

Application of epidemiological findings to individuals*

PAOLO BOFFETTA^{1,2}, ANDREA FARIOLI², EMANUELE RIZZELLO²¹Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York NY, USA²Department of Medical and Surgical Sciences, University of Bologna**KEY WORDS:** Epidemiology; external validity; clinical medicine**PAROLE CHIAVE:** Epidemiologia; validità esterna; medicina clinica

SUMMARY

Three types of issues need to be considered in the application of epidemiology results to individuals. First, epidemiology results are subject to random error, and can be applied only to an ideal subject with average values of all variables under study, including potential confounders included in the regression models. Second, the observational nature of epidemiology makes it susceptible to systematic error, and any extrapolation to individuals would mirror the validity of the original results. Quantitative bias analysis has been proposed to assess the likelihood, direction and magnitude of bias, but this has not yet become part of the normal practice of epidemiology. Finally, external validity of the results (i.e., their application to individuals and populations other than those included in the underlying studies) needs to be addressed, including population-based factors, such as heterogeneity in exposure or disease circumstances, and individual-based factors, such as interaction of the risk factors of interest with other determinants of the disease. Similar considerations apply to the application of results of clinical trials to individual patients, although in these studies sources of systematic error are better controlled.

RIASSUNTO

«**Applicazioni di risultati epidemiologici al singolo individuo**». Quando si attribuiscono i risultati dell'epidemiologia al singolo individuo bisogna considerare tre differenti aspetti. In primo luogo, i risultati dell'epidemiologia sono soggetti a errori casuali e possono essere applicati solo ad un soggetto ideale con valori medi per tutte le variabili studiate, compresi i potenziali fattori di confondimento inclusi nei modelli di regressione. In secondo luogo, la natura osservazionale dell'epidemiologia la rende suscettibile di errori sistematici (bias) e qualsiasi attribuzione dei risultati agli individui rispecchierebbe la validità dei risultati originali. Sono stati proposti modelli di analisi quantitativa dei bias per valutare la probabilità, la direzione e l'entità dell'errore, ma queste metodologie non sono di uso comune. Infine, deve essere considerata anche la validità esterna dei risultati (ossia la loro applicabilità a individui e popolazioni diverse da quelle incluse negli studi originali), includendo fattori caratteristici della popolazione, come l'eterogeneità delle condizioni di esposizione o di malattia, e fattori caratteristici dell'individuo, come l'interazione dei fattori di rischio di interesse con altri determinanti della malattia. Considerazioni simili si applicano all'attribuzione dei risultati da trial clinici ai singoli pazienti, sebbene in questi studi le fonti di errore sistematico siano meglio controllabili.

Received 26.11.2019 - Accepted 10.1.2020

Corresponding author: Paolo Boffetta, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1130, New York NY 10029 USA - E-mail paolo.boffetta@mssm.edu

* A preliminary version of this manuscript was presented at the 82nd Annual Meeting of the Italian Society of Occupational Medicine (SIML), Trieste, September 2019 / Una versione preliminare di questo manoscritto è stata presentata all'82° Congresso Annuale della Società Italiana di Medicina del Lavoro (SIML), Trieste, settembre 2019.

INTRODUCTION

It is customary to consider a separation between clinical medicine, aimed at improving the health of individuals through prevention or treatment, and public health, aimed at improving health at the population level. Epidemiology provides research tools to both public health and clinical medicine: the implications of epidemiology results oftentimes clearly address one of these two domains, but in numerous instances they are applicable to both. Moving from these considerations, Rogawski et al. (23) distinguish between public health epidemiology, which “*informs interventions that are applied to populations or that confer benefits beyond the individual*” and medical epidemiology, which “*informs interventions that improve the health of treated individual*”. Based on this distinction, they argue in favor of public health epidemiology, which, in their opinion, has been neglected in favor of individual-oriented approaches. We would like to opine that such dichotomy is epistemologically incorrect, and to provide a framework to apply epidemiology results to both populations and individuals. It has been a long time since Rose highlighted the link between “sick individuals and sick populations” (24).

THE NATURE OF EPIDEMIOLOGIC RESULTS

Epidemiology measures health-related conditions and events in groups of individuals, and compares them to derive inferences on possible determinants. They therefore represent averages of the likelihood that the condition or event occur (or, in case of continuous variables, that they take a particular value or range) in the different groups under study: at the individual level, the corresponding likelihood is just zero or one. For example, a measure of incidence indicates the number of individuals in which the event of interest (e.g., diagnosis of a disease) occurs over the person-time of observation: while the measure can be interpreted as a hypothetical average likelihood that the event occurred in each individual under study, the actual individual likelihood of occurrence was one for cases and zero for non-cases. Analogously, a comparative measure such a ratio of incidences between two groups would indicate the

ratio of the average likelihoods of the individuals in the two groups.

The application of group-based likelihoods, and their comparative measures, to individuals is particularly helpful to make prediction regarding individuals outside the population-time under study: in practice, we predict the risk of an individual to die over a given period of time based on the most recent mortality rates of their population, or, if these are not available, of a similar population, and we apply comparative measures to predict the risk of an individual with a given characteristics relative to their counterfactual without that characteristics.

It is important to note that these considerations apply to results of studies based on both observational (epidemiology studies) and experimental design (so called clinical trials), although there are differences in their interpretation as discussed below.

ISSUES IN THE APPLICATION OF GROUP-DERIVED MEASURES TO INDIVIDUALS

Precision

All biological variables are subject to random error, which is operationalized using probability distribution models derived from frequentist statistics: measures aimed at quantifying the variability, such as the standard error and the confidence interval, are customarily reported in clinical and epidemiological studies. The notion of random error and its quantification are familiar to most medical researchers: a simple interpretation is that the central measure of the parameter represents the value in the “average” individual or patient, and that the distribution of all individuals and patients in real life is described by the measures of variability. If the measures of interest are conditional to the distribution of other variables, as in the case of adjustment for potential confounders in stratified or regression analysis, the latter will also be considered in determining the evasive “average” subject.

