

Systematic reviews of effectiveness of Public Health practice

Fabrizio Faggiano^{1,2}; Federica Vigna-Taglianti^{1,2}

¹Dipartimento di Medicina Clinica e Sperimentale, Università "Amedeo Avogadro", Novara, Italia;

²Osservatorio Epidemiologico delle Dipendenze (OED) della regione Piemonte, Grugliasco (Torino), Italia

Correspondence to: Fabrizio Faggiano, Dipartimento di Medicina Clinica e Sperimentale, Università Avogadro, Via Solaroli 17, 28100, Novara, Italia. E-mail: fabrizio.faggiano@med.unipmn.it

Abstract

Background: Public Health is the main discipline involved in the prevention of avoidable deaths. To implement interventions aimed at eliminating causes of death, and to take decisions based on evidence of effectiveness, research summaries are needed.

Methods: To discuss the critical points for synthesizing the evidence on the effectiveness of public health interventions, the inconsistencies between the results produced by different epidemiological study designs and the methodological issues related to the quality of the summaries are underlined.

Results: The Randomized Controlled Trial is recognized as the best study design for the assessment of the effectiveness of clinical practices; however, the integration of the results from non-randomized and randomized interventions has been suggested recently. In particular this involves Public Health, where the randomization of individuals is often impossible, and interventions are administered at different levels (individual, group, population).

Conclusions: Systematic reviews are not common in Public Health. Methodological tools to write rigorous summaries of the evidence from public health interventions, taking into account all possible biases, appear to be a priority.

Key words: public health, systematic reviews, evidence-based practice, prevention, avoidable mortality

Introduction

The fraction of avoidable mortality attributable to preventable factors remains high in Europe, and the gap between industrialized and developing countries is wide [1]. Several million lives could be saved if the mortality rate due to chronic diseases could be reduced by 2% annually. Most of the effort to achieve such a gain should be concentrated in developing countries, where general structural strategies are required; whereas in developed countries, the resources allocated to Public Health within European Health Systems should be re-directed. In fact, there is huge variation among industrialized countries in the amount and extent of prevention programs carried out by Health Authorities. Major risk factors are often neglected (smoking, road traffic, violence), while others are sometimes over-represented, both in terms of the financial resources allocated and the public alarm raised (e.g. bird flu). Moreover, most of the recommended interventions are not based on scientific evidence [2].

In Italy, 69,000 avoidable deaths occurred in 2002 [3]. Based on Rutstein's list [4], and applying a conservative definition of avoidable death [3], 42.2% of deaths in men from 5-64 years of age were

likely to have been preventable through primary prevention interventions [5]. Considering as a goal the elimination of risk exposures, the number of avoidable deaths could be even higher. For example, 65,000 deaths could be avoided with primary prevention of smoking, which corresponds to 12.1% of all deaths over the age of 35 [6].

Knowledge of the risk factors related to the main causes of death is not enough to ensure their elimination. Public Health is the main field involved in the prevention of avoidable deaths. Public Health is broadly defined as the "promotion of health and prevention of disease through the organized efforts of society" [7]. Its theoretical functions includes: the development of interventions aimed at eliminating causes of death as identified by epidemiology, the assessment of program effectiveness, and the transfer of effective programs into practice [8].

The aim of this paper is to discuss the main methodological issues involved in synthesizing the evidence on the effectiveness of public health interventions. It will focus on two key issues in the process of producing evidence: i) study designs for the measurement of the effect size in prevention interventions; ii) unbiased methods to summarize the evidence.



Study designs for the measurement of the effect size of prevention interventions

The development of the evidence-based medicine approach was based on the identification of the Randomised Controlled Trial (RCT) as the best study design for the assessment of the effectiveness of clinical practices [9]. Although this is due to the unquestioned superiority of randomization in controlling for confounding factors, a large debate has arisen in recent years on the integration of the results from non-randomised studies with the evidence from randomised interventions [10]. Although several reviews have agreed with the practice of using non-randomised results at least in support of randomised evidence [11-13], caution is needed in adopting such a practice. There are relevant examples in the scientific literature on the inconsistencies between experimental and observational results. The case of beta-carotene is one of the most interesting.

