MAYO CLINIC

To Screen or Not to Screen: Reconciling Individual and Population Perspectives on Screening

L. Rachid Salmi, MD, PhD; Gaëlle Coureau, MD, PhD; Marion Bailhache, MD, MSc; and Simone Mathoulin-Pélissier, MD, PhD

Abstract

Screening is the early detection of a latent disorder by a test to allow early intervention with the aim of improving prognosis. Individual and population perspectives on screening are perceived as opposing interests of patients and the population. In this article, we try to reconcile these perspectives. The individual perspective is based on the clinical experience of a better prognosis at early stages and patients with missed opportunities. In the population perspective, screening is based on a population-oriented, evidence-based model and addresses the acceptability and possible negative effects, including for people without the disorder. Known possible obstacles to a positive effect of screening include a short latent stage, lead time, overdiagnosis, lack of acceptability, poor performance of tests, and misclassification of outcome. Randomized trials of screening are challenging and need an adaptation of standards such as the Consolidated Standards of Reporting Trials (CONSORT). Simulating the effects of screening can allow the consideration of complex screening strategies and other options to help avoid biases related to treatment improvement and prevention success. Reconciling both perspectives is possible by considering that hypotheses underlying the former are prerequisites for the latter. From an evidence-based medicine and policy perspective, we suggest that recommending screening or prescribing a test is unethical if all possible obstacles are not documented by providing the best available evidence.

 ■ Mayo Clin Proc. 2016;91(11):1594-1605

From the University of Bordeaux (L.R.S., G.C., M.B., S.M.-P.) and INSERM (L.R.S., G.C., M.B., S.M.-P.), ISPED, Centre INSERM U-1219 Bordeaux Population Health, Bordeaux, France; CHU de Bordeaux, Pole de Sante Publique, Service d'Information Medicale, Bordeaux, France (L.R.S., G.C.); CHU de Bordeaux, Pole de Pediatrie, Place Amélie Raba Léon, Bordeaux, France (M.B.); and Institut Bergonié, Unité de Recherche et d'Épidémiologie Cliniques, Inserm CIC1401, Bordeaux, France (S.M.-P.).

Screening is the early detection of a latent disorder with the goal of starting treatment early to improve prognosis or intervening to avoid unwanted consequences of the disorder.[1] For example, screening for breast cancer with mammography is proposed in many developed countries to initiate treatment of localized tumors and hopefully achieve cure.[2] Screening of blood donors for infections is performed to avoid contamination of transfusion recipients.[3] Screening is fully described as a test proposed for a target population, and an organized process including a frequency of screening, a confirmed diagnostic procedure for those who tested positive, an early intervention for those confirmed, and an outcome to improve.

When targeted at a complete population, screening is called systematic screening.[4] Screening can be limited to an explicitly defined subgroup based on criteria such as age or other markers of high risk of developing the disorder or a poor prognosis; it is then called targeted or selective screening.[4] Systematic and targeted screening are 2 forms of what is also called mass screening.[4] Screening can also be proposed outside of any program to patients consulting a physician for an unrelated problem; this opportunistic approach is called case finding.[5] Whereas systematic and selective screening are usually performed within programs developed with a population-based, public health perspective, case finding depends on the physician's clinical, individual perspective on what is best for his or her patient.

The individual and population perspectives are perceived as opposing interests of patients and populations. Briefly, screening may seem like an obviously good idea from a clinical experience perspective because a perfect test would allow either the chance for reassurance or the possibility of improving prognosis with early intervention. Nevertheless, screening is a complex intervention from the population

Mayo Clin Proc. ■ November 2016;91(11):1594-1605 ■ http://dx.doi.org/10.1016/j.mayocp.2016.07.017
www.mayoclinicproceedings.org ■ © 2016 Mayo Foundation for Medical Education and Research

perspective,[6] therefore implying specific evaluations of the advantages and disadvantages for all individuals to whom the test will be proposed, including people with and without the disorder.

Herein, using examples of screening for chronic diseases (eg, cancer, human immunodeficiency virus infection, and psychosis) and in specific populations (eg, children and the elderly), we illustrate how screening is even more complex than usually perceived, and we try to reconcile the individual and population perspectives. We define and illustrate issues and biases that are potential obstacles to the implementation or positive effects of screening and that need to be addressed in clinical trials and modeling of the effects of screening; we expect to promote adequate methods to document the effects of screening and to improve evidence-based decisions regarding screening.

