

RIVISTA QQ

La Qualità e le Qualità in Medicina Generale

Periodico di Audit e Ricerca in Medicina Generale fondato nel 1996

**ANNO 2026
NUMERO 2**

EDITORIAL BOARD

Editor-in-Chief

Roberto Mora

Senior Editors

Mario Baruchello,

**Francesco Del Zotti (Science
Editor)**

Managing Editor

Giulio Rigon

Senior Assistant Editors

**Davide Luppi(MO), Attilio Dalla
Via(PD), Ferdinando**

Petrazzuoli(CE), Sandro

Giroto(VR), Giobatta Gottardi(VI),

Giovanni Merlino(PA), Giorgio

Visentin(VI), Gianluigi

Passerini(SO), Angelica

Salvadori(TO), Renato Tulino(GR),

Andrea Virga(PA), Sonia Zenari(VR)

International Committee

Julian Tudor Hart † (UK), Paul

Wallace † (UK), Hans Thulesius

(SW), Christos Lionis (GR)



Direzione: Via dell'Artigliere, 16 Legnago (Verona)
Redazione: c/o OdM di Vicenza,
Via Paolo Lioy, 13, 36100 Vicenza
Iscrizione Tribunale di Verona n. 1187 del 12.12.95
Proprietario ed Editore: Associazione Qualità Medica
Contatti : info@rivistaqq.org
In collaborazione con <http://www.netaudit.org>

INDICE RIVISTA QQ NUMERO 2 2026

ANNIVERSARIO RIVISTA QQ TRENTENNALE.....	2
EDITORIALE: AI MEDICINA GENERALE E OPPORTUNITÀ E RISCHI.....	7
INTELLIGENZA ARTIFICIALE, COMPLESSITÀ DOCUMENTALE E MODELLO HUMAN-IN-THE-LOOP.....	8
L'INTELLIGENZA ARTIFICIALE E IL RAGIONAMENTO MEDICO: LA DOMANDA NON È SE INDOVINA, MA COME SBAGLIA.....	10
MEDICAL SCRIBES, AI SCRIBES E CLINICAL DECISION SUPPORT SYSTEMS NELLA PRATICA CLINICA REVISIONE DELLA LETTERATURA.....	14
DAL SINTOMO AL PROTOTIPO: UN MEDICO DI FAMIGLIA TRA UK E ITALIA, L'INTELLIGENZA DEGLI ECOSISTEMI E UN VIDEOGIOCO PER RESPIRARE MEGLIO	24
NETAUDIT NT-PRO-BNP, BNP E SCOMPENSO CARDIACO	28
CONGRESSO EGPRN VERONA MAGGIO 2026.....	38

ANNIVERSARIO RIVISTA QQ TRENTENNALE

Una storia di qualità, ricerca e pratica nella Medicina Generale

Un trentennio di Ricerca sulla Qualità di indicatori quantitativi rigorosi e sulle varie Qualità delle varie pratiche e relazioni non-quantitative nella Medicina Generale italiana

A cura del Comitato di Redazione

Nel 1996 un gruppo di colleghi Medici di Medicina Generale (MG) di Verona e Vicenza, provenienti da varie aree culturali (Ordini dei Medici delle due province; Simg; membri della Società Italiana per la Qualità dell'Assistenza Sanitaria – VRQ; membri del Centro Studi FIMMG Verona) constatarono che vi era uno spazio per una rivista culturale della MG italiana che concretamente si occupasse di Qualità in MG e si collegasse alla cultura europea della General Practice. Infatti, sin dall'inizio ci fu l'incoraggiamento del famoso guru della GP internazionale, l'inglese dr Julian Tudor Hart che lavorò in Galles prima come epidemiologo con Cochrane e poi come straordinario GP nel Galles. Inoltre forte fu l'incoraggiamento del GP e professore di primary care a Londra, Prof. Wallace. Del Zotti e Baruchello frequentarono a lungo questi due straordinari leader che diedero una sorta di imprinting alla nostra iniziativa. La SIQUAS e Baruchello offrirono un appoggio organizzativo importante per la registrazione della testata; ed egualmente importante fu l'aiuto del Dr Roberto Mora, che tra noi è l'unico ad essere giornalista accreditato; divenne (ed è) Direttore responsabile: non possiamo che rafforzare il nostro grazie a Roberto.

In un periodo in cui erano limitati gli spazi editoriali dedicati alla medicina generale come disciplina autonoma, il progetto si propose fin dall'inizio come un luogo di confronto professionale, indipendente e orientato alla qualità delle cure. Nel tempo la rivista ha sempre di più acquisito un ruolo nazionale e rapporti internazionali. E negli anni è divenuta sempre più indipendente, non sponsorizzata, senza pubblicità.

L'impostazione editoriale si è basata su una scelta chiara: valorizzare il contributo diretto dei medici di famiglia alla produzione di conoscenza. In particolare fu deciso il lancio di Audit clinici, lancio suggellato dalla nascita 25 anni fa del gruppo di prassi Netaudit (www.netaudit.org). La Rivista QQ in questi 30 anni ha altresì promosso ricerche di servizio, studi osservazionali di piccole dimensioni, riflessioni qualitative e strumenti operativi. Non a caso la rivista ha due Q: la prima si riferisce agli studi di quantitativi in cui con gli Audit ci compariamo ad indicatori quantitativi accreditati (da serie linee guida o da Ebm). La seconda Q (le qualità) richiama la necessità di analisi Qualitative che vadano oltre i numeri e affrontino mediante la cultura biopsicosociale, sistemica, motivazionale le tante complessità non enumerabili facilmente. La preferenza per articoli brevi e leggibili ha avuto l'obiettivo di favorire la partecipazione, ridurre le barriere alla scrittura e sostenere una cultura della condivisione tra pari. Un altro filone di riflessione è quello organizzativo / formativo nel tentativo (non ancora riuscito in Italia) di portare la medicina generale a livello di riconosciuta specialità clinica, entrando alla Università dalla porta principale

come è successo alle Medicine Generali di tutta Europa. Nello stesso tempo la Rivista ha mostrato anche al suo esterno (l'università) i numerosi e svariati contributi para-accademici della MG italiana (articoli, libri, consensus conference tra MG esperti, pubblicazioni nazionali ed internazionali)

Il primo numero della rivista costituisce ancora oggi un riferimento identitario. I temi affrontati – l'organizzazione del lavoro, la qualità delle cure, gli audit clinici, i processi decisionali della medicina generale – anticipavano questioni che sarebbero diventate centrali negli anni successivi. La ricerca sui sistemi di accesso su appuntamento (SAP), allora poco diffusi in Italia, rappresenta un esempio di approccio pragmatico e basato sui dati, attento ai vincoli strutturali e organizzativi della pratica reale. È particolarmente significativo che emerga come la mancata diffusione del SAP non dipenda da una sua inutilità percepita, ma da vincoli sistemici (assenza di supporto organizzativo, costi non riconosciuti, isolamento professionale). Un messaggio che, a distanza di quasi trent'anni, mantiene una sorprendente attualità. Allo stesso modo, la proposta della griglia "modificabilità–importanza" offriva uno strumento semplice e concreto per orientare le priorità negli audit clinici. In un contesto spesso segnato da riunioni inconcludenti e dispersione di energie, la proposta di una griglia decisionale rappresenta un antidoto alla casualità nella scelta dei temi di audit, un supporto alla motivazione dei gruppi, un esempio di metodo "leggero" ma rigoroso.

Sin dall'esordio, la rivista ha affiancato al radicamento territoriale un'apertura al confronto internazionale. Il commento, sin dal primo numero, del Professor Paul Wallace (mancato da pochi mesi), docente di Primary Health Care all'Università di Londra, ha rappresentato un riconoscimento autorevole della qualità metodologica e della coerenza culturale del progetto, inserendolo in un dialogo europeo sulla medicina generale.

A distanza di quasi trent'anni, la rivista è ancora attiva e fedele ai principi che ne hanno guidato la nascita. La sua continuità nel tempo testimonia la validità di una visione che considera la qualità della medicina generale come un processo dinamico, costruito nella pratica quotidiana, alimentato dalla riflessione critica e rafforzato dal confronto tra colleghi. Una visione che continua a dare senso al lavoro del medico di famiglia e a sostenerne lo sviluppo professionale.

IL COMITATO EDITORIALE

Roberto Mora, Mario Baruchello, Francesco Del Zotti,
Giulio Rigon, Ferdinando Petrazzuoli, Attilio Dalla Via,
Sandro Giroto, Gio batta Gottardi, Davide Luppi Giovanni
Merlino, Gianluigi Passerini, Renato Tulino, Andrea Virga,
Diego Sabbi Angelica Salvadori, Sonia Zenari

Bibliografia

Alcuni articoli o capitoli che spiegano la nostra storia e metodi

-Del Zotti F. Brizio E. Netaudit: un sistema telematico collettivo per una medicina generale di qualità . Rivista QA Marzo 2006 , pagine 33-39

-Presentazione internazionale di Netaudit nel libro di Equip (società europea che analizza la qualità in MG) "Improving Patient care in primary care in Europe (editore Equip, 2004)."

Il nostro capitolo, nel suddetto libro, è stato :

Campanini A. , Del Zotti Francesco: "Netaudit": an original online tool for research and quality improvement in general practice (Italy), pagine 93-97

Sabbi D, Salvadori A, Del Zotti F. Tecniche di scavo: audit sugli esiti in medicina generale. Recent Prog Med. 2024 Nov;115(11):552-553. Italian. doi: 10.1701/4365.43598. PMID: 39550659. (articolo indicizzato su Pubmed)

EDITORIALE

MEDICINA GENERALE E OPPORTUNITÀ E RISCHI DALL'USO DI INTELLIGENZA ARTIFICIALE

A cura di G.B. Gottardi e M. Baruchello

L'intelligenza artificiale (IA) è oggi sempre più integrata nella ricerca e nella pratica della medicina di famiglia (FM). Si ritiene che possa contribuire al miglioramento della diagnostica, dell'analisi dei dati ed erogazione dell'assistenza. La sua rapidissima adozione, spinta dalla coincidenza fra la proposta ubiquitaria di fornitura, anche gratuita, da parte delle grandi company mondiali, ha superato la condivisione di riflessioni etiche sulle risultanze del loro largo uso.

Nel campo della FM, l'IA è diventata una risorsa sempre più preziosa, potendo guidare progressi nella pratica clinica e nella ricerca. In ambito clinico, l'IA è utilizzata per migliorare l'accuratezza diagnostica, prevedere gli esiti dei pazienti e personalizzare i piani di cura. Strumenti di IA possono divenire strumenti abituali a supporto della gestione di malattie croniche, in telemedicina e prevenzione. Non parliamo poi dell'uso facilitante nella ricerca sulla FM, come l'analisi di grandi set di dati, con l'automazione estrattiva di grandi set di dati per le revisioni. Tuttavia, "man mano che l'IA si integra nella ricerca, solleva sfide in termini di trasparenza, etica e riproducibilità. Queste sfide includono... un'inadeguata divulgazione dell'utilizzo dell'IA nella progettazione o nella stesura di studi, ... distorsioni algoritmiche che possono perpetuare le disuguaglianze nella ricerca e nell'editoria, e barriere alla riproducibilità derivanti dall'accesso limitato a codici, dati e parametri del modello di ricerca".

Rimandiamo ad approfondimenti successivi l'analisi sulle principali tensioni etiche, talora dilemmatiche, quando si implementano strumenti di IA nei flussi di lavoro in FM (come all'interno dei piani di cura, formazione medica continua, revisione di pratiche, attività di ricerca). Tra questi citiamo i problemi di equità per la discriminazione nell'accesso agli strumenti di AI, l'impatto nel lavoro del medico, la privacy, la responsabilità medica nelle scelte operative, gli errori sistematici.

Qui accenniamo ad un esempio di utile approfondimento: le riviste di FM hanno politiche editoriali esplicite riguardo all'uso di strumenti di IA generativa nella redazione, scrittura di manoscritti, generazione di contenuti e immagini e nella divulgazione? Anche il solo sviluppo di politiche editoriali standardizzate può sollevare preoccupazioni in materia di trasparenza, etica e riproducibilità. Sono urgentemente necessarie, infatti, linee guida chiare da parte delle riviste per garantire un uso responsabile dell'IA nella ricerca e nell'editoria. Esempi ne esistono già: linee guida come CONSORT-AI (Consolidated Standards of Reporting Trials involving Artificial Intelligence) e SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials involving Artificial Intelligence) promuovono la trasparenza, la riproducibilità e gli standard etici offrendo quadri strutturati per il reporting dei metodi di IA.

INTELLIGENZA ARTIFICIALE, COMPLESSITÀ DOCUMENTALE E MODELLO HUMAN-IN-THE-LOOP: UNA NUOVA FRONTIERA PER LA MITIGAZIONE DEL RISCHIO IN SANITÀ

Maurizio Murino - Fondatore e Ceo di Epitomea

Negli ultimi anni il sistema sanitario ha assistito a una crescita costante della complessità informativa. Cartelle cliniche, lettere di dimissione, procedure aziendali, linee guida, requisiti di accreditamento, standard di qualità, normative nazionali ed europee, protocolli interni e documentazione di compliance costituiscono oggi un ecosistema documentale vastissimo e in continua evoluzione. Questa trasformazione ha prodotto benefici importanti in termini di qualità, tracciabilità e sicurezza delle cure, ma ha anche generato una sfida crescente per professionisti e organizzazioni sanitarie. Sempre più spesso il problema non è la mancanza di informazioni, bensì la difficoltà di individuarle, comprenderle, verificarle e mantenerle coerenti all'interno di processi sempre più articolati.

In questo contesto, l'intelligenza artificiale generativa e i moderni sistemi di elaborazione del linguaggio naturale stanno aprendo nuove possibilità. Tuttavia, la loro applicazione in sanità richiede particolare cautela. L'obiettivo non può essere quello di delegare decisioni cliniche a una macchina, ma di costruire strumenti capaci di assistere il professionista nell'analisi di grandi quantità di documentazione. Per questo motivo riteniamo che il paradigma più promettente sia quello dell'Human-in-the-Loop, nel quale l'intelligenza artificiale opera come supporto e non come sostituto dell'essere umano. In questo modello il sistema analizza, estrae, confronta e organizza le informazioni, mentre il professionista mantiene il controllo del processo decisionale e la responsabilità finale delle valutazioni effettuate. È proprio questa la filosofia che guida il lavoro dell'azienda Epitomea e lo sviluppo della piattaforma Library Med.

Library Med è stato il primo prodotto sviluppato dal nostro team ed era focalizzato sull'analisi e sulla sintesi della letteratura scientifica. Attraverso questo progetto abbiamo sviluppato metodologie e tecnologie per l'interpretazione di documenti complessi mediante modelli di intelligenza artificiale. L'esperienza maturata con Library Med ha successivamente portato Epitomea, la società, a sviluppare soluzioni di analisi documentale applicabili a contesti molto più ampi. Pur mantenendo la letteratura scientifica come uno degli ambiti di interesse, Epitomea ha esteso il proprio raggio d'azione alla documentazione clinica, ai processi di qualità, alla compliance normativa, ai sistemi di certificazione e, più in generale, a tutti quei contesti in cui la crescente complessità documentale rappresenta una fonte di rischio operativo, clinico o organizzativo.

