Preventive Medicine 83 (2016) 22–25

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Commentary

Recommendation without experts? Epistemological implications in the development of screening guidelines

Paolo Giorgi Rossi *

Servizio Interaziendale di Epidemiologia, AUSL, Reggio Emilia, Italy
Arcispedale S. Maria Nuova, IRCCS, Reggio Emilia, Italy

ARTICLE INFO

Available online 7 December 2015

Keywords:
Guidelines
Evidence-based prevention
Cancer screening

ABSTRACT

Controversies concerning mammographic and cervical cancer screening with HPV-DNA recommendations lead to an analysis of the role played by a knowledge of disease epidemiology, natural history and pathogenesis in producing sound recommendations.
This analysis calls into question the decision to exclude experts on the specific topic from guideline and recommendation development because such experts may bring prejudices or even conflicts of interest to the debate. According to this approach, methodology is the only factor that guarantees the soundness of evidence assessment. The assumption underlying such an epistemological point of view is that evidence is "absolute," i.e. not linked to any interpretative model or conjecture. Actually, any form of scientific knowledge includes conjectures, which by definition are not demonstrable, in order to interpret evidence. Even as we assess evidence, we need to select or formulate conjectures that explain most of the evidence available. In order to decide on such conjectures, we require individuals who are familiar with the epidemiology and the aetiology of the disease, as well as with the rationale behind the technologies or interventions proposed. Finally, we need individuals who know the strengths and the weaknesses of alternative conjectures; in other words, we also require content experts.

New controversies regarding the interpretation and assessment of important trials continue to emerge despite increasing recognition of evidence-based medicine by clinicians and the scientific community.

Cancer screening presents two exemplary cases: a controversy on breast cancer screening, which has continued for 15 years without reaching any universally accepted conclusion regarding the balance between risk and benefit (Independent UK Panel on Breast Cancer Screening, 2012; Lauby-Secretan et al., 2015), and a less acrimonious controversy on using a test for human papillomavirus (HPV) DNA as a primary screening test for cervical cancer (Whitlock et al., 2011; Moyer, 2012; Saslow et al., 2012; Canadian Task Force on Preventive Health Care et al., 2013; The SOGC, 2013; Ronco et al., 2015).

In both instances, systematic reviews mainly based on the same trial results have produced different and sometimes completely opposite recommendations.

In the realm of mammographic screening, a number of large panels have recently worked on assessing risks and benefits, and on providing a painstaking estimate of mortality reduction and over-diagnosis (Independent UK Panel on Breast Cancer Screening, 2012; Lauby-Secretan et al., 2015). The estimates of cause-specific mortality reduction in women aged 50–69 are consistent in all meta-analyses (Independent UK Panel on Breast Cancer Screening, 2012; Lauby-Secretan et al., 2015; EUROSCREEN Working Group, 2012; Gotzsche and Jorgensen, 2013; Nelson et al., 2009; Canadian Task Force on Preventive Health Care, 2011).[1] The current debate mainly centres on the extent of over-diagnosis, which none of the trials was specifically designed to measure. Nevertheless, over-diagnosis can be estimated using the results of only a few trials and of multiple observational studies with different designs and different methodological quality and soundness. Indeed, the systematic review and assessment of the results of these observational studies is challenging, because bias and methodological errors in such studies are more problematic (Puliti et al., 2011; Biesheuvel et al., 2007).

* Servizio Interaziendale di Epidemiologia, AUSL Reggio Emilia, Via Amendola 2, I42122 Reggio Emilia, Italy. Fax: +39 0522335460.
E-mail address: Paolo.giorgirossi@ausl.re.it.

[1] The estimates of cause-specific mortality reduction for invited women aged 50–69 are: 20% according to the UK independent panel (Independent UK Panel on Breast Cancer Screening, 2012), 19% for the USPSTF (Nelson et al., 2009), and 21% for the Canadian Task Force (Canadian Task Force on Preventive Health Care, 2011) (these three included only randomised trials in their reviews); 23% according to the IARC document (2), which also included observational studies; and 25% for the EUROSCREEN group, which included only observational studies (EUROSCREEN Working Group, 2012), since its goal was to estimate the effectiveness of organised screening programs in Europe. The Cochrane review estimates a 15% reduction for the 39–74 age group, while it specifies a 23% reduction for the over-50 group, but underscores that the reduction is 6% for "adequately randomised trials" and 30% for trials with "suboptimal randomization" (Gotzsche and Jorgensen, 2013).

http://dx.doi.org/10.1016/j.ypmed.2015.11.025
0091-7435/

When determining the scope of this paper, I focused only on the most recent recommendations; i.e., those by the UK independent panel and by the IARC. In particular, I was interested in the method used by the two working groups for selecting the studies to be included in their reviews. I was not concerned about the results that were obtained, because there are many sources of variability in over-diagnosis estimates. Such variability is in part due to inconsistent ways of presenting results (Independent UK Panel on Breast Cancer Screening, 2012; Lauby-Secretan et al., 2015).[2]

When the UK independent panel was constituted, all those who had already published on the topic were excluded, with the aim of avoiding the inclusion of researchers with a prejudice for or against screening (Independent UK Panel on Breast Cancer Screening, 2012). The IARC panel also included experts, radiologists, epidemiologists and public health workers who had previously been involved in screening evaluation or implementation (Lauby-Secretan et al., 2015). Faced with the complex process of selecting observational study designs that would give unbiased (or less biased) estimates, the UK independent panel decided to include only data from the two available trials.