INDIVIDUAL PERSPECTIVE ON SCREENING

The Obvious Seduction of Early Detection

The individual perspective considers that screening anticipates an intervention for patients with a poor prognosis: (1) for many disorders, patients have a better prognosis or response to treatment at early compared with later stages[7]; (2) some patients seen late in the disorder process have a history of previous contacts with health services for manifestations possibly linked to early stages of the disorder.[8] For example, proposals of early detection of child abuse have been triggered by the observation of children seen at the stage of serious or lethal injuries but who had been previously seen by social services or in emergency departments.[9,10] In such contexts, it is logical to hypothesize that an active search for early stages of the disorder should improve the prognosis (Figure 1A).

This logical hypothesis explains why screening is so "popular" among physicians dealing in their practice with patients with a late diagnosis or a poor prognosis. This popularity is reflected by the US Preventive Services Task Force recommendations: as of July 2015, of 150 active recommendations, 106 (70.7%) are about screening. Still, the difficulty of screening is reflected by the many instances in which screening is contraindicated (23 grade D recommendations) or the evidence is insufficient (39 grade I recommendations) (for details, see Supplemental Table 1, available online at http://www.mayoclinicproceedings.org).

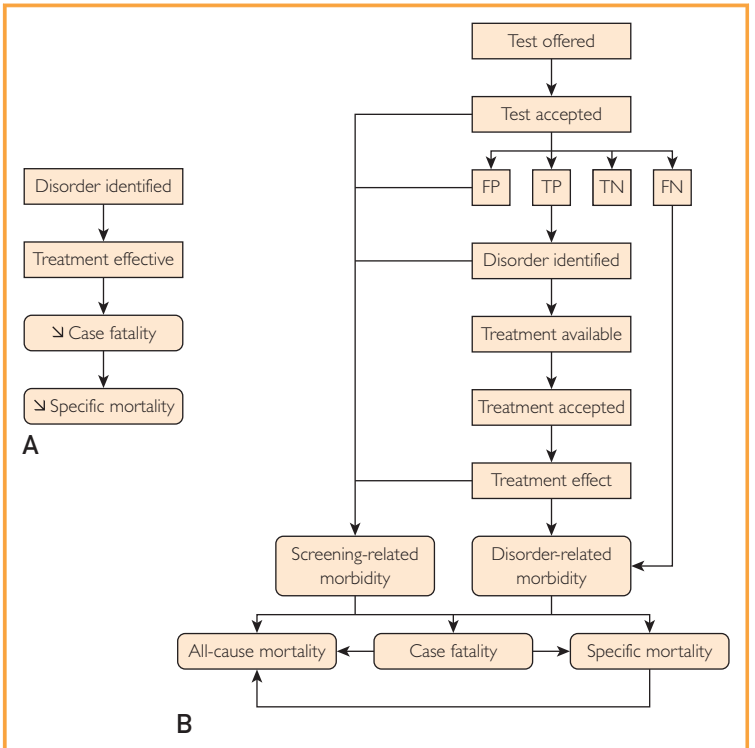

FIGURE 1. Hypotheses underlying screening according to the individual (A) and population (B) perspectives. FN = false-negative; FP = false-positive; TN = true-negative; TP = true-positive.

Difficulties in Judging the Clinical Effects of Screening

The hypothesis underlying the individual perspective is logical, but it can be flawed in three ways. First, early detection is possible only if there is a preclinical stage (obstacle 1) and if this stage is long enough for patients with a poor prognosis, if untreated, to have time to benefit from earlier treatment.[1] The preclinical stage is the period when the disorder is present but without any of the manifestations that usually trigger diagnosis. Many cancers have a potentially long preclinical phase, making screening relevant and implying repeated testing. In other instances, the preclinical phase is always short (bacterial contamination by blood transfusion)[11] or too poorly defined or documented to guarantee its existence (some forms of child abuse).[12]

Second, the anticipation of diagnosis related to the application of a test before any

symptom occurrence will seem to improve prognosis even if treatment is not better early compared with later. For example, a patient with a severe form of lung cancer who would have died rapidly after diagnosis[13] would seem to survive longer if diagnosis is performed a few months earlier as part of case finding. Because this lead time (obstacle 2)[14] is usually imperceptible at the individual level,[14] it is the main justification for conducting randomized trials of early vs delayed treatment.[15]