I sistemi sviluppati da Epitomea sono progettati per leggere e interpretare documenti complessi, identificare informazioni rilevanti, confrontarle con requisiti normativi o procedurali e segnalare eventuali criticità, omissioni, incoerenze o aree che meritano un approfondimento umano. Non forniscono verdetti definitivi né sostituiscono il giudizio professionale; producono invece elementi di supporto che consentono una revisione più rapida, sistematica e documentata. L'obiettivo principale è la mitigazione del rischio.

Dal punto di vista clinico, il rischio deriva spesso dalla possibilità che informazioni rilevanti vengano trascurate all'interno di una mole crescente di dati e documenti. Una documentazione incompleta, incoerente o difficilmente consultabile può incidere negativamente sulla continuità assistenziale e, nei casi più critici, sulla sicurezza del paziente. Dal punto di vista professionale e organizzativo, esiste inoltre un rischio legale e reputazionale sempre più rilevante. Le strutture sanitarie operano all'interno di un contesto normativo complesso, nel quale la capacità di dimostrare conformità, tracciabilità e correttezza documentale assume un'importanza crescente. Errori formali, mancanze documentali o difformità procedurali possono generare conseguenze significative anche quando l'attività clinica è stata svolta correttamente.

In tale scenario, l'intelligenza artificiale può essere considerata come un nuovo livello di controllo e verifica, capace di supportare gli operatori nell'individuazione preventiva di potenziali criticità. Il suo valore non risiede nella sostituzione delle competenze umane, ma nell'aumentarne la capacità operativa in contesti caratterizzati da elevata complessità informativa. La vera innovazione non consiste quindi nell'automazione completa dei processi decisionali, bensì nella costruzione di sistemi collaborativi in cui uomo e macchina operano insieme. L'intelligenza artificiale contribuisce con velocità di analisi, capacità di elaborazione e uniformità di controllo; il professionista apporta esperienza, responsabilità, comprensione del contesto e capacità di giudizio.

Riteniamo che questa integrazione rappresenti una delle direzioni più promettenti per il futuro della sanità digitale. In un settore nel quale la complessità continuerà inevitabilmente ad aumentare, gli strumenti capaci di supportare la qualità documentale e la gestione del rischio diventeranno sempre più centrali. Non per sostituire il medico o l'organizzazione sanitaria, ma per consentire loro di operare in modo più sicuro, consapevole ed efficace.

L'intelligenza artificiale non deve essere vista come un decisore autonomo, ma come un alleato nella gestione della complessità. È in questa prospettiva che *Epitomea* continua a sviluppare le proprie soluzioni: utilizzare la tecnologia per ridurre il rischio, aumentare la trasparenza dei processi e supportare il lavoro di chi, ogni giorno, è chiamato a prendere decisioni che hanno un impatto diretto sulla salute delle persone.

L'INTELLIGENZA ARTIFICIALE E IL RAGIONAMENTO MEDICO: LA DOMANDA NON È SE INDOVINA, MA COME SBAGLIA

Giampaolo Collecchia – Medicina Generale Massa Carrara , autore del libro "Intelligenza artificiale e medicine digitale. Una guida critica"

Fino a poco tempo fa una domanda dominava il dibattito sull'intelligenza artificiale in medicina: i grandi modelli linguistici possono imitare il ragionamento clinico dei medici? La risposta sembrava prudente ma incoraggiante. Un modello come GPT-4, opportunamente guidato da prompt diagnostici, riusciva a costruire spiegazioni in forma di diagnosi differenziale, ragionamento analitico, intuizione clinica o inferenza bayesiana senza perdere accuratezza rispetto al chain-of-thought tradizionale[1]. Non era la prova che il modello ragionasse davvero come un medico, ma mostrava qualcosa di rilevante: l'LLM poteva rendere visibile una parte del proprio percorso, offrendo al clinico una traccia da verificare. Attualmente la domanda è cambiata. Non ci chiediamo più soltanto se l'intelligenza artificiale sappia arrivare alla diagnosi corretta. Ci chiediamo se sappia attraversare l'intero processo clinico: generare una diagnosi differenziale ampia e sensata, scegliere gli esami appropriati, rivedere le ipotesi alla luce di nuovi dati, gestire l'incertezza, proporre un piano terapeutico coerente. In altre parole: non basta più il risultato finale ma il percorso.

La diagnosi differenziale

I modelli più avanzati possono essere molto efficaci quando il caso è già sufficientemente definito e quando devono produrre una diagnosi finale. Sono meno affidabili nelle fasi iniziali del ragionamento, quelle in cui il medico deve ancora definire il problema. L'LLM può fornire didatticamente una lista di diverse ipotesi diagnostiche ma senza aiutare molto nella diagnosi differenziale[2],[3]. Questo è un dato cruciale, perché nella pratica clinica la diagnosi spesso non nasce come risposta secca ma come ventaglio di possibilità, confronto tra ipotesi, progressiva riduzione dell'incertezza, un ambito nel quale i medici sono esperti e apportano il massimo valore. I dati osservati assumono infatti significato diverso in funzione del contesto clinico, delle loro caratteristiche, della cronologia di comparsa, della frequenza nella malattia ipotizzabile. Di fronte a sintomi-segni aspecifici e poco definiti, l'algoritmo tende a proporre una grande quantità di esami ed accertamenti, una sorta di "rete a strascico" che, utile ai fini di trovare la diagnosi corretta, non tiene conto dell'appropriatezza delle risposte. L'IA lavora infatti in maniera efficace in situazioni che favoriscono l'approccio statistico, ma non può accedere ai dati non verbali e para-verbali del colloquio con il paziente. Invece, soprattutto nelle fasi iniziali della cura, il curante dispone di informazioni spesso indistinte, dove prevalgono elementi soggettivi e sono necessarie qualità tacite, come intuizione, occhio clinico, empatia per interpretare il

volto del paziente, il suo modo di parlare, i silenzi significativi, la postura, le incertezze, la sequenza espositiva, il detto e il non detto, la sofferenza, la ricchezza e le sfumature che servono a 'profilare' correttamente il paziente e che invece possono rappresentare un bias (di profilazione) per l'IA. Ciò che ancora distingue il ragionamento del medico dalle procedure dell'IA si colloca proprio nell'ambito della condivisione di valori, credenze e cultura, nel valore aggiunto della conoscenza continuativa dei pazienti, nella possibilità di 'vedere e toccare' i corpi che cura. I modelli generativi imparano dai dati di testo e (per ora?) non hanno modo di interagire con mondo fisico. Manca il cosiddetto "grounding", il "mettere i piedi per terra", l'acquisizione del senso della realtà circostante e della socialità che gli umani imparano già poco dopo la nascita, la capacità di collegare i simboli agli oggetti o ai concetti del mondo reale per capire il contesto e il significato dei simboli che stanno manipolando. Il ragionamento del medico, che tiene conto di tutto quel "sapere di sfondo" di cui parla Popper, è ancora indispensabile nel processo diagnostico-terapeutico[4]. In questi casi appare evidente il limite dell'imitazione linguistica. Un modello può proporre una spiegazione elegante, ma questa non è necessariamente un ragionamento robusto. Può arrivare alla diagnosi giusta per vie "fragili", oppure formulare una diagnosi plausibile escludendo troppo presto alternative importanti. La chiusura prematura del ragionamento è peraltro una delle fonti classiche di errore diagnostico anche umano.

Il triage

Ciò emerge in particolare nel contesto del triage. Alcuni modelli di nuova generazione possono mostrare buone prestazioni nel ragionamento diagnostico su casi di medicina d'urgenza, ma questo non equivale a una capacità affidabile. Uno studio pubblicato su Science nel 2026 ha mostrato il lato più promettente: un LLM ha identificato diagnosi esatte o molto vicine nel 67,1% dei casi al triage iniziale di pronto soccorso, con risultati che suggeriscono un ruolo possibile come "secondo parere" nei casi complessi, non come sostituto autonomo del medico[5]. Il triage non consiste peraltro soltanto nel fornire una diagnosi possibile: significa decidere quanto rapidamente un paziente debba essere visto, riconoscere le emergenze anche quando i sintomi sono incompleti o ambigui. Questo appare particolarmente evidente nei casi di modelli rivolti direttamente alla popolazione. Uno studio pubblicato su Nature Medicine ha riportato i risultati di ChatGPT Health, la funzione di OpenAI lanciata a gennaio per offrire consigli sanitari collegando cartelle cliniche e app di benessere. I ricercatori intendevano verificare se di fronte a una vera emergenza ChatGPT Health fosse in grado di indirizzare l'utente al pronto soccorso. Il team ha costruito 60 scenari clinici realistici (da malattie lievi a emergenze acute) ottenendo quasi 1.000 risposte dalla piattaforma, e le ha confrontate poi con le valutazioni di tre medici indipendenti. Se nelle emergenze "da manuale" le performance sono state adeguate, in situazioni più ambigue, come crisi asmatiche o chetoacidosi diabetica, l'IA ha fornito indicazioni inappropriate. Ad esempio, nel 51,6% dei casi che richiedevano accesso immediato al pronto soccorso, il sistema ha sottovalutato l'urgenza, suggerendo di restare a casa o fissare una visita ordinaria. In pratica il sistema tendeva a sottostimare le emergenze e a sovrastimare i casi lievi, con rischi di sicurezza nei casi estremi[6]. Tenendo conto del fatto che secondo OpenAI oltre 40 milioni di persone ogni giorno consultano ChatGPT per questioni riguardanti la salute, la posta in gioco è altissima. Questo è infatti il tipo di errore più pericoloso in medicina:

non semplicemente sbagliare una diagnosi, ma assicurare quando sarebbe necessario agire subito. Per questo è importante distinguere tra “supporto al ragionamento diagnostico” e “strumento di triage”. Un LLM può essere utile come secondo parere, come generatore di ipotesi o come aiuto alla diagnosi differenziale, ma i dati disponibili non supportano un uso autonomo. La direzione più promettente non è quindi l'autonomia, ma la collaborazione controllata.

Il LLM ideale

Un LLM clinicamente utile dovrebbe fare tre cose. Primo: formulare una diagnosi differenziale pertinente e coerente, non solo una diagnosi finale. Secondo: dichiarare quali dati supportano o indeboliscono ciascuna ipotesi. Terzo: indicare quali informazioni mancano per ridurre l'incertezza. In questo modo il modello non sarebbe una sorta di oracolo, ma uno strumento di lavoro che produce ipotesi, espone passaggi, accetta verifiche, in pratica fornisce una interpretabilità clinica, cioè una spiegazione che sia non solo fluente e coerente, ma anche verificabile, contestabile e integrata nel flusso reale della decisione medica. L'obiettivo è rendere il ragionamento non solo leggibile, ma “auditabile”. Non basta che l'AI dica “penso che sia questa diagnosi”; deve mostrare quali dati, regole, evidenze, passaggi logici hanno portato a quel risultato[7]. Questo vale ancora di più al di fuori della lingua inglese. Le valutazioni più recenti indicano che la lingua del prompt può influenzare sia l'accuratezza diagnostica sia la qualità del ragionamento. È un punto essenziale per l'Europa e per i sistemi sanitari multilingue: un modello che ragiona bene in inglese non è automaticamente affidabile in italiano, francese o altre lingue[8].

Conclusioni

L'intelligenza artificiale in ambito medico non va giudicata soltanto sulla capacità di “indovinare” ma su quella di sbagliare in modo riconoscibile, perché un errore visibile può essere corretto dal medico. Per questo il futuro non sarà probabilmente quello di una macchina che sostituisce il clinico, ma di una relazione duale. Il medico resta responsabile della decisione, ma può usare il modello come interlocutore diagnostico, generatore di ipotesi e strumento di controllo versus omissioni e bias cognitivi. La vera promessa non è un'intelligenza artificiale che “fa diagnosi” ma che aiuta a chiarire la qualità del percorso per arrivare a farla. E la vera domanda non è se l'AI sappia imitare il ragionamento medico ma se sapremo costruire sistemi, metriche e responsabilità capaci di distinguere un ragionamento clinico autenticamente utile da una spiegazione soltanto convincente.

Bibliografia

- [1] Savage T, Nayak A, Gallo R, Rangan E, Chen JH. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. *npj Digital Medicine*. 2024;7:20. doi:10.1038/s41746-024-01010-1
- [2] Rao AS, Esmail KP, Lee RS, et al. Large Language Model Performance and Clinical Reasoning Tasks. *JAMA Netw Open*. 2026;9(4):e264003. doi:10.1001/jamanetworkopen.2026.4003
- [3] Tordjman M, Mei X. Limitations of Large Language Models in Clinical Diagnostic Reasoning. *JAMA Netw Open*. 2026;9(4):e264014. doi:10.1001/jamanetworkopen.2026.4014
- [4] Rossi RL, Collecchia G, De Gobbi R, Fassina R, Zamperini D. L'errore medico – Un percorso fra bias cognitivi, aspetti etici e conseguenze legali. Il Pensiero Scientifico Editore, 2025
- [5] Peter G, Brodeur et al. Performance of a large language model on the reasoning tasks of a physician. *Science* 392,524-527(2026).DOI:10.1126/science.adz4433
- [6] Ramaswamy, A., Tyagi, A., Hugo, H. et al. ChatGPT Health performance in a structured test of triage recommendations. *Nat Med* 32, 1671–1675 (2026). <https://doi.org/10.1038/s41591-026-04297-7>
- [7] Fan X, Cai Y, Hou Z, Dong JS. Uncertainty Reasoning with Large Language Models for Explainable Disease Diagnosis. <https://arxiv.org/abs/2605.25566> 2026.
- [8] Bazoge A, Corvellec J, Djillali Sid-Ahmed S, Gourraud PA. Prompting language influences diagnostic reasoning and accuracy of large language models. <https://arxiv.org/abs/2605.19173> 2026.

Lettura consigliata

Collecchia G, De Gobbi R, Fassina R, Ressa G, Rossi RL. La diagnosi ritrovata - Le basi del ragionamento clinico Il Pensiero Scientifico Editore, 2021

MEDICAL SCRIBES, AI SCRIBES E CLINICAL DECISION SUPPORT SYSTEMS NELLA PRATICA CLINICA REVISIONE DELLA LETTERATURA

Clementino Stefanetti, MMG - Casorezzo (MI)

Introduzione

La trasformazione digitale della pratica clinica ha reso la documentazione elettronica un elemento centrale dell'assistenza, ma anche una fonte crescente di carico amministrativo. Nel contesto ambulatoriale, la compilazione della cartella, la chiusura delle note, la produzione di lettere e sintesi, la gestione delle prescrizioni e l'aggiornamento dei dati clinici possono sottrarre tempo alla relazione medico-paziente e contribuire al lavoro fuori orario. Per questo motivo medical scribes, digital scribes e AI scribes sono stati presentati come possibili strumenti per recuperare tempo clinico e ridurre il burden documentale. L'entusiasmo tecnologico, tuttavia, deve essere interpretato criticamente. Ridurre il tempo di scrittura non equivale automaticamente a migliorare la pratica clinica. La qualità della visita dipende anche da accuratezza diagnostica, appropriatezza prescrittiva, continuità assistenziale, sicurezza farmacologica, comprensibilità della nota, responsabilità medico-legale e protezione dei dati sanitari. Un sistema che produce rapidamente una nota plausibile ma incompleta o non integrata con la cartella può addirittura aumentare il rischio clinico-documentale.

