

La calprotectina fecale con cut-off a 100 µg/g aumenta l'accuratezza diagnostica per le MICI nelle cure primarie pediatriche. Uno studio di coorte

Walker GJ, Chanchlani N, Thomas A, et al.

Primary care faecal calprotectin testing in children with suspected inflammatory bowel disease: a diagnostic accuracy study

Arch Dis Child. 2020;105(10):957-963

L'accuratezza diagnostica della calprotectina fecale - esame di solito richiesto dallo specialista - è risultata elevata con il cut-off di 100 µg/g, in un contesto di cure primarie inglesi in cui sono stati valutati bambini 4-18 anni con disturbi gastrointestinali sospetti per MICI; l'elevato valore predittivo negativo del test suggerisce il suo utilizzo nelle cure primarie, all'interno di un protocollo diagnostico condiviso, per ridurre in modo significativo gli invii allo specialista ed esami di approfondimento diagnostico più invasivi. La numerosità non elevata del campione e il fatto che solo il 30% delle diagnosi di MICI sono state effettuate in pazienti valutati inizialmente nelle cure primarie, fa concludere gli autori stessi che il dato trovato andrà verificato in successivi studi multicentrici.

Fecal calprotectin with cut-off at 100 µg/g increases the diagnostic accuracy for Inflammatory Bowel Disease (IBD) in primary pediatric care. A cohort study

The diagnostic accuracy of fecal calprotectin - usually required by the specialist - was found to be high with the cut-off of 100 µg/g, in a British primary care setting in which children aged 4-18 years with suspected gastrointestinal disorders were evaluated for IBD. The high negative predictive value of the test suggests its use in primary care, within a shared diagnostic protocol, to significantly reduce referrals to the specialist and more invasive diagnostic examinations. The small sample number and the fact that only 30% of IBD diagnoses were made in patients initially evaluated in primary care, makes the authors conclude that the data found need to be verified in subsequent multicentre studies.

Metodo

Obiettivo (con tipo studio)

Studio di coorte prospettico osservazionale che valuta: 1. l'accuratezza diagnostica della calprotectina fecale, con cut-off posto a 100 µg/g, sulla base delle linee guida NICE del 2013, nei bambini con sospetta MICI in un setting di cure primarie; 2. l'eventuale cambiamento nell'invio a cure di secondo livello e riduzione del tempo di diagnosi.

Popolazione

195 bambini età media 15.1 anni (12-17.1), parte di 75.000 residenti 0-18 anni di una zona della regione del Devon in Gran Bretagna. Nel gennaio 2014 è stato condiviso un nuovo percorso per la richiesta della calprotectina fecale sulla base delle linee

guida NICE del 2013, al quale hanno aderito 48/49 dei pediatri della regione.

Criteri di inclusione:

- età: 4-18 anni;
- sintomi gastrointestinali ed inserimento nel nuovo protocollo a discrezione del curante.

Criteri di esclusione:

- pregressa diagnosi di malattia infiammatoria cronica intestinale (MICI);
- sospetto di neoplasia;
- uso di antiinfiammatori non steroidei nelle precedenti 6 settimane.

Test

Calprotectina fecale utilizzata nel contesto delle cure primarie, cioè richiesta dai pediatri del territorio nel sospetto di MICI nei bambini 4-18 anni, valutata con metodologia quantitativa ELISA con range 6-2.100 µg /g e con cut off posto a 100 µg/g.

Gold Standard

Follow-up per 12 mesi e successiva diagnosi di MICI secondo la classificazione di Parigi e i criteri di Porto che utilizza criteri clinici, radiologici e istopatologici.

Tempo

Da gennaio 2014 ad agosto 2017, con follow-up di almeno 12 mesi.

