

Sapere cosa chiedere, sapere come chiedere

Michele Nichelatti, Maurizio Nordio, Umberto Maggiore, Maurizio Postorino, Aurelio Limido

a nome del Comitato Scientifico SIN-RIDT

KNOWING WHAT TO ASK, KNOWING HOW TO ASK

Abstract. The paper presents a brief exposition about the false and potentially dangerous results which may arise from a wrong sample choice or a biased interpretation of even a single question, in particular when the privacy of the patient is involved.

Key words: Bias, Surveys, Questions, Privacy

Conflict of interest: None.

Ricevuto: 9 Luglio 2013; Accettato: 22 Luglio 2013

I sondaggi

I sondaggi sono un diffusissimo strumento di indagine: sono utilizzati per chiedere delle informazioni al pubblico per ottenere dati sulle tipologie e sui determinanti di acquisto di un certo bene, cioè sui fattori che influenzano la decisione di acquistare un detersivo X o un'automobile Y. Sono utilizzati (spesso a sproposito) anche come messaggio pubblicitario di rinforzo, quando, per esempio, si dice che l'80% delle donne che l'hanno provato è soddisfatto del prodotto Z. Qui, se volessimo agire con correttezza, dovremmo obbligare il produttore a dichiarare la dimensione del campione, nonché la metodica di ricerca e la copia del *log* dell'analisi. Infatti, a parità di stima puntuale della percentuale, se il campione è costituito da 5 donne di cui 4 soddisfatte, la percentuale vera, con il 95% di confidenza, è inclusa tra il 28.4% e il 99.5%, mentre, se il campione è formato da 10 donne (8 soddisfatte), la percentuale vera, sempre con la medesima confidenza è inclusa tra il 44.4% e il 97.5%; se, invece, il campione è di 100 o di 1000 donne, la percentuale vera (sempre con il 95% di confidenza) è inclusa rispettivamente tra il 70.8% e l'87.3% e tra il 77.4% e l'82.4%. Diciamo che più piccolo è il campione, meno è il caso di ricorrere ad affermazioni trionfistiche relative alla soddisfazione di chi fa uso di cibi per gatti, carta igienica e balsamo per capelli con riccioli ribelli.

Il problema vero dei sondaggi non è ottenere le informazioni (visto che molta gente si sente gratificata quando le viene richiesto un parere su un certo bene di consumo), bensì ottenere delle informazioni non distorte, cioè unbiased, come dicono gli inglesi.

Il padre dei sondaggi moderni è George Gallup, uno statistico americano nato nel 1901, che, nel 1935, aveva fondato

l'American Institute of Public Opinion, che era giunto agli onori della ribalta l'anno successivo, quando era riuscito a prevedere, utilizzando un questionario cartaceo somministrato a poche migliaia di persone, la vittoria di Roosevelt nelle elezioni presidenziali (sbagliando, comunque, la previsione della percentuale dei voti che avrebbe ottenuto: 54% anziché l'effettivo 61%). La massima parte dei media, invece, aveva previsto una vittoria trionfale del candidato repubblicano, Landon, sulla base di un enorme numero di interviste telefoniche (quasi due milioni).

Che cos'era successo? Come è stato possibile che un questionario utilizzato su poche migliaia di persone avesse un'efficienza migliore di 2 milioni di telefonate nel prevedere il risultato delle elezioni presidenziali? La risposta è molto semplice: nel 1936, il telefono, negli Stati Uniti, era un bene posseduto da relativamente poche famiglie, le più benestanti in generale. Quindi, il sondaggio effettuato telefonicamente intercettava soltanto la popolazione con la fascia di reddito più elevata, per la quale non era difficile attribuire delle simpatie per il candidato repubblicano. Ma il sondaggio telefonico ignorava completamente la parte della popolazione di reddito medio-basso, che, quindi, non era stata interpellata. In pratica, il sondaggio telefonico aveva dato dei risultati, se vogliamo, attendibili, ma tali risultati erano stati ottenuti da un campione distorto, cioè non rappresentativo della popolazione, ma selezionato solamente all'interno di una fascia di reddito elevata.