In this context, a multivariable relative risk of lung cancer among smokers equal to 10 can be interpreted as the ratio of the likelihood to develop lung cancer of the “average” smoker in the study population to that of their non-smoking counter-

factual, where “average” refers not only to the variables capturing the carcinogenic effect of tobacco smoking (amount, duration, time since quitting, age at start, etc.), but also to other variables included in the regression models. Such ideal average individual, and their counterfactual, are useful simplifications to explain the implications of group-based results (in the example above, “Our study shows that the risk of lung cancer of a smoker is 10-times higher than that of a non-smoker”); these results, however, cannot be applied with certainty to any real individual.

Internal validity

Random error is not the only factor complicating the application of group-derived results of epidemiology studies to individuals. Well-designed and conducted epidemiologic studies provide the best risk estimates when experimental approaches are not applicable. The observational nature of epidemiologic research, however, makes it susceptible to systematic error. Complete control of bias and confounding can seldom be achieved due to:

- residual and unmeasured confounding;
- selection and information bias;
- publication bias.

Although the effect of known and measurable confounders can be controlled – at least in part – by including appropriate terms in regression models, control of bias requires appropriate provisions in the design, conduct and analysis of the study (25). In addition, quantitative bias analysis has been increasingly used to assess the possible effect of selected sources of bias (14). This represents a formal approach to provide a quantitative estimate of the likelihood, direction, and magnitude of the error introduced by one or multiple sources of bias. Several types of quantitative bias analysis have been described, depending on whether one or multiple types of bias are addressed, and whether a fixed value or a range of values are assigned to the bias parameter (15). Steps in the bias analysis include (i) to identify potential source of bias, (ii) to identify sources of information on bias parameters, (iii) to derive alternative values to the original study variables, and (iv) to quantify their effect on the origi-

nal results. Recommendations have been developed, that quantitative bias analysis should accompany any presentation of results of observational studies (7), however, most investigators simply ignore them, resulting in an unknown amount of bias affecting results of such studies.

External validity

External validity concerns the applicability of the results of a study to a population other than that under study. It is also referred to as ‘generalizability’ of the results. Lack of external validity does not reduce the ability of a study to contribute to causal inference, and failure to recognize this fact is one of the most common mistakes in the interpretation of clinical and epidemiological studies. However, external validity becomes an important issue in the context of use of group-based results to individuals.

The considerations made above on the need to identify the “average” study subject to account for random error, and to control sources of bias to generate valid results apply only to the populations from which the results were generated. Any application of results of epidemiological or clinical research to individuals outside the populations under study should address factors that may differ between the two. These factors can be operationally divided in two groups.

The first group comprises factors external to the individuals to whom results are to be applied. Differences in exposure circumstances is one such factor. Table 1 illustrates this phenomenon in the case of the lower risk of lung cancer from tobacco smok-

Table 1 - Levels of cotinine and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL) in urine samples of smoking lung cancer cases from Shanghai and Singapore (28)

Tabella 1 - Livelli di cotinina e di 4-(metilnitrosamino)-1-(3-piridil)-1-butanolo (NNAL) in campioni di urine di soggetti fumatori con tumore del polmone urine nelle città di Shanghai e Singapore (28)

	Shanghai (N=155)	Singapore (N=91)
Cotinine (ng/mg creat.)	3,033	2,873
NNAL (pmol/mg creat.)	0.23	0.89

ing in Chinese smokers compared to European and American smokers (13). In a series of elegant studies conducted in two populations of Chinese smokers from Shanghai and Singapore, Yuan et al. (28) showed that the lower risk was likely due to the characteristics of the cigarettes consumed in China vs. Singapore: although the level of urinary cotinine (a marker of amount of tobacco smoking) was comparable in the two groups, the levels of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), a tobacco-specific nitrosamine and a markers of carcinogenicity of tobacco smoke, was significantly lower in smokers from Shanghai, likely due to differences in curing and manufacturing processes of traditional local Chinese cigarettes smoked in Shanghai compared to standard industrial cigarettes smoked in Singapore.

Most occupational epidemiological studies are of retrospective design, and address the health effects of exposure that occurred in the past. In many industries there have been changes in technology and industrial hygiene, which have resulted in important changes in exposure circumstances and levels. Although in some cases detailed dose-response and other data are available, that help transferring results of these studies to other populations of workers in the same industries or jobs, in many instances these data are not available. Use of results of studies conducted in other countries, where technological and industrial hygiene conditions might differ, is another potential source of lack of generalizability.

Another factor is the need to consider absolute, rather than relative measures of occurrence. Exposure to a risk factor increases the number of cases of the disease of interest in the population, i.e., its incidence; however, the relative measure of incidence depends also on the incidence in the unexposed group. In the absence of effect modification, the relative measure of incidence will therefore be lower in a population with higher incidence in the unexposed compared to a population with lower incidence, as shown in table 2. A well-known example of this phenomenon is the apparent stronger association between tobacco smoking and lung cancer in women compared to men: although various explanations have been proposed, such as a role of hormonal factors (21), the most likely explanation is

the higher rate of lung cancer among men for causes other than tobacco smoking (e.g., occupational exposures): a reduction of the role of these other factors in men explain why the gender gap in relative risks observed in the past has disappeared in recent studies (22). The presence of interaction between the risk factor of interest and the determinants of the incidence among the unexposed would further affect the relative risk.