Risultati principali

Il 7% dei testati (13/195) ha ricevuto una diagnosi di MICI. Nello stesso periodo le diagnosi totali di MICI sono state 42; 29 diagnosi cioè non sono passate attraverso le cure primarie; in 19 di questi pazienti è stata valutata la calprotectina prima della diagnosi che è risultata superiore a 100 µg/g. Nelle cure primarie, utilizzando un cut-off di 100 µg/g, la calprotectina presenta una accuratezza diagnostica del 91% (IC 95% 86%, 95%), con sensibilità 100% (IC 95% 75%, 100%) e specificità 91% (IC 95% 85%, 94%), il valore predittivo positivo (VPP) è 43% (IC 95% 25%, 63%) mentre il valore predittivo negativo (VPN) è 100% (IC 95% 98%, 100%). La valutazione della calprotectina fecale non influenza la tempistica della diagnosi ma il riscontro di un valore nella norma riduce l'invio allo specialista e gli esami diagnostici

secondari. L'aumento del cut-off della calprotectina da 50 µg/g a 100 µg/g ha permesso di raddoppiare il VPP senza ridurre il VPN. La calprotectina ha mostrato una curva ROC, che valuta l'accuratezza diagnostica, significativamente migliore rispetto al confronto con albumina ($p=0.010$), emoglobina ($p = 0.004$), globuli bianchi ($p < 0.001$) e piastrine ($p < 0.001$) ma non con la PCR ($p=0.16$), anche se ha permesso di diagnosticare 3 pazienti con MICI che avevano PCR nella norma. Utilizzando il metodo di Youden per massimizzare la differenza tra il tasso di veri positivi e falsi positivi sopra tutti i possibili valori di cut-off, la soglia di cut-off ottimale per distinguere pazienti con MICI verso pazienti senza MICI è risultato di 110 µg/g. In assenza della possibilità di dosare la calprotectina fecale i pediatri riferivano l'intenzione di invio allo specialista per 104/195 pazienti (primo gruppo), il non invio per 11/195 (secondo gruppo) e incertezza per i restanti 64/195 pazienti (terzo gruppo); per 16/195 dato non disponibile. La valutazione della calprotectina fecale da parte del pediatra ha risparmiato 64 invii allo specialista (31 nel primo gruppo e 33 nel terzo gruppo), mentre ne ha decisi 9 (1 nel primo gruppo e 8 nel terzo gruppo).

Conclusioni

La valutazione, in un setting di cure primarie, della calprotectina fecale con cut-off 100 µg/g nei bambini con sintomatologia suggestiva permette di distinguere con accuratezza gli affetti da MICI dai pazienti con disturbi gastrointestinali funzionali, riducendo gli invii allo specialista e gli esami di approfondimento diagnostico.

Altri studi sull'argomento

Uno studio di Health Technology Assessment ha eseguito una revisione sistematica allo scopo di definire il valore di cut-off della calprotectina per permettere di distinguere una MICI da un colon irritabile, raccogliendo 28 studi. Negli adulti la sensibilità è 93% (IC 95% 83%, 100%) e la specificità 94% (IC 95% 60%, 100%) con un cut-off di 50 µg/g. Nelle popolazioni pediatriche valutate la sensibilità varia da 95% a 100% e la specificità da 44% a 93% con lo stesso valore di cut-off dell'adulto. Le aree di incertezza includono la gestione delle persone con risultati borderline (50-150 µg/g) di calprotectina fecale, la maggior parte delle quali potrebbe non avere una MICI e rischia di doversi sottoporre a indagini invasive come una colonscopia [1]. Uno studio sul campo olandese ha valutato due coorti di bambini (4-18 anni) con sintomi intestinali (diarrea cronica e/o dolore addominale ricorrente) per 12 mesi; la prima coorte, 114 bambini, è stata seguita dal medico delle cure primarie e la seconda coorte, 90 bambini, da specialisti, 24 dei quali ad essi inviati dal medico delle cure primarie per sintomi/segni di allarme (perdita di peso, rettoragia, storia familiare di MICI, deficit di crescita, sintomi extraintestinali, anemia, PCR o VES elevata, piastrine aumentate). Le analisi della calprotectina fecale (cut-off di 50 µg/g) hanno indicato una specificità nel primo gruppo di 0.87 (IC 95% 0.8, 0.92), nel secondo gruppo sensibilità 0.99 (IC 95% 0.81, 1.0) e specificità 0.84 (IC 95% 0.74, 0.91), VPP 0.60 (IC 95% 0.42, 0.76), VPN 1.00 (IC 95% 0.94, 1.00). Nel caso fosse stato considerato un cut-off di 100 µg/g, la sensibilità sarebbe scesa a 0.87 e VPN 97%, e la