The case of Beta-Carotene: the chemoprevention hypothesis suggested by Peto and Doll for beta-carotene in the case of cancer, based on a large review of cohort studies [14], has been refuted by some large double-blind, randomised trials of dietary supplementation. In the cohort studies, individuals whose consumption level of dietary beta-carotene was in the upper tertile showed a relative risk (RR) of developing cancer of about 0.30 compared to individuals whose consumption level ranked in the lower tertile; furthermore, evidence from laboratory research was highly consistent with this result. The empirical results from experimental studies, however, showed the opposite effect. Two large randomised prevention studies were conducted: the ATBC Study involving 29,000 smokers started in 1986 in Finland [15], and the CARET study involving 18,000 subjects, not only smokers, started in 1992 in the US. Both studies were randomised, double-blind, placebo controlled studies and were funded by National Cancer Institute (NCI). After a few years of follow-up, both studies were stopped due to a suspicion of harm in the intervention groups: the RR of cancer among those receiving supplements of beta-carotene and alpha-tocopherol was significantly higher than 1, nearly 1.3 [16].

Another well-known example is hormone replacement therapy (HRT).

The case of Hormone Replacement Therapy: several cohort studies conducted during the 70s and the 80s showed a substantial effect of HRT in

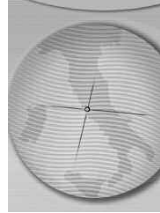
reducing cardiovascular diseases and osteoporosis [17]. The first RCT carried out during the late 90s refuted these results, and even showed a slight increase in the risk of developing coronary heart disease, breast cancer and stroke among women taking HRT for many years [18].

The common explanation for these inconsistencies is the inadequate control for confounding factors in cohort studies. In food, beta-carotene likely drives the true preventive factors, which act as preventive factors alone or when linked to beta-carotene. Hormone replacement therapy is used more frequently by women of high social class who have healthier lifestyles, including diet, preventive medical visits and physical exercise.

Despite potential inconsistencies between observational and experimental evidence, Public Health needs to overcome the paradigm of the RCT. While some public health interventions, such as vaccinations, can be evaluated using RCTs, many others cannot because the intervention is not administered at the individual level, for example, prevention programs in schools and the fluoridation of municipal water supplies, or because it is impossible to identify an unexposed population, as with information campaigns and regional policies. In these cases, it is essential to define a specific methodological approach that includes, as best evidence, studies other than RCTs, such as Cluster Randomised Trials, Before-After Studies and Interrupted Time Series [19,20].

The appropriate study design for evaluation is determined by the population to whom the intervention will be administered; thus, we can identify three types of interventions: *i) projects*, targeted at the individual level, e.g. counselling, vaccination, chemoprevention; *ii) programs*, targeted to population groups, e.g. screening programs, media campaigns; and *iii) policies*, targeted to the entire population, e.g. health policies and national laws. Table 1 summarizes the methodological constraints on public health interventions.

Projects can be evaluated using RCTs, since the intervention can be randomised at the individual level and an unexposed reference group can easily be identified. Programs are provided at the group level and, therefore, individual randomization cannot be carried out. To allow for group randomization and individual outcome assessment, a cluster randomised study is the required evaluation tool [19]. The crucial problem is the evaluation of policies, as these interventions usually involve the entire population: randomization cannot be used and neither can an unexposed reference

**Table 1. Methodological constraints on Public Health interventions and appropriate study designs**

Intervention	Individual Randomization	Reference Group	Examples	Appropriate design
Project	YES	YES	Vaccination	Randomised controlled trial
Program	NO	YES	School-based prevention	Cluster randomised trial
Policy	NO	NO	Anti-smoking national laws	Before-after trial Controlled before-after trial Interrupted time series

group be identified. To evaluate policies, observational studies are required, such as controlled or uncontrolled before-after trials, and interrupted time series [20]. The inclusion of observational studies among the designs able to provide the best evidence entails a high risk of bias and, therefore, requires the fulfilment of methodological quality criteria that are not as yet available. Evidence-based medicine has, in fact, devoted considerable effort in the development of tools for the quality appraisal of RCTs; however, these tools are only partly transferable to observational studies, mainly due to the problem of controlling for confounding factors. An agreement within the scientific community on how to evaluate the quality of observational studies is still lacking.