Characteristics of the individuals represent the second group of factors that may affect external validity. The simplest form they can take is that of modifiers of the effect of the exposure of interest (interaction), which is presented (in the form of a positive interaction) in table 3. The incidence among those unexposed to either factor is 10/1000. In the absence of factor B, exposure to factor A increases the incidence by 10/1000; in the absence of factor A, exposure to factor B increases the incidence by 20/1000. In the absence of interaction, the incidence among those exposed to both factors should

Table 2 - Effect of incidence in unexposed on the relative risk – Hypothetical example of two populations with 1000 exposed and 1000 unexposed subjects each, and higher incidence in the unexposed in one population. The incidence due to the exposure is set to 40/1000 in both populations

Tabella 2 - Effetto di incidenza tra i non esposti sul rischio relativo – esempio ipotetico di due popolazioni di soggetti con 1000 esposti e 1000 non esposti con una maggiore incidenza nei non esposti in una delle popolazioni. L'incidenza dovuta all'esposizione è fissata a 40/1000 in entrambe le popolazioni

	Population 1	Population 2
Incidence rate in unexposed	10/1000	20/1000
Incidence rate in exposed	50/1000	60/1000
Rate ratio	5	3

Table 3 - Hypothetical example of positive interaction between two risk factors on the incidence of a disease

Tabella 3 - Esempio ipotetico di interazione positiva tra due fattori di rischio sull'incidenza della malattia

	Incidence of the disease	
	Unexposed to A	Exposed to A
Unexposed to B	10/1000	20/1000
Exposed to B	30/1000	50/1000

be $(10+10+20) = 40/1000$; in the example in the table, the incidence among those exposed to both is $50/1000$, suggesting a positive interaction between the two exposures (for sake of simplicity no consideration is given to the statistical significance of the interaction term). When rate ratios are used instead of incidence rates, and the group unexposed to both factors is taken as reference, the interaction is described by the formula:

$$RR_{ab} \neq RR_a + RR_b - 1.$$

Interaction is conceptually similar to the problem of difference in background incidence across populations described above; however, it applies to the characteristics of the individuals, irrespective of the distribution of the two risk factors in the population. Several examples of interaction have been identified among causes of chronic diseases, both genetic and environmental (in broad sense). Although their effect on the risk of disease at the individual level can in principle be accounted for, a precise estimate of their magnitude is available only for a fraction of them, such tobacco smoking and asbestos for lung cancer (20) and tobacco smoking and alcohol drinking for head and neck cancer (9).

ADEQUACY OF STATISTICAL MODELS AIMED AT MEASURING ASSOCIATIONS

In current epidemiological practice, observational studies are designed to identify or confirm an association between an exposure and the occurrence of a certain disease. The focus is posed on the measures of association (such as relative risk, when measuring on the multiplicative scale, or risk difference, when quantifying the absolute risk); little attention is devoted to the overall performance of the regression model. Sometimes researchers present direct comparison between two or more regression models by applying statistical testing (e.g. maximum likelihood ratio test (11)) or information criteria which are largely based on the likelihood function of the model (e.g. Bayesian Information Criterion or Akaike Information Criterion) (2, 3, 26). Of note, such comparisons inform whether a specific model adapts to sample data better than a few others; how-

ever, they do not convey information on the absolute goodness-of-fit of the models. A perhaps even worse practice is testing the goodness-of-fit (for instance, through the Hosmer-Lemeshow goodness-of-fit test (10)) and interpreting a low *p-value* as an indication that the model is performing well; in fact, these tests only inform that introducing a specific variable in the regression model contributes to improving the goodness-of-fit, but do not provide a meaningful measure of the overall performance of the regression model.

People with a quantitative background (e.g., industrial hygienists) should be familiar with measuring and reporting the proportion of the variance in the dependent variable (disease status) that is predictable from the independent variable (exposure); in the context of simple linear regression, this can be achieved by the coefficient of determination (usually reported as R^2). Outside linear regression, similar measures have been proposed, like the McFadden pseudo- R^2 for logistic regression (17, 18). This index assumes a value of 0 in the empty model (no predictive value) and a value of 1 in case of perfect prediction. A conceptually similar index is the Harrell's C index of concordance estimated after fitting a Cox proportional hazards regression models (8). Describing the properties and the (several) limitations of these indices goes beyond the intents of this paper. However, a consideration is worthwhile: how often is the reader of an epidemiological paper informed about the absolute goodness of fit of a regression model whose results are reported in the classic form of one or several relative risks (and corresponding confidence intervals and *p-values*)? It has been shown that in most observational studies the absolute goodness of fit of regression models is usually rather low (e.g. odds ratios from case-control studies with a McFadden pseudo- R^2 not higher than 0.3) (19). This circumstance might not a limitation if the purpose of the analysis is to demonstrate the effect of a certain exposure in increasing (or decreasing) the risk relative or absolute risk of a specific condition; indeed, estimates of association will be valid as far as bias, including confounding, can be excluded (see above), independent from the overall goodness of fit of the regression model. Conversely, knowledge of the overall model performance

measured through an absolute goodness-of-fit index is fundamental if the goal is to answer to the following questions:

- 1) Did a specific subject in the study population develop the condition under investigation due to a specific exposure (*in-sample prediction*)?
- 2) Will a specific subject develop the condition under investigation, and when (*out-of-sample prediction*)?

To be answered, these questions need an extremely high predictive value from the underlying regression models. Outside the clinical context (e.g. prediction of tumor response based on treatment protocols), this condition is seldom achieved. A worthwhile example is the calculation of the risk of cardiovascular events based on the few strong determinants highly prevalent in the general population. The most known example is the so called “Framingham score”, which consists in a series of formulas derived from Cox proportional hazards regression models applied to a prospective population-based cohort study (4). The authors were able to adapt models with an Harrel’s $C > 0.7$ – a conventional, somehow questionable, threshold that identifies models with good predictive value – based on a few variables: gender (the models were actually gender-specific), age, diabetes status, tobacco smoking, treated and untreated systolic blood pressure, total and high-density lipoprotein cholesterol (or body mass index, as a surrogate measure). Knowledge of these few data is used in current clinical practice to predict the 10-year risk of cardiovascular disease. However, a large body of literature suggested that the external validity of the formula might be limited (e.g. (27)); in particular, an overestimation of the risk has been observed in certain populations (5). This could occur because of improvements in the treatment and control of predisposing conditions (such as hypertension and diabetes) or due to a different baseline risk determined by lifestyle (including diet) and genetic factors.

In synthesis, the use of estimates from observational studies to predict individual events is a complex process often hampered by the lack of fundamental knowledge on the disease process and, hence, a limited predictive value of the multivariable regression models used to generate the results.

