

La ricerca in un mare di dati

Generare nuove ipotesi da verificare con gli strumenti dell'ebm

quest'ultimo ambito, il cambiamento potrebbe consistere nell'integrare differenti soggetti con un approccio più ampio, nel quale tutte le risorse (competenze, dati, finanziamenti) concorrono nel completare il profilo di evidenze delle nuove tecnologie e ridurre i rischi³. Il paradigma delle produzioni di informazioni sui farmaci è molto cambiato negli ultimi 10 anni: se prima l'intero pacchetto informativo veniva prodotto dall'azienda e valutato dal decisore, oggi ogni stakeholder è promotore e produttore di dati ed evidenze.

I decisori e pagatori italiani trattano questi temi a livello internazionale anche se nella realtà la loro applicazione ai fini decisionali (prezzo e rimborso, gestione del ciclo di vita del prodotto) è complessa e poco regolamentata. Tuttavia la produzione di big data è potenzialmente positiva sotto molteplici aspetti e per diversi attori: stimola l'appropriatezza d'uso dei farmaci e il governo della spesa; tutela tutti i portatori di interesse (agenzie regolatorie, pagatori, pazienti, industria, clinici, accademia); alimenta la collaborazione tra istituzioni, industria, accademia, esperti e decisori; incoraggia la cooperazione pubblico-privato sulla generazione di evidenze e la diffusione di quanto è già stato raccolto; sostiene un processo dinamico di valutazione che include diversi portatori di conoscenze. Le aziende, dal canto loro, prevedono già investimenti per i big data, finanziando studi pre- e post-registrativi, sostenendo il fondo per la ricerca indipendente e i registri di monitoraggio Aifa, erogando contributi per creazione e sostegno di grandi registri di patologia e alimentando in modo diretto e indiretto attività di ricerca su database amministrativi.

Si devono sostenere una produzione e una metodologia di interrogazione dei dati di qualità. ... Ma manca un sistema a "doppio senso" premiante sia per il pubblico sia per il privato.

Servono figure preparate, esperte in epidemiologia, sociologia, statistica. Si devono sostenere una produzione e una metodologia di interrogazione dei dati di qualità. Il passaggio dal dato all'evidenza non può prescindere dalla fase di validazione e di discussione critica. Le aziende farmaceutiche sono sempre più consapevoli che il potenziale di questo strumento si realizza solo attraverso la trasparenza e la collaborazione, ma manca un sistema a "doppio senso" premiante sia per il pubblico sia per il privato.

Le problematiche che si incontrano sono note e, purtroppo, la loro risoluzione non sembra immediata: ostacoli in Italia per gli investimenti, costi e "burocrazia", e scarsa chiarezza su regole di accesso ai data base, in primis, fanno emergere la difficoltà di gestire gli intenti comuni e i (legittimi) conflitti di interesse che si possono superare solo con la trasparenza e la chiarezza degli obiettivi. Si percepisce, sempre più, la necessità di adottare un linguaggio comune, di avere accesso ai dati già disponibili, di adottare una metodologia rigorosa e condivisa per avere evidenze riproducibili, di stabilire una partnership trasparente con i decisori per la definizione di come e perché produrre big data, di avere una legislazione dedicata e aggiornata. Privacy, etica, raccolta, proprietà, capacità di interrogare correttamente i dati e di trasformarli in evidenze sono alcuni dei temi ricorrenti che vengono sollevati da più parti e che difficilmente trovano ora una risposta valida per tutti.

Il futuro, o forse sarebbe più corretto dire il presente, richiede quindi una collaborazione pubblico-privato sull'uso dei big data e la creazione di un network di eccellenza per la generazione di evidenze, in cui tutti gli stakeholder (Aifa, Istituto superiore di sanità, il mondo accademico, le società scientifiche, l'industria farmaceutica e dei device, le associazioni dei pazienti, le Regioni, il garante della privacy) trovino un consenso sulle migliori soluzioni a queste problematiche. Riconoscerne la priorità deve essere il passo successivo perché il fattore tempo è vitale e impetuoso. F

I big data hanno assunto un ruolo sempre più centrale nella raccolta, elaborazione e interpretazione di dati.

