

## Da dove vengono e dove vanno le classificazioni internazionali

Prolegomeni per una riflessione in occasione della pubblicazione italiana della Classificazione Internazionale delle Malattie per l'Oncologia, 3<sup>a</sup> edizione

### Origine delle standardizzazioni internazionali

Gran secolo, il XIX. Lungo e pieno di cose nuove. Produzione industriale e trasporto dell'energia, ferrovie dappertutto, navi autopropulse, chimica industriale, grandi macchine, l'acciaio a sostituzione di legno e pietra nella costruzione, e tante altre novità, quasi tutte ancora in uso. Per far girare ed espandere industria e commercio occorre accordo di unificazione e normalizzazione (oggi sembra preferito il termine standardizzazione), e gli accordi furono fatti su molte cose: pesi e misure (1875, con innesto sulla lontana radice del primo metro campione, 1799), poste, scartamenti (dove l'uguale era la norma e il diverso la difesa militare) e sagome ferroviari, grandezze elettriche e meccaniche, nomenclatura chimica e biologica, eccetera. Non essendoci organismi internazionali, gli accordi si facevano per Conferenze *ad hoc*, cui i paesi che volevano essere nel giro partecipavano. Dal grande al piccolo, e dal materiale all'astratto, gli accordi riguardarono anche la Classificazione delle malattie, entro gli sforzi di unificare indicatori demografici e statistici, e anzi la Classificazione internazionale delle malattie (CIM)<sup>1,2</sup> fu tra i primi esempi in questo campo. Le Conferenze erano strumenti multilaterali, non organismi internazionali nel senso attuale. Questi ultimi, dopo l'esperienza pilota della Società delle nazioni del 1919, vennero sviluppati solo a partire dal 1945, quando la sfera internazionale fu regolata in modo nuovo: per le cose che contano (guerra, soldi, commercio) furono create agenzie/conferenze oligarchiche (NATO/SEATO, FMI, GATT, eccetera), mentre le cose che contano meno (salute, lavoro, istruzione) furono affidate a organismi democratici (almeno in quanto regolati da un suffragio universale, un paese un voto), cioè all'ONU e ai suoi satelliti (FAO, OMS, BIT, UNESCO, eccetera). Come si sa, la Classificazione internazionale delle malattie non andò assegnata all'ufficio studi della Banca Mondiale (che pure ne è un vivace utente), ma fu messa a fiorire in punta ad uno dei mille rami del tronco OMS.

Rispetto agli organismi internazionali, se non è qui la sede per una riflessione sul tutto (che è tanto facile criticare quanto difficile immaginare diverso), merita forse osservare per la parte (la gestione OMS delle statistiche sanitarie) che l'istituzione di un livello internazionale (che non è conferenza ma apparato, con un suo piccolo o grande diritto e potere) ha creato una dualità rispetto alle strutture nazionali. Per alcuni paesi, particolarmente gli USA, la dualità si è espressa in competizione culturale, per altri, come il nostro, si è tradotta in una progressiva reciproca passività.

### Della tendenza delle statistiche correnti in generale

I sistemi pubblici di dati correnti e di rappresentazione statistica nazionale e internazionale ci paiono quasi dovunque in crisi. Crisi di mandato, conseguente alla perdita di potere e di capacità progettuale dei loro referenti (i governi e gli apparati statali a livello nazionale, gli organismi internazionali a livello sovranazionale) in rapporto al potere crescente dei centri economici; crisi culturale, per il progressivo affievolirsi del loro rapporto con il mondo della ricerca metodologica e con quello dell'utilizzo dei dati; crisi operativa, forse inevitabile conseguenza delle due precedenti, testimoniata dalla scarsità di introduzione di nuovi strumenti, dalla caduta di qualità e confrontabilità dei dati, dalla ingravescenza dei ritardi. Un esempio per tutti, la deriva dei censimenti.