Questo problema si ripropone puntualmente quando ci sono le elezioni, più o meno in tutti i Paesi del mondo: conosciamo tutti le tecniche che si utilizzano, a partire dagli *exit poll*, per finire con le proiezioni dei risultati e con le attribuzioni dei seggi in Parlamento. Il problema della bassa attendibilità degli *exit poll* è comune a gran parte dei Paesi che li uti-

lizzano: nel Regno Unito, per esempio, è noto che chi vota per il partito conservatore è più restio a dichiararlo, mentre chi vota laburista lo fa con un certo malcelato orgoglio e, quindi, i sondaggisti e gli statistici inglesi sono a conoscenza della necessità di correggere i risultati crudi ottenuti negli *exit poll* utilizzando delle tecniche che minimizzino questo errore (anzi, questa distorsione), che ha una chiara origine sociologica e psicologica.

Certo, si potrebbe dire, ma cosa c'entra tutto questo con la statistica in medicina? I sondaggi servono per capire le preferenze sugli *shampoo*, le caratteristiche che dovrebbero avere i nuovi telefonini e, persino, che cosa si chiede a un partito che si presenti alle elezioni, ma, in medicina, i sondaggi sono una cosa utilizzata molto raramente, magari giusto per chiedere il grado di soddisfazione e la qualità dei servizi percepita da un paziente che ha subito un determinato trattamento medico o chirurgico in un ospedale.

Come fare domande antipatiche

In realtà, il problema dei sondaggi, cioè il problema di ricavare le informazioni da un campione non distorto, ovvero da un campione rappresentativo dell'intera popolazione che vogliamo studiare, è un problema tangibile anche nelle sperimentazioni e nelle elaborazioni dei dati di origine biomedica.

Assumiamo di dover effettuare un'indagine esplorativa per capire quale sia la percentuale della popolazione giovanile che fa consumo di droghe, leggere o pesanti che siano, oppure che ci venga richiesta una stima del numero dei soggetti a rischio per una certa malattia trasmessa per via sessuale o che, comunque, si debba effettuare un'indagine che richiede informazioni sensibili o semplicemente personali. È ovvio che non si possa approcciare ogni individuo del campione ponendogli delle domande dirette che potrebbero essere imbarazzanti (per esempio, il numero di *partner* negli ultimi tre anni) o che potrebbero, in caso di risposta positiva, avere delle ripercussioni di carattere penale o, semplicemente, invadere la *privacy* del soggetto, per esempio, chiedendogli le sue abitudini di consumo di un'eventuale droga leggera o pesante o chiedendogli informazioni sull'uso di farmaci specifici che lo individuerebbero immediatamente come soggetto sofferente di una determinata patologia.

Come si deve fare, in questo caso? Quali sono le tecniche che si possono utilizzare per minimizzare gli errori e le distorsioni che possono derivare da risposte non esatte o volutamente errate, date a domande invadenti?

Partiamo, allora, dall'ipotesi di voler sapere quale sia il consumo di droghe fra gli studenti di una qualsiasi Università. Si dovrà partire dalla scelta del campione in modo che sia omogeneamente distribuito tra le varie facoltà (non necessariamente chi studia ingegneria deve consumare droghe con la stessa frequenza di chi studia filosofia o di chi studia medicina) e in base alla numerosità degli iscritti e poi si dovrà individuare il sistema di porgere la domanda in modo da essere certi che la risposta sia quanto più possibile sincera. A questo proposito, vale la pena di ricordare un metodo molto ingegnoso utilizzato da alcuni ricercatori americani che erano stati incaricati di stimare il consumo di droga fra le

truppe impegnate nella guerra del Vietnam e che potrebbe dimostrarsi adatto anche per il nostro problema.

Quale *escamotage* avevano ideato i ricercatori americani? Avevano preparato un sacco in cui erano state messe tre carte (concettualmente simili a quelle da gioco). La prima riportava stampata la domanda "hai mai fatto uso di droghe?", la seconda riportava la domanda "c'è un triangolo nero stampato su questa carta?", con sotto un triangolo nero, mentre la terza aveva stampata la stessa domanda della seconda, ma senza il triangolo nero sotto la domanda. In pratica ogni soldato facente parte del campione selezionato per l'indagine doveva pescare una delle tre carte e, senza essere visto, leggere la domanda stampata sulla carta e rispondere. Il metodo riusciva a garantire una buona riservatezza, perché nessuno avrebbe potuto sapere a quale delle tre domande stesse effettivamente rispondendo il soggetto. Una risposta negativa avrebbe potuto essere sia la risposta alla domanda stampata sulla prima carta sia la risposta alla domanda stampata sulla terza carta, mentre una risposta positiva avrebbe potuto essere la risposta sia alla domanda stampata sulla prima carta sia alla domanda stampata sulla seconda: nessuno avrebbe potuto saperlo con certezza.

