

# LA CREAZIONE DI UN CAMPIONE VALIDATO DI MEDICI DI MEDICINA GENERALE NEL DATABASE DI HEALTH SEARCH

EMILIANO SESSA\*, FABIO SAMANI\*\*, CARLO NICCOLAI\*\*\*, SERENA PECCHIOLI\*; GIUSEPPE VENTRIGLIA\*\*\*\*, GIAMPIERO MAZZAGLIA\*\*\*\*\*

\* Statistici Health Search; \*\* Responsabile Nazionale, Area Ricerca SIMG;

\*\*\* General Manager Health Search; \*\*\*\* Responsabile Nazionale, Area Formazione SIMG;

\*\*\*\*\* Responsabile Servizi Epidemiologici Health Search

## BACKGROUND

Nel 1998 è stato costituito *Health Search*, Istituto di ricerca della Società Italiana di Medicina Generale basato su una rete di Medici di Medicina Generale (MMG) ricercatori, una scuola presso la quale i MMG sono formati e un database in cui sono raccolti i dati derivanti dalla pratica clinica quotidiana dei MMG ricercatori. Essi hanno aderito al progetto a titolo volontario, impegnandosi all'uso di un software (Millewin®) di gestione dei dati clinici dei pazienti e alla trasmissione periodica dei dati. Millewin® è una cartella clinica informatica nata per aiutare i MMG nella loro pratica quotidiana tramite moduli, procedure e funzioni. È una cartella clinica elettronica orientata per problemi. Le informazioni registrate da ogni MMG sono tutte quelle concernenti la sua pratica clinica quotidiana e variano dalle informazioni demografiche, alle informazioni cliniche (es. diagnosi, indagini di laboratorio con rispettivi risultati, ricoveri, ecc.), ai dati di prescrizione, a quelli relativi a fattori di rischio e a determinanti di salute (es. pressione arteriosa, BMI, fumo, ecc.). Tutti questi dati sono collegati ad un codice anonimo ed univoco per l'identificazione del paziente, secondo i criteri classici di un database relazionale<sup>1</sup>. Gran parte dei dati è codificata secondo standard comunemente utilizzati. La registrazione della prescrizione farmaceutica avviene per mezzo di un elenco di prodotti codificati per nome commerciale e relativa forma farmaceutica, codice ATC e ministeriale, principio attivo. Per gli accertamenti, la codifica avviene in accordo con il Nomenclatore Tariffario come da Gazzetta Ufficiale. I problemi sono inseriti secondo la codifica ICD-9. Attualmente il database di *Health Search* contiene dati relativi a quasi un milione di pazienti ed è una fonte unica d'informazioni per la ricerca scientifica<sup>2,3</sup>.

## OBBIETTIVO

L'obiettivo di questo lavoro è la descrizione delle procedure utilizzate per la costruzione di un campione di MMG selezionati ed affidabili sui quali condurre studi sperimentali e osservazionali<sup>4</sup>. Il campione deve garantire una numerosità campionaria ed un'affidabilità sufficiente per condurre *trial* clinici, studi di coorte, studi caso-controllo e studi di prevalenza anche su *malattie rare*. Inoltre, per

rispondere a criteri di rappresentatività della popolazione generale, il campione deve avere caratteristiche di distribuzione della popolazione rappresentata, per sesso, fasce d'età e per macroaree geografiche, tale da essere sovrapponibile alla distribuzione della popolazione italiana di età adulta, secondo quanto descritto dall'ISTAT.

## METODOLOGIA

### Calcolo indice

Per ogni medico sono state calcolate, sul totale dei propri assistiti, le seguenti percentuali:

- pazienti con dato contemporaneo di fumo, peso, altezza (FPA);
- terapie associate a problema (TAP);
- problemi codificati (PC);
- accertamenti con valore associato (AVA).

Una volta ottenuti i valori descritti abbiamo creato un indice di completezza (IC) calcolato come media pesata di essi.

$$IC = 0,35 * FPA + 0,30 * TAP + 0,25 * PC + 0,1 * AVA$$

I MMG migliori sono quelli che hanno il valore dell'indice di completezza più alto, cioè coloro che hanno le percentuali più alte di registrazione di FPA, terapie associate a problema, problemi codificati, accertamenti con valore associato. In base a tale indice si è ottenuta una graduatoria di tutti i MMG presenti nel database che chiameremo GC.