In medicina generale italiana la questione assume caratteristiche specifiche. Il gestionale ambulatoriale rimane il sistema di record e il punto di raccordo operativo con prescrizione, certificazioni, esenzioni, presa in carico dei cronici, Patient Summary e Fascicolo Sanitario Elettronico. Uno scribe isolato, non integrato e privo di audit trail, può essere comodo nella fase di scrittura ma debole nella filiera assistenziale. Al contrario, uno strumento integrato, validabile e tracciabile può contribuire alla qualità documentale e alla continuità di cura.

Definizioni operative

Per evitare ambiguità, è utile distinguere quattro categorie. Il medical scribe umano è un operatore non clinico che assiste il medico nella documentazione in tempo reale o subito dopo la visita. Il digital scribe comprende strumenti di dettatura, riconoscimento vocale e trascrizione, non necessariamente generativi. L'ambient AI scribe utilizza riconoscimento vocale, natural language processing e modelli generativi per ascoltare o processare la visita e produrre una bozza strutturata di nota clinica. Il CDSS, invece, non ha come funzione principale la scrittura della nota, ma il supporto decisionale attraverso regole, alert, calcolatori, knowledge base, modelli di rischio o sistemi di telemonitoraggio.

La distinzione è cruciale. Uno scribe può aiutare a documentare ciò che è avvenuto; un CDSS può aiutare a decidere cosa fare. Entrambi possono coesistere, ma hanno profili di rischio diversi: omissioni, allucinazioni e privacy opaca per gli scribes; alert fatigue, regole rigide e dipendenza da dati incompleti per i CDSS.

Metodi della sintesi narrativa

Questo documento è una sintesi narrativa accademica, non una revisione sistematica indipendente. Integra tre livelli di evidenza: studi primari recenti su ambient AI scribes, revisioni sistematiche/metanalisi su scribes umani e strumenti AI per la documentazione, e considerazioni operative per il contesto italiano della medicina generale. Sono stati privilegiati studi con disegno randomizzato o multicentrico, revisioni sistematiche, metanalisi e fonti istituzionali o regolatorie pertinenti alla governance dei dati sanitari.

Gli outcome sono stati interpretati separando quattro dimensioni: efficienza documentale, produttività operativa, benessere del clinico e valore clinico-assistenziale. Questa separazione è necessaria perché un miglioramento del tempo di documentazione o degli RVU non dimostra automaticamente un miglioramento di qualità, sicurezza o outcome del paziente.

Evidenze su medical scribes umani

La letteratura sugli scribes umani è più matura rispetto a quella sugli AI scribes, ma resta metodologicamente eterogenea. La revisione sistematica e metanalisi di Heaton et al. ha rappresentato una delle prime sintesi quantitative, focalizzandosi su throughput, revenue e soddisfazione. Il segnale complessivo suggeriva benefici possibili su alcuni indicatori operativi e di soddisfazione, ma senza prova forte di riduzione della length of stay in pronto soccorso. La qualità degli studi inclusi era limitata, con molti disegni osservazionali e contesti prevalentemente di emergenza-urgenza. La metanalisi aggiornata di Gottlieb et al. ha incluso un numero maggiore di studi e oltre mezzo milione di encounter. Il risultato più concreto è un incremento medio di circa 0,30 pazienti per ora, con aumento di RVU per encounter e per ora e miglioramento della soddisfazione del provider. Anche in questa revisione, tuttavia, non emerge un miglioramento chiaro della length of stay in pronto soccorso. Il dato è quindi coerente con un effetto produttivo-operativo plausibile, ma non necessariamente con un miglioramento dell'efficienza globale del sistema o della qualità clinica. La revisione del Veterans Affairs Evidence Synthesis Program ha aggiunto un elemento importante: l'effetto degli scribes cambia molto a seconda del setting. In emergenza i benefici su pazienti/ora e alcuni parametri di flusso appaiono più plausibili; in cardiologia il dato è incerto; in ortopedia l'evidenza utile è scarsa. Inoltre, lo scribe umano non è una semplice tecnologia ma un intervento organizzativo: richiede selezione, training, supervisione, copertura delle assenze, gestione del turnover e valutazione economica.

Categoria	Funzione primaria	Output tipico	Beneficio atteso	Rischio prevalente
Medical scribe umano	Documentazione delegata o assistita	Nota clinica, sintesi, dati EHR inseriti da operatore	Riduzione del carico di inserimento dati	Costi, turnover, privacy, variabilità della qualità
Digital scribe / speech recognition	Dettatura e trascrizione	Testo clinico da correggere	Velocità di redazione	Errori di riconoscimento, accenti, rumore
AI / ambient scribe	Generazione di bozza clinica da conversazione	SOAP, anamnesi, assessment, piano, lettere	Riduzione burden documentale e potenziale miglioramento workflow	Omissioni, inferenze non dette, hallucinations, governance dati
CDSS prescrittivo	Supporto decisionale farmacologico	Alert interazioni, dose, duplicazioni, controindicazioni	Sicurezza prescrittiva	Alert fatigue, dati incompleti
CDSS guideline-based / cronicità	Promemoria e percorsi di cura	Checklist, richiami, target, follow-up	Appropriatezza e continuità assistenziale	Rigidità, sovraccarico informativo

Evidenze su AI/ambient scribes

Gli ambient AI scribes sono tecnologie più recenti. La promessa è più ambiziosa rispetto agli scribes umani: non solo trascrivere, ma sintetizzare la visita, strutturare la nota e talora proporre lettere, piani e task. La letteratura recente è promettente ma ancora non conclusiva. L'editoriale di Kim, Liu e Singh in NEJM AI ha fornito una cornice interpretativa prudente: gli AI scribes non dovrebbero essere considerati automaticamente strumenti maturi di produttività. Possono ridurre una parte della documentazione, ma richiedono revisione, correzione e adattamento al workflow. La metrica "minuti risparmiati" va quindi integrata con qualità della nota, sicurezza, accuratezza, esperienza del paziente e lavoro fuori orario. Il trial randomizzato pragmatico di Lukac et al. ha incluso 238 medici ambulatoriali di 14 specialità e ha confrontato due ambient AI scribes con la pratica abituale. Uno dei due strumenti, Nabla, ha ridotto il time-in-note di circa il 9,5%; DAX non ha prodotto una riduzione statisticamente significativa nel confronto principale. Il dato è importante perché mostra sia la plausibilità del beneficio sia la sua dipendenza dal prodotto, dal contesto e dal modo in cui lo strumento viene incorporato nella pratica. Il trial pragmatico stepped-wedge di Afshar et al. ha valutato l'impatto sul benessere professionale. L'ambient AI ha ridotto work exhaustion e interpersonal disengagement, ma non ha aumentato significativamente il professional fulfillment. Questo risultato suggerisce che il razionale più solido possa essere la riduzione di alcune componenti negative del carico documentale, più che un miglioramento globale dell'esperienza professionale o della produttività.

Lo studio multicentrico JAMA di Rotenstein et al. ha valutato l'adozione di AI scribes in cinque academic medical centers. L'adozione è stata associata a una riduzione modesta del tempo EHR e del tempo di documentazione, e a un piccolo incremento del volume di visite settimanali; non è stata osservata una riduzione significativa del tempo EHR fuori orario. La natura osservazionale impone cautela causale, ma lo studio è utile perché documenta un beneficio reale ma contenuto in contesti real-world.

Sintesi delle revisioni sistematiche e metanalisi su AI e documentazione

Le revisioni sistematiche sugli AI scribes confermano il potenziale documentale ma evidenziano limiti importanti. La revisione di Sasseville et al. ha incluso otto studi specifici sugli AI scribes, con miglioramenti in documentazione e workflow ma impatto limitato e non uniforme su burnout, outcome del paziente ed efficienza di sistema. La revisione ha sottolineato accuratezza variabile, necessità di editing, problemi di integrazione EHR e questioni etiche legate a consenso, privacy e data governance. La metanalisi di Zhao et al. sugli strumenti AI per la documentazione clinica è oggi una delle più rilevanti per il tema. Ha incluso 23 studi, con un effetto complessivo favorevole sulla riduzione del burden documentale e del burnout correlato alla documentazione. La stima aggregata è promettente, ma deve essere interpretata con prudenza: eterogeneità elevata, qualità metodologica generalmente bassa, molti studi before-after, pochi trial randomizzati e possibile publication bias per outcome temporali. La revisione di Ng et al. sulla speech recognition clinica illumina la componente tecnologica di base di molti

Revisione	Intervento	Risultati principali	Lettura critica
Heaton et al., 2016	Scribes umani	Possibili benefici su throughput, revenue e soddisfazione; nessun segnale forte su length of stay.	Evidenza preliminare, eterogenea e soprattutto osservazionale.
Gottlieb et al., 2021	Scribes umani	+0,30 pazienti/ora; aumento RVU/encounter e RVU/ora; soddisfazione provider migliore; nessuna riduzione ED length of stay.	Beneficio operativo plausibile, ma dipendente da setting, volume e costi.
Ullman et al., 2020	Scribes in cardiologia, ortopedia, ED	Benefici più plausibili in ED; evidenza incerta in cardiologia; insufficiente in ortopedia.	Sottolinea la natura organizzativa e non puramente tecnologica dell'intervento.

scribes digitali. Il riconoscimento vocale può migliorare velocità e precisione in alcuni contesti, ma la performance varia in funzione di rumore, qualità audio, accenti, terminologia specialistica, turn-taking conversazionale e complessità clinica. Questo è particolarmente rilevante in medicina generale italiana, dove dialetti, linguaggio

colloquiale, pazienti anziani, ipoacusia e presenza di familiari possono alterare la qualità della trascrizione.

Medical scribes versus CDSS: complementarità e confini

Medical scribes e CDSS non sono sinonimi. Lo scribe opera sulla documentazione; il CDSS opera sulla decisione o sull’organizzazione del percorso. Questa distinzione è essenziale per non attribuire allo scribe un beneficio che appartiene a sistemi di supporto prescrittivo o di gestione della cronicità. Un AI scribe può produrre una nota ben scritta ma non verificare automaticamente la dose di un farmaco nell’insufficienza renale, non richiamare vaccini mancanti e non stratificare correttamente un rischio cardiovascolare se non integrato con dati strutturati e regole validate.

Il modello più promettente per la medicina generale non è quindi lo scribe isolato, ma una filiera in cui documentazione assistita, campi strutturati, codifiche, CDSS prescrittivi e telemonitoraggio dialogano dentro un sistema governato. In questa prospettiva lo scribe riduce il carico redazionale, mentre il CDSS intercetta errori o omissioni decisionali. Tuttavia, l’integrazione deve essere progettata per evitare alert fatigue, duplicazioni, falso senso di sicurezza e aumento del carico cognitivo.

Punto chiave

Uno scribe migliora soprattutto la produzione del testo clinico; un CDSS migliora potenzialmente la decisione clinica. Il valore massimo si ottiene quando la nota generata diventa dato validato, strutturato e utilizzabile da sistemi di supporto decisionale, senza perdere controllo umano e tracciabilità.

Studio	Disegno e setting	Risultato principale	Interpretazione
Kim et al., 2025	Editoriale NEJM AI su trial recenti	Gli AI scribes non sono ancora strumenti maturi di produttività.	Cornice critica: distinguere documentazione, produttività e valore clinico.
Lukac et al., 2025	RCT pragmatico; 238 medici; 14 specialità	Nabla riduce time-in-note di circa 9,5%; DAX non significativo nel confronto principale.	Effetto modesto e prodotto-dipendente.
Afshar et al., 2025	RCT pragmatico stepped-wedge; 66 operatori; 24 settimane	Riduzione di work exhaustion e interpersonal disengagement; professional fulfillment non significativamente aumentato.	Più convincente come intervento anti-burden che come leva produttiva.
Rotenstein et al., 2026	Cohort multicentrica; 5 centri; 8581 clinici; 1809 adottatori	-13,4 min EHR/8 ore cliniche; -16,0 min documentazione; +0,49 visite/settimana; non ridotto EHR fuori orario.	Benefici misurabili ma modesti; cautela causale.

Implicazioni per il MMG: cartella clinica, privacy e FSE

Nel contesto italiano il gestionale del MMG è il vero nodo operativo. L'integrazione con la cartella clinica è il discrimine tra uno strumento utile e un'applicazione suggestiva. Un AI scribe può produrre un testo eccellente, ma generare poco valore se obbliga il medico a copia-incolla, non aggiorna campi strutturati, non produce audit log o non consente un flusso affidabile verso il Patient Summary e il Fascicolo Sanitario Elettronico.

L'integrazione può essere pensata su quattro livelli progressivi. Il livello minimo è la dettatura o inserimento nel campo attivo, semplice ma poco strutturato. Il secondo livello è una sidebar con bozza validabile e incolla selettivo. Il terzo è un middleware documentale con import/export strutturato. Il quarto, più maturo, è l'integrazione nativa tramite plugin o API del gestionale, con campi strutturati, ruoli, audit trail e interoperabilità più robusta.

Riferimento	Tipo di evidenza	Messaggio principale	Criticità
Sasseville et al., 2025	Revisione sistematica narrativa su AI scribes	Promessa su efficienza documentale e workflow.	Evidenza limitata, eterogenea; pochi dati su outcome paziente e sistema.
Zhao et al., 2025/2026	Revisione sistematica e metanalisi su AI tools per documentazione	Riduzione moderata di burden/burnout documentale e tempo di documentazione.	Qualità bassa, I2 elevato, pochi RCT, rischio publication bias e obsolescenza tecnologica.
Ng et al., 2025	Revisione sistematica su speech recognition clinica	Potenziale miglioramento di velocità/precisione in contesti selezionati.	Errori con accenti, rumore, terminologia specialistica, workflow complessi.

La privacy non si esaurisce nel consenso. Gli ambient scribes possono trattare audio, trascrizioni grezze, bozze, log tecnici e dati sanitari identificabili. I punti da valutare includono luogo di processing, subfornitori, tempi di conservazione, riuso per training o analitici, cifratura, ruoli autorizzativi, minimizzazione, informativa al paziente, possibilità di rifiuto e DPIA (Data Protection Impact Assessment) quando necessaria. Il principio operativo deve essere: registrare meno possibile, conservare solo ciò che serve, vietare riusi non controllati, validare sempre prima del salvataggio.

Il rapporto con il FSE deve essere interpretato realisticamente. Nella filiera attuale, lo scribe non dovrebbe essere considerato un canale autonomo verso il Fascicolo. La sequenza più sicura è: lo scribe produce una bozza; il medico valida; il gestionale struttura, firma e trasmette secondo le regole regionali e nazionali. Il Patient Summary richiede qualità, sintesi e aggiornamento: una nota generata automaticamente senza controllo può compromettere continuità e sicurezza.

Criticità metodologiche e cliniche

Outcome surrogati. Molti studi misurano tempo di documentazione, tempo EHR, RVU, pazienti/ora o soddisfazione. Questi outcome sono utili ma non dimostrano automaticamente migliore qualità clinica, appropriatezza diagnostica, sicurezza o esiti del paziente.

Eterogeneità. Studi e revisioni aggregano pronto soccorso, specialistica, primary care, contesti ospedalieri, strumenti stand-alone, sistemi EHR-integrati e prodotti diversi. Il risultato medio è quindi poco trasferibile al singolo ambulatorio.

Disegni osservazionali e pre-post. Molti studi sono esposti a confondimento, effetto novità, selezione degli early adopters, regressione alla media e cambiamenti organizzativi concomitanti.

Revisione umana e tempo di editing. Il beneficio reale dipende da quanto tempo il medico impiega a correggere la nota. Una bozza formalmente elegante ma clinicamente incompleta può richiedere revisione estesa.

