli studi osservazionali raramente forniscono evidenze sufficientemente forti che possano incidere sulla pratica clinica o sulle decisioni di politiche della salute. Per questo motivo è

sempre preferibile che i risultati di uno studio osservazionale siano supportati da altre evidenze empiriche in modo tale da essere maggiormente utilizzabili e generalizzabili.

Il questionario di valutazione degli studi di natura qualitativa

Il questionario finalizzato all'analisi critica degli studi di natura qualitativa ha l'obiettivo di valutare il rigore, la credibilità e la rilevanza dei risultati presentati. Si tratta di considerare, rispettivamente, l'appropriatezza e la puntualità scientifica della metodologia di ricerca, la chiarezza delle conclusioni e la significatività dello studio rispetto agli scopi iniziali dell'indagine bibliografica.

Il questionario si articola in dieci domande. Le prime due (*screening*) sono rivolte alla valutazione della chiarezza nell'enunciazione degli scopi della ricerca (domanda 1) e della appropriatezza della metodologia qualitativa (domanda 2).

Le otto domande specifiche sono finalizzate a discutere ed indagare le ragioni che hanno sostenuto le scelte del ricercatore. Esse prevedono un'iniziale riflessione sulle scelte del ricercatore relative all'appropriatezza del disegno (domanda 3 "appropriatezza disegno di ricerca"), alle scelte metodologiche compiute rispetto alla selezione del campione (domanda 4 "campionamento": come e perché sono stati selezionati i partecipanti/criteri di inclusione ed esclusione), alle modalità di raccolta dati (domanda 5 "raccolta dati": *setting*, descrizione analitica delle procedure impiegate ed eventuali modificazioni apportate durante il corso dello studio, formato dati scelto) e all'analisi dati (domanda 8 "analisi dati": descrizione del processo di analisi e considerazione critica di eventuali *biases* da parte dello sperimentatore, discussione dei dati contraddittori). Segue l'analisi critica della discussione del rapporto tra ricercatore e partecipanti, con particolare riferimento ad eventuali *bias* dello sperimentatore durante la formulazione delle ipotesi e la raccolta dati (domanda 6 "riflessività"), oltre che ai problemi etici implicati nella ricerca (domanda 7 "problemi etici": domande approvazione comitato etico, consenso informato, confidenzialità, restituzione risultati). Le ultime 2 domande sono finalizzate alla valutazione della descrizione dei risultati (domanda 9 "risultati": chiarezza espositiva, discussione dei dati a favore e contro le ipotesi formulate, credibilità) e alla discussione dei potenziali contributi forniti dalle conclusioni allo stato delle conoscenze disponibili sulla tematica affrontata (domanda 10 "valore della ricerca").

Il questionario di valutazione dei test diagnostici

I *tests* diagnostici sono strumenti di orientamento utilizzati nel processo di *assessment* che hanno come scopo la valutazione obiettiva di un soggetto che ne permetta di prevedere il comportamento in situazione analoghe.

Il questionario volto all'analisi critica degli articoli scientifici che discutono l'efficacia diagnostica degli strumenti di valutazione clinica è costituito da dodici domande. Gli ambiti indagati investono la tipologia dei risultati ottenuti, il rigore metodologico con conseguente validità dei dati e l'utilità clinica dello strumento presentato. Le domande di *screening* valutano la chiarezza di esposizione degli obiettivi della ricerca (in termini di popolazione presa in esame, *tests*, *setting*, conclusioni) e l'appropriatezza degli strumenti diagnostici utilizzati come termine di paragone (*best reference standard*). Le successive *Detailed Questions* analizzano gli aspetti più strettamente metodologici come l'indipendenza di risultati tra test in esame e stru-

menti di valutazione standard (*review bias e blinding*), l'accuratezza nella descrizione del disturbo preso in esame (*symptoms, disease stage o severity, co-morbidity, spectrum bias*), la descrizione del protocollo di applicazione del test. È possibile continuare la valutazione, con la sezione B "*what are the result?*", solo se la maggior parte delle risposte date nelle sezioni precedenti sono positive; manca tuttavia un *cut off* di punteggi; nelle domande successive si valutano i risultati, la loro sensibilità al costrutto, la capacità rappresentativa ed il livello di affidabilità (analisi livelli di confidenza). L'analisi critica dell'articolo prosegue con la valutazione dell'utilità clinica dello strumento in base alla specifica popolazione di pazienti trattati. Si indaga la generalizzabilità dei dati (analisi di caratteristiche quali sesso, età, differenze etniche e disturbo della popolazione presa in esame), l'utilità del test con i propri pazienti (analisi dei costi e benefici, livello di preparazione richiesto per interpretare i risultati), l'utilità delle conclusioni in termini di miglioramento della capacità di gestione dei pazienti e della loro qualità di vita.