Third, the screening test might detect preclinical forms that would never have been diagnosed had the test not been performed. This is the phenomenon of overdiagnosis (obstacle 3),[16,17] that is, the identification of forms of the disorder with a spontaneous good prognosis. Overdiagnosis is likely to be very important, for instance, when screening for prostate cancer using prostate-specific antigen because many of the detected lesions, although meeting the pathologic criteria for invasive cancer, would never progress to severe or lethal stages.[18] Overdiagnosis, the extreme of what was called length bias,[14] is a major clinical problem because persons detected with these forms will have no benefit from an early treatment that they do not need (what is called overtreatment) and will experience the adverse effects of the test, the treatment, and of being labeled with the diagnosis.[19] Length bias is also accentuated by the fact that cases with a short preclinical phase, corresponding to forms of the disorder with the poorest prognosis, are most likely to be missed by a single test and, therefore, are more likely to be detected through clinical presentation than after screening.[14] Both lead time and overdiagnosis are major issues when evaluating screening at the population level because they tend to make screening seem better than it actually is.

POPULATION PERSPECTIVE ON SCREENING

The Obvious Complexity of Screening a Target Population

The population perspective considers that screening not only is an issue of early intervention for patients with a poor prognosis but also is proposed to a population that includes people with all forms of the disorder and a majority of people without the disorder.[4] A screening recommendation, systematic or case finding, should consider the acceptability and possible negative effects of the process and the perspective of people without the disorder (Figure 1B).[20]

Acceptability (obstacle 4) is a key issue in the individual and population perspectives because most people in the target population either do not need screening (they do not have the disorder) or do not feel that they need screening (they have the disorder but no manifestations). The latter individuals, who will be detected early (the true-positives), might not feel the need for a potentially aggressive treatment (screening for cancer)[20] or stigmatizing intervention (screening for decreased ability to drive an automobile in the elderly).[21] In the former group of people without the disorder, adverse effects of tests and, potentially, being false-positive are likely to be unacceptable.[22] Conversely, being falsely reassured as "in good health" after a first test (false-negative) can also negatively affect the acceptability of a later-round test and its results.[23]

Further Difficulties in Judging the Population Effects of Screening

From the population perspective, the recommendation to develop a screening program or policy or to introduce screening in clinical practice should occur only after demonstrating that the advantages for the population are greater than the disadvantages.[4] Lead time and overdiagnosis, already discussed from the individual perspective, are major issues in the evaluation of the population effects of screening and have to be considered in any comparison of consequences in populations with and without the program, policy, or practice.[24] Furthermore, overdiagnosis also occurs because persons without the disorder may be diagnosed as having incidental findings apart from the target condition; for example, screening for lung cancer by chest radiography or computed tomography also identifies asymptomatic nonevolutive lesions.[25]

Selection (obstacle 5), another issue, can occur at several steps of the process. The ability of screening programs to reach the target population has been documented to vary across socioeconomic strata.[26] For example, the populations included in randomized trials

Mayo Clin Proc. ■ November 2016;91(11):1594-1605 ■ http://dx.doi.org/10.1016/j.mayocp.2016.07.017
www.mayoclinicproceedings.org

of screening for lung cancer by low-dose computed tomography in heavy smokers were of higher socioeconomic levels and smoked less than the general population in the United States,[27] Denmark,[28] and the United Kingdom.[29] In established screening programs, acceptability depends on the type of test (higher for mammography[30] than for fecal occult blood tests[31]), and it is lower in high- than in low-risk groups (screening for human immunodeficiency virus infection).[32]

Detection issues (obstacle 6) are related to the performance of tests, reflected by their accuracy[20] and reliability.[33] A test developed and demonstrated to be accurate (ie, correctly classifying both people with and without the disorder) and reliable (ie, providing stable results when repeated in the same individuals or across observers) for individual diagnosis in the clinical setting might not be appropriate for screening. For a fixed sensitivity and specificity, the positive predictive value will be much lower and the negative predictive value higher in a screening context than in clinical diagnostic context. Indeed, the prevalence of the disorder is lower in populations targeted by screening programs or case finding than in a diagnostic context.[34] Thus, tests will usually be better to rule out the disorder than to identify early stages. Even with a high specificity, false-positive results and their negative consequences are likely to be considerably frequent. Favoring the specificity to reduce false-positive results is likely to decrease the sensitivity and, thus, the effectiveness of screening by increasing false-negative results. Moreover, achieving the same level of reliability and accuracy across all structures (eg, laboratories) and professionals (eg, radiologists) involved in a screening program is hard to achieve.[35] Reliability in real-life conditions needs to be assessed before implementation of a screening program and must be monitored in a quality assurance program.[36] Other issues, such as selecting a population with an adequate spectrum of preclinical disease and patients with competing disorders that mimic the target condition,[37] and considering the possible variation of sensitivity and specificity with prevalence,[34] are also important to consider when assessing the performance of tests in a screening context.