Un ulteriore indice creato è l'indice delle connessioni calcolato per ogni medico, dividendo il numero di connessioni totali del 2002 per 50 (numero di connessioni annue al server di raccolta dati ritenuto ottimale). In base a tale indice è stata creata una graduatoria dei MMG: i MMG migliori sono i soggetti che hanno il valore dell'indice più elevato cioè quelli che si sono connessi più frequentemente nel 2002. Tale graduatoria sarà di seguito indicata con GCT.

Inoltre abbiamo calcolato il numero medio di visite/die standardizzate per l'anno 2002 in questo modo: VDS = numero medio di visite/die del medico nel 2002 \* (numero medio di assistiti in carico presenti nel database nel 2002 / numero medio di assistiti del medico in carico nel 2002). Anche in questo caso abbiamo creato una gra-

duatoria (GV) per i MMG mettendo in ordine decrescente le visite/die standardizzate. Per ogni medico sono state inoltre calcolate, sul totale dei propri assistiti, le seguenti prevalenze:

- pazienti ipertesi;
- pazienti affetti da BPCO;
- pazienti diabetici;
- pazienti affetti da ulcera;
- pazienti asmatici.

Una volta calcolate tali prevalenze, per ogni patologia viene calcolata la media aritmetica delle prevalenze (rispettivamente  $M_i$ ,  $M_b$ ,  $M_d$ ,  $M_u$ ) e per ciascun medico si calcola il rapporto tra il valore assoluto della differenza tra percentuale e media della distribuzione considerata, e la media stessa:

$$S_i = \frac{|p_i - M_i|}{M_i}; S_b = \frac{|p_b - M_b|}{M_b}; S_d = \frac{|p_d - M_d|}{M_d}; S_u = \frac{|p_u - M_u|}{M_u}$$

dove con  $p$  si è indicata, per ogni medico, la percentuale di pazienti affetti dalle patologie considerate rispetto al totale degli assistiti.

Una volta ottenuti i valori descritti, per ogni medico è possibile calcolare la seguente somma:

$$S = S_i + S_b + S_d + S_u$$

Quanto più  $S$  è basso tanto migliore si presume essere il lavoro di registrazione dei dati da parte del medico, perché i suoi valori di prevalenza si avvicinano a quelli medi del database. Questo valore permette di ordinare i MMG in modo crescente, così da ottenere una graduatoria, dal migliore al peggiore, per quanto riguarda la segnalazione delle malattie croniche dei propri pazienti. In seguito tale graduatoria sarà chiamata GS.

Infine, come indicatore di attenzione alla costanza di aggiornamento del proprio database, abbiamo calcolato per ogni medico la percentuale di deceduti e il relativo scostamento dal valore di mortalità ISTAT (0,95%). In questo caso la graduatoria dei MMG (GM) è stata creata in base a tale scostamento. Abbiamo assunto che i

migliori MMG sono coloro che hanno il dato di mortalità più vicino al valore medio; inoltre i MMG con uno scostamento positivo sono stati ritenuti migliori di quelli con uno scostamento negativo.

Si sono così ottenute cinque distinte graduatorie (GP, GC, GV, GS, GM) dei MMG presenti nel database. Per arrivare ad un'unica graduatoria si è creato un indice finale (IF) come media ponderata delle sopraindicate graduatorie:

$$IF = 0,4*GC + 0,1*GCT + 0,15*GV + 0,25*GS + 0,1*GM$$

Nella Tabella I viene riportato il numero di pazienti dei MMG inclusi nel campione dei migliori 320 MMG, il tutto distinto per area geografica, con le rispettive percentuali. Notiamo che tali percentuali sono praticamente sovrapponibili a quelle ISTAT. Tramite lo stesso indice è stato creato anche un campione dei migliori 50 MMG.