CONSIDERATIONS ABOUT CLINICAL TRIALS

The above discussion was formulated with respect to observational research. One can argue that these considerations do not apply to experimental studies, in which the determinant under investigation (exposure) is assigned to study subjects. In this respect, the results of trials, and in particular clinical trials, are directly applicable to individual patients with the same conditions as those included in the trials. After all, when clinicians prescribe a new drug to their patients based on the results of a trial, they do so because they expect in the patients the same effect shown in the trial.

If clinical and other medical trials are well designed and executed, they can prevent bias from affecting their results. However, the other two sources of error in applying results from populations to individuals, that were described above for observational studies, also apply to trials. Results of trials are affected by random error, and their results would precisely apply only to a hypothetical “average” patient. In practice, the clinicians mentioned in the previous paragraph would not be so naïve to expect in each of their patients exactly the result reported in the trial: they prescribe the new drug with the expectation to see in their patients, on average, the effect observed in the trial, but they recognize that there might be plenty of individual variation in the response.

More important, however, is the issue of external validity of results of experimental studies. The problem that trials, in particular treatment trials, include selected samples of patients who might in principle benefit from the treatments under has been increasingly recognized in the medical literature, in particular with respect to sociodemographic characteristics such as age (e.g., underrepresentation of elderly patients in clinical trials (16)) and race/ethnicity (e.g., overrepresentation of non-Hispanic Whites (6)).

CONCLUSIONS

Considerations about the applicability of results of epidemiology studies to individuals are analogous to those developed within the framework of personalized medicine. The goal of personalize medicine

is to describe all individual characteristics that determine the response of the individual patient to a given treatment, and select the most effective one (1). An analogous approach can be invoked for epidemiology, although issues of internal validity would complicate the process, as discussed above. Although an exhaustive description of all relevant individual factors remains elusive, steps can be taken in this direction.

Systematic reviews, meta-analyses and umbrella reviews (12) help improving the precision of risk estimates and offer opportunity for stratified analysis to address sources of heterogeneity of results across populations. Routine application of quantitative bias analysis (15), as discussed above, would improve the validity of inferences at the individual level. Integration of biology and epidemiology would contribute to reducing uncertainties on the external validity of the results.

In conclusion, epidemiology results can be applied to individuals under the stringent framework we outlined here. As in most instances sources of random error, internal validity, and external validity are only partially controlled, extrapolation to individuals remains tentative at best. One case in which extrapolation to individuals may be justified is that of high-penetrance susceptibility genes: in which results of clinical or epidemiological have shown such a high risk in carriers that consideration about random and systematic have less relevance, and it may be justified to assume external validity even in the absence of direct evidence supporting it.

NO POTENTIAL CONFLICT OF INTEREST RELEVANT TO THIS ARTICLE WAS REPORTED BY THE AUTHORS

PB is an associate editor of the journal, but this article was reviewed by an anonymous reviewer, who provided useful suggestions to improve it

REFERENCES

1. Academy of Medical Sciences: Stratified, Personalised or P4 Medicine: A New Direction for Placing the Patient at the Centre of Healthcare and Health Education (Technical Report). London: Academy of Medical Sciences, 2015
2. Akaike H: A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974; 19: 716-723
3. Clayton D, Hills M: *Statistical Models in Epidemiology*. Oxford, UK: Oxford University Press, 1993
4. D'Agostino RB, Vasan RS, Pencina MJ, et al: General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008; 117: 743-753
5. Damen JA, Pajouheshnia R, Heus P, et al: Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. *BMC Med*. 2019; 17: 109
6. Ford JG, H M, Lai GY, et al: Barriers to recruiting underrepresented population to cancer clinical trials: a systematic review. *Cancer* 2008; 112: 228-242
7. Fox MP, Lash TL: On the need for quantitative bias analysis in the peer-review process. *Am J Epidemiol* 2017; 185: 865-868
8. Harrell FE, Califf RM, Pryor DB, et al: Evaluating the yield of medical tests. *JAMA* 1982; 247: 2543-2546
9. Hashibe M, Brennan P, Chuang SC, et al: Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol Biomarkers Prev* 2009; 18: 541-550
10. Hosmer DW Jr, Lemeshow SA: Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics—Theory and Methods* 1980; 9: 1043-1069
11. Hosmer DW Jr, Lemeshow SA, Sturdivant RX: *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley, 2013
12. Ioannidis JP: Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *CMAJ* 2009; 181: 488-493
13. Jung KJ, Jeon C, Jee SH: The effect of smoking on lung cancer: ethnic differences and the smoking paradox. *Epidemiol Health* 2016; 38: e2016060
14. Lash TL, Fox MP, Fink AK: *Applying quantitative bias analysis to epidemiologic data*. New York, NY: Springer, 2009
15. Lash TL, Fox MP, MacLehose RF, et al: Good practices for quantitative bias analysis. *Int J Epidemiol* 2014; 43: 1969-1985
16. Le Saux O, Falandry C, Gan HK, et al: Inclusion of elderly patients in oncology clinical trials. *Ann Oncol* 2016; 27: 1799-1804
17. Lemeshow S, Hosmer DW: A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982; 115: 92-106
18. Mc Fadden D: Conditional logit analysis of qualitative choice behavior. In: Zarembka P, Ed. *Frontiers in Econometrics*. Cambridge, MA: Academic Press, 1974, p. 105-142

19. Menard S: Coefficients of determination for multiple logistic regression analysis. *Am Stat* 2000; 54: 17-24
20. Ngamwong Y, Tangamornsuksan W, Lohitnavy O, et al: Additive synergism between asbestos and smoking in lung cancer risk: A systematic review and meta-analysis. *PLoS One* 2015; 10: e0135798
21. O’Keeffe LM, Taylor G, Huxley RR, et al: Smoking as a risk factor for lung cancer in women and men: a systematic review and meta-analysis. *BMJ Open* 2018; 8: e021611
22. Pauk N, Kubík A, Zatloukal P, Krepela E: Lung cancer in women. *Lung Cancer* 2005; 48: 1-9
23. Rogawski ET, Gray CL, Poole C: An argument for renewed focus on epidemiology for public health. *Ann Epidemiol* 2016; 26: 729-733
24. Rose G: Sick individuals and sick populations. *Int J Epidemiol* 1985; 14: 32-38
25. Rothman KJ, Greenland S, Lash TL: Design strategies to improve study accuracy. In: Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*, Third Ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2008, pp. 168-182
26. Schwarz G: Estimating the dimension of a model. *Annals of Statistics* 1978; 6: 461-464
27. Siontis GC, Tzoulaki I, Siontis KC, Ioannidis JP: Comparisons of established risk prediction models for cardiovascular disease: systematic review. *BMJ* 2012; 344: e3318
28. Yuan JM, Koh WP, Murphy SE, et al : Urinary levels of tobacco-specific nitrosamine metabolites in relation to lung cancer development in two prospective cohorts of cigarette smokers. *Cancer Res* 2009; 69: 2990-2995