specificità salita a 0.93 con VPP 0.74 [2 casi persi di MICI (12%), risparmiati 69 (77%) invii allo specialista]; infine con un cut-off di calprotectina fecale a 250 µg/g, la sensibilità sarebbe stata 0.81 con VPN 0.92 e la specificità 0.98 con VPP 0.96, [3 casi persi di MICI (18%), risparmiati 74 (82%) invii allo specialista]; analizzando i diversi cut-off, gli autori commentano che innalzando la soglia della calprotectina fecale a 250 µg/g aumenta il VPP (da 0.60 a 0.96) con una riduzione di invii di falsi positivi ma aumentano i falsi negativi, con una riduzione del VPN (da 1 a 0.92), e quindi i casi persi di MICI (con un aumento di casi persi da 0 al 18%). Non viene valutato qual è il cut-off ottimale per massimizzare la specificità mantenendo il VPN a 1. Gli autori concludono che una calprotectina fecale positiva nei bambini con sintomi gastrointestinali cronici non è indicativa di MICI, mentre un valore negativo permette di escludere con buona sicurezza una MICI evitando l'invio allo specialista [2]. Gli stessi autori hanno confrontato per 90 pazienti (65 valutati dal medico di medicina generale, 25 dal pediatra, 17 diagnosticati con MICI), l'accuratezza diagnostica per MICI di: 1) soli sintomi d'allarme, 2) sintomi d'allarme e PCR, 3) sintomi d'allarme e calprotectina fecale. L'aggiunta della calprotectina fecale ai sintomi di allarme ha aumentato l'area sotto la curva (AUC) in modo significativo da 0.80 (0.67–0.92) a 0.97 (0.93–1.00), al contrario l'aggiunta della PCR non ha aumentato l'AUC in modo significativo ($p > 0.05$). Gli autori concludono che nelle cure primarie, quando i bambini vengono identificati come ad alto rischio di MICI, l'aggiunta del test della calprotectina fecale ai sintomi è la strategia ottimale per migliorare la stratificazione del rischio [3]. Una revisione sistematica ha cercato di stabilire quali siano gli esami laboratoristici caratteristici della MICI nella popolazione pediatrica. Sono stati selezionati 8 studi dove sono stati presi in esame: calprotectina, VES, PCR, Hb, albumina, piastrine; il miglior indicatore di è risultata essere la FC (AUC 0.26, IC 05% 0.21, 0.31); il secondo miglior indicatore è risultato essere la VES (AUC 0.16, IC 95% 0.11, 0.21) [4]. Una revisione sistematica del 2018 ha raccolto 38 studi sull'utilizzo della calprotectina fecale nella MICI. I criteri di inclusione degli studi erano la misura della calprotectina fecale per individuare in pazienti adulti con sintomi addominali cronici (almeno 6-8 settimane) sia in cure primarie che secondarie. La sensibilità variava in base al tipo di analisi da 0.85 (IC 95% 0.75, 0.92) a 0.94 (IC 95% 0.75, 0.90) e la specificità da 0.67 (IC 95% 0.56, 0.76) a 0.88 (IC 95% 0.77, 0.94); i 5 studi effettuati nelle cure primarie hanno rilevato una sensibilità di 0.72 e una specificità di 0.65 con un cut-off di calprotectina fecale a 50 µg/g. La sensibilità così bassa è stata imputata dagli autori della revisione a una selezione di popolazione (alcuni pazienti sono stati inviati direttamente dallo specialista a causa della importante sintomatologia) oppure alla probabile presenza di MICI non gravi nelle quali il dosaggio di FC risulta falsamente negativo [5].

Che cosa aggiunge questo studio

Questo studio è il primo che dimostra in un setting di cure primarie pediatrico l'ottima accuratezza diagnostica della calprotectina fecale utilizzando come cut-off 100 µg/g.