The summary of evidence

A single assessment, even from an appropriate experimental study, can hardly be considered a sufficient basis for undertaking a public health intervention. Apart from the statistical issues related to frequent inadequate sample size, there are many other considerations limiting the direct transfer of study results into practice, the most relevant being the variation of the effects in different contexts, and the need for a systematic assessment of the quality of the study.

Systematic reviews, summarizing the results of research in a particular area, are intended to be used mainly as a tool to keep professionals up-to-date in a world where the scientific literature is exploding and there are more and more articles to be read every day [21]. Beyond this generic aim, systematic reviews are also useful in avoiding publication bias [22] and in detecting inconsistencies between studies by the analysis of heterogeneity. In the Public Health domain, summary methods are available for systematic reviews of etiology. Studies of this kind tend to be limited in size, especially for weak associations, but have a large impact at the population level. Examples include the health effects of dietary intake of beta-carotene [14], Vitamin D [23] and antioxidants [24]. Systematic reviews of studies assessing the effectiveness of public health interventions are less common. Apart from considerations about the general delay in the use of evidence, a possible reason for the lack of

systematic reviews in this field is the difficulty in evaluating interventions such as programs or policies, which cannot be allocated randomly at the individual level. In fact, there are many more reviews of public health interventions targeting the individual rather than the population. Vaccination is a good example of this imbalance: many systematic reviews have assessed the effectiveness of vaccines, such as pneumococcus, haemophilus influenzae, hepatitis B and influenza, some of which are included in the Cochrane Library (www.cochrane.org) [25]. Many other examples can be found, for example, antioxidant supplementation for prevention of cancer [26].

There is extensive evidence on programs targeting groups of individuals and reviews of the summarizing data are now available. Concerning school-based interventions, for instance, the effectiveness of prevention programs for smoking [27], alcohol [28] and drugs [29] have been systematically reviewed and useful data are available to make decisions in these fields. On the contrary, the amount of available evidence on interventions, such as policies, targeting populations as a whole is poor. On occasion, some Health Technology Agencies have produced reviews on the effects of policies, such as the dissemination of practice guidelines in clinical settings [30], policies for immigration and health [31], and population-based interventions aimed at reducing sexually transmitted infections [32]. These reviews differ widely from those on projects or programs. In some cases, they are systematic but limited to interventions targeted at groups, and include only cluster randomized studies [32]; in others, they include observational studies without rigorous criteria for selection and quality assessment.

The matching of the type of intervention and the study design produces the need for developing methods to write sound summaries of the evidence from public health interventions. For this purpose, several methodological issues have to be taken into consideration.

The first is *validity*: the validity of any summary method depends on the quality of the primary studies. Even the most sophisticated method to produce a meta-analysis does not correct for confounding or bias. This is particularly true when

**Table 2. Studies assessed for inclusion by the 18 Cochrane Drug & Alcohol Group's reviews, stratified by study design (08/2005)**

	Studies				
	Included		Excluded		Total
	n	%	n	%	N
RCT	280	78.0	79	22.0	359
CCT	13	41.9	18	58.1	31
CPS	20	20.2	79	79.8	99
Other / not specified	38	8.7	400	91.3	438
Total	351	37.9	576	62.1	927

In two reviews, the number and type of excluded studies were not specified

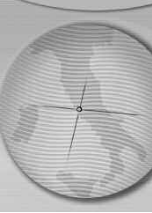
the authors are willing to include non-randomised studies, such as those useful in assessing public health policies [33-34]. Without taking into consideration the main sources of bias in the original studies, a meta-analysis of observational studies may well be simply producing *tight confidence intervals around spurious results*. An example of the differential quality between randomised and non-randomised studies is presented in Table 2. It shows the analysis of the studies assessed for inclusion in 18 systematic reviews led by the Cochrane Drugs and Alcohol Group in 2005. The authors of these reviews assessed 927 studies overall: the percentage of studies included, out of the total number assessed of each type was 20.2% of Controlled Prospective Studies (CPS), 8.7% of other non-randomised studies and 78.0% of RCTs. The main exclusion criterion was the inadequate or lacking of control for confounding factors at the analysis stage.