Applicazione di risultati epidemiologici al singolo individuo

INTRODUZIONE

È consuetudine considerare come separate la medicina clinica, volta a migliorare la salute delle persone attraverso programmi di prevenzione o terapeutici, e la salute pubblica, che ha lo scopo di migliorare la salute a livello di popolazione. L’epidemiologia fornisce strumenti di ricerca utili sia in sanità pubblica che in clinica: le implicazioni dei risultati epidemiologici spesso sono attribuibili ad uno di questi due settori, ma in numerosi casi sono applicabili ad entrambi. Rogawski et al. (23) fanno una distinzione tra epidemiologia della salute pubblica, che «affronta gli interventi che vengono applicati alle popolazioni o che conferiscono benefici oltre l’individuo» e l’epidemiologia medica, che «affronta gli interventi che migliorano la salute del singolo individuo». Sulla base di questa distinzione, questi autori sostengono l’epidemiologia della salute pubblica, che, a loro avviso, è stata trascurata per concentrarsi su approcci orientati all’individuo. È nostra intenzione affermare che tale dicotomia è epistemologicamente scorretta e fornire una metodologia per applicare i risultati epidemiologici sia alle popolazioni che agli individui. Dopotutto, è passato molto tempo da quando Rose mise in evidenza il legame tra «*sick individuals and sick populations*» (24).

LA NATURA DEI RISULTATI EPIDEMIOLOGICI

L’epidemiologia misura condizioni e eventi relativi alla salute di gruppi di individui e li confronta per ricavare “inferenze” su possibili determinanti. Questi risultati rappresentano quindi una probabilità che si verifichi la condizione o l’evento (o, nel caso di variabili continue, che assumano un determinato valore o un intervallo di valori) nei diversi gruppi oggetto dello studio: a livello individuale, la probabilità può essere solo zero o uno. Ad esempio, una misura di incidenza indica il numero di individui in cui si verifica l’evento di interesse (es. diagnosi di una malattia) nel tempo di osservazione della persona: mentre la misura può essere interpretata come un’ipotetica probabilità media che l’evento si sia verificato in ogni individuo in studio, l’effettiva probabilità individuale di occorrenza può essere unicamente uno per i casi e zero per i non casi. Analogamente, una misura comparativa di un tale rapporto di incidenze tra due gruppi indicherebbe il rapporto delle probabilità medie degli individui nei due gruppi.

L’applicazione ai singoli individui delle probabilità basate sul gruppo e delle loro misure di associazione è particolarmente utile per fare previsioni su individui al di fuori del periodo di studio della popolazione: in pratica, prevediamo

il rischio che un individuo muoia in un determinato periodo di tempo in base ai tassi di mortalità più recenti della popolazione di appartenenza o, se non sono disponibili, di una popolazione simile e applichiamo misure di associazione per prevedere il rischio di un individuo con una data caratteristica rispetto al suo controfattuale senza quelle caratteristiche.

È importante notare che queste considerazioni si applicano a risultati di studi basati sia su disegni osservazionali (studi epidemiologici) sia sperimentali (trial clinici), sebbene vi siano delle differenze nell'interpretazione come riportato di seguito.

PROBLEMI DERIVANTI DALLE APPLICAZIONI DI MISURE BASATE SU UN GRUPPO AD UN INDIVIDUO

Precisione

Tutte le variabili biologiche sono soggette ad un errore casuale, che è valutabile utilizzando modelli di distribuzione della probabilità derivante dalla statistica frequentista: le misure con lo scopo di quantificare la variabilità, come l'errore standard e gli intervalli di confidenza, sono comunemente riportate sia negli studi clinici che negli studi epidemiologici. Il concetto di errore casuale e la sua quantificazione sono familiari alla maggior parte dei ricercatori: un'interpretazione semplicistica consiste nel ritenere che la misura centrale di un parametro rappresenti il valore della media degli individui o dei pazienti e che la distribuzione di tutti gli individui e i pazienti nella vita reale sia descritta da misure di variabilità. Se le misure di interesse sono condizionate dalla distribuzione di altre variabili, come nel caso dell'aggiustamento per confondenti potenziali in analisi stratificate o di regressione, anche queste saranno prese in considerazione nel determinare il soggetto "medio".

In questo contesto, un rischio relativo multivariabile di sviluppare un carcinoma polmonare uguale a 10 tra i fumatori può essere interpretato come il rapporto tra la probabilità del fumatore "medio" di sviluppare il carcinoma polmonare nella popolazione oggetto di studio e quella del suo controfattuale non fumatore; il termine "medio" si riferisce non solo alle variabili che catturano l'effetto cancerogeno del fumo di sigaretta (quantità, durata, tempo dalla cessazione, età all'inizio, ecc.), ma anche altre variabili incluse nei modelli di regressione. Tale individuo medio ideale, e il suo termine di paragone, sono delle utili semplificazioni per spiegare le implicazioni dei risultati all'interno di un gruppo (nell'esempio sopra, "Il nostro studio mostra che il rischio di cancro ai polmoni di un fumatore è 10 volte superiore a quello di un non-fumatore"); questi risultati, tuttavia, non possono essere applicati con certezza ad individuo reale.