### Delle statistiche sanitarie in particolare

Le statistiche sanitarie sono più difficili da fare di quelle di molti altri campi. Innanzi tutto per la vastità del dominio di interesse (si pensi solo alla dimensione della matrice malattie-test diagnostici-farmaci-tecnologie); poi per la dimensione della popolazione in studio (in un paese povero nessun uomo-anno è esente da malattia, in un paese ricco nessun uomo-anno è esente da un consumo medico personale e da un'intervento sanitario collettivo). Poi sono onerose da fare per la circostanza che la maggior parte dei dati sanitari sono eloquenti solo se riferiti all'individuo, o se sono calcolati come valori medi di dati individuali. Infatti i dati generali si formano per aggregazione di singoli casi, e non invece (come è frequente nelle statistiche economiche) i singoli casi si stimano per divisione di maggiori grandezze. Infine le statistiche sanitarie sono impegnative da fare perché la cultura medica, meritoriamente, ambisce a capire il perché mentre conosce il quanto (di nuovo, a differenza delle statistiche economiche, più interessate al risultato e alla sua tendenza che alla sua eziologia) e ciò arricchisce il lavoro ma complica le cose per lo statistico sanitario. Tutto questo fa sì che nei paesi poveri (o neo-poveri, come nell'Est europeo) sia sempre più difficile contare perfino gli eventi più gravi e rari (per esempio la mortalità), mentre nei paesi ricchi, a fronte di una disponibilità crescente di dati su singoli eventi minuti sia sempre più difficile comporre quadri informativi d'insieme (per esempio è facile stimare quanti esami PET vengono fatti in un anno negli USA, mentre è impossibile sapere quante e quali donne *non* fanno il pap-test). In generale, la disponibilità di dati sanitari predisposti per una lettura su scala mondiale oggi non è maggiore di 40 anni fa.

### Della CIM-O in effimero

La CIM-O è uno strumento per fare le statistiche sui tumori.<sup>3</sup> Si compone di due parti: una prima parte, cosiddetta Topografica, non è altro che la CIM, alla quale 150 anni di Comitati di revisione non hanno aggiunto un solo spunto di ideazione all'impostazione tassonomica originale di W. Farr (1853) e poi di J. Bertillon (1911);<sup>4</sup> una seconda parte, cosiddetta Morfologica, dove il *corpus* nomenclativo dell'anatomia patologica confrontato alla ordinalità gerarchica dei numeri di codice rivela tutta la propria indefinitezza logica formale (restando inteso che l'anatomia patologica deve sicuramente essere dotata di una sua *logica sostanziale*, visto che in molti casi essa si rivela utile per fare le diagnosi, decidere i trattamenti e formulare le prognosi). Mentre la parte Topografica è molto utile per redigere le statistiche (e i tumori sono il settore della patologia che dispone delle più ricche e valide statistiche), la parte Morfologica non ha finora mostrato che rarissimi esempi di utilità pratica. Questo problema crediamo fosse lucidamente presente a due epidemiologi, F. Berrino e J. Estève, che circa 20 anni fa idearono e realizzarono il sistema EuroCIM.<sup>5</sup> Il sistema EuroCIM tentava di individuare quelle combinazioni (chiamate entità) della classe Topografica con la 'classe' Morfologica che aggiungessero, appunto per effetto della loro combinazione, capacità informativa ai dati, soprattutto nel senso di potenziare i differenziali spaziali e temporali osservabili dagli studi descrittivi. Il sistema era pensato per sperimentarsi su di una grandissima base di dati (la somma dei dati dei Registri Tumori europei, da cui il nome) cosicché si potesse valutare quali combinazioni fossero molto arricchenti, quali meno, e quali punto. Purtroppo questo modo di guardare ai dati e alle classificazioni non si è sviluppato e non si è imposto. I dati sui tumori continuano ad essere pubblicati e letti quasi solo secondo la sola classe Topografica, mentre continuano ad essere predisposti anche secondo la 'classe' Morfologica; quest'ultima usanza richiede, nel mondo, il lavoro di non meno di 200 persone-anno all'anno. Inoltre questo imponente (sulla scala delle risorse disponibili per l'epidemiologia) lavoro di codifica delle diagnosi morfologiche si svolge con limitati (alla fase finale del processo dei dati) controlli di qualità; mentre la riproducibilità dei dati sulle diagnosi topografiche è intrinsecamente garantita, quando non da analisi *ad hoc*, dalla pluralità delle fonti di segnalazione e dalla molteplicità dei test diagnostici considerati. A nostro avviso il copioso e in gran parte inutilizzato lavoro di codificazione morfologica potrebbe essere sostituito dalla semplice archiviazione dei testi diagnostici originali (poco costosa se in formato elettronico) e dal loro processamento al bisogno, previo controllo di qualità. Oppure dalla adozione di sistemi automatici di co-