## RISULTATI

### Indice per regione e macroaree

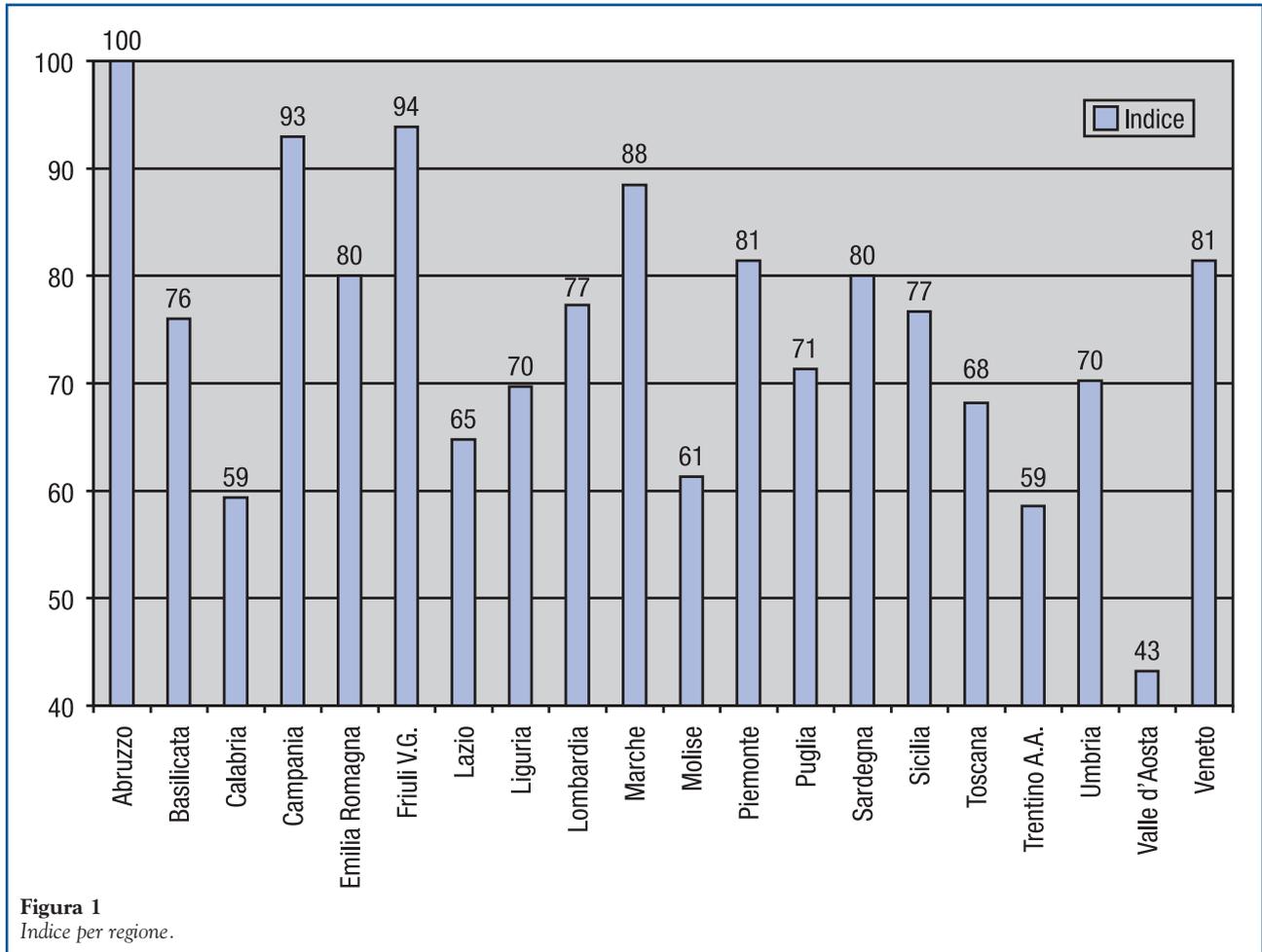
Nella Figura 1 viene riportato l'indice per regione; per una maggiore facilità di interpretazione dei dati abbiamo attribuito un valore uguale a 100 alla regione con la migliore media degli indici relativi ad ogni medico di detta regione, rapportando a tale valore i valori medi delle altre regioni. Come si può vedere, la regione con i dati migliori risulta essere l'Abruzzo, seguito da Friuli Venezia Giulia, Campania e Marche; agli ultimi posti di questa particolare classifica troviamo Valle D'Aosta, Trentino, Basilicata e Molise. Nella Figura 2 vengono riportati i dati aggregati per macroaree; anche in questo caso abbiamo utilizzato la medesima tecnica. Possiamo notare come i dati migliori siano quelli relativi ai medici dell'Italia Nord-Orientale e del Meridione.

