

## Selezione e armonizzazione delle variabili della base dati Health-Search e dell'indagine Istat sulle condizioni di salute

V. Atella, F. Belotti, A. Piano Mortari - Centre for Economic and International Studies, Università di Roma Tor Vergata

Obiettivo: Selezione e armonizzazione delle variabili della base dati Health-Search al fine di renderle compatibili con quelle rilevate nell'indagine ISTAT sulle condizioni di salute. Costruzione delle variabili di spesa per profilo.

Il problema principale che si incontra nell'effettuare il *linkage* tra le due basi di dati considerate è l'assenza di un codice identificativo univoco che permetta di legare le informazioni ivi contenute. In questi casi, la procedura standard consiste nel fare riferimento a un insieme di variabili (variabili comuni) presenti in ambedue le basi di dati che siano in grado di identificare congiuntamente e univocamente ciascuna unità. La nostra proposta consiste quindi nel procedere a quello che possiamo definire un *profile matching*, ovvero al *merge* di gruppi di unità che presentano una serie di caratteristiche in comune.

Al fine di assicurare un corretto *profile matching*, le due basi di dati sono state rese omogenee in termini di età dei rispondenti/pazienti. Si è proceduto quindi all'eliminazione dalla base di dati Istat dei rispondenti con un'età inferiore ai 20 e superiore ai 95 anni, e la stessa operazione è stata fatta per la base di dati Health Search. La base di dati Istat risultante contiene 103,144 osservazioni la cui somma dei pesi (*coefi2*) è di 46,806,400 (circa l'80% della popolazione italiana del 2005).

Attraverso un'analisi dettagliata dai tracciati record delle due basi di dati, è stata definita una strategia per la creazione di una serie di variabili categoriali attraverso le quali definire una serie di profili omogenei per il *merge*. Un primo gruppo di variabili che potremmo definire "di base" sono rappresentate dal *sex*, dalla *classe di età* e dall'*area geografica di residenza*. Chiaramente, le modalità di tali variabili sono state rese omogenee fra le due basi di dati. Visto che le informazioni raccolte attraverso l'Indagine Condizioni di Salute e Ricorso ai Servizi Sanitari sono riferibili all'ultimo mese del trimestre di rilevazione (Dicembre 2004, Marzo 2005, Giugno 2005 e Settembre 2005) mentre nella base di dati Health Search le informazioni sono raccolte dai Medici di Medicina Generale (MMG) su base giornaliera e descrivono in modo costante l'attività dei MMG nei loro rapporti con ciascun paziente, un secondo gruppo di variabili che potremmo definire "specifiche" è stato creato tenendo conto di queste differenze, prendendo in considerazione per la base di dati Health Search tutte e solo le informazioni rilevanti riferibili agli stessi periodi di rilevazione che caratterizzano l'indagine Istat. Le variabili chiave "specifiche" considerate sono:

- Malattie croniche. Variabile dummy che indica la presenza o meno di almeno una delle seguenti malattie croniche (diagnosticate in Health Search, dichiarate in Istat): diabete, enfisema, morbo di Parkinson, cirrosi epatica, asma, ipertensione, ictus, tumore, calcolosi, tiroide;
- Malattie del cuore. Variabile dummy che indica la presenza o meno di almeno una delle seguenti malattie croniche (diagnosticate in Health Search, dichiarate in Istat): infarto, angina pectoris, ed altre patologie cardiovascolari
- Invalidità. Variabile dummy che indica la presenza di almeno una delle seguenti invalidità: cecità, sordomutismo e sordità;
- Accertamenti diagnostici. La variabile indica se il paziente si è sottoposto nelle ultime 4 settimane ad almeno uno dei seguenti esami: esami del sangue, esami delle urine o altri accertamenti diagnostici (Rx, diagnostica per immagini, esami di laboratorio).

Una volta create le variabili comuni, la variabile chiave su cui basare il *merge* è stata costruita considerando tutte le possibili combinazioni tra le modalità delle variabili caratterizzanti. Quindi, la variabile chiave è a sua volta una variabile categoriale con  $k$  modalità, dove  $k$  è il numero di profili ottenibile dalle combinazioni delle modalità delle variabili caratterizzanti considerate.

Inoltre a partire dal database Health Search, integrato con i tariffari regionali e nazionali, sono state costruite le seguenti variabili di spesa per i diversi profili:

- Spesa per farmaci: è la spesa calcolata a partire dalle prescrizione dei MMG (Classe A – SSN) al lordo dello sconto SSN
- Spesa per visite: è la spesa calcolata a partire dalle prescrizione di visite specialistiche da parte dei MMG
- Spesa per accertamenti: è la spesa calcolata a partire dalle prescrizioni di accertamenti (e.g. laboratorio, diagnostica per immagini) effettuate dai MMG
- Spesa totale: è la somma delle voci precedenti