

“Potenzialità e prospettive derivanti dall’integrazione di dati su sanità e salute”

A.Burgio, R.Crialesi, P.Di Filippo, L.Gargiulo, L.Iannucci, G.Sebastiani, A.Solipaca – ISTAT

La comunicazione ha come obiettivo la descrizione dei risultati preliminari ottenuti mediante l’applicazione dei metodi per l’integrazione descritti nella comunicazione di Marco Di Zio dal titolo “Metodi per l’integrazione tra la base dati Health Search e l’indagine Istat sulle condizioni di salute”.

Per effettuare l’integrazione, le due fonti di dati devono essere rappresentative della stessa popolazione. Tale requisito è rispettato dalla fonte IS in quanto è basata su un campione probabilistico rappresentativo della popolazione residente, ma non è necessariamente rispettato dalla fonte HS che deriva da campione non probabilistico. Pertanto per utilizzare congiuntamente le due fonti è stata effettuata una post-stratificazione dei dati di HS per età e sesso.

Le variabili comuni devono inoltre avere lo stesso significato nelle due fonti. Tale requisito non è rispettato per le malattie croniche disponibili in entrambe le fonti, in quanto in IS la presenza di una malattia cronica è dichiarata dall’intervistato (percezione soggettiva), mentre in HS è diagnosticata dal medico (condizione “oggettiva”).

Pertanto sono state selezionate due malattie (ipertensione e diabete) ed è stata calcolata la probabilità di un individuo appartenente al profilo j di dichiarare di avere la malattia dato che un individuo dello stesso profilo j in HS ha una diagnosi medica per la stessa malattia. Tale probabilità per l’ipertensione è pari a 0,77 e per il diabete è pari a 0,69. Nelle donne risulta sempre più elevata che negli uomini per entrambe le patologie.

Questa probabilità è stata utilizzata per modificare la variabile “presenza di almeno una malattia cronica” nell’indagine IS al fine di migliorare il suo utilizzo come variabile comune per la successiva operazione di matching statistico tra i dati dell’indagine Istat e i dati HS. In questo modo è come se si riconducesse la variabile di matching ad avere lo stesso significato nelle due fonti.

Come prima applicazione, il matching statistico tra le due fonti è stato effettuato, sotto l’ipotesi di indipendenza incondizionata, considerando la regione Lombardia. Le altre variabili comuni sono il sesso, l’età, la presenza di almeno una malattia cronica e la presenza di almeno una malattia del cuore.

Il dataset ottenuto con il matching per la regione Lombardia ha 10.417 record (che riportati all’universo sono riferiti a 7.659.228 individui di 20 anni e più).

L’analisi della spesa in relazione alla cattiva salute percepita mostra come la spesa media totale risulti molto più elevata per chi dichiara di stare male o molto male; la spesa è maggiore per gli uomini (anche per età) e il gap di genere è massimo nella classe di età 60-69 anni.

Un altro risultato analizza la spesa media totale in relazione al titolo di studio. La spesa media totale è più elevata per chi ha un basso titolo di studio, ma ciò dipende dalla struttura per età più anziana di questo gruppo di popolazione, difatti la distribuzione per età non mostra differenze significative rispetto a chi ha un titolo di studio medio o alto.

Il terzo risultato considera la spesa media totale SSN per un malato di diabete, che è pari a circa 700 euro. Tale valore sale a circa 950 euro se la persona affetta da diabete dichiara di stare male o molto male e a 1.150 euro se ha almeno altre 2 malattie croniche.

Infine si presentano i risultati ottenuti con “l’analisi dell’incertezza” (quindi svincolandosi dall’ipotesi dell’indipendenza condizionata). In questo caso il risultato è un intervallo di frequenza ottenuto per le classi di spesa in relazione alle variabili di studio (salute percepita e titolo di studio). I risultati ottenuti con questa analisi sono simili a quelli ottenuti con il matching sotto l’ipotesi di indipendenza condizionata. Tuttavia considerando l’ampiezza media degli intervalli di frequenze, possiamo affermare che le relazioni che si vanno a studiare tra titolo di studio e spesa totale SSN sotto l’ipotesi di indipendenza condizionata sono affette un grado di incertezza maggiore.

In conclusione le prospettive per questo tipo di analisi vanno nella direzione: di estendere i risultati ad altre regioni o a tutta l’Italia; di studiare altre possibili variabili comuni per aumentare la qualità del matching statistico; di effettuare l’integrazione per diverse repliche dell’indagine salute (2004-2005, 2012-2013); di utilizzare informazioni aggiuntive presenti nell’indagine salute 2012-2013, sia perché in caso di forte relazione con le variabili di spesa di HS migliorerebbe la qualità dell’integrazione sotto l’ipotesi di indipendenza condizionata, sia per poter ricostruire la spesa complessiva (spesa SSN di HS e spesa out of pocket di IS).