Commento

Validità interna

:: Il nuovo test diagnostico (index) è stato valutato su uno spettro appropriato di pazienti (quelli su cui sarebbe stato utilizzato nella pratica clinica; ad esempio: selezione casuale di pazienti consecutivi (senza esclusioni ingiustificate) sospettati di avere la condizione indagata)?	SI NO NA
:: Se NO alla precedente: Il test diagnostico è stato valutato in un gruppo di pazienti con malattia/condizione confermata o severa e confrontato con soggetti senza malattia/condizione in studio (approccio caso-controllo)?	SI NO NA
:: Il gold standard era validato e appropriato (adatto a classificare correttamente la condizione indagata e costruito in modo indipendente dal test index)?	SI NO NA
:: Il confronto con il gold standard era realizzato in maniera indipendente (in cieco: chi applica/valuta il gold standard non è a conoscenza del risultato del test index e viceversa)?	SI NO NA
:: Il gold standard è stato applicato a tutti i pazienti, indipendentemente dal risultato del test diagnostico in studio (il risultato del test diagnostico in studio può avere influenzato la decisione di applicare il gold standard)?	SI NO NA

Disegno dello studio: questo studio è stato condotto non in un setting di ricerca ma nel mondo reale coinvolgendo quasi tutti i pediatri (48/49) delle cure primarie afferenti alle stesse cure secondarie. Questo rappresenta il suo punto di forza, anche se il campione considerato non è molto ampio. Potrebbe essere presente un *verification bias* a causa dei diversi criteri utilizzati per la diagnosi, tuttavia il periodo di follow-up di 12 mesi per tutti i pazienti arruolati riduce questo rischio; infatti il tempo di follow-up considerato è più del doppio del tempo medio riportato in UK per la diagnosi di MICI - 5 mesi- e nello studio il tempo medio per arrivare alla diagnosi è stato poco più di 2 mesi. Tra i pazienti diagnosticati per MICI solo una minoranza è passata per le cure primarie (13/42); tuttavia il 73% dei 29 diagnosticati nelle cure secondarie ha dosato la calprotectina fecale che è risultata per tutti superiore a 100 µg/g; sembrerebbe quindi che non sia presente uno *spectrum bias*, tuttavia gli autori stessi concludono che i risultati ottenuti andranno riverificati in uno studio multicentrico.

Esiti: rilevanti e ben definiti.

Conflitto di interesse: gli autori dichiarano assenza di conflitti d'interesse.

Trasferibilità

Popolazione studiata: sovrapponibile a quella che accede ai nostri ambulatori.

Tipo di intervento: la calprotectina fecale è un esame facilmente eseguibile anche in Italia, fondamentale nello screening delle MICI; tuttavia, in alcune regioni rimane appannaggio dello spe-

cialista non essendo richiedibile nell'ambito delle cure primarie. Essendo un esame di screening dall'alto VPN, sarà fondamentale che ne venga estesa la prescrivibilità per ridurre l'inappropriato invio a centri di secondo livello.

1. Waugh N, Cummins E, Royle P, et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: systematic review and economic evaluation. *Health Technol Assess.* 2013;17(55):xv-xix, 1-211.
2. Holtman GA, Lisman-van Leeuwen Y, Kollen BJ, et al. Diagnostic Accuracy of Fecal Calprotectin for Pediatric Inflammatory Bowel Disease in Primary Care: A Prospective Cohort Study. *Ann Fam Med.* 2016;14(5):437-45
3. Holtman GA, Lisman-van Leeuwen Y, Kollen BJ, et al. Diagnostic test strategies in children at increased risk of inflammatory bowel disease in primary care. *PLoS One.* 2017; 12(12): e0189111.
4. Holtman GA, Lisman-van Leeuwen Y, Day AS, et al. Use of Laboratory Markers in Addition to Symptoms for Diagnosis of Inflammatory Bowel Disease in Children: A Meta-analysis of Individual Patient Data. *JAMA Pediatr.* 2017 ;171(10):984-991.
5. Freeman K, Willis BH, Fraser H, et al. Faecal calprotectin to detect inflammatory bowel disease: a systematic review and exploratory meta-analysis of test accuracy. *BMJ Open.* 2019;9(3):e027428.
6. O'Sullivan JW, Banerjee A, Heneghan C, et al. Verification bias. *BMJ Evid Based Med.* 2018;23(2):54-55.
8. S.A. Mulherin, W.C. Miller. Spectrum Bias or Spectrum Effect? Subgroup Variation in Diagnostic Test Evaluation. *Ann Intern Med.* 2002;137:598-602
9. D'Arrigo G, Provenzano F, Torino C, et al. I test diagnostici e l'analisi della curva ROC. *G Ital Nefrol* 2011;28(6):642-647