Validità interna

L'errore casuale non è l'unico fattore che complica l'attribuzione di risultati di studi epidemiologici derivati da gruppi agli individui. Quando gli approcci sperimentali non sono possibili, sono gli studi epidemiologici ben progettati a fornire le migliori stime di rischio. La natura osservazionale della ricerca epidemiologica, tuttavia, la rende suscettibile anche agli errori sistematici. Raramente è possibile ottenere un controllo completo dei *bias* e del confondimento a causa di:

- confondimento residuo e non misurato;
- *bias* di selezione e/o di informazione;
- *bias* di pubblicazione.

Sebbene l'effetto dei confondenti noti e misurabili possa essere controllato - almeno in parte - includendo termini appropriati nei modelli di regressione, il controllo dei *bias* richiede appropriate disposizioni nel disegno, conduzione e analisi dello studio (25). Inoltre, l'analisi quantitativa dei *bias* è stata via via più utilizzata per valutare il possibile effetto di determinate fonti di *bias* (14). Ciò rappresenta un approccio formale per fornire una stima quantitativa della probabilità, della direzione e dell'entità dell'errore introdotto da una o più fonti di *bias*. Sono stati descritti diversi tipi di analisi quantitative di *bias*, a seconda se uno o più tipi di *bias* siano attribuibili e se si possa assegnare al *bias* un valore puntuale o un intervallo di valori (15).

I vari passaggi per condurre un'analisi quantitativa dei *bias* sono i seguenti (i) identificare la potenziale fonte di *bias*, (ii) identificare fonti di informazione sui parametri di *bias*, (iii) derivare valori alternativi alle variabili di studio originali e (iv) quantificare il loro effetto sui risultati originali. Recenti raccomandazioni suggeriscono che l'analisi quantitativa dei *bias* dovrebbe essere sempre effettuata nella presentazione dei risultati degli studi osservazionali (7); tuttavia, la maggior parte dei ricercatori non quantifica i *bias*, e il loro effetto sui risultati di tali studi rimane sconosciuto.

Validità esterna

La validità esterna riguarda l'applicabilità dei risultati di uno studio a una popolazione diversa da quella originariamente studiata. Questo processo viene anche definito "generalizzazione" dei risultati. La mancanza di validità esterna non riduce la capacità di uno studio di contribuire all'inferenza causale e l'incapacità di riconoscere questo fatto è uno degli errori più comuni nell'interpretazione degli studi clinici ed epidemiologici. Tuttavia, la validità esterna diventa importante nel contesto della traslazione dei risultati basati sul gruppo agli individui.

Le considerazioni espone precedentemente sulla necessità di identificare il soggetto "medio" tenuto conto dell'errore

casuale e di controllare le fonti di *bias* per generare risultati validi si applicano solo alle popolazioni da cui sono scaturiti i risultati. Qualsiasi applicazione dei risultati della ricerca epidemiologica o clinica a soggetti esterni alle popolazioni oggetto di studio dovrebbe tener conto dei fattori che possono differire. Questi fattori possono essere divisi operativamente in due gruppi.

Il primo gruppo comprende fattori esterni agli individui ai quali devono essere applicati i risultati. Le differenze nell'esposizione sono uno di questi fattori. La Tabella 1 illustra questo fenomeno nel caso di un minor rischio di cancro ai polmoni dovuto al fumo di sigaretta nei fumatori cinesi rispetto ai fumatori europei e americani (13). In una serie di interessanti studi condotti in due popolazioni di fumatori asiatici, a Shanghai e Singapore, Yuan et al. (28) hanno dimostrato che il rischio più basso era probabilmente dovuto alle caratteristiche delle sigarette utilizzate in Cina rispetto a quelle fumate a Singapore: sebbene il livello di cotinina urinaria (indicatore della quantità di fumo di sigaretta) fosse comparabile nei due gruppi, i livelli di 4 - (metilnitrosamino) -1- (3-piridil) -1-butanolo (NNAL), una nitrosamina specifica per il fumo di sigaretta e un marcatore di cancerogenicità, era significativamente più bassa nei fumatori di Shanghai, probabilmente a causa delle differenze nei processi di trattamento del tabacco e di produzione di sigarette cinesi tradizionali fumate a Shanghai rispetto alle sigarette industriali standard fumate a Singapore.

La maggior parte degli studi epidemiologici occupazionali sono studi retrospettivi e attribuiscono alla salute gli effetti di una determinata esposizione avvenuta nel passato. In molti settori produttivi ci sono stati cambiamenti, sia nella tecnologia che nell'igiene industriale, che hanno portato a variazioni importanti nelle modalità e nei livelli di esposizione. Sebbene in alcuni casi siano disponibili dati dettagliati sull'effetto dose-risposta o altri dati utili, che aiutano ad attribuire i risultati di questi studi ad altre popolazioni di lavoratori negli stessi settori o lavori, in molti casi questi dati non sono disponibili. L'uso dei risultati di studi condotti in altri Paesi, in cui le condizioni igieniche tecnologiche e industriali potrebbero differire, è un altro potenziale motivo di mancanza di validità esterna.

Un altro fattore è la necessità di considerare la probabilità assoluta piuttosto che relativa che si verifichi un evento. L'esposizione a un fattore di rischio aumenta il numero di casi della malattia di interesse nella popolazione, cioè la sua incidenza; tuttavia, la misura relativa dell'incidenza dipende anche dall'incidenza nel gruppo dei non esposti. In assenza di modifiche dell'effetto, la misura relativa dell'incidenza sarà quindi più bassa in una popolazione con un'incidenza nei non esposti più elevata rispetto a una popolazione con incidenza più bassa, come mostrato nella Tabella 2. Un esempio ben noto di questo fenomeno è l'apparente associazione più

forte tra fumo di sigaretta e cancro ai polmoni nelle donne rispetto agli uomini: sebbene siano state proposte varie spiegazioni, tra cui il ruolo di fattori ormonali (21), la spiegazione più probabile è rappresentata dal fatto che il tasso più elevato di cancro ai polmoni tra gli uomini sia dovuto a cause diverse dal fumo (es. esposizioni professionali): una riduzione del ruolo di questi ulteriori fattori negli uomini spiega perché il divario di genere nei rischi relativi osservati in passato è scomparso in studi recenti (22). La presenza di interazione tra il fattore di rischio di interesse e i determinanti dell'incidenza tra i non esposti influirebbe ulteriormente sul rischio relativo.