Come funziona, allora, questo metodo? Poniamo di aver reclutato un campione di 3000 studenti e di aver utilizzato le stesse tre carte dell'indagine dell'esercito USA: in media, ci aspettiamo che circa 1000 studenti peschino la prima carta, altrettanti la seconda e altrettanti ancora la terza, per cui ci si aspetterà sicuramente di avere 1000 risposte che contengono un sì (risposta di chi ha pescato la seconda carta) e altrettante 1000 che contengono un no (risposta di chi ha pescato la terza carta). Il numero di sì in eccesso rispetto ai 1000 attesi (cioè le 1000 risposte di chi ha sollevato la seconda carta) corrisponderanno alle risposte di chi, estratta la prima carta, ammetterà di fare uso di droga e questo consentirà di stimare la percentuale di studenti che utilizzano queste sostanze. Per esempio, se, al termine dell'indagine, trovassimo 1800 risposte negative e 1200 risposte positive, potremmo concludere che 200 dei 1000 studenti che avevano pescato la prima carta facevano uso di droga e che, quindi, la percentuale è stimabile intorno al 20%. Si noti che, al denominatore, non deve comparire il numero complessivo degli studenti arruolati (cioè 3000), bensì il numero di quelli che si stima abbiano risposto alla prima domanda (cioè 1000), pertanto sarebbero circa 200 studenti su 1000 (il 20%, appunto) quelli a fare effettivamente uso di droga.

Se non avessimo utilizzato questo metodo già sperimentato negli anni '60 sui soldati americani che combattevano in Vietnam, avremmo potuto essere tentati di porre la domanda in modo anonimo, per esempio somministrando un questionario a un campione più o meno grande di studenti e poi raccogliendo i questionari compilati senza il nome del soggetto. Anche questa avrebbe potuto essere un'opzione valida, ma, probabilmente, pochi avrebbero potuto credere a un'effettiva anonimata e avrebbero potuto ripararsi dietro una risposta falsa e, quindi, questo metodo avrebbe potuto portare a risultati distorti da un numero consistente di falsi negativi.

Attenzione a come si pongono e a come ci si pone le domande

Porre le domande in modo corretto è importante quanto ottenere le risposte corrette. È, quindi, necessario, prima di raccogliere le informazioni che ci interessano, in qualsiasi modo si vengano a raccogliere (questionari scritti, colloqui, interviste telefoniche, ecc.), decidere la modalità con cui le domande saranno poste, facendo in modo che l'intervistato sia "costretto" a rispondere in modo sincero, dopo aver compreso perfettamente la domanda stessa. Questo problema è tale che, quando si raccolgono informazioni da soggetti di etnicità diversa dalla nostra, è sempre bene utilizzare un mediatore culturale, che assicuri la corretta trasmissione del senso della domanda e la corretta trasmissione della risposta. Un'altra possibile garanzia relativa alla correttezza delle risposte e alla buona interpretazione delle domande è l'introduzione nel questionario di una o più "domande di controllo", che abbiano una struttura simile ad alcune delle domande già poste, in modo tale che si possa verificare la concordanza tra le risposte stesse: una concordanza scarsa o nulla sarà indice di scarsa sincerità nella risposta.

Riassunto

Il lavoro presenta una breve discussione dei risultati falsi, e talvolta dannosi, che possono derivare da una scelta errata del campione, o dall'interpretazione distorta anche di una singola domanda, in particolare quando questa coinvolge la sfera privata e personale del soggetto.

Parole chiave: Distorsione, Sondaggi, Domande, Privacy

Dichiarazione di conflitto di interessi: Gli Autori dichiarano di non avere conflitto di interessi.

Indirizzo degli Autori:

Dr. Michele Nichelatti
Servizio di Biostatistica
Dipartimento di Ematologia e Oncologia
Ospedale Niguarda Ca' Granda
Piazza Ospedale Maggiore 3
20162 Milano
michele.nichelatti@ospedaleniguarda.it