Accuratezza e allucinazioni. Gli AI scribes possono omettere elementi, sintetizzare in modo fuorviante o generare dettagli non espressi. La responsabilità finale rimane del clinico.

Produttività non uguale a qualità. Aumentare pazienti/ora o visite/settimana può migliorare metriche economiche, ma non dimostra migliore presa in carico. In alcuni contesti può aumentare pressione produttiva.

Cost-effectiveness insufficiente. Sono spesso sottostimati licenze, training, supporto tecnico, integrazione EHR, gestione privacy, audit e tempo di revisione.

Equità e inclusione. Pazienti anziani, ipoacusici, multilingue, con dialetti, linguaggio colloquiale, deficit cognitivi o presenza di familiari possono essere più esposti a errori o esclusione digitale.

Obsolescenza rapida. Gli strumenti AI cambiano rapidamente; studi su versioni precedenti possono diventare rapidamente poco rappresentativi. Serve valutazione continua.

Rischio di normalizzazione dell'errore. Se il clinico si abitua a fidarsi della bozza, può ridurre la sorveglianza critica. Il workflow deve mantenere momenti espliciti di validazione.

Livello	Modalità	Vantaggio	Criticità	Rapporto con FSE
1	Dettatura/testo nel campo attivo	Semplice da adottare	Dato poco strutturato	Indiretto; dipende dal gestionale
2	Sidebar con bozza validabile	Maggiore controllo umano	Copia-incolla ancora presente	Contributo indiretto ai contenuti trasmessi
3	Middleware o documento strutturato	Migliore riuso del contenuto	Rischio di creare solo allegati	Possibile supporto al Patient Summary se mappato
4	Plugin/API native EHR	Campi strutturati, audit migliore	Richiede sviluppo con vendor	È il livello più coerente con interazioni robuste FSE

Proposta di implementazione locale per uno studio di medicina generale

L' introduzione di uno scribe o di un ecosistema scribe-CDSS dovrebbe essere trattata come un progetto di miglioramento organizzativo. È opportuno partire da un periodo basale di 4-6 settimane, misurando tempo EHR, tempo di documentazione, lavoro fuori orario, numero di visite, qualità delle note e percezione del carico. La fase pilota dovrebbe includere visite brevi, follow-up cronici, prime visite complesse e situazioni medico-legali, evitando di generalizzare dai soli early adopters.

Area di rischio	Descrizione	Impatto	Misure di mitigazione
Flussi verso fornitori esterni	Audio, trascrizioni, bozze e log verso cloud o sub-responsabili	Alto	DPA, data mapping, cifratura, localizzazione, policy di conservazione
Riuso secondario dei dati	Training, analitici, debugging umano non chiaramente delimitati	Alto	Clausole esplicite, opt-out, limitazione finalità, valutazione DPIA
Eccesso di dati	Registrazione di contenuti non pertinenti alla cura	Medio-alto	Minimizzazione, niente audio persistente salvo necessità, cancellazione automatica
Errore documentale	Omissioni, attribuzioni errate, frasi plausibili ma non dette	Alto	Human in the loop, audit a campione, salvataggio solo dopo conferma
Accessi indebiti	Accesso tecnico o interno non necessario	Alto	Least privilege, audit log, ruoli granulari, segregazione ambienti

Durante il pilota, un campione casuale di note dovrebbe essere sottoposto ad audit per omissioni, errori, inferenze non supportate, coerenza del piano e corretta attribuzione di farmaci, allergie e sintomi. Il beneficio netto va calcolato sottraendo dal tempo di scrittura risparmiato il tempo di revisione e correggendo per costi di licenza, formazione, integrazione, supporto tecnico e governance privacy.

Fase	Obiettivo	Indicatori minimi
1. Baseline	Misurare il carico prima dell'intervento	Tempo EHR, tempo in nota, lavoro fuori orario, visite/settimana, qualità note
2. Pilota controllato	Testare lo strumento in contesti rappresentativi	Quota visite con scribe, tempo editing, soddisfazione clinico/paziente
3. Audit di accuratezza	Verificare sicurezza e qualità	Omissioni, errori, hallucinations, farmaci/allergie, coerenza piano
4. Valutazione economica	Stimare valore netto	Costo licenza, formazione, integrazione, tempo risparmiato, retention
5. Decisione	Proseguire solo se beneficio documentato	Soglie predefinite di accuratezza, tempo, privacy e sostenibilità

Conclusioni

Gli scribes umani e gli AI scribes non sono inutili: le evidenze mostrano un segnale favorevole su burden documentale, soddisfazione del clinico e, in alcuni setting, produttività operativa. Tuttavia, non sono ancora dimostrati come strumenti maturi e generalizzabili di miglioramento della pratica clinica complessiva. La letteratura misura prevalentemente outcome documentali o organizzativi, mentre restano meno studiati qualità clinica, sicurezza, errori, appropriatezza, outcome dei pazienti, continuità assistenziale, privacy e valore economico netto.

Per il MMG italiano, il criterio decisionale non dovrebbe essere l'effetto promesso, ma la capacità dello strumento di inserirsi in una filiera sicura: bozza generata, revisione umana, campi strutturati, integrazione con gestionale, audit trail, protezione dei dati, eventuale collegamento al Patient Summary/FSE e monitoraggio continuo. In questo quadro, lo scribe può diventare un supporto concreto per recuperare tempo clinico; fuori da questo quadro, rischia di essere solo una nuova interfaccia da controllare.

Concludendo, allo stato attuale, gli AI/medical scribes possono ridurre parte del carico documentale e migliorare alcuni domini del benessere professionale, ma non hanno ancora dimostrato in modo robusto di migliorare, da soli, la qualità complessiva della pratica clinica. La loro adozione deve essere progressiva, misurata e subordinata a sicurezza, accuratezza, governance dei dati e reale integrazione nel workflow.

Bibliografia

1. Heaton HA, Castaneda-Guarderas A, Trotter ER, Erwin PJ, Bellolio MF. Effect of scribes on patient throughput, revenue, and patient and provider satisfaction: a systematic review and meta-analysis. *Am J Emerg Med*. 2016;34(10):2018-2028. doi:10.1016/j.ajem.2016.07.056. Link
2. Gottlieb M, Palter J, Westrick J, Peksa GD. Effect of Medical Scribes on Throughput, Revenue, and Patient and Provider Satisfaction: A Systematic Review and Meta-analysis. *Ann Emerg Med*. 2021;77(2):180-189. doi:10.1016/j.annemergmed.2020.07.031. Link
3. Ullman K, McKenzie L, Bart B, et al. The Effect of Medical Scribes in Cardiology, Orthopedic, and Emergency Departments: A Systematic Review. Washington (DC): Department of Veterans Affairs; 2020. Link
4. Sasdeville M, Yousefi F, Ouellet S, et al. The Impact of AI Scribes on Streamlining Clinical Documentation: A Systematic Review. *Healthcare*. 2025;13(12):1447. doi:10.3390/healthcare13121447. Link
5. Zhao J, Liu H, Chen Y, Song F. Application of artificial intelligence tools and clinical documentation burden: a systematic review and meta-analysis. *BMC Med Inform Decis Mak*. 2026;26:29. doi:10.1186/s12911-025-03324-w. Link
6. Ng JJW, Wang E, Zhou X, et al. Evaluating the performance of artificial intelligence-based speech recognition for clinical documentation: a systematic review. *BMC Med Inform Decis Mak*. 2025;25:236. doi:10.1186/s12911-025-03061-0. Link
7. Kim E, Liu VX, Singh K. AI Scribes Are Not Productivity Tools (Yet). *NEJM AI*. 2025;2(12). doi:10.1056/Ale2501051. Link
8. Lukac PJ, Turner W, Vangala S, Chin AT, Khalili J, Shih YT, Sarkisian C, Cheng EM, Mafi JN. Ambient AI Scribes in Clinical Practice: A Randomized Trial. *NEJM AI*. 2025;2(12). doi:10.1056/Aloa2501000. PMID:41497288. Link
9. Afshar M, Baumann BM, Morris A, et al. A Pragmatic Randomized Controlled Trial of Ambient Artificial Intelligence to Improve Health Practitioner Well-Being. *NEJM AI*. 2025;2(12). doi:10.1056/Aloa2500945. PMID:41625485. Link
10. Rotenstein LS, Holmgren AJ, Thombly R, et al. Changes in Clinician Time Expenditure and Visit Quantity With Adoption of Artificial Intelligence-Powered Scribes: A Multisite Study. *JAMA*. 2026;335(16):1408-1417. doi:10.1001/jama.2026.2253. Link
11. Garcia Sanchez C, et al. Use of ambient AI scribe in physicians' clinical documentation: a protocol for a systematic review on effectiveness, efficiency and satisfaction. *BMJ Open*. 2026;16:e115562. Link
12. Garante per la protezione dei dati personali. Compendio sul trattamento dei dati personali attraverso piattaforme volte a mettere in contatto i pazienti con i professionisti sanitari accessibili via web e app. Marzo 2024. Link
13. Garante per la protezione dei dati personali. Dossier sanitario: provvedimento del 6 marzo 2025. Link
14. Portale Fascicolo Sanitario Elettronico. Profilo Sanitario Sintetico / Patient Summary e specifiche di interoperabilità FSE. Aggiornamenti 2025-2026. Link
15. European Commission. European Health Data Space (EHDS): regulation and materials on EHR systems, interoperability and logging. 2025-2026. Link

DAL SINTOMO AL PROTOTIPO

UN MEDICO DI FAMIGLIA TRA UK E ITALIA, L'INTELLIGENZA DEGLI ECOSISTEMI E UN VIDEOGIOCO PER RESPIRARE MEGLIO

Dr Andrea Virga · GP, North East London – NHS; membro comitato di redazione di Rivista la Qualità e le qualità in MG (www.rivistaqq.org)

«Non bisogna dimenticare che l'uomo è al centro delle cure e che questi strumenti devono essere ad uso dell'uomo.» *Francesco Del Zotti*

I. Il sintomo

È lunedì mattina, e l'ambulatorio è pieno. Il paziente successivo è un bambino di sette anni, accompagnato dalla madre: respira a fatica, con un sibilo, wheezy, e ha quell'aria di chi non sa bene cosa gli stia succedendo nel petto. Un sospetto di asma, nato in pronto soccorso qualche mese prima, quando un episodio come questo aveva spaventato tutta la famiglia: «asma», le avevano detto, e due inalatori, prescritti lì. Ma in quel momento di paura non c'era stato il tempo di capire bene cosa stesse succedendo, né quando, come e perché usare quei due inalatori. Così usa solo quello blu, quando il bambino sta male; il marrone, quello da dare mattina e sera, le sfugge. Ha altri due figli, le giornate corrono, e «tanto, quando sta bene, che senso ha?». Il medico lo visita, ascolta il respiro: è un'esacerbazione moderata, e decide la terapia. Questa volta, però, torna a casa con un breve ciclo di prednisolone, con istruzioni chiare su come e quando usare gli inalatori, e con un link per scaricare un videogioco.

Quello che è accaduto non è un caso isolato: è la fotografia di un bisogno che il sistema, allo stato, non copre. Le linee guida britanniche aggiornate nel novembre 2024 (BTS/NICE/SIGN, NG245) hanno spostato adolescenti e adulti verso un unico inalatore di combinazione, ma per i bambini fra i cinque e gli undici anni il trattamento di prima linea resta lo schema a due inalatori: corticosteroide inalatorio di mantenimento due volte al giorno, più broncodilatatore al bisogno. Il MART pediatrico (la Maintenance And Reliever Therapy, ossia un unico inalatore di combinazione usato sia come mantenimento sia al bisogno) esiste come opzione di intensificazione, ma è rimasto a lungo off-label per limiti di licenza nei più piccoli. Tradotto: la cura di questi bambini poggia ancora interamente sull'aderenza quotidiana al preventer (il farmaco di fondo, un corticosteroide inalatorio da assumere ogni giorno per controllare l'infiammazione), il marrone, quello che «non senti che fa niente», proprio l'anello che le famiglie faticano di più a tenere. E per quel preciso percorso, in quella precisa fascia d'età, non esiste uno strumento digitale costruito apposta. È l'elemento mancante. Da un lato la terapia, che va presa ogni giorno anche quando il bambino sta bene; dall'altro la famiglia, che a tenere quella costanza fatica. Due sponde che nessuno strumento mette in collegamento, ed è lì che ho provato a gettare un ponte.

II. L'ecosistema

Mentre lavoravo a tutto questo mi è capitato fra le mani **A spasso con il cane Luna** di Giorgio Vallortigara, e mi ha colpito un passaggio sui pesci pulitori, quei piccoli pesci che ripuliscono gli altri dai parassiti. Scelgono i loro clienti con una logica quasi economica e tendono a dare la precedenza a quello nuovo, perché sanno (scommettono!) che l'abituale resterà lì comunque. Da quella lettura mi è rimasto impresso il concetto di intelligenza, o capacità adattiva, di un ecosistema, che ho poi scoperto essere già esplorato a fondo (in fondo, siamo nani sulle spalle dei giganti, per dirla con Bernardo di Chartres). L'idea è semplice: un sistema nel suo insieme premia certi comportamenti e ne scoraggia altri, e da quella somma di interazioni emerge, o non emerge, la capacità di adattarsi al cambiamento.

La mia esperienza con l'ecosistema britannico mi ha fatto pensare a quanto sia diverso da quello italiano. Quello britannico premia il pesce pulitore che punta sul cliente nuovo: il rischio, il prototipo, l'idea ancora grezza. Quello italiano sembra premiare chi punta sul cliente noto, sulle relazioni consolidate, sulle posizioni sedimentate. Non che un sistema prudente sia in sé peggiore: il problema è il momento storico, perché in un mondo a cambiamento rapido chi resta fermo non conserva, rischia di farsi travolgere. È da qui che nasce la domanda concreta: perché un gioco come Puff Puff è nato nel Regno Unito, e non in Italia?

III. Il prototipo

La risposta sta tutta in quello scarto fra i due ecosistemi. Ma prima dell'ecosistema viene il prototipo, ed è qui che è successo qualcosa che dieci anni fa non sarebbe stato pensabile. Gli strumenti di intelligenza artificiale, i modelli linguistici, la generazione di immagini e di asset, il codice scritto in dialogo con la macchina, hanno abbassato la barriera d'ingresso al punto che un singolo medico generalista, curioso e non esperto, può oggi non solo immaginare una soluzione, ma costruirne un prototipo. "Puff Puff" (da me sviluppato) è nato proprio così, in un clinical game jam britannico, un evento di prototipazione rapida di videogiochi a finalità clinica: partito dal mio studio di medicina generale e culminato nella sede di Unity a Brighton. È un endless runner educativo (un videogioco in cui il personaggio corre senza sosta schivando ostacoli): la barra dell'energia è la funzione respiratoria, gli ostacoli sono i trigger reali, l'uso regolare dell'inalatore riduce l'infiammazione, e il principio è severo, la meccanica deve premiare l'aderenza al preventer (il corticosteroide inalatorio) e mai l'iper-affidamento al reliever (il farmaco sintomatico al bisogno, il broncodilatatore a rapida azione). Puff-puff è gratuito, pensato come Community Interest Company, e nasce per scelta come strumento educativo e non come dispositivo medico, perché possa raggiungere i bambini in fretta. Lo dico con l'umiltà del dilettante: non sono uno sviluppatore. Sono un medico di famiglia che ha avuto una curiosità e, per la prima volta, gli strumenti per inseguirla.