### Valori riassuntivi medi degli indici di qualità sui campioni e sul totale

Nella Tabella II sono descritti i vari indici di qualità considerati per la selezione del nostro campione dei 320 MMG. Inoltre, viene effettuato un confronto interno tra i migliori 320 ed il campione totale, rispetto ad un *gold*

**TABELLA I**  
Dati riassuntivi sul campione dei migliori 320 medici.

AREA GEOGRAFICA	NUMERO DI MEDICI	NUMERO DI SOGGETTI ATTIVI	PERCENTUALE DI SOGGETTI NEL CAMPIONE	PERCENTUALE DI SOGGETTI DATI ISTAT 2002
NORD-ORIENTALE	59	85.381	18,5	18,4
NORD-OCCIDENTALE	84	119.534	25,9	26,2
CENTRALE	62	90.921	19,7	19,2
MERIDIONALE	78	113.073	24,5	24,5
INSULARE	37	52.614	11,4	11,7
TOTALE	320	461.523	100	100



standard formato dai migliori 50 MMG. A fronte di una media di assistiti in carico di poco superiore (per i top 50) o di poco inferiore (320 e campione totale) ai 1.400 assistiti/ricercatore, si registrano incrementi nel numero medio di deceduti (104,7 per ricercatore nei top 50, 78,9 nei 320, 70,7 nel campione totale) e nel numero medio di revocati (235,8 nei top 50, 213,2 nei top 320, 179,3 sul totale).

Tali dati indicherebbero una crescente attenzione alla necessità di tenere puliti e aggiornati gli archivi, quale presupposto imprescindibile per qualsiasi stima di preva-

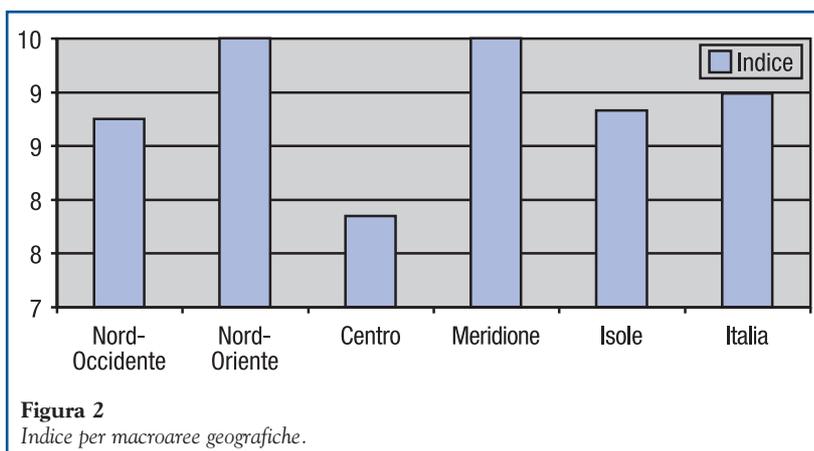
lenza e di incidenza attendibile. Il basso gradiente tra il campione 320 e quello dei migliori 50 fa supporre che questa attenzione giochi un ruolo fondamentale nella determinazione di parametri di qualità nel complesso delle attività di registrazione dei dati.

Complessivamente, possiamo notare una notevole differenza per quanto riguarda la completezza della registrazione di fumo, peso e altezza se consideriamo il campione dei 320 (35,4%) e ancor di più considerando il campione dei migliori 50 (68,5%). Lo stesso trend si può verificare anche considerando le singole voci di fumo,

peso e altezza, laddove il campione dei top 50 è sempre abbondantemente oltre il 70% di completezza per ogni singola voce, quello dei top 320 è oltre il 40% e il campione totale è intorno al 30%.