The second is *heterogeneity*: some characteristics of the studies can produce heterogeneity. Any differences in selecting the study population, as well as in measuring the outcomes will be integrated into the result and will produce heterogeneity [35]. Heterogeneity analysis can also be considered an important tool in detecting differences and bias among the studies included in a meta-analysis, leading to the exclusion of biased studies and thereby improving the quality of the review. Pooling data, as an alternative to producing summaries of results, can be considered an efficient method to control for heterogeneity under some conditions.

The third is *publication bias*: it is the differential probability of publication of *positive* (expected) results compared to *negative* ones. All summary methods can be victims of publication bias [22]. This also likely applies to observational studies. A pooled analysis can reduce publication bias by using data from studies that have not produced useful measures for meta-analysis, but cannot correct for the bias due to unpublished studies. Some methods, based on systematic review, have been recently developed to diagnose bias and estimate the true RR associated with the relationship [36]. Since

Table 3. Reporting checklist for meta-analyses of observational studies (37)

Background	Problem definition Hypothesis statement Description of study outcomes Intervention or exposure Study designs Study population
Search strategy	Qualification of searchers Search strategy, time period and keywords Efforts to include all available studies Databases searched Search software Hand searching Citations located and excluded Methods of addressing non-english articles Methods of handling abstracts and unpublished studies Contact with authors
Methods	Relevance and appropriateness of studies assembled Rationale for selection and coding of data Classification and coding of data Assessment of confounding Assessment of study quality Assessment of heterogeneity Statistical methods Tables and graphs
Results	Graphic summarizing individual study and overall estimates Table giving information on individual studies Results of sensitivity testing Indication of statistical uncertainty of findings
Discussion	Quantitative assessment of bias Justification for exclusion Assessment of quality of included studies
Conclusions	Alternative explanations of results Generalization of the conclusions Guidelines for future research Disclosure of funding source



publication bias can increase the combined RR, some attempt should be made to control for it. In an attempt to reduce the risk of bias in systematic reviews that include non-randomised studies, a check-list has been proposed by the editors of the main scientific journals (Table 3) [37].

Conclusions

There is a delay in building up the evidence on the effectiveness of interventions to tackle population health problems in Public Health. To overcome this important limitation, it is imperative to increase the amount of evaluative research in this field. It is also time to develop sound methodological tools for the evaluation of interventions in which randomization is not possible.