Ma un prototipo, da solo, muore in un cassetto. Serve un ecosistema che lo raccolga, e qui il confronto diventa concreto. Nel Regno Unito esiste un'infrastruttura pubblica, il NIHR con il suo programma Invention for Innovation, che finanzia lo sviluppo di dispositivi e tecnologie sanitarie digitali a partire dal proof-of-concept(*) e dal prototipo, in progetti che mettono insieme università, sistema sanitario e industria; inoltre offre consulenza di

business oltre al denaro, e ha l'obiettivo dichiarato di ridurre il rischio d'innovazione per renderla attraente agli investitori privati successivi, coprendo il 100% dei costi della ricerca senza chiedere cofinanziamento. In altre parole, mette denaro pubblico nella fase più precoce e rischiosa, quella in cui il venture capital non entrerebbe mai, e accompagna l'innovatore verso il mercato. Il percorso è chiaro e impersonale: un medico sa dove andare, e non gli serve conoscere il ministro.

In Italia il denaro per la ricerca c'è. La Ricerca finalizzata del Ministero è un bando competitivo annuale, valutato per peer review, aperto a tutti gli operatori del Servizio sanitario, e ha persino una linea dedicata ai giovani ricercatori under quaranta. Ma è modellata come finanziamento alla ricerca, alla produzione di conoscenza, non come sviluppo di prodotto; manca il ponte verso l'investimento privato, manca la consulenza, manca soprattutto un percorso unico e visibile. Le risorse sono frammentate fra Ministero, Regioni, IRCCS, università, PNRR, e un medico con un'idea semplicemente non sa dove portarla. Non è vero che l'ambiente non investa: è vero che non è costruito per prendere per mano il prototipo e accompagnarlo dal banco dell'ambulatorio al mercato.

IV. Un sentiero possibile

Quello che servirebbe, allora, non è inventare tutto da capo. È usare ciò che già abbiamo, immaginare che un ente come il CNR, o una linea nuova del Ministero, possa diventare una sorta di NIHR italiano: una porta unica che supporta il primo prototipo, fornisce consulenza regolatoria e di business, e connette esplicitamente l'innovatore agli investitori. Una riforma piccola nella forma, radicale nella logica: non nuove strutture, ma un modo nuovo di far circolare le risorse, che dica al pesce pulitore curioso «qui il cliente nuovo è premiato».

E qui torno alla medicina generale, ma da un'altra porta, lasciando da parte la querelle contrattuale. In un sistema strutturato come il nostro, l'accesso all'innovazione passa quasi solo per un canale: quello universitario, con le sue scuole di specializzazione, i suoi istituti, i suoi bandi di ricerca. È lì che si impara a fare ricerca e si incontrano i finanziamenti, ed è lì che un'idea trova il suo metodo. E la medicina generale, in Italia, è l'unica grande disciplina che resta fuori da quel canale: non è una scuola di specializzazione universitaria, ma un corso di formazione triennale gestito dalle Regioni.

Non è una questione di prestigio. È che così escludiamo dalla ricerca proprio la disciplina con il maggior potenziale di innovazione applicata: quella che comprende la persona nel suo insieme, che conosce il territorio e le sue specificità, che vede i problemi prima e altrove rispetto all'ospedale. È la prospettiva da cui nascono le idee di cui le cure primarie hanno davvero bisogno, e la teniamo lontana dall'unico luogo in cui, da noi, le idee diventano ricerca. Non a caso nel Regno Unito l'academic general practice abita le università, e il NIHR ha persino una School for Primary Care Research: la ricerca delle cure primarie ha un tetto sotto cui crescere, con un metodo e un finanziamento propri.

Agire sull'intero ecosistema dell'innovazione italiana è difficile e richiede tempo. Ma una cosa si può, e si deve, fare prima e a costo contenuto: portare la medicina generale dentro l'università, come specializzazione e come materia, perché possa finalmente fare ricerca e innovazione invece di subirla. È il modo più rapido di alzare la

capacità adattiva del sistema proprio dal lato oggi più trascurato, e che più conta per le persone.

Resta il punto da cui Del Zotti partiva, e che ho messo in cima a queste righe: l'uomo al centro, la tecnica al suo servizio. **Puff Puff** non è nato perché io lo implementi come un genio isolato.

Esiste perché un ecosistema, quello britannico, premia chi prova: i game jam clinici, gli acceleratori, i bandi, gli investitori disposti a scommettere su un prototipo grezzo. Se lo stesso ecosistema esistesse in Italia, un gioco come questo sarebbe nato a Roma o a Palermo, e non avrebbe dovuto attraversare la Manica. È questa la posta in gioco quando parliamo di adattabilità: non un'astrazione, ma la differenza fra un'idea che diventa cura e un'idea che resta nel cassetto.

Ec'è un ultimo aspetto, che mi riguarda da vicino. L'innovazione ha bisogno di propensione al rischio, e il rischio è in parte una prerogativa della gioventù, per una semplice ragione di tempo: chi è giovane può permettersi di fallire e ricominciare. Un ecosistema adattabile investe sui giovani non per generosità, ma perché sono loro a potersi assumere il costo evolutivo dell'innovazione. L'Italia tende a fare il contrario, e così seleziona per fuga: i talenti, come i pesci pulitori, migrano verso l'ecosistema che premia chi tenta. La chiamiamo fuga di cervelli e la raccontiamo come una questione morale; è invece, banalmente, selezione darwiniana di habitat. Nella speranza che questo cambiamento avvenga il prima possibile, concludo con un invito, lo stesso con cui ho chiuso il mio intervento al congresso europeo EGPRN a Verona, di un mese fa: torniamo a casa e progettate i vostri strumenti di cura.

**da wikipedia: proof-of-concept è una realizzazione incompleta o abbozzata (sinossi) di un determinato progetto o metodo per dimostrare la fattibilità o confermare la validità di alcuni principi o concetti fondamentali.*

Bibliografia

1. BTS/NICE/SIGN. Asthma: diagnosis, monitoring and chronic asthma management (NG245), novembre 2024.
2. G. Vallortigara. A spasso con il cane Luna. Adelphi, 2025.
3. A. Virga, M. Nardelli. NHS & SSN: guardare oltremania per vedere il futuro dell'Italia? (articolo precedente di questa serie).
4. NIHR. Invention for Innovation (i4i) programme.
5. Ministero della Salute. Bando Ricerca finalizzata.
6. Playpuffpuff.com

NETAUDIT NT-PRO-BNP, BNP E SCOMPENSO CARDIACO

Ideatrici e Ideatori : MMG : Melania Cagnazzo e Domenico Pasculli (Molfetta Bari); Lista Netaudit; Lista Coordinatori di Netaudit (www.netaudit.org)

Direzione: G.S. Blengio, D. Pasculli, F. Del Zotti

MMG partecipanti; Antista Giuseppe, Biaggioni Michela, Cagnazzo Melania, Dalla Benetta Elisa, De Luigi Giovanni, Lippa Luciano, Lupi, Lorenzo, Luppi Davide, Merlino Giovanni, Salvadori Angela, Tangorra Francesco, Tedeschi Luca, Zenari Sonia

Motivazioni (Background)

La prevalenza di scompenso cardiaco è in aumento: in Italia, al II posto nel mondo per quota di anziani, raggiunge il 1-2% ed è destinata a crescere. È necessario valutare la prevalenza e qualità della diagnosi, che, secondo le ultime Linee Guida ESC, si basa non solo sulla presenza di segni, sintomi e fattori di rischio, ma anche sul risultato degli esami NT-proBNP o BNP. Questi ultimi, se con risultato sotto soglia definita (vedi **Figura 1**), abbassano notevolmente la probabilità di diagnosi. Nelle Linee-Guida ESC il valore soglia dell'NT-proBNP o del BNP è quindi un punto essenziale.

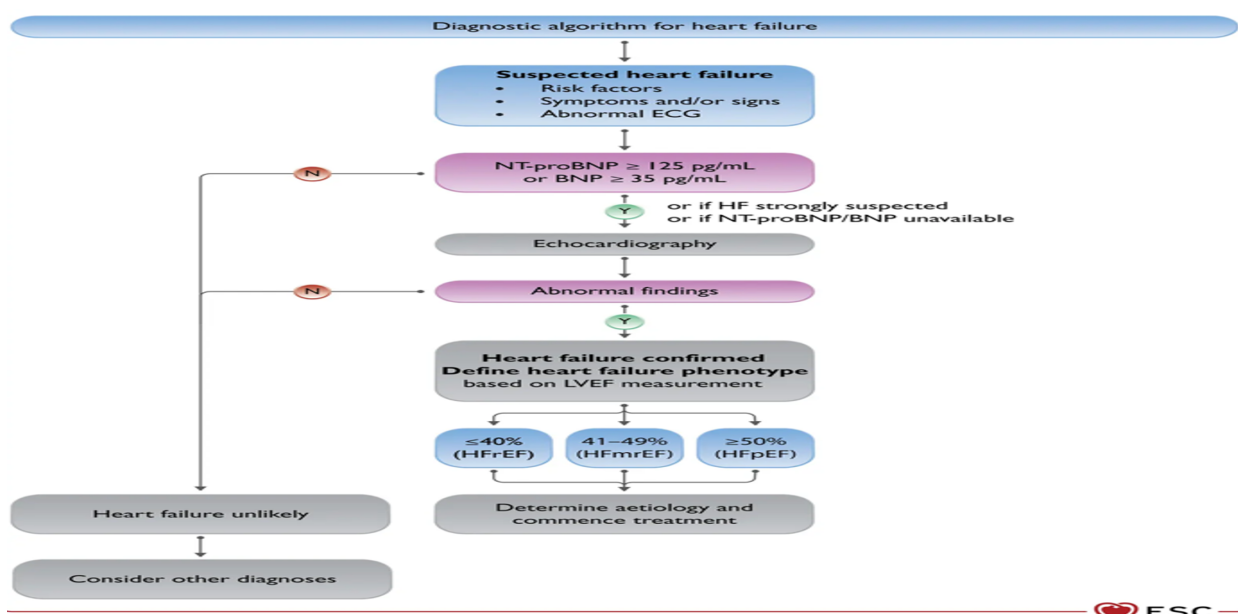
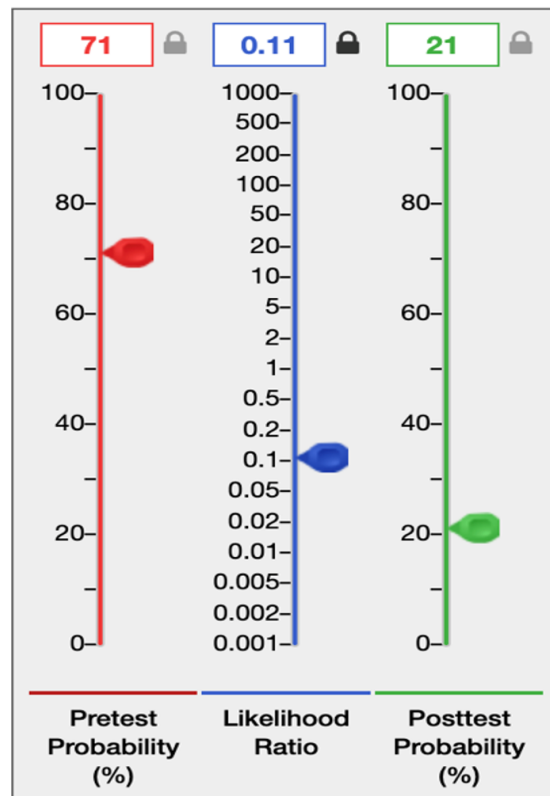


Figura 1

L'importanza di questi due esami deriva da una caratteristica epidemiologico-clinica. Nel precedente numero della nostra rivista, gennaio 2026, abbiamo pubblicato un articolo dal titolo "Likelihood Ratio (LR) per la Diagnosi e per Netaudit" (<https://rivistaqq.org/gennaio-2026/>). Le LR di un esame di laboratorio possono essere positive (LR+) o negative (LR-). Le LR- dei suddetti esami, nello scompenso cardiaco, hanno un valore molto basso (0,1). Se abbiamo raccolto segni e sintomi e abbiamo codificato con ICD9 lo scompenso (senza aver richiesto uno dei due esami) e abbiamo, con tali sintomi e segni clinici, circa il 70% di possibilità di scompenso, in realtà con risultati una LR- di circa 0,1 si determina una riduzione importante della probabilità di diagnosi di "scompenso" verso il 20%.

Come si ottiene questa riduzione importante di probabilità? Innanzitutto, la suddetta LG ESC del 2025 fornisce i valori soglia compatibili con il codice ICD9 di scompenso cardiaco: sono per l'NT-proBNP maggiore o uguale a 125 pg/ml; per il BNP un valore maggiore o uguale a 35 pg/ml. Quindi se i valori riscontrati di questi esami, in pazienti codificati con ICD9 per scompenso, sono sotto queste soglie si deve adoperare la suddetta LR- (intorno allo 0,1) e la diagnosi da noi fatta deve esser ri-esaminata. In effetti, adoperando il semplice e fondamentale nomogramma Fagan (vedi **Figura 2** e link: https://jamaevidence.mhmedical.com/data/calculators/LR_nomogram.html) e usando un righello con un test con valore sotto-soglia per il sospetto (e quindi negativo, ove usare le LR-), andremo, come già detto, dal 70% a circa il 20%. Ciò può provocare una crisi del codice ICD9 di "scompenso" da noi assegnato e la necessità di ulteriori raccolte di sintomi, segni ed esami (ad es. Rx torace ed ecocardiogramma, ad es. basati su Boston Score dello scompenso).



J. Altareios. R. Havward

Figura 2: Nomogramma Fagan

Obiettivi del net-audit

Criterio Principale Frequenza di richiesta di dosaggio NT-proBNP o del BNP e valutazione del valore nei pazienti codificati con ICD9 per scompenso cardiaco. In buona sostanza bisogna valutare: la proporzione della discrepanza tra Codice ICD9 già inserito in cartella e, nei pazienti selezionati; la proporzione di pazienti con riduzione del risultato sotto la soglia indicata dalla LG ESC dei due esami (il che fa scattare l'effetto controdiagnostico della LR- bassa). Nella sostanza abbiamo bisogno di estrarre i pazienti con Diagnosi ICD9 e successivamente verificare se essi hanno ricevuto la richiesta di uno o entrambi i suddetti esami e se per essi esistono uno o più valori numerici di NT-proBNP o BNP. In maniera più estensiva, vi riferiamo che gli obiettivi principali dello studio sono stati triplici: a) Valutare, per ognuno dei 14 MMG, nel totale dei pazienti assistiti, in quanti pazienti sono stati registrati i codici ICD9 per lo scompenso cardiaco. b) Nei pazienti selezionati (sulle modalità di selezione si veda dopo): a quanti pazienti con ICD per scompenso cardiaco è stato almeno una volta richiesto NT-proBNP o BNP o entrambi. c) Nei pazienti selezionati, valutare in quanti esisteva almeno un risultato in valore numerico di uno o entrambi i due esami suddetti; quindi valutare il valore rispetto alle soglie definite dalla Società Europea dei Cardiologi (ESC).

Criteri aggiuntivi

Richiesta dell'ecocardiogramma nei pazienti selezionati.

Valutare se è stata registrata la presenza di storia di ricovero o pronto soccorso a causa della patologia scompenso cardiaco.