Il questionario di valutazione dei randomized controlled trials

Il disegno sperimentale *Randomised Controlled* è il piano di ricerca più appropriato per la verifica dei trattamenti. Il procedimento prevede che i soggetti scelti da una popolazione con procedimento di campionamento casuale vengano assegnati in modo randomizzato al gruppo sperimentale o al gruppo di controllo; si mantengono costanti tutte le condizioni per ambedue i gruppi ad eccezione di quella del trattamento e si verifica, successivamente, la significatività della differenza tra i due gruppi.

Il questionario utilizzato per la valutazione critica degli articoli scientifici che utilizzano questa metodologia è composto da dieci domande suddivise in tre sezioni che indagano rispettivamente la validità dello studio (*Section A*), la validità dei risultati (*Section B*) e l'utilità per il clinico (*Section C*).

Nella prima sezione, dunque, si analizza la solidità metodologica dell'articolo; nello specifico, le prime due domande di *screening* indagano quanto l'ipotesi di ricerca sia specificata in termini di popolazione indagata, trattamento erogato, considerazioni finali e se il piano di ricerca è il più indicato rispetto alle finalità dello studio. Solo se le risposte alle prime due domande non sono negative è possibile proseguire nell'analisi dell'articolo.

Le successive *Detailed Questions* indagano l'appropriatezza dell'assegnazione dei pazienti ai gruppi (3), l'utilizzo della metodologia del "doppio-cieco" (4), l'analisi del *follow-up* (5) la coerenza della raccolta dati nel tempo (6) e la numerosità del campione (7). Le successive due domande (8-9) indagano l'accuratezza nella presentazione dei dati e nella loro discussione. Infine, l'ultima domanda analizza l'utilità clinica dei risultati.

CASP: potenzialità e limiti

Dall'esperienza pratica di gruppo condotta sull'applicazione di CASP ad un prescelto ambito di letteratura clinica, è possibile trarre alcune iniziali considerazioni.

La praticità, la maneggevolezza e la rapidità con cui consentono di selezionare il materiale scientifico rendono i questionari sufficientemente fruibili da parte di chi non possiede una conoscenza specifica della metodologia della ricerca.

CASP potrebbe fornire criteri comuni in base ai quali selezionare il materiale scientifico, indirizzando in modo strutturato e condiviso il processo di *decision making*.

Nonostante la praticità e la puntualità descrittiva dei questionari, la loro selezione implica competenze metodologiche di base e non rappresenta una soluzione esaustiva alla molteplicità dei possibili disegni di ricerca.

L'adozione rigida del protocollo CASP, inoltre, potrebbe comportare l'esclusione di gran parte della letteratura a disposizione, esponendo al rischio di sottovalutare interessanti spunti teorici. L'ideazione di un *cut-off* totale, non confinato esclusivamente allo *screening*, potrebbe rappresentare una modalità meno rigida di selezione, almeno in termini quantitativi.

Il lieve ma evidente sbilanciamento a favore di teoria e linguaggio utilizzati nella pratica della ricerca potrebbe comportare, inoltre, il restringimento del *target* di clinici ai quali lo strumento si rivolge. Per questa ragione una semplificazione linguistica, potrebbe risultare funzionale ad una maggiore spendibilità e ad una più ampia diffusione del CASP.

Bibliografia

- Bradley PMA & Hill A (2001). Critical appraisal skills programme international network: making sense of the evidence. *European Journal of Public Health* 11, 2, 238
- Davies HT & Crombie IK (1998) Getting to grips with systematic reviews and meta-analyses. *Hospital Medicine* 59, 12, 955-8.
- Hill A & Spittlehouse C (2001). What is critical appraisal? *What is... series*, 3, 2, 1-8.
- Oxman AD & Cook DJ, Guyatt GH (1994) Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *Journal of the American Medical Association* 2, 17,1367-71.
- Sackett DL, Rosenberg WM, Gray JA, Haynes RB & Richardson WS(1996). Evidence based medicine: what it is and what it isn't *British Medical Journal* 312, 71-72

www.evidence-based-medicine.co.uk
www.phru-nhs-uk
www.redcaspe.org
<http://www.cebm.net/index.aspx?o=1914>