La considerazione emersa sulla registrazione del dato di morte trova conferma nel dato di prevalenza che mostra differenze non significative tra un campione e l'altro, ed indica inoltre un uso sufficientemente diffuso della codifica ICD9 per le principali patologie considerate.

Le differenze osservabili in indicatori di attività professionale, come il numero di colesterolemie e di pressioni arteriose



**TABELLA II**  
**Dati riassuntivi campione 50, campione 320, database totale.**

VARIABILI	CAMPIONE 50	CAMPIONE 320	DATABASE TOTALE
Pazienti attivi	69.172	461.523	777.183
Pazienti deceduti	5.235	25.242	39.679
Pazienti revocati	11.789	68.220	100.603
Deceduti (%)	0,8	0,7	0,5
Pazienti con dato di fumo (%)	72,1	41,3	28,9
Pazienti con dato di peso (%)	74,2	42,1	31,2
Pazienti con dato di altezza (%)	73,8	41,5	30,3
Pazienti con dato di FPA (%)	68,5	35,4	24,0
Problemi codificati (%)	96,7	94,8	84,6
Terapie associate a problema (%)	99,2	97,2	79,1
Accertamenti con valore associato (%)	79,7	70,1	60,0
Ricoveri (%)	0,9	0,8	0,9
Prevalenza di ipertensione (%)	19,7	19,0	17,4
Prevalenza di diabete (%)	5,5	5,4	4,9
Prevalenza di BPCO (%)	3,6	3,7	3,2
Prevalenza di ulcera (%)	2,5	2,6	2,3
Prevalenza asma (%)	3,1	3,0	2,7
Pazienti con dato di colesterolo (%) 2002	23,6	21,2	17,5
Pazienti con dato di colesterolo (%) 1999-2002	43,2	39,6	36,8
Pazienti con dato di pressione (%) 2002	30,4	22,1	16,9
Pazienti con dato di pressione (%) 1999-2002	51,0	37,2	31,6
Numero medio visite giornaliere standardizzate	24,8	20,4	16,3

**TABELLA III**  
**Dati sul Campione 320, divisi per macroaree geografiche.**

VARIABILI	AREA GEOGRAFICA				
	NORD-ORIENTALE	NORD-OCCIDENTALE	CENTRALE	SUD	ISOLE
Pazienti attivi	85.381	119.534	90.921	113.073	52.614
Pazienti deceduti	5.526	6.443	6.140	5.052	2.081
Pazienti revocati	13.550	18.624	14.944	14.840	6.262
Pazienti con dato di fumo (%)	55,1	36,0	29,2	47,9	37,2
Pazienti con dato di peso (%)	57,1	36,6	29,3	49,7	36,6
Pazienti con dato di altezza (%)	55,9	36,0	28,7	49,2	36,3
Pazienti con dato di FPA (%)	48,3	30,5	23,3	42,9	30,4
Problemi codificati (%)	94,6	94,7	92,5	96,2	96,4
Terapie associate a problema (%)	99,1	98,3	93,3	97,7	96,9
Accertamenti con valore associato (%)	80,4	70,2	64,4	66,8	69,9
Ricoveri (%)	1,0	0,9	0,6	0,9	0,9
Deceduti (%)	0,7	0,7	0,6	0,6	0,6
Prevalenza di ipertensione (%)	20,6	19,3	17,7	18,8	18,1
Prevalenza di diabete (%)	5,2	4,8	5,4	5,8	6,1
Prevalenza di BPCO (%)	3,5	3,1	3,4	4,7	3,5
Prevalenza di ulcera (%)	2,6	2,3	2,7	3,1	2,1
Prevalenza asma (%)	2,9	2,8	2,6	4,4	3,3
Pazienti con dato di colesterolo (%) 2002	22,4	19,7	19,7	22,6	22,0
Pazienti con dato di colesterolo (%) 1999-2002	43,2	37,5	40,6	39,6	36,8
Pazienti con dato di pressione (%) 2002	27,1	20,8	17,4	24,1	21,2
Pazienti con dato di pressione (%) 1999-2002	44,9	35,6	32,1	38,4	34,2
Numero medio visite giornaliere standardizzate	21,4	18,1	18,3	22,3	24,0