dificazione, cui si allude nel paragrafo successivo.

### La potenza dell'automazione elettronica: per unire o per disperdere?

Classificare consente di ridurre la molteplicità e fare generalizzazioni, quando possibile (e possibile lo è nella misura in cui la natura è disposta gerarchicamente) e utile (ed è utile quando vi è interesse per i diversi livelli della gerarchia). Codificare consente di operare le classificazioni secondo logiche numeriche elementari, e di farlo con strumenti artificiali (ieri meccanici, oggi elettronici) risparmiatori di lavoro intellettuale umano. All'epoca di Bertillon non c'erano strumenti artificiali, e lo statistico doveva 1. istituire o adottare le classi; 2. assegnare loro i soggetti; 3. fare, disfare e contare pile di documenti per ogni possibile lettura delle classi. Con l'avvento degli strumenti artificiali, il passo 3 ha potuto essere fatto dalle macchine, previa una codificazione (passo 2, umano) compiuta *una tantum*. Dunque sono restati al lavoro umano i passi 1 e 2. Ora, entrambi questi passi, se e per quanto razionali, sono passibili di automazione, per giunta a costi ragionevolmente bassi da quando la potenza di calcolo è diffusamente disponibile. Perché, mentre il censimento statunitense del 1911 ha propulso lo sviluppo delle macchine per il passo 1, invece il moderno mondo della statistica non è committente di prodotti software che aiutino l'uomo nel passo 1 e lo sostituiscano nel passo 2? Non troviamo a questa domanda risposte diverse dalla crisi di mandato statale per la statistica, cui abbiamo fatto allusione più sopra, e della attenzione crescente degli statistici al particolare e ai metodi più che al generale e ai risultati. Nell'ambito della automazione del lavoro statistico ancillare vi sono poche ma significative realizzazioni, alcune in commercio, altre allo stadio prototipo.<sup>6-7</sup> Auspicheremmo che gli epidemiologi ne fossero tanto curiosi quanto sembrano esserlo delle 'novità' delle successive revisioni della CIM-O, che sono dell'ordine del 2% ogni 10 anni.

Roberto Zanetti, Stefano Rosso

### Bibliografia

1. <http://isi.cbs.nl/>. E' il sito di ISI, International Statistic Institute
2. [www.ined.fr/bdd/causfra/bib-cim.html](http://www.ined.fr/bdd/causfra/bib-cim.html). E' il sito del francese Institut National d'Etudes Démographiques
3. referenza della versione italiana della CIM-O, nella traduzione di S. Ferretti e A. Giacomini. Completare a cura della redazione di E&CP
4. [www.who.int/classification/icd/en/](http://www.who.int/classification/icd/en/). E' il sito della Organizzazione Mondiale della Sanità
5. vedere alla voce EuroCIM nel sito [www.iarc.fr/](http://www.iarc.fr/)
6. AIM, Artificial Intelligence in Medicine. AutoCode Semantic Analysis Engine. [www.aim.on.ca](http://www.aim.on.ca)
7. Patriarca S, Vicari P, Favero C, Segnan N, Zanetti R. Sviluppo di un codificatore automatico per le diagnosi istologiche. XXVI Réunion du Groupe pour l'Epidémiologie et l'Enregistrement du Cancer dans les Pays de Langue Latine. Neuchâtel, Suisse, 24-25.05.01