References

- 1) Strong K, Mathers C, Leeder S, Beaglehole R. Preventing chronic diseases: how many lives can we save? *Lancet* 2005;366 (9496):1578-82.
- 2) Baldasseroni A, Bernhardt S, Ciani Passeri A. Valutazione dell'efficacia dell'intervento libretto di idoneità sanitaria per alimentaristi [lisa] all'interno dei programmi di salute pubblica per la salubrità e la sicurezza degli alimenti. [Efficacy evaluation of the health idoneity card for the workers in the food industry (lisa) in the public health and food security programs] Firenze: ArsToscana, ASL 10 Firenze, 2003.
- 3) ERA. Atlante 2006. Mortalità evitabile e contesto demografico per ASL. [Atlas 2006, avoidable mortality and demographic contest for ASL] Roma: Istituto Superiore di Sanità, 2006.
- 4) Rutstein DD, Berenberg W, Chalmers TC, et al. Measuring the quality of medical care. A clinical method. *N Engl J Med* 1976;294:582-8.
- 5) Simonato L, Ballard T, Bellini P, Winkelman R. Avoidable mortality in Europe 1955-1994: a plea for prevention. *J Epidemiol Community Health* 1998;52:624-30.
- 6) Faggiano F, editor. Fatti e cifre sul fumo di tabacco in Piemonte - 2005 [Facts and figure about tobacco smoking in Piemonte - 2005]. Torino: Regione Piemonte, 2005.
- 7) WHO. European health care reform - analysis of current strategies. Copenhagen: Regional Publication, European Series, No. 72, WHO, 1997.
- 8) Beaglehole R, Bonita R, Kjellstrom T. Epidemiology, health services and health policy. In: *Basic Epidemiology*. Geneva: World Health Organization, 2000.
- 9) Cochrane AL. Effectiveness and efficiency: random reflections on health services. London: Nuffield Provincial Hospital Trust, 1974.
- 10) Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.
- 11) Concato J, Shah N, Horwitz RI. Randomised, controlled trials, observational studies, and the hierarchy of research design. *N Engl J Med* 2000;342:1887-92.
- 12) Benson K, Hartz AJ. A comparison of observational studies and randomised, controlled trials. *N Engl J Med* 2000;342:1878-86.
- 13) Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.
- 14) Peto R, Doll R, Buckley JD et al. Can dietary beta-carotene materially reduce human cancer rates? *Science* 1981;290:201-8.
- 15) The ATBC study group. Incidence of cancer and mortality following alpha-tocopherol and beta-carotene supplementation. *JAMA* 2003;290:476-85.
- 16) Omenn GS. Chemoprevention of lung cancer: the rise and demise of beta-carotene. *Ann Rev Public Health* 1998;19:73-99.
- 17) Stampfer MJ, Colditz GA. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med* 1991;20:47-63.
- 18) Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, et al. Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomised controlled trial. *JAMA* 2002;288:321-33.
- 19) Murray DM. Design and analysis of Group-randomised trials. New York: Oxford University Press, 1998.
- 20) Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin Co, 2002.
- 21) Dickersin K. Systematic reviews in epidemiology: why are we so far behind? *Eur J Epidemiol* 2002;31:6-12.
- 22) Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;323:101-5.
- 23) Garland CF, Garland FC, Gorham ED et al. The role of Vitamin D in cancer prevention. *Am J Public Health* 2006;96:252-61.
- 24) Knekt P, Ritz J, Pereira MA et al. Antioxidant vitamins and coronary heart disease risk: a pooled analysis of 9 cohorts. *Am J Clin Nutr* 2004;80:1508-20.
- 25) Jefferson T, Rivetti D, Rivetti A, Rudin M, Di Pietrantonj C, Demicheli V. Efficacy and effectiveness of influenza vaccines in elderly people: a systematic review. *Lancet* 2005;366:1165-74.
- 26) Bjelakovic G, Nikolova D, Simonetti RG, Gluud C. Antioxidant supplements for prevention of gastrointestinal cancers: a systematic review and meta-analysis. *Lancet* 2004;364:1219-28.
- 27) Thomas R. School-based programmes for preventing smoking (Cochrane Review). Oxford: The Cochrane Library, Update Software, 2004; Issue 2.
- 28) Foxcroft DR, Ireland D, Lister-Sharp DJ, Lowe G, Breen R. Primary prevention for alcohol misuse in young people (Cochrane Review). Oxford: The Cochrane Library, Update Software. 2004; Issue 2.
- 29) Faggiano F, Vigna-Taglianti F, Versino E, Zamboni A, Borraccino A, Lemma P. School-based prevention for illicit drugs' use (Cochrane Review). *Cochrane Database Syst Rev* 2005, Issue 2.
- 30) Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;8(6):iii-iv,1-72.
- 31) Catalan Agency for Health Technology Assessment and Research (CAHTAR). Experiences in research and policies related to health and migration in European countries: the case of the Netherlands, the United Kingdom and Switzerland. Barcelona: Catalan Agency for Health Technology Assessment and Research (CAHTA), 2004:48.
- 32) Sangani P, Rutherford G, Wilkinson D. Population-based interventions for reducing sexually transmitted infections, including HIV infection. *The Cochrane Database of Systematic Reviews* 2004, Issue 3.
- 33) MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. *Lancet* 2001;357:455-61.
- 34) McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999;319:312-5.
- 35) Egger M, Schneider M, Smith GD. Meta-analysis spurious precision? Meta-analysis of observational studies. *BMJ* 1998;316:140-4.
- 36) Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ* 1997;315:629-34.
- 37) Stroup DF, Berlin JA, Morton SC et al. Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. *JAMA* 2000;283:2008-12.