Partiamo da alcuni numeri: l'80% dei dati generati quotidianamente non è strutturato, cioè non proviene da archivi e database predefiniti e codificati ma piuttosto dal web, dai blog, dai social network e dai forum. C'è addirittura chi sostiene che nel 2020 tale tipologia di dati rappresenterà il 100%. A ciò occorre aggiungere il fatto che la diffusione di smartphone, wearable e, in generale, dell'*internet of things* sta producendo una quantità enorme di dati, anche in ambito sanitario. Si stima che ogni giorno, sommando a questi dati quelli provenienti dalle conversazioni su internet, si sfiori la cifra di 2,5 miliardi di miliardi di byte. La crescita è stata impetuosa soprattutto nel corso degli ultimi due anni, durante i quali è stato generato il 90% dei dati disponibili oggi nel mondo. Addirittura si stima che ogni persona, attraverso le proprie appendici elettroniche e le operazioni che compie nella quotidianità, contribuisca a produrre al minuto 1,7 megabyte di dati.

Non c'è dubbio che siano necessarie tecniche di analisi in grado di interpretare questa grossa mole di dati, così da poter attivare



Eugenio Santoro

Laboratorio di Informatica medica Irccs - Istituto di ricerche farmacologiche "Mario Negri"

naturale e di proporre risposte basate sulle evidenze scientifiche), al fine di individuare nuovi approcci sulle cause, sui trattamenti e sulla prevenzione delle malattie cardiovascolari. E c'è già chi, come il Memorial Sloan Kettering cancer center di New York, ha integrato queste soluzioni all'interno del percorso assistenziale dove Ibm Watson collabora con i gli oncologi di questa struttura per diagnosticare un tumore e identificare la cura migliore.

Dal punto di vista della ricerca clinica, l'impiego dei big data potrebbe aggiungere molto agli strumenti oggi impiegati. Per esempio, potrebbero essere utilizzati per analizzare in modo integrato i dati genomici, medico-sanitari, nutrizionali ed economico-sociale, con l'obiettivo di individuare possibili cure per diverse patologie (questo

Disporre di più dati non significa necessariamente disporre di più conoscenza.

tutta una serie d'iniziative che in ambito sanitario potrebbe portare a un miglioramento dell'assistenza sanitaria, alla riduzione dei costi per l'erogazione dei servizi, alla fruizione di servizi che aiutino a supportare le decisioni cliniche, alla realizzazione di sistemi di sorveglianza delle malattie e di farmacovigilanza e, in generale, alla gestione della salute pubblica.

In questa direzione vanno numerose iniziative intraprese nel corso degli ultimi anni, soprattutto da parte dei colossi dell'informatica, spesso in collaborazione con le multinazionali del farmaco. Alcune recenti iniziative vedono coinvolti diverse alleanze: Sanofi e Google Life Science (la divisione di Google che si occupa di sviluppare nuove tecnologie in ambito medico) per sviluppare nuovi strumenti che integrino i vari approcci terapeutici e di controllo nei pazienti con diabete di tipo 1 e di tipo 2; Medtronic e Samsung per il monitoraggio e l'analisi di dati ottenuti da nuovi strumenti di monitoraggio; Novo Nordisk e Ibm per migliorare la gestione dei pazienti diabetici.

Anche le società scientifiche negli Stati Uniti stanno entrando in questo settore, come dimostrano gli accordi tra l'American heart association e Google Life Science, e quelli tra la stessa associazione scientifica e l'Ibm Watson Health (la divisione di Ibm che gestisce il super computer Watson in grado, attraverso tecniche di intelligenza artificiale e di *cognitive computing*, di analizzare elevati volumi di dati, di comprendere domande complesse poste in linguaggio

è proprio uno degli obiettivi di Human technopole elaborato dall'Istituto italiano di tecnologia). Oppure, semplicemente monitorando le conversazioni sui social media, possono essere usati per raccogliere ed elaborare dati legati alla frequenza dei sintomi di una malattia, agli effetti collaterali registrati dal paziente in seguito alla assunzione di un dato farmaco, o alla diffusione di specifiche malattie infettive. O ancora, elaborando i campi note di una cartella clinica elettronica, potrebbero aiutare a catturare in profondità dettagli che racchiudono informazioni di valore sulla salute del paziente.

Tutto ciò rappresenta un'alternativa agli strumenti oggi in uso per fare ricerca clinica? Non è detto che ciò sia vero. Concetti come la rappresentatività, la riproducibilità dei risultati, l'evidence based medicine sono piuttosto distanti dalle tecniche di big data che, sebbene lavorino su grandi mole di dati, possono contenere al loro interno delle distorsioni. Il loro impiego potrebbe invece risultare prezioso nel generare ipotesi che devono poi essere studiate e verificate attraverso gli strumenti (come le sperimentazioni cliniche controllate, gli studi osservazionali e quelli epidemiologici) della medicina basata sulle prove.

D'altra parte, disporre di più dati non significa necessariamente disporre di più conoscenza. La disponibilità di dati più complessi non dovrebbe farci dimenticare l'importanza che ha il programma di ricerca e l'interpretazione appropriata dei risultati. F