Le caratteristiche degli individui rappresentano il secondo gruppo di fattori che influenza la validità esterna. La forma più semplice che possono assumere è quella dei modificatori d'effetto dell'esposizione di interesse (interazione), che è rappresentato nella Tabella 3 (secondo la formula di interazione positiva). Nell'esempio proposto, l'incidenza tra coloro che non sono esposti a nessuno dei due fattori è 10/1000. In assenza del fattore B, l'esposizione al fattore A aumenta l'incidenza di 10/1000; in assenza del fattore A, l'esposizione al fattore B aumenta l'incidenza di 20/1000. In assenza di interazione, l'incidenza tra coloro che sono esposti a entrambi i fattori dovrebbe essere $(10 + 10 + 20) = 40/1000$; nell'esempio nella tabella, l'incidenza tra le persone esposte ad entrambe è 50/1000, il che suggerisce un'interazione positiva tra le due esposizioni (per semplicità non viene presa in considerazione la significatività statistica del termine di interazione). Quando i rapporti di tasso vengono utilizzati al posto dei tassi di incidenza e il gruppo non esposto a entrambi i fattori viene preso come riferimento, l'interazione è descritta dalla formula:

$$RR_{ab} \neq RR_a + RR_b - 1.$$

L'interazione è concettualmente simile al problema della differenza nell'incidenza di fondo tra le popolazioni sopra descritte; tuttavia, si applica alle caratteristiche degli individui, indipendentemente dalla distribuzione dei due fattori di rischio nella popolazione. Numerosi esempi di interazione sono stati identificati tra le cause delle malattie croniche, sia genetiche che ambientali (in senso lato). Sebbene i loro effetti sul rischio di sviluppare la malattia a livello individuale possano, in linea di principio, essere giustificati, una stima precisa della loro entità è disponibile solo per una parte limitatissima di essi, come il fumo di sigaretta e l'amianto per il cancro ai polmoni (20) e il fumo di sigaretta e il consumo di alcool per i tumori del distretto testa/collo (9).

ADEGUATEZZA DEI MODELLI STATISTICI VOLTI A MISURARE LE ASSOCIAZIONI

Nella pratica epidemiologica attuale, gli studi osservazionali sono prevalentemente progettati per identificare o confermare un'associazione tra un'esposizione e l'insorgenza di una determinata malattia. L'attenzione è posta sulle misure di associazione (come il rischio relativo, quando si misura sulla scala moltiplicativa, o la differenza di rischio, quando si quantifica il rischio assoluto); si pone meno attenzione alle prestazioni complessive del modello di regressione. A volte i ricercatori presentano un confronto diretto tra due o più modelli di regressione applicando test statistici (ad es. test del rapporto di massima verosimiglianza (11)) o criteri di informazione che sono in gran parte basati sulla funzione di massima verosimiglianza del modello (ad es. Criteri di informazione Bayesiana e Criteri di Informazione di Akaike) (2, 3, 26). Da notare che tali confronti ci informano se uno specifico modello si adatta ai dati osservati meglio di un altro; tuttavia, non forniscono informazioni sull'assoluta bontà di adattamento dei modelli. Probabilmente una pratica ancora peggiore è testare la bontà di adattamento (ad esempio, attraverso il test di Hosmer-Lemeshow (10)) e interpretare un basso valore di p come indicazione di una buona capacità del modello di adattarsi ai dati; in effetti, questi test evidenziano solo il fatto che l'introduzione di una variabile specifica nel modello di regressione contribuisce a migliorare la bontà di adattamento, ma non sono in grado di fornire una misura delle prestazioni complessive del modello di regressione.

Tecnici e scienziati in possesso di buone conoscenze nell'ambito dell'analisi quantitativa (ad esempio, igienisti industriali) dovrebbero avere familiarità nel misurare e riportare la percentuale della varianza nella variabile dipendente (es. stato della malattia) che è prevedibile dalla variabile indipendente (esposizione); nel contesto della regressione lineare semplice, ciò può essere ottenuto mediante il coefficiente di determinazione (generalmente riportato come R^2). Al di fuori della regressione lineare, sono state proposte misure simili, come la pseudo- R^2 di McFadden per la regressione logistica (17, 18). Questo indice assume un valore 0 nel modello vuoto (nessun valore predittivo) e un valore 1 in caso di previsione perfetta. Un indice concettualmente simile è l'indice di concordanza C di Harrel, stimato dopo aver adottato i modelli di regressione per rischi proporzionali di Cox (8). Descrivere le proprietà e le (diverse) limitazioni di questi indici va oltre le intenzioni di questo documento. Tuttavia, vale la pena porsi il seguente quesito: quanto spesso il lettore di un studio epidemiologico viene informato dell'assoluta bontà di adattamento di un modello di regressione i cui risultati sono riportati nella forma classica di uno o più rischi relativi (e corrispondenti intervalli di confidenza e valori di p)?

È stato osservato che nella maggior parte degli studi osservazionali l'assoluta bontà di adattamento dei modelli di regressione è generalmente piuttosto bassa (ad es. *odds ratio* da studi caso-controllo con un pseudo- R^2 di McFadden non superiore a 0,3) (19). Questa circostanza può non essere una limitazione se lo scopo dell'analisi è dimostrare che l'effetto di una determinata esposizione aumenta (o diminuisce) il rischio relativo o assoluto di una condizione specifica; in effetti, le stime dell'associazione saranno valide tenendo in considerazione i *bias*, incluso il confondimento (vedi sopra), indipendentemente dalla bontà complessiva di adattamento del modello di regressione. Al contrario, la conoscenza delle prestazioni complessive del modello misurate attraverso un indice assoluto di bontà di adattamento è fondamentale se l'obiettivo è rispondere alle seguenti domande:

1. un soggetto specifico della popolazione in studio ha sviluppato la condizione in esame a causa di un'esposizione specifica (previsione *in-sample*)?
2. un soggetto specifico svilupperà la condizione in esame e quando la svilupperà (previsione *out of sample*)?