Criteri di esclusione Sono stati esclusi dalla valutazione pazienti con scompenso cardiaco iscritti nella lista dei nuovi pazienti del MMG da meno di 6 mesi; sono stati altresì esclusi scompensati terminali e scompensati con demenza grave.

Le variabili- Campi richiesti per Data Entry

Ecco i campi che sono stati richiesti per i MMG aderenti:

- Età in anni del paziente
- Sesso del paziente
- Valutazione in cartella in merito alla presenza dell'esame NT-proBNP o dell'esame BNP

Se il risultato dei suddetti esami è stato registrato, è stato richiesto di inserire il primo valore dei suddetti esami in cartella (se esistevano sia BNP sia NT-proBNP, è stato chiesto di inserire solo il valore dell'NT-proBNP).

È stato chiesto di valutare se nella cartella esistesse una ricetta di ecocardiogramma.

Se esisteva una richiesta di ecocardiogramma, è stato domandato se esistesse un valore (l'ultimo) della Frazione di Eiezione e, in caso di FE uguale o maggiore di 50%, se esistesse una descrizione significativa (in particolare: disfunzione diastolica).

L'ultima domanda ha riguardato la presenza o assenza di ricovero dello scompensato al pronto soccorso o in reparto.

Periodo di analisi È stato analizzato il database delle 176 cartelle per tutto il periodo di presa in carico dei pazienti da parte del MMG.

Metodo e Modalità della randomizzazione

Prima di tutto, ci preme sottolineare il modo di partecipazione dei 14 MMG aderenti. I MMG che hanno sino a 15 casi di scompenso (con codice ICD9 rispettivo) hanno analizzato tutti i casi con scompenso, paziente per paziente; e quindi per loro non è servita la randomizzazione. Se invece il MMG ha avuto più di 15 casi ha seguito la solita procedura di randomizzazione di Netaudit, che da tempo è utile per ridurre e uniformare il carico di lavoro dei volontari (video esplicativo disponibile). Questi MMG hanno alla fine analizzato 15 casi randomizzati.

Dopo queste regole sono restati, rispetto ai 325 scompensati iniziali, 176 scompensati da valutare. Seguendo la statistica del sample size e adoperando una proporzione cautelativa del 50% (conservativa se non si conosce la proporzione cercata), l'usuale margine di errore del 5% e infine un livello di confidenza del 90% (indicata quando si eseguono audit e non trial), abbiamo trovato come numero minimo da analizzare: 149, il che è molto al di sotto del nostro numero di pazienti analizzati: 176. E se anche volessimo usare un intervallo di confidenza del 95%, avremmo 177 (molto vicino al nostro 176).

L'analisi dell'obiettivo principale e delle variabili correlate è stata quindi eseguita solo sui 176 scompensati reclutati.

Totale dei pazienti reclutati dopo la randomizzazione: 176/325 scompensati.

L'analisi delle variabili è stata quindi eseguita solo sui 176 scompensati.

Risultati

I 14 MMG partecipanti, delle 3 macroregioni italiane (Centro, Nord, Sud), hanno in carico 325 pazienti con ICD9 di scompenso su totale dei loro 19.399 pazienti (1,7%).

Valutazione preliminare su rimborsabilità dei due esami. Al momento dell'adesione ai MMG è stato richiesto se ci fosse nella regione di appartenenza la prescrivibilità dei due esami (NT-proBNP oppure BNP). Si è ottenuta questa risposta: 11 MMG riferiscono concedibilità regionale SSN di NT-proBNP o BNP; 3 non concedibilità.

MMG invianti dati e loro sesso: Hanno inviato i dati relativi a questo Netaudit 14 MMG delle 3 macroregioni italiane (Centro, Nord, Sud), che hanno in totale 19.399 pazienti. Di questi 14 MMG: 9 sono stati i MMG maschi e 5 le MMG femmine.

Analisi dei risultati sui 14 MMG che hanno inviato i Dati

Sono stati estratti (con logica booleana) dai database dei 14 MMG 325 pazienti con ICD9 "scompenso cardiaco" rispetto al totale pazienti 19.399, il che equivale a una prevalenza dell'1,7% (media 1,45%, mediana 1,365%, DS $\pm 1,08$). Il dato rappresenta una prevalenza in linea con i dati nazionali ed europei.

Sesso dei 176 pazienti da analizzare con scompenso: Le femmine con scompenso sono 75, pari al 42%; i maschi 101 (58%).

Età dei Pazienti

- Categoria
- Età media
- Età media complessiva
- 79,7 anni (Popolazione molto anziana)
- Età media maschi
- 78,0 anni
- Età media femmine
- 81,9 anni (Femmine più anziane di 4 anni)

Abbiamo usato il test di Student per verificare se questa differenza di età tra i due sessi è significativa. Lo è, in quanto $p < 0,01$, con intervallo della differenza che varia da +1 anno a +6,6 anni rispetto all'età dei maschi scompensati. Durata media dello scompenso cardiaco nella popolazione in esame: 3,8 anni.

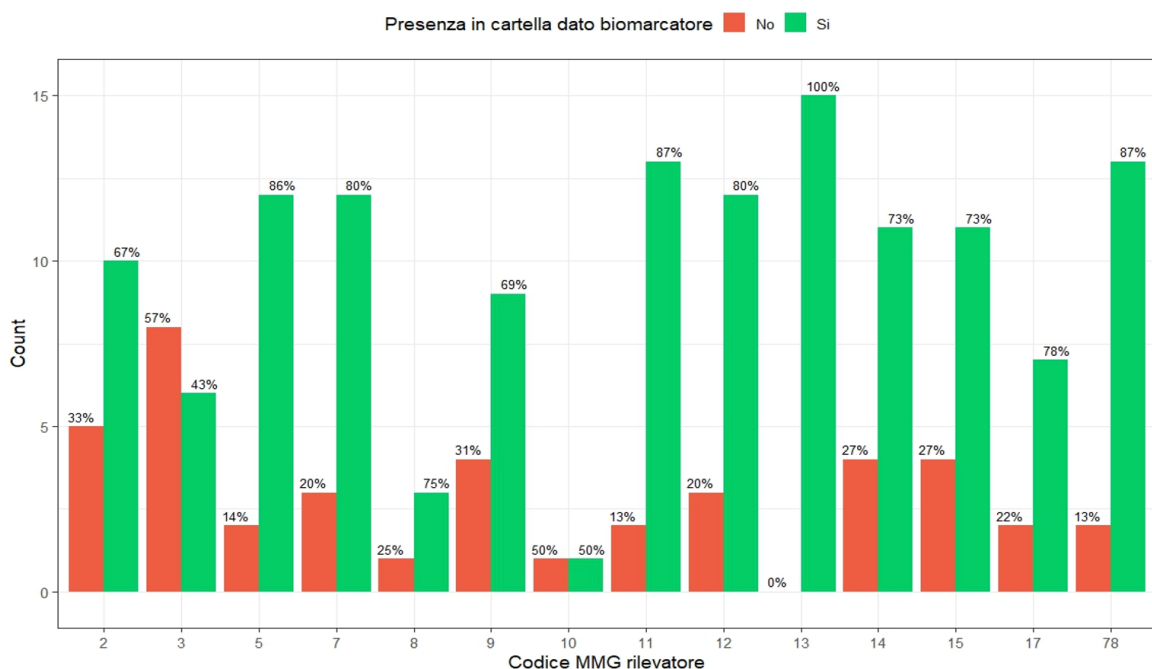


Figura 3: Variabilità tra i 14 MMG, con Diagramma a Barre, della presenza di richiesta del biomarcatore di scompenso

Criterio Principale: Scompensati e Richiesta di NT-proBNP O BNP

Dei 176 pazienti totali con scompenso, 135 (77%) hanno almeno uno dei due esami in cartella; 41 (23%) non hanno in cartella nessuno dei due esami.

Presenza del valore numerico di NT-proBNP

Su 135 pazienti con NT-proBNP, 88 pazienti (65%) hanno un valore di NT-proBNP registrato. Di questi 88 pazienti, 78 (88,6%) hanno un valore superiore a 125 pg/ml. Solo 10 pazienti (11,4%) hanno un valore inferiore a 125 pg/ml. Il valore soglia di 125 pg/ml è comunemente utilizzato come cut-off: se il valore è inferiore a 125, lo scompenso è meno probabile, a causa dell'importante Likelihood Ratio Negativa (0,1).

In buona sostanza, nei pazienti ove è presente il valore NT-proBNP, nella maggioranza (88,6%) si constata un risultato del test che ben collima con il Codice ICD9 immesso dal MMG. Invece, nella piccola minoranza di pazienti il cui valore dell'NT-proBNP è basso e sotto-soglia, il risultato non collima con la Diagnosi ICD9 (circa l'11%). Per essi, i partecipanti si sono impegnati nella rivalutazione delle altre variabili connesse alla diagnosi (segni, sintomi, ecocardiogramma, ECG, ecc.).

Comunque, ripetiamo, nei pazienti in cui è presente il valore, l'alta percentuale di pazienti con valori di NT-proBNP > 125 pg/ml (circa il 90%) è coerente con una popolazione di pazienti con il codice ICD9 di scompenso assegnato dal MMG.

Scompensati e valore di BNP

Intanto, vi è da notare che i pazienti con il solo BNP (41) hanno tutti in cartella un valore numerico di BNP (invece di NT-proBNP); in questi pazienti 40/41 hanno un valore uguale o superiore a 35 pg/ml e quindi coerente con codice di scompenso (in particolare per lo scompenso cronico), mentre solo 1 ha un valore inferiore e perciò in contrasto con il Codice ICD9 di scompenso. Anche se si vuole seguire un valore più conservativo, 100 pg/ml (fornito da alcune fonti, che si riferiscono anche allo scompenso cardiaco acuto), solo 4 pazienti su 40 hanno valore inferiore a 100.

Variabilità tra MMG nella richiesta di uno dei due esami

Fermo restando il dato complessivo alto di richieste dei due esami nei pazienti nel gruppo dei 176 pazienti analizzati, resta una discreta variabilità tra i 14 MMG, come testimoniato dalle **Figure 3 e 4**. In particolare, la proporzione di richiesta di almeno uno dei due esami tra i 14 MMG varia in questo modo:

Minimo: 43%

Media: 74,8%

Mediana: 76,4%

Massimo: 100%

I quartile (25° percentile): 69,2%

Il quartile (mediana): 76,4%

III quartile (75° percentile): 85,7%

Pur constatando una certa distanza tra il minimo di 43% e massimo di 100%, constatiamo che il coefficiente di variazione del 19,7% indica comunque una sufficiente compattezza positiva del gruppo verso la richiesta dei due esami.

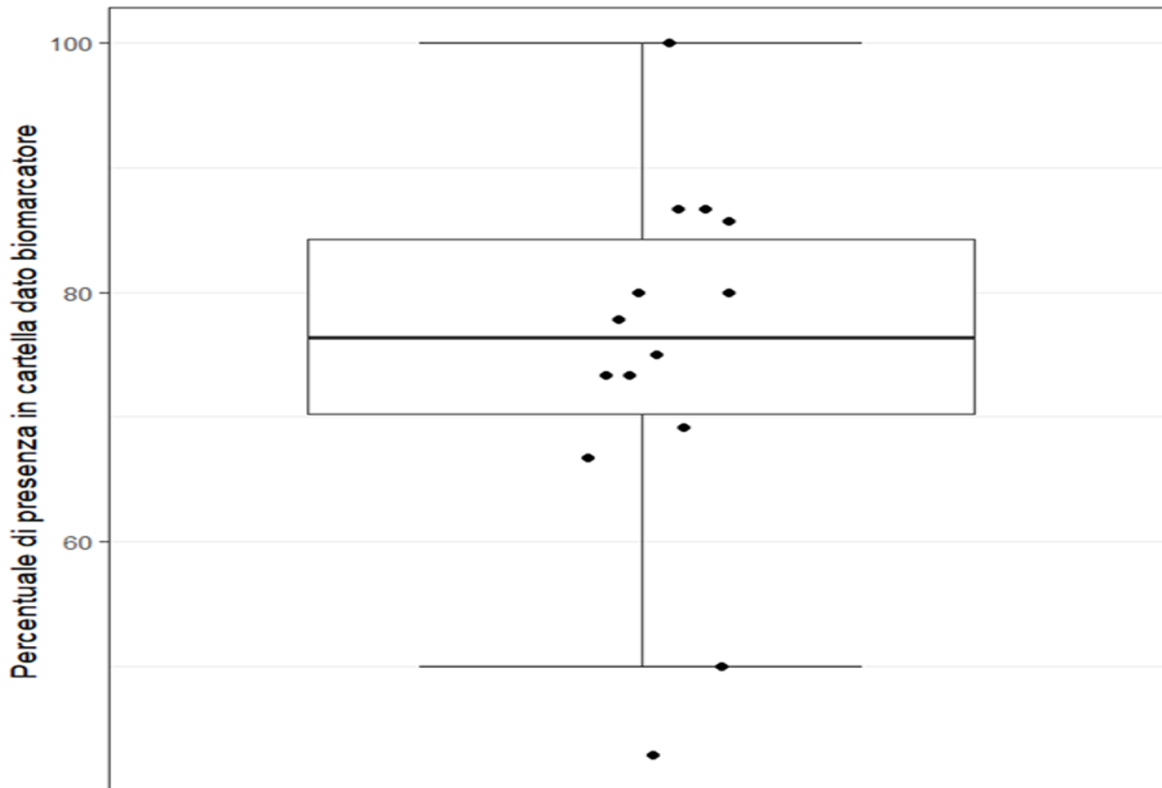


Figura 4: BOXPLOT, variabilità tra i 14 MMG rispetto alla proporzione di presenza in cartella di richiesta di almeno uno dei due test e totale dei pazienti selezionati

Scompenso e Richiesta dell'ecocardiogramma

In 158/176 (il 90%) pazienti si constata la registrazione in cartella sia di un ecocardiogramma sia di almeno una frazione di eiezione (FE). Il restante 10% dei pazienti scompensati o ha un ecocardiogramma registrato (13/176) ma non ha la FE registrata, o, in una minuscola parte dei pazienti con ICD9 di scompenso, non si ha alcuna registrazione di ecocardiogramma (5/176).

Più in particolare, se ci limitiamo ai pazienti con ecocardiogramma e FE registrata:

Sì, con FE maggiore o uguale a 50%, senza altre chiare descrizioni: 39 pazienti.

Sì, con frazione di eiezione (FE) inferiore a 40%: 43.

Sì, con FE tra 41 e 49%: 33.

Sì, con FE maggiore o uguale a 50% e disfunzione diastolica del ventricolo sinistro: 43.

Considerando la severità dello scompenso, i 176 pazienti sono così stratificati:

HFrEF (FE < 40%): 43 pazienti (25,1%) - Scompenso con FE ridotta

HFmrEF (FE 41-49%): 33 pazienti (19,3%) - Scompenso con FE moderatamente ridotta

HFpEF (FE ≥50%): 82 pazienti (47,9%) - Scompenso con FE preservata

Di questi 82 pazienti, 43 (25,2%) con HFpEF (FE ≥50%): hanno disfunzione diastolica documentata, 39 (22,8%) senza altre descrizioni.

Il valore della FE non è disponibile in soli 13 pazienti (7,6%).

