thebmj

BMJ 2018;360:k757 doi: 10.1136/bmj.k757 (Published 22 February 2018)

RESEARCH METHODS & REPORTING

STATISTICS NOTES

Population attributable fraction

Mohammad Ali Mansournia *assistant professor of epidemiology*[1], Douglas G Altman *professor of statistics in medicine*[2]

[1]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran; [2]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK

Much statistical analysis seeks to identify associations between exposures and outcomes. The population attributable fraction (PAF) is an epidemiologic measure widely used to assess the public health impact of exposures in populations. PAF is defined as the fraction of all cases of a particular disease or other adverse condition in a population that is attributable to a specific exposure; PAF equals (O – E)/O, where O and E refer to the observed number of cases and the expected number of cases under no exposure, respectively. The term "attributable" has a causal interpretation: PAF is the estimated fraction of all cases that would not have occurred if there had been no exposure.[1] As an example, in early 1950,[2] Doll derived O = 11189 and E = 1875 using the Doll and Hill case-control study of smoking and lung cancer deaths throughout England and Wales,[3] so the smoking PAF for lung cancer deaths was (11189 – 1875)/11189 = 83%.

Using a cohort study, following Miettinen, we can estimate the PAF from the estimated relative risk (RR) for the exposure and the prevalence of exposure among cases (p_c), as PAF = $p_c(1 - 1/RR)$.[4] Suppose that a particular exposure doubles the risk of a certain outcome (that is, RR = 2). If the prevalence of exposure among cases is 0.6, then PAF = 0.6(1 – 0.5) = 0.3 (that is, 30%). PAF depends not only on the increased risk associated with the exposure but is also directly related to the prevalence of exposure. PAF is usually expressed as a percentage.

Cohort studies are observational and thus liable to confounding,[5 6] so crude (unadjusted) RR should not be used. An adjusted RR can be used in the Miettinen formula to estimate a valid PAF. Alternatively, one can directly estimate the PAF original formula "(O – E)/O" using results from a multivariable logistic regression model.[7]

As an example of the latter approach, the authors of a recent *BMJ* paper[8] calculated the population attributable fraction (PAF) of concurrent benzodiazepine/opioid use for the risk of opioid overdose in a retrospective analysis of claim data. This fraction represents opioid overdose case reduction in the population that would occur if concurrent benzodiazepine/opioid use could be eliminated entirely. The PAF estimate was 15% (95% confidence interval 14 to 16%).[8] Valid 95% confidence intervals for PAF should take into account the uncertainty in both the observed and expected number of cases.[7]

The PAF formula with adjusted RR is easily generalised to exposures with more than two levels.[9] In a cohort study the PAF for the effect of maternal overweight and obesity on infant mortality in relation to normal weight was estimated as 11%.[10] Similarly, we can calculate PAF for the joint effects of two or more exposures. Such a PAF is expected to be less than the sum of the PAF for each exposure because people exposed to both exposures should not be counted twice. Finally, for preventive exposures one can reverse the coding: RR is now the adjusted risk ratio for no exposure and p_c is the prevalence of no exposure among cases. The result is known as preventable fraction: the fraction of all cases that would be prevented if the whole population were exposed.

We can use valid estimates of hazard ratio (or rate ratio) from cohort studies or odds ratio from case-control studies instead of RR in the Miettinen PAF formula if the outcome is uncommon. Here we assume that removing an exposure does not affect the person-time at risk, which may not be true. For example, omitting smoking expands person-year at risk of coronary deaths by removing other competing risks for deaths such as lung cancer.[11]

Other important assumptions underlie the PAF. As usual, we make the strong assumptions that there is no bias in the study design and data analysis; in particular, that the estimated effect is adjusted for all confounders. In addition, we assume that removing the exposure does not affect other risk factors. This assumption may not be true in practice; for example, removing smoking may decrease alcohol consumption, making interpretation of smoking PAF for coronary deaths difficult.

Also, PAF assumes that there is a perfect intervention which eradicates the exposure. However, complete removal of an exposure is often unrealistic; even with legal restrictions and cessation programmes, many people will continue to smoke. A

Correspondence to: M A Mansournia mansournia_ma@yahoo.com

measure that allows for these realities is the generalised impact fraction, which is the fractional reduction of cases that would result from changing the current level of exposure in the population to some modified (partially removed) level.[12] More technical issues about PAF, including its difference from aetiologic fraction, can be found elsewhere.[13]

Contributors: MAM and DGA jointly wrote and agreed the text.

Competing interests: We have read and understood the BMJ Group policy on declaration of interests and have no relevant interests to declare.

Provenance and peer review: Not commissioned; not externally peer reviewed.

1 Greenland S, Rothman KJ, Lash TL. Measures of effect and measures of association. In: Rothman KJ, Greenland S, Lash T, eds. *Modern epidemiology*. 3rd ed. Lippincott Williams & Wilkins, 2008: 51-70.
2 Doll R. Mortality from lung cancer among non-smokers. *Br J Cancer* 1953;7:303-12. 10.1038/bjc.1953.29 13106197
3 Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. *Br Med J* 1952;2:1271-86. 10.1136/bmj.2.4797.1271 12997741
4 Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am J Epidemiol* 1974;99:325-32. 10.1093/oxfordjournals.aje.a121617 4825599
5 Mansournia MA, Altman DG. Inverse probability weighting. *BMJ* 2016;352:i189. 10.1136/bmj.i189 26773001
6 Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. *BMJ* 2017;359:j4587. 10.1136/bmj.j4587 29038130
7 Brady A. Adjusted population attributable fractions from logistic regression. *Stata Tech Bull* 1998;42:8-12.
8 Sun EC, Dixit A, Humphreys K, Darnall BD, Baker LC, Mackey S. Association between concurrent use of prescription opioids and benzodiazepines and overdose: retrospective analysis. *BMJ* 2017;356:j760. 10.1136/bmj.j760 28292769
9 Hanley JA. A heuristic approach to the formulas for population attributable fraction. *J Epidemiol Community Health* 2001;55:508-14. 10.1136/jech.55.7.508 11413183
10 Johansson S, Villamor E, Altman M, Bonamy AK, Granath F, Cnattingius S. Maternal overweight and obesity in early pregnancy and risk of infant mortality: a population based cohort study in Sweden. *BMJ* 2014;349:g6572. 10.1136/bmj.g6572 25467170
11 Greenland S. Applications of stratified analysis methods. In: Rothman KJ, Greenland S, Lash T, eds. *Modern epidemiology.* 3rd ed. Lippincott Williams & Wilkins, 2008: 283-302.
12 Morgenstern H, Bursic ES. A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. *J Community Health* 1982;7:292-309. 10.1007/BF01318961 7130448
13 Greenland S. Concepts and pitfalls in measuring and interpreting attributable fractions, prevented fractions, and causation probabilities. *Ann Epidemiol* 2015;25:155-61. 10.1016/j.annepidem.2014.11.005 25498918