registrate, possono invece indicare una maggiore attenzione alla registrazione da parte del campione dei 50, ma anche un diverso comportamento professionale, leggibile anche nel diverso carico di lavoro giornaliero desumibile dal numero medio di visite giornaliere standardizzate. In conclusione, è possibile che un ricercatore più attento nella registrazione dei propri dati sia anche un professionista più attento ai processi di cura.

In conclusione, la risposta da parte di tutti i ricercatori è stata molto buona, con l'evidenza di uno "sviluppo sostenibile" da parte dei migliori, tale da far ritenere possibile l'obiettivo di raggiungere nell'arco di un paio d'anni il *gold standard* dei migliori 50 MMG per tutti i MMG partecipanti a *Health Search*.

### Valori riassuntivi medi degli indici di qualità per aree geografiche

Anche la lettura degli stessi indicatori di qualità per macroarea geografica sul campione dei 320 (Tabella III) evidenzia differenze molto ridotte in termini di numero medio di visite giornaliere, così come anche gli altri indicatori non mostrano evidenze sostanziali tra un'area e l'altra, se non nel rimarcare la necessità, per i ricercatori del Centro e in parte per quelli del Nordovest e delle Isole, di intensificare la registrazione dei dati di fumo, peso e altezza. La registrazione del valore degli accertamenti sembra una peculiarità del Nordest, che si discosta significativamente da tutte le altre aree ad indicare, una volta di più, una linea di "sviluppo sostenibile" anche per i ricercatori delle altre regioni.

### CONCLUSIONI

L'applicazione di un modello matematico basato su indicatori espliciti e ponderati, per quanto arbitrari, ha consentito di creare una "graduatoria temporanea di qualità" che consente di identificare un *pool* di ricercatori i cui dati possono essere considerati completi, affidabili, rappresentativi della popolazione generale italiana di età adulta. Tali caratteristiche hanno reso possibili i primi studi di validazione del database di *Health Search* e le prime pubblicazioni scientifiche di rilievo su riviste scientifiche internazionali ad alto *impact factor*<sup>2-4</sup>. Questo fatto rappresenta la vera "validazione" delle attività dell'Istituto di ricerca (e dei suoi ricercatori), secondo parametri comunemente accettati da quella comunità scientifica internazionale di cui *Health Search* ed i suoi ricercatori si onorano di far parte.

Infine, le caratteristiche di elasticità applicativa dell'algoritmo descritto nel presente lavoro sono tali da costituire un ulteriore incentivo per tutti i ricercatori, in particolar modo per quanti non siano stati finora inclusi nel *pool* dei 320 validati. Il nostro auspicio è che in tempi ragionevoli la qualità complessiva dei dati di *Health Search* sia tale da poter considerare "validato" ogni ricercatore che vi operi.

### Bibliografia

- <sup>1</sup> Health Search, Istituto di ricerca della Società Italiana di Medicina Generale. *Report 2002*. Firenze, giugno 2003.
- <sup>2</sup> Cricelli C, Mazzaglia G, Samani F, Marchi M, Sabatini A, Nardi R, et al. *Prevalence estimates for chronic diseases in Italy: estimating the differences between self report and primary care databases*. J Pub Health Med 2003;25:254-7.
- <sup>3</sup> Filippi A, Bignamini AA, Sessa E, Samani F, Mazzaglia G. *Secondary prevention of stroke in Italy: a cross-sectional survey in family practice*. Stroke 2003;34:1010-4.
- <sup>4</sup> Filippi A, Sabatini A, Badioli L, Samani F, Mazzaglia G, Catapano A, et al. *Effects of an automated electronic reminder in changing prescribing behaviour among General Practitioners in Italy: an intervention trial*. Diabetes Care 2003;26:1497-500.