Scheda redatta dal gruppo di lettura di Parma:

Maria Teresa Bersini, Maria Cristina Cantù, Maddalena Marchesi, Sandra Mari, Francesca Manusia, Manuela Musetti, Costantino Panza, Luisa Seletti.

Glossario

Sensibilità (di un test diagnostico): la percentuale dei pazienti affetti da una malattia che risulta positiva al test. Un test con un'alta sensibilità riduce il rischio di falsi negativi (detto in altre parole: riconosce gli ammalati)

Specificità (di un test diagnostico): La percentuale di pazienti sani, che non hanno una malattia e che è risultata negativa al test. Un test con un'alta specificità è importante per confermare una diagnosi, perché i falsi positivi saranno pochi.

Valore predittivo negativo: esprime la percentuale di persone con un test negativo che è libera da malattia, ovvero la probabilità di una persona con un test negativo di essere veramente sana.

Valore predittivo positivo: esprime la percentuale di persone con un test positivo che ha la malattia, ovvero la probabilità di una persona con un test positivo di essere veramente ammalata.

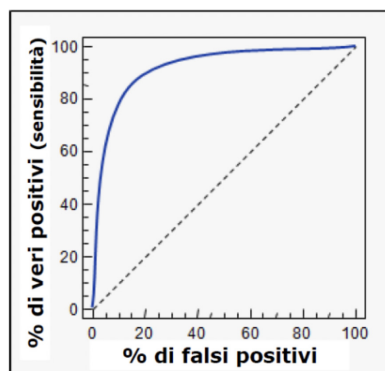
Il Valore predittivo, oltre ad essere legato alla sensibilità e specificità e quindi alle caratteristiche intrinseche del test, dipende anche dalla prevalenza della condizione studiata (a differenza della sensibilità e della specificità che dipendono solo dalle proprietà del test). All'aumentare della prevalenza della condizione studiata il VPP del test aumenta e il VPN diminuisce; al contrario, in caso di malattie rare il VPP di un test si riduce e il VPN aumenta. Occorre quindi fare molta attenzione nel riporre "fiducia" nei risultati di un test senza conoscere la prevalenza della patologia nella popolazione in cui il test viene applicato; il VP di un test osservato in uno studio non può essere applicato a contesti differenti. Se la prevalenza di una malattia è molto bassa il VPP non sarà quasi mai prossimo al 100% anche se la sensibilità e la specificità del test sono alti. Dunque in caso di screening sulla popolazione generale di una patologia rara è inevitabile che molte persone con il test positivo siano falsamente positivi. È questa la ragione per cui alcuni test sono più affidabili se usati su popolazioni a rischio, in cui è più probabile la presenza della patologia.

Verification bias

Gli studi sull'accuratezza diagnostica determinano l'abilità di un nuovo test per confermare o escludere una malattia. Occorre sottoporre tutti i partecipanti allo studio a un test indice (il nuovo test) e un test standard di riferimento (di solito il test considerato il migliore per la diagnosi della condizione che viene indicato come "Gold standard"). I risultati del test in esame vengono quindi confrontati con il test standard di riferimento, e vengono determinati il numero di pazienti che sono risultati veri positivi (VP), veri negativi (VN), falsi positivi (FP) e falsi negativi (FN) e può quindi essere calcolata la sensibilità e la specificità del test indice [sensibilità = $VP / (VP + FN)$, specificità = $VN / (VN + FP)$]. Esistono due tipi di bias di verifica: **bias di verifica parziale** in cui solo alcuni pazienti ricevono il test standard di riferimento e altri no; e **bias di verifica differenziale**, dove vengono utilizzati due diversi test standard di riferimento, tipicamente alternati a seconda che il test indice fosse positivo o negativo. Molti test di riferimento sono invasivi, costosi o comportano un rischio procedurale (p. es., angiografia, biopsia e chirurgia), e quindi in molti studi il bias di verifica è un pregiudizio inevitabile; i ricercatori dovrebbero sempre discutere chiaramente il potenziale impatto di questo pregiudizio sui risultati, nonché sul potenziale delle conseguenze cliniche [6].