The UK Panel reviewed the observational studies (trends, incidence-based and case control studies) to assess efficacy, but considered these kind of study affected by residual bias that could inflate the estimate of benefit. About the estimates of over-diagnosis provided by observational studies, the panel states: "(…) by varying the assumptions and statistical methods underlying these studies, using the same datasets, estimates of over-diagnosis rates were found to vary across the range of 0% to 36% of invasive breast cancers diagnosed during the screening period. The Panel had no reason to favour one set of estimates over another and concluded that this method could give no reliable estimate of the extent of over-diagnosis". It is disconcerting that a group of experts cannot judge which studies are less probably biased than others.

On the other hand, the IARC panel carried out an in-depth evaluation of the worldwide estimate of over-diagnosis based on observational studies and, in particular, of the summary estimate from bias-adjusted studies evaluating the breast cancer screening programs in Europe. The panel ultimately decided to include observational studies with sound designs that would distinguish over-diagnosis from anticipation of diagnosis (Puliti et al., 2012).

In the case of HPV-based screening, the controversy has been quelled by the facts, but the dynamic of developing recommendations is emblematic. The US Preventive Services Task Force published a systematic review in November 2011 that had been commissioned to a group of researchers at Kaiser Permanente (Whitlock et al., 2011). The review concluded that there was no evidence supporting the introduction of HPV as a primary screening test. In March 2012, the USPSTF itself published new guidelines (Moyer, 2012) that were in substantial agreement with joint guidelines published simultaneously by major US scientific societies (Saslow et al., 2012). These guidelines recommend the use of HPV as a primary screening test, together with a Pap test, every 5 years in women aged 30 or older.

The 2013 recommendations by the Canadian Task Force on Preventive Health Care had a similar history. The relative document concluded that there was insufficient evidence for adopting HPV as primary screening (Canadian Task Force on Preventive Health Care et al., 2013). The principal Canadian scientific societies immediately reacted to the recommendation and announced new joint guidelines (The SOGC, 2013)

Interestingly enough, the attempt to block the introduction of HPV was mostly justified by the risks of increasing the number of unnecessary colposcopies and of overtreating regressive lesions. In the end, however, the protocol finally proposed by the USPSTF recommendation (i.e. Pap test plus HPV test — so called "co-testing") is much more intensive than the protocols suggested in many European countries, which are based on the sequential use of the two tests; i.e. first the HPV test, and then a Pap test only for those who tested positive — the so called "triage strategy" (Saslow et al., 2012).

Even if guidelines are specifically designed to reduce variability in the way patients are treated, they are not immune to subjective interpretation and inevitably are a source of controversy. The acrimony of the controversies is even greater in the public health sphere than in clinical guidelines; in fact, public health recommendations often become laws or fixed protocols for interventions, and these laws and protocols cannot be modified or interpreted by the clinician when treating individual cases. Such is the case for screening programs in countries with universal-type public health services.

There are many sources of variability and error in the way in which evidence is assessed and recommendations are formulated: over-valuation of experimental data and of surrogate end-points (Shrier et al., 2007; Hernán and Taubman, 2008), misinterpretation of significance testing (Stang et al., 2010), misconduction of meta-analyses (Greenland, 1994; Thompson and Pocock, 1991; Eysenck, 1994; Lau et al., 1998) and suboptimal decision-making processes (Mattingly and Ponsonby, 2014). Finally, the same evidence can be observed by different subjects and yield divergent conclusions that are largely influenced by background knowledge and experience (Bate et al., 2012); this last point highlights the fact that the inclusion of authors with diverse expertise will probably influence final recommendations. A multidisciplinary approach has therefore been considered to be indispensable for developing any sort of guideline or recommendation (AGREE Collaboration, 2003; Brouwers et al., 2010; Schünemann et al., 2009).

In this paper, I attempt to clarify some of the assumptions underlying exclusion criteria for several important expert panels whose task was to develop screening guidelines.

More and more often, governmental agencies decide to approach or prevent possible controversies by applying a method that resembles a criminal trial; i.e. they assemble a working group that has never been involved in the specific research question and has never published on the topic. Members of the group must be experts in methodology and in the disease in general, but it is therefore very difficult that they have expertise on the intervention that is going to be evaluated. After examining all the available evidence, they act as the jury in a trial. In this way, the members of working group should be free from prejudice and competing interest.

This was the decision explicitly made by the Canadian Task Force (Canadian Task Force on Preventive Health Care et al., 2013). In the case of the first systematic review by the USPSTF (Whitlock et al., 2011), on the other hand, the decision was not explicit, yet no cervical cancer screening experts were included in the group. The archetypal case is the UK Independent Panel on breast cancer screening (Independent UK Panel on Breast Cancer Screening, 2012).

From this standpoint, the absence of prejudice seems to be more important to a correct assessment of the evidence than specific knowledge of the disease and of the technology or intervention to be evaluated; indeed, overly deep knowledge appears to be an obstacle to objectivity. According to this approach, it is the methodology – i.e. how to evaluate evidence and how to assess it – that guarantees whether the process is correct.

[2] The proportion of overdiagnosed cancer can be shown in a different way: although the numerator remains the same (the numerator being excess cancers in the screening group compared to the non-screening group, subsequent to a sufficiently long follow-up period after end of screening, thus allowing over-diagnosis to be distinguished from anticipation of diagnosis), different denominators can be used; i.e., all cancers observed during the period (including follow-up), or only cancers that occurred during the screening period. As a result, the same estimate of over-diagnosis in terms of incidence or numbers can yield substantially different results when reported in terms of percentage: 10.7% of cancers occurring throughout the period of study, or 19% of cancers occurring in the screening period (Independent UK Panel on Breast Cancer Screening, 2012).