In sintesi: Quasi la metà dei pazienti (47,9%) presenta scompenso cardiaco con frazione di eiezione preservata (HFpEF), tipico di una popolazione anziana. Un quarto dei pazienti (25,1%) ha FE ridotta < 40%. Nel 7,6% dei casi l'ecocardiografia è stata eseguita ma il valore della FE non è stato riportato in cartella.

Correlazione tra valori di BNP/NT-proBNP e Severità dello scompenso

Sulla base dei dati rilevati è possibile rilevare una chiara correlazione tra valori di BNP/NT-proBNP e severità dello scompenso:

Classe

Media NT-proBNP

Media BNP

HFrEF (FE < 40%) - Scompenso Severo

~2.614 pg/ml (la più alta)

~2.427 pg/ml

HFmrEF (FE 41-49%) - Scompenso Moderato

~1.333 pg/ml (valori intermedi)

HFpEF (FE ≥50%) - FE Preservata

~1.679-1.825 pg/ml

I dati confermano la correlazione inversa tra FE e livelli di BNP/NT-proBNP: più bassa è la FE, più alti sono i valori dei peptidi natriuretici, in accordo con la letteratura scientifica sullo scompenso cardiaco.

Richiami di Ospedalizzazione e Severità

Abbiamo cercato anche di correlare la severità dello scompenso al numero e frequenza dei ricoveri:

HFrEF (FE < 40%): Tasso di ricovero più alto (86%) - conferma che lo scompenso severo porta a più ospedalizzazioni.

HFmrEF (FE 41-49%): Tasso intermedio (73%).

HFpEF con disfunzione diastolica (FE ≥50%): Tasso 65% - comunque elevato.

HFpEF senza altre descrizioni: Tasso più basso (44%) - forma meno severa.

Il 75,9% dei ricoveri è avvenuto negli ultimi 5 anni; solo il 6% risale a oltre 10 anni fa.

Questo suggerisce che:

La maggior parte dei pazienti ha una diagnosi/storia recente e i ricoveri recenti sono probabilmente legati alla progressione della malattia.

Esiste dunque correlazione tra severità dello scompenso (misurata tramite FE) e tasso di ospedalizzazione: più bassa è la FE, maggiore è la probabilità di ricovero. Inoltre, i ricoveri sono concentrati negli ultimi 5 anni, indicando una popolazione con storia clinica recente o progressione della patologia.

Conclusioni

In merito all'obiettivo principale (la presenza di richiesta di uno o due esami nei pazienti già definiti "scompensati dai 14 MMG) il gruppo ha mostrato positività: il 77% dei 176 pazienti presenta la richiesta in cartella. Restano 41 pazienti (circa il 23% del totale in esame 176) che non hanno alcun valore registrato di BNP né di NT-pro-BNP nella loro cartella clinica. E' sicuramente un dato da migliorare sia da parte dei MMG sia da parte delle autorità regionali (3 dei 14 MMG hanno dichiarato che nella loro realtà questi due esami non sono rimborsabili dal SSN). Resta da migliorare anche il recupero del risultato: mancano 91 risultati numerici su 176 casi analizzati (il 51%). Non sono chiari i motivi di tale mancanza, importante per valutare le suddette LR- e quindi per potenziare o depotenziare, in modo deciso, la probabilità ed esattezza della nostra diagnosi e relativo codice ICD9. Resta inoltre non chiaro un dato: nei pazienti con solo BNP hanno tutti il risultato in cartella e ciò diversamente dai pazienti con NTporBNP. Forse nelle ASL di questi pazienti si pratica solo il BNP e magari negli stessi territori vige il passaggio diretto del risultato dal laboratorio alla casella giusta della cartella dei MMG. E' per ora solo un'ipotesi. Certamente non aiuta l'inserimento del risultato in cartella la sempre più frequente spedizione di statici pdf-risultati, simil-foto, da parte delle ASL e dagli stessi pazienti. Esiste una discreta variabilità nella prescrizione di questi due esami tra i 14 MMG: si va da un minimo di circa il 40% al massimo del 100%. Ci auguriamo che chi ha

percentuali più basse possa migliorare in poco tempo: basta aggiungere questi esami alla prossima prescrizione. Ci confortano, comunque i dati della prevalenza dello Scompenso che, seppure in un campione piccolo di MMG, sono sovrapponibili ai dati nazionali (1,7%). E' anche confortante la prescrizione massiccia di Ecocardiogramma con frequente immissione del dato Frazione di Eiezione, altro caposaldo dell'algoritmo dell'ESC, a valle dei due esami in oggetto. Probabilmente questo pervasivo uso dell'ecocardiogramma è un dato ormai acquisito tra i MMG.

Lo spunto più interessante da questa prima valutazione è che per i due biomarcatori di scompenso vale sia la necessità di maggiore formazione continua sia l'approfondimento di un concetto metodologico: nel "costruire una diagnosi" come quella dello scompenso non vale solo il tradizionale invito a raccogliere sintomi e segni, ma anche a conoscere per ogni elemento, compreso quello del laboratorio (vedi NtproBNP e BNP), lo studio ed utilizzo di un concetto di epidemiologia clinica che sempre più si impone in medicina: il concetto della clinimetrics (del grande metodologo Alan Feinstein), e cioè l'imperativo di registrare in maniera definita o "operativa" i sintomi e segni ed i test; il che include il loro peso e quindi le loro Likelihood Ratio.

Bibliografia e Sitografia

- AHA_ACC_HFSA 2022
- Eur Heart J. 2023 Oct 1;44(37):3627-3639. <https://doi.org/10.1093/eurheartj/ehad195>
<https://onlinelibrary.wiley.com/doi/10.1002/ejhf.3024>
- Eur Heart J. 2021 Dec 21;42(48):4901. <https://doi.org/10.1093/eurheartj/ehab368>
- Del Zotti F, Blengio GS. Likelihood Ratio per la diagnosi e per Netaudit.
<https://rivistaqq.org/sito/wp-content/uploads/2026/01/RIVISTA-QQ-1-2026.pdf>

CONRGEESO EGPRN VERONA MAGGIO 2026

Giulio Rigon*, Sonia Zenari, MMG Verona

*Italian EGPRN Representative

*Qui di seguito una breve sintesi del congresso e i tre poster presentati all'evento di Verona di maggio 2026 sul tema **Continuity of Care now and in Future.***

Continuità delle cure: la Medicina Generale e il ruolo dell'EGPRN La continuità delle cure è un pilastro della medicina generale: assicura assistenza coordinata nei diversi servizi e lungo tutto il percorso di vita. In un'Europa che invecchia e convive sempre più con le malattie croniche, è anche una condizione essenziale per garantire equità nell'accesso alle cure. **Le sfide in Italia e in Europa** Entro il 2050 oltre 129 milioni di europei avranno più di 65 anni, con conseguenze rilevanti per i sistemi sanitari. L'Italia, tra i Paesi con la maggiore aspettativa di vita, sarà particolarmente esposta alla multimorbilità cronica, destinata a diventare uno dei principali nodi assistenziali. **La centralità della Medicina Generale** Il medico di medicina generale è il primo punto di contatto e il riferimento con cui il paziente costruisce una relazione di fiducia. Cure longitudinali e personalizzate, coordinamento dei team multidisciplinari e gestione delle cronicità sono i principali punti di forza dell'assistenza primaria. **Riforma e innovazione digitale In Italia** il PNRR mira a rafforzare i team di cure primarie e i percorsi integrati, spostando il baricentro dall'ospedale al territorio. Telemedicina, fascicolo sanitario elettronico ed European Health Data Space possono rendere più semplici le transizioni tra servizi e più sicura la condivisione dei dati clinici. **Ricerca e qualità delle cure Gran parte dei bisogni di salute può essere gestita efficacemente nell'assistenza primaria.** Prevenzione, diagnosi differenziale, gestione delle cronicità e della politerapia richiedono un approccio complessivo alla persona. La ricerca offre le evidenze necessarie per migliorare le linee guida e ridurre la variabilità delle cure, integrando qualità clinica e organizzativa.

Il contributo dell'EGPRN

L'esperienza del congresso EGPRN di Verona ha rappresentato un'importante occasione, dopo molti anni, di ospitare nuovamente questo evento in Italia, che ha visto la partecipazione di circa 200 medici di medicina generale provenienti da tutta Europa.

L'intero gruppo organizzativo di MMG coinvolti ha fornito un contributo determinante, affiancato dal sindacato FIMMG Verona, che ha sostenuto e promosso attivamente il congresso. All'iniziativa ha inoltre aderito l'Università di Verona, concedendo il proprio patrocinio all'evento.

Tra i momenti scientifici più rilevanti, il congresso ha ospitato due keynote lecture dedicate al tema centrale della continuità delle cure. La keynote internazionale è stata tenuta dal Prof. Steinar Hunskaar, che ha affrontato una delle sfide principali per le cure primarie europee: il difficile equilibrio tra accesso diretto e continuità assistenziale, sottolineando come quest'ultima — relazionale, informativa e gestionale — migliori la soddisfazione dei pazienti, riduca ospedalizzazioni e costi, ma sia spesso messa a rischio dalla crescente pressione verso un accesso più rapido. La keynote nazionale è stata affidata al collega italiano Dr. Ferdinando Petrazzuoli, dal titolo *"Continuity of Care: an essential characteristic of the discipline of general practice/family medicine"*: forte di oltre 35 anni di esperienza in medicina di famiglia, ha evidenziato come, nel contesto italiano, l'accesso migliorato non si traduca automaticamente in maggiore continuità, portando come esempio il dato secondo cui gli studi medici più piccoli, con uno o due professionisti, registrano il 33% in meno di ricoveri ospedalieri prevenibili rispetto ai modelli organizzativi più grandi.

In collaborazione con i colleghi di Modena e Parigi, è stato inoltre organizzato un workshop dedicato al ruolo educativo dei pazienti esperti.

Per tutti i partecipanti si è trattato di un momento di confronto e scambio di opinioni e punti di vista sulle molteplici sfaccettature della medicina generale, conclusosi con la practice visit presso alcuni studi di MMG veronesi: un'occasione che ha permesso uno scambio più diretto e ravvicinato tra i partecipanti.

EGPRN Verona 2026 gruppo organizzativo:

G.Rigon, S. Zenari, F.Del Zotti, F.Petrazzuoli, P.Benini, F.Altimari, R.Cattano, L.Rizzotto, L.Luzi Crivellini, A. Furnari, B.Mozzo

I poster che seguono sono stati presentati dal Dott. Luca Raccagni (Verona) dalla Dott.ssa Federica Ferrara (Soliera, Modena) e dalla Dott.ssa angelica Salvadori (Torino).

PREVALENCE, CLINICAL CHARACTERISTICS AND MANAGEMENT OF CHRONIC KIDNEY DISEASE IN PRIMARY CARE

Giulio Rigon, Luca Raccagni, Sonia Zenari, Ferraro Pietro Manuel
Primary Care, University of Verona School of Medicine

BACKGROUND

Chronic kidney disease (CKD) is a silent and frequently underrecognized condition in primary care. Early identification of at-risk patients is essential to enable timely intervention and reduce disease progression, complications, and healthcare costs.

RESEARCH QUESTION

To evaluate the adequacy of the diagnostic approach used by a general practitioner for early CKD detection and to describe the prevalence, clinical characteristics, and management of CKD in a real-world primary care population

METHODS

Patients managed at a Primary Care Center- Casa di Salute Verona- by a single general practitioner were enrolled in this observational study, which spanned January 2020 through December 2024. Eligibility required the presence of metabolic or cardiovascular risk factors and/or a prior diagnosis of kidney disease. Among the variables recorded were age, sex, ICD-9-coded comorbidities, serum creatinine, urinary markers of kidney damage, and use of RAAS inhibitors and SGLT2 inhibitors.

RESULTS

A total of 482 patients were included (30% of the GP's population), and 410 patients with at least one creatinine measurement were analysed and stratified according to CKD severity (GFR). In this subgroup, hypertension remained the most frequent comorbidity (90.7%), diabetes was present in 18.3%, and any CKD-related condition was recorded in 13.2%.

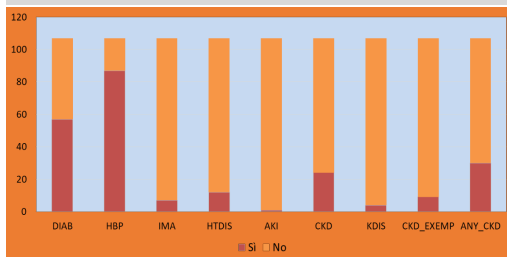
CKD-related coding increased across GFR-based severity strata, reaching 30.9% in the 45<GFR<59 group and 52.6% in the GFR<45 group.

CHARACTERISTICS	VALUE
Number of participants	482 (30% GP's patients)
GENDER DISTRIBUTION	
Female	226 (46,9%)
Male	256 (53%)
AGE-yr	72,6 (11.2)

At least «ONE CREATININE» value (GFR),
and stratification by severity of CKD
and by Sex and Diseases at risk for CKD

	GFR > 60 (N=336)	45 < GFR < 59 (N=55)	GFR < 45 (N=19)	OVERALL (N=410)
SEX				
Female	157 (46,7%)	27 (49,1%)	13 (68,4%)	197 (48%)
Male	179 (53,3%)	28 (50,9%)	6 (31,6%)	213 (52%)
DIABETES				
YES	57 (17%)	14 (25,5%)	4 (21,1%)	75 (18,3%)
HIGH BLOOD PRESSURE				
YES	306 (91,1%)	50 (90,9%)	16 (84,2%)	372 (90,7%)
ACUTE MYOCARDIAL INFARCTION				
YES	16 (4,8%)	3 (5,5%)	2 (10,5%)	21 (5,1%)
ANY CKD (AKI + KDIS + CKD exemption)				
YES	27 (8%)	17 (30,9%)	10 (52,6%)	54 (13,2%)

PERCENTAGE OF AT LEAST ONE PROTEINURIA DETECTION FOR EACH DISEASE



KIDNEY DAMAGE ASSESMENT (URINARY MARKERS)		
	OVERALL (N=107)	
	YES	NO
MICROALB_status	58 (54,2%)	49 (45,8%)
PCR_status	60 (56%)	47 (44%)
UPE_status	7 (6,5%)	100 (93,5%)
ACR_status	62 (58%)	45 (42%)
UALB_status	7 (6,5%)	100 (93,5%)
UCREAT_status	22 (20,5%)	85 (79,5%)
UPROT_status	1 (1%)	106 (99%)

TREATMENT PATTERNS: RAAS inhibitor prescription was more common in patients with type 2 diabetes than in those without diabetes (82.7% vs 63.3% overall); SGLT2 inhibitor use was concentrated in diabetic patients (29.3% vs 1.8% in non-diabetic patients).

CONCLUSIONS

- ❑ CKD remains underestimated in primary care
- ❑ The current diagnostic approach appears insufficiently standardized to support consistent early detection.
- ❑ Current treatment guidelines suggest there is margin for further optimization of nephroprotective therapies in primary care, particularly in patients at higher risk of chronic Kidney disease

POINT OF DISCUSSION

- ❖ Could multidisciplinary co-management improve early diagnosis, treatment optimization, patient quality of life, and sustainability of the healthcare system? How?
- ❖ Could a larger-scale study help outline critical issues and best practice, promoting more uniform and effective management at the national level? Perhaps by providing more adequate diagnostic tools to support general practitioners?

Burnout-associated factors in a sample of General Practitioners from the Province of Modena, Italy

Federica Ferrara^{*1}, Giada Zecchi¹, Michele Granata², Davide Luppi³

¹General Practitioner in Soliera, Province of Modena, Italy

²Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia

³General Practitioner in Bomporto, Province of Modena, Italy; tutor Vocational Training GP, ER Region; teacher in family medicine, Unimore.