Spectrum bias

Lo spectrum bias o l'effetto spettro descrive la variazione tra le impostazioni nelle prestazioni dei test utilizzati per prevedere, esaminare e diagnosticare una malattia. In particolare, l'uso predittivo di un test può essere diverso quando viene applicato in una popolazione generale piuttosto che nel campione di studio in cui è stato sviluppato e testato. L'effetto dello spettro pertanto descrive la variazione nelle prestazioni dei test per la previsione, lo screening e la diagnosi della malattia tra i diversi sottogruppi di popolazione. Ad esempio un test sviluppato in una popolazione con una maggiore prevalenza di malattia (o a rischio più elevato) avrà tipicamente una sensibilità inferiore e una specificità più elevata se applicato in una popolazione con una prevalenza di malattia inferiore (o a rischio inferiore). I medici dovrebbe considerare attentamente se il campione studiato permette di applicare i risultati ottenuti sul test nella loro pratica specifica [7].



La curva ROC e l'area sotto la curva (Area Under the Curve - AUC)

La curva ROC (acronimo di Receiver Operating Characteristic) è una tecnica statistica che misura l'accuratezza di un test diagnostico lungo tutto il range dei valori possibili. Questo strumento, visibile graficamente, permette di individuare per un test di screening il miglior valore di un cut-off e di valutare il grado di accuratezza del test in esame ed esprime una correlazione tra la probabilità di ottenere un vero positivo nei malati veri (sensibilità del test) e la probabilità di ottenere un falso positivo nei non malati (1-specificità); in altre parole permette di verificare graficamente se i veri positivi al test sono maggiori dei falsi positivi. L'estensione dell'area sotto la curva è la misura dell'accuratezza del test. Più la curva è spostata a sinistra, maggiore è l'area da essa sottesa. In questa figura la curva ROC è in blu e l'area sottesa al di sotto della curva indica la accuratezza del test diagnostico. Ogni punto sulla curva ROC rappresenta una coppia sensibilità/1-specificità corrispondente a una particolare soglia o cut-off. Il test ideale, ossia con sensibilità e specificità del 100%, sarebbe indicato da una curva che passa all'angolo sinistro superiore, con l'area che diventa l'intero quadrato.

Se un ipotetico nuovo test discriminasse perfettamente i malati dai sani, l'area della curva ROC avrebbe valore 1 (corrispondente all'intero quadrato), cioè il 100% di accuratezza. Nel caso in cui il nuovo test non discriminasse per niente i malati dai sani, la curva ROC avrebbe un'area di 0.5 (o 50%) che coinciderebbe con l'area sottostante la diagonale del grafico (linea tratteggiata). Nella realtà, si considera adeguato un test diagnostico con un'area sotto la curva $\geq 80\%$. L'area sotto la curva può assumere valori compresi tra 0.5 e 1.0. Tanto maggiore è l'area sotto la curva (cioè tanto più la curva si avvicina al vertice del grafico) tanto maggiore è il potere discriminante del test [8].

Per l'interpretazione dei valori dell'area sottostante la curva ROC è possibile riferirsi alla classificazione proposta da Swets:

1. $AUC=0.5$ il test non è informativo;
2. $0.5 < AUC \leq 0.7$ il test è poco accurato;
3. $0.7 < AUC \leq 0.9$ il test è moderatamente accurato;
4. $0.9 < AUC < 1$ è altamente accurato;
5. $AUC = 1$ il test è perfetto.

Nello studio in esame viene calcolata una AUC di 0.97 cioè significa che in un ipotetico esperimento in cui si sceglie in 100 diverse prove, in modo random, una coppia di pazienti di cui uno con MICI e uno sano, nel 97% dei casi i livelli di FC sono più alti nel soggetto con MICI rispetto a quello sano.