Per rispondere a queste domande è necessario un valore predittivo estremamente elevato dai modelli di regressione sottostanti. Al di fuori del contesto clinico (ad es. previsione della risposta del tumore in base ai protocolli di trattamento), questa condizione viene raramente raggiunta. Un esempio utile è il calcolo del rischio di eventi cardiovascolari maggiori basato sui pochi forti determinanti ampiamente prevalenti nella popolazione generale. L'esempio più noto è il cosiddetto "indice di Framingham", che consiste in una serie di formule derivate da modelli di regressione di Cox applicati a uno studio di popolazione prospettico (4). Gli autori sono stati in grado di adattare modelli con una C di Harrel > 0.7 - una soglia convenzionale, in qualche modo discutibile, che identifica i modelli con un buon valore predittivo - sulla base di alcune variabili: genere (i modelli erano in realtà specifici per genere), età, presenza di diabete, fumo di sigaretta, ipertensione arteriosa trattata e non trattata, colesterolo totale e HDL (o indice di massa corporea, come misura surrogata). La conoscenza di questi pochi dati viene utilizzata, nell'attuale pratica clinica, per prevedere il rischio a dieci anni di malattie cardiovascolari. Tuttavia, un'ampia letteratura ha suggerito che la validità esterna della formula potrebbe essere limitata (ad esempio (27)); in particolare, in alcune popolazioni è stata osservata una sovrastima del rischio (5). Ciò potrebbe verificarsi per merito di miglioramenti nel trattamento e nel controllo delle condizioni predisponenti (come ipertensione e diabete) o a causa di un diverso rischio di base determinato dallo stile di vita (compresa la dieta) e da fattori genetici.

In sintesi, l'uso di stime tratte da studi osservazionali per prevedere singoli eventi è un processo complesso spesso ostacolato dalla mancanza di conoscenze fondamentali sulla

storia naturale della malattia e, quindi, da un valore predittivo limitato dei modelli di regressione multivariabile utilizzati per generare i risultati.

TRIAL CLINICI: CONSIDERAZIONI SPECIFICHE

Le informazioni riportate precedentemente sono relative all'epidemiologia osservazionale.

Si potrebbe sostenere che queste considerazioni non siano applicabili agli studi sperimentali, in cui una determinata esposizione è assegnata ai soggetti secondo il disegno dello studio stesso. A questo proposito, i risultati degli studi, e in particolare degli studi clinici, sono direttamente applicabili ai singoli pazienti con le stesse caratteristiche di quelli inclusi negli studi. Dopotutto, quando i medici prescrivono un nuovo farmaco ai loro pazienti sulla base dei risultati di uno studio, lo fanno perché si aspettano lo stesso effetto mostrato nello studio.

Se i trial clinici e altri studi sono ben disegnati ed eseguiti, si può evitare che i risultati siano inficiati dalla presenza di *bias*. Tuttavia, le altre due fonti di errore che si riscontrano nell'applicazione dei risultati ricavati dalle popolazioni agli individui, descritte per gli studi osservazionali, si possono applicare anche ai trial clinici. I risultati dei trial sono influenzati da errori casuali e si applicano precisamente solo a un ipotetico paziente "medio". Infatti, i clinici menzionati nel paragrafo precedente non sono così ingenui da aspettarsi esattamente il risultato riportato nello studio in ciascuno dei loro pazienti: prescrivono il nuovo farmaco con l'aspettativa di vedere nei loro pazienti, in media, l'effetto osservato nel trial, ma sono consapevoli del fatto che la risposta potrebbe essere molto variabile.

Tuttavia è molto importante il problema della validità esterna dei risultati degli studi sperimentali. Gli studi clinici, in particolare quelli riguardanti nuove terapie, includono campioni selezionati di pazienti che potrebbero, in linea di principio, beneficiare dei trattamenti in questione: si mette sempre più in evidenza, anche nella letteratura scientifica, il problema di una selezione di particolari classi di soggetti con determinate caratteristiche sociodemografiche come l'età (ad es. sotto rappresentazione di pazienti anziani negli studi cli-

nici (16)) e razza / etnia (ad esempio, sovra rappresentazione di bianchi non ispanici (6)).

CONCLUSIONI

Le considerazioni sull'applicabilità dei risultati degli studi epidemiologici a livello individuale sono analoghe a quelle sviluppate nell'ambito della "medicina personalizzata". L'obiettivo della medicina personalizzata è descrivere tutte le caratteristiche individuali che determinano la risposta del singolo paziente a un determinato trattamento e selezionare quella più efficace (1). Un approccio analogo può essere auspicabile anche in epidemiologia, nonostante il problema della validità interna complichino il processo, così come riportato prima. Sebbene una descrizione esaustiva di tutti i singoli fattori rilevanti rimanga un obiettivo difficile da raggiungere, è possibile compiere passi in questa direzione.

Revisioni sistematiche, meta-analisi e revisioni ombrello (12) aiutano a migliorare la precisione delle stime dei rischi e offrono opportunità di analisi stratificate per gestire le fonti di eterogeneità dei risultati all'interno delle popolazioni. L'applicazione dell'analisi quantitativa dei *bias* (15), come illustrato in precedenza, migliorerebbe la validità delle inferenze a livello individuale. L'integrazione di biologia ed epidemiologia potrebbe contribuire a ridurre le incertezze sulla validità esterna dei risultati.

In conclusione, i risultati epidemiologici possono essere applicati al singolo individuo solo rispettando il quadro rigoroso delineato. Quando, come nella maggior parte dei casi, le fonti di errore casuale, la validità interna e la validità esterna possono essere controllate solo parzialmente, l'estrapolazione dei risultati al livello individuale è incerta nella migliore delle ipotesi.

Un caso particolare, in cui l'estrapolazione a livello individuale può essere giustificata, è quello dei geni di suscettibilità ad alta penetranza: i risultati di carattere clinico o epidemiologico hanno mostrato un rischio così elevato nei portatori, che le considerazioni su errore casuale e sistematico hanno meno rilevanza e può essere giustificabile assumere validità esterna anche in assenza di prove dirette a supporto.