Background

Burnout among general practitioners (GPs) is a common issue associated with reduced quality of care and adverse outcomes for both physicians and patients, yet evidence on its prevalence among Italian GPs remains limited.

Research questions

This study aimed to assess burnout prevalence among Italian GPs in Carpi District (province of Modena) and to evaluate associated socio-demographic and managerial risk and protective factors.

Methods

An analytical observational study was conducted among general practitioners in the same district using an online questionnaire. It included a shortened, adapted Maslach Burnout Inventory along with socio-demographic and work-related variables to assess burnout and professional well-being.

Results

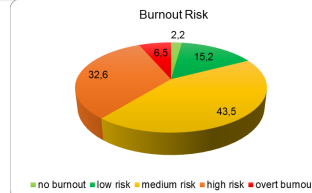
High burnout risk was found in most of the 46 GPs, with 6.5% showing overt burnout. No clear age-related trends were observed, but overt burnout occurred only in female GPs, and less experienced physicians showed higher risk. Workload factors showed no consistent patterns, despite extreme workloads being present in both overt burnout and no-risk groups.

Conclusions

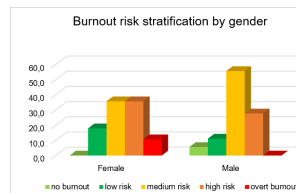
Burnout risk is widespread, with female gender and early career stage as potential contributing factors, while its association with workload remains unclear, suggesting a multifactorial and systemic nature influenced by organizational aspects. Effective prevention therefore requires organizational redesign to improve sustainability, efficiency, and physicians' well-being.

 ***Dr. Federica Ferrara**
dottfedericaferrara@gmail.com

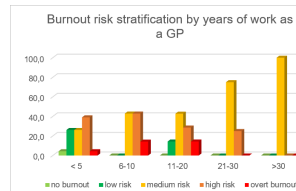
Analysis



Graph 1: distribution of burnout risk in the study sample.



Graph 2: burnout risk stratification by gender.



Graph 3: burnout risk stratification by years of work as a GP.

	Burnout Risk					Total
	No burnout risk	Low risk	Medium risk	High risk	Overt burnout	
	1 (2.2)	7 (15.2)	20 (43.5)	15 (32.6)	3 (6.5)	46 (100.0)
Ambulatory organization						
Work setting						
Community Health	0 (0.0)	1 (16.7)	1 (16.7)	4 (66.6)	0 (0.0)	6 (100.0)
Single-physician practice	0 (0.0)	0 (0.0)	2 (100.0)	0 (0.0)	0 (0.0)	2 (100.0)
Group practice (shared site with other GPs)	1 (2.6)	6 (15.8)	17 (44.7)	11 (29.0)	3 (7.9)	38 (100.0)
Average daily outpatient contacts (no visits)						
<15	0 (0.0)	1 (50)	1 (50)	0 (0.0)	0 (0.0)	2 (100.0)
15-35	0 (0.0)	1 (14.3)	5 (71.4)	1 (14.3)	0 (0.0)	7 (100.0)
36-55	1 (3.8)	4 (15.4)	9 (34.6)	10 (38.5)	2 (7.7)	26 (100.0)
>55	0 (0.0)	1 (9.1)	5 (45.4)	4 (36.4)	1 (9.1)	11 (100.0)
Average daily outpatient ambulatory visits						
<10	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100)	1 (100.0)
10-15	1 (3.7)	5 (18.5)	12 (44.5)	8 (29.6)	1 (3.7)	27 (100.0)
15-20	0 (0.0)	2 (13.3)	6 (40.0)	6 (40.0)	1 (6.7)	15 (100.0)
>20	0 (0.0)	0 (0.0)	2 (66.7)	1 (33.3)	0 (0.0)	3 (100.0)
Daily hours dedicated to outpatient ambulatory visits						
2-4	0 (0.0)	6 (28.6)	7 (33.3)	6 (28.6)	2 (9.9)	21 (100.0)
5-7	1 (5.3)	1 (5.3)	9 (47.3)	7 (36.8)	1 (5.3)	19 (100.0)
>7	0 (0.0)	0 (0.0)	4 (66.7)	2 (33.3)	0 (0.0)	6 (100.0)
Daily hours dedicated to home visits						
<1	0 (0.0)	3 (14.3)	10 (47.6)	7 (33.3)	1 (4.8)	21 (100)
1-3	0 (0.0)	1 (12.5)	7 (87.5)	0 (0.0)	2 (25.0)	20 (100)
>3	1 (20.0)	1 (20.0)	3 (60.0)	0 (0.0)	0 (0.0)	5 (100)
Daily hours dedicated to back-office work						
<2	0 (0.0)	1 (33.3)	2 (66.7)	0 (0.0)	0 (0.0)	3 (100.0)
2-4	1 (3.4)	4 (13.8)	14 (48.3)	4 (13.8)	0 (0.0)	29 (100.0)
5-7	0 (0.0)	2 (20.0)	3 (30.0)	4 (40.0)	1 (10.0)	10 (100.0)
>7	0 (0.0)	0 (0.0)	1 (25.0)	1 (25.0)	2 (50.0)	4 (100.0)

Table 1: distribution of burnout risk and ambulatory organization in the study sample.

Supplementary material available scanning the QR-code below.



Angelica Salvadori^{1,2}, Enrico d'Alessandro², Valerio Andretta², Valentina Chianale², Ilaria Gazzaneo², Matteo Bova³, Niccolò Fabbri³, Alessia Barcellini⁴, Claudia Mantica⁴, Ornella Di Carlo⁵, Giovanni De Luigi¹, Giulio Rigon², Lorenzo Rizzotto², Lucia Luzi Crivellini², Pietro Benini², Maria Francesca Altissimi², Federico Gelain², Giacomo Mattioli², Alessandra Madera², Stefano Bortoletti², Claudia Negri², Maddalena Sarti², Francesco Del Zotti², Sonia Zenari²

¹ CFSMG Piemonte; ² ASL Torino3-Borgaretto-Beinasco; ³ ASL Torino; ⁴ ASL Novara; ⁵ Primary Care, University of Verona School of Medicine, Verona; Italy

BACKGROUND

Home visits will continue to represent a historical component of primary care. In the current context, the Italian National Recovery and Resilience Plan (NRRP) aims to strengthen home-based care, with the target of reaching at least 10% of the population aged 65 years and over. However, despite population ageing and the expected increase in the need for home care, the number of home visits performed by general practitioners (GPs) seems to be declining in Italy, in line with the trend observed in other European countries.

RESEARCH QUESTION

This study aims to explore, through a qualitative approach, the main factors that may contribute to the reduction in home visits, by investigating GPs' perceptions, experiences, and representations of the role and meaning of home visits in the current context.

METHODS

The study is a qualitative research based on 2 comparable focus groups, conducted in Turin and Verona, each including GPs selected through purposive sampling to ensure representation by sex, age group, and practice setting. The conduction was based on a shared semi-structured topic guide, moderated by experienced researchers. The analysis was carried out through the identification of themes and relevant categories and interpretative dimensions.

FOCUS GROUP	TOTAL GP No.	AGE			SEX		Workload		Working area		Type of association	
		<35yrs	35-45yrs	45-60yrs	M	F	Practice only	Practice + training	Urban	Rural	Solo Practice	Group Practice
VERONA	9	3	2	4	4	5	7	2	5	4	3	6
TURIN	10	5	1	4	5	5	5	5	7	3	1	9

RESULTS

WHY ARE HOME VISITS DECLINING?

EXCESSIVE ORGANIZATIONAL DIFFICULTIES OR EASE OF ACCESS TO THE SURGERY

- Lack of time and heavy workload: many GPs have a caseload ranging from 1,500 to 1,800 patients, making it difficult to fit in appointments that take on average 45 to 60 minutes, including travel time and the consultation itself.
- Better practice organization (administrative staff, nurses and a booking system, with reduced waiting times) in contrast to Home Care, which has remained unchanged for the last 20-30 years – lacking technological and equipment support – a disadvantage for home visits! Definitely discouraging requests for home visits for reasons of convenience.

"The recall actually allows me to also reduce that feeling of going to sleep, fearing that I have missed something and in numerical terms the recall allows me to significantly reduce the number of clinical visits"

REPLACEMENT OF HOME VISITS WITH OTHER METHODS

- Impact of COVID-19: the pandemic has changed the attitude of many frail patients who, for fear of infection, have begun to refuse to allow doctors into their homes, leading to the suspension of many scheduled visits that have not been resumed.
- Risk management via telephone: the experience of COVID-19 has taught doctors to refine telephone triage and to use 'recall' (calling the patient back after a few hours or days) to monitor symptoms, reducing the need for immediate visits.
- Furthermore, digitalization has provided alternative communication tools.

"What really worries me when I get a call about a home visit isn't so much what I might find there, but rather, where on earth is this place I'm supposed to go to?"

REDUCED NEED FOR HOME VISITS THANKS TO GREATER INDEPENDENCE AMONG THE ELDERLY AND BETTER PATIENT EDUCATION

- Cultural change and independence: today's elderly are often more independent; they drive and prefer to visit the surgery to feel they are still active.
- Patient education: a cultural shift is underway, influenced in part by the paediatricians' model, which has taught families to bring patients (even those with a fever) to the clinic for a more comprehensive assessment.

"So many families no longer trust the GP who turns up with a stethoscope to listen to their lungs, they want the data!"

INSUFFICIENCY OF HOME VISITS TO REASSURE PATIENTS AND THEIR FAMILIES

- The challenge of 'definitive data': patients and their families are often no longer satisfied with a clinical assessment (stethoscope) and immediately seek 'instrumental data', turning independently to A&E or private facilities.
- Clinical and technological vulnerability: doctors, especially younger ones, feel anxious when visiting a patient's home 'with hammer and scalpel' (i.e. only a stethoscope), feeling deprived of the technological and diagnostic support available in the surgery.

"For me, a home visit is a clinical procedure that must have its reasons and be carried out in a certain way. Perhaps they are on the decline if we count those carried out in a rather haphazard manner! Quantity has given way to quality!"

REDUCTION IN INAPPROPRIATE OR INEFFECTIVE HOME VISITS, GIVEN THE HISTORICAL MOMENT AND SOCIAL CONTEXT

- Evolution of the doctor's role: there has been a shift from the 'doctor popping round' in a friendly manner (for a coffee or a glass of wine) to a view of the visit as a professional clinical act based on necessity, thereby reducing purely 'courtesy' visits.
- Self-protection against burnout: some doctors admit to limiting their visits so as not to become entangled in situations of severe social deprivation or loneliness that they would not know how to resolve, in the absence of a support network (healthcare assistants, social workers).

LACK OF TIME AND UNDER-REPORTING DUE TO BUREAUCRATIC BURDEN

- Bureaucratic burden: the reporting of visits is described as complex, subject to nit-picking by the local health authority, and often unpaid. Some doctors prefer not to record them at all to avoid bureaucratic frustrations, which leads to potentially distorted official data.

"I might as well do it for free! ... Half the Home Visits I do, I end up forgetting to report on them..."

ARE THEY STILL NEEDED TODAY?

A HOLISTIC APPROACH AND 'ENVIRONMENTAL ASSESSMENT': the home is regarded as an entity that 'speaks' more than the patient themselves!

About the actual relational dynamics that would remain hidden in a clinical setting, or the adequacy of the care provided by carers and family members. But also ...the state of the fridge, hygiene, architectural barriers.

IDENTIFYING INVISIBLE STRESS: a home visit can reveal non-clinical causes of ill health (stress arising from caring for a sick partner or situations of social hardship or loneliness that affect physical health). The home visit becomes the only means of identifying needs that would otherwise remain unseen.

"For me, the most challenging cases are the social ones, because I feel powerless and without the skills required to deal with those situations..."

"For me, home visits are mainly about... end-of-life care... the fact that it's just me and the family"

MANAGING GENUINE FRAILTY: for patients who are genuinely unable to travel and are alone, the GP often remains the only healthcare service truly present in the local area. *Actively scheduled* home visits are considered *essential in end-of-life and palliative care* to support the patient and family in a safe environment, and for patients with *chronic conditions* who cannot be transported (heart failure, COPD, the very elderly) and who would not benefit from hospital admission.

EDUCATIONAL VALUE AND STRENGTHENING OF THE RELATIONSHIP: entering the patient's home is described as a *moment of the utmost humanization of care and empathy*, which allows one to 'close the circle' of trust; the doctor can take the opportunity to *educate the patient and family members* on the future management of symptoms, thereby improving the appropriateness of subsequent requests. *The most 'noble' act of the profession.*

"Sometimes I really enjoy going there... just for fun, to open the fridge of some of the elderly residents and see what's inside. These are things you wouldn't discover unless you went into their homes"

DIFFERENCES BETWEEN URBAN AND RURAL SETTINGS?

The geographical context influences the FREQUENCY OF VISITS:

- In rural areas, where services are further apart, doctors tend to make more visits (10-15 a week).
- In urban areas, the proximity to hospitals and A&E departments reduces demand, as patients often 'make do' or prefer the speed of hospital treatment. (2-3 a week). (Some urban doctors note that, following the Covid pandemic, there has been a return to demand for home visits to avoid the long waiting times at A&E departments)

RURAL SETTING GP:

"I don't refuse to make a courtesy visit; I do go, for a brief visit, a brief, objective assessment, and that's it. Even if it could be a bit of a hassle, in some other respects I'm actually quite happy to do it... I realise also that in certain contexts it's difficult to handle..."

COMPARISON OF THEMES BETWEEN THE TWO FOCUS GROUPS:

- Total Convergence on «what is an appropriate HV»:** The Home Visit should be a clinical act, rather than a defensive or bureaucratic one. On the other side, inappropriateness is often determined in hindsight, and so the GP has to carefully evaluate the clinical risk of refusing to carry out a HV.
- Convergence on «the need to protect themselves from social issues»:** In Turin, there is talk of not wanting to 'get too involved in the situation' so as not to carry it into one's bad dreams; in Verona, they describe to feel themselves as 'powerless' in psycho-social cases.
- Divergence on Telemedicine:** whilst in Turin there is hope that telemedicine will bridge the technological gap and it will allow to share psycho-social information also at distance; in Verona scepticism prevails: "I need to touch the patient", and at present telemedicine is considered useful mainly for teleconsultations with a specialist.
- Divergence on management:** in Turin, telephone monitoring is seen as a clinical risk management strategy; in Verona, greater emphasis is placed on trust in the caregiver and on the trust already built up over the years.

POINT OF DISCUSSION/FUTURE PERSPECTIVES

- What is the appropriate role of GPs' home visits within the evolving models of primary care, increasingly based on AI and algorithms? Could HVs be considered an opportunity to train our relational soft skills and complexity management?
- Would prioritising GPs' Home Visits for patient groups with greater clinical and functional needs lead to greater effectiveness? Using what method?
- Extending the study to include patients' perspectives could allow for a comparison between professionals' views and users' expectations and needs, thereby improving healthcare?

CONCLUSIONS

Considering the data emerging from the two focus groups in Turin (TO) and Verona (VR), both groups highlight a form of General Practice under strong pressure, caught between the historical legacy of the "Home Visit" and the challenges of a modern system that is bureaucratized and technologically advanced.

Contact: angelicasalvadori@gmail.com