RESEARCH IMPLICATIONS

The Challenging Randomized Screening Trial

The previously mentioned issues justify the need to adequately assess both the positive and negative effects of screening on populations, not only on individuals with the disorder. Ideally, this assessment must be based on a properly conducted randomized community trial: usually, screening is proposed to half the target population, and the other half is allocated to what is considered usual care in the absence of screening.[38] Such trials must consider all issues expected in reports of randomized trials, as defined by the Consolidated Standards of Reporting Trials (CONSORT).[39] Many of the CONSORT items, however, are considered difficult to address[40] and, in our opinion, need to be adapted: these items concern the objective of the trial; the choice of an appropriate control arm; the choice of relevant yet measurable outcomes, including all aspects of the benefit to harm balance; and sample size calculation (Supplemental Table 2, available online at http://www.mayoclinicproceedings.org).

The first issue concerns the objective of the trial (obstacle 7). The challenge is to demonstrate that screening, considered in all its dimensions, is superior to usual care.[38] This implies that the compared strategies are defined and operationally described in (Table, items 5a and 5b) (1) the screening group (initial test, confirmation strategy, early intervention for those with the disorder, process of dealing with false-positive and false-negative results) and (2) the control group of usual care in the health care system where the study is conducted (accessibility to diagnosis and treatment). Demonstrating a favorable advantages to disadvantages ratio also implies that all relevant positive and negative effects be documented, in the methods and results, similarly for all participants, whether they have the disorder or not (Table, items 6a-6e) and in the screening and control groups.

The Controversial Definition of Relevant Outcomes

One major issue is what main outcome (obstacle 8) should be used to document the possible advantages of screening.[49] From

TABLE. Main Specific Methodological Issues to Be Addressed in Trials Assessing the Effects of Screening[a,b]

Section/topic	Item No.[c]	Checklist item	Examples of observed limitations or variations
Introduction			
Objective	2b	Identify the study as assessing whether the proposed screening program is superior to usual care in the absence of screening in terms of mortality	None of the 2 major trials of screening for prostate cancer with PSA expressed the notion of superiority in the objective[41,42]
Methods			
Interventions	5a	All steps of screening are defined and operationally described: initial test, confirmation strategy, early intervention, process of dealing with false-positive and false-negative results	In the NLST assessing CT screening for lung cancer in heavy smokers, the description of screening only provided information on the test; modalities of confirmation and treatment were left to the referring physician's initiative[27]
	5b	The comparison strategy of usual care in the absence of screening and specificities of the health care system are defined	In the NLST,[27] the control strategy included a chest radiograph, whereas the NELSON trial, the largest European trial, compared CT screening with usual care in the Netherlands and Belgium[43]
	5c	Any changes in screening or comparison interventions after the trial commenced, and reasons	In prostate cancer screening, the threshold for considering a PSA positive has been lowered over time[41,42]
Outcomes	6a	The prespecified main outcome is all-cause mortality or, if not possible, disorder-specific or incidence-based mortality[d]	The use of suicide ideation as the main outcome in a trial of screening for risk of suicide precludes concluding on the actual effects on risk[44]
	6b	Deaths and their causes, and all other relevant outcomes, are assessed similarly in all groups	In the ERSPC,[42] verification of causes of death in patient records was recommended but varied across countries
	6c	The prespecified length of follow-up is long enough	In screening for many cancers, it is estimated that the full benefit of a program would be observed only after 20 y of follow-up[45]; the use of surrogate end points might help shorten the follow-up needed[46]
	6d	Practical modalities of follow-up are similar in all groups	No standardization within or across studies of screening for asymptomatic carotid artery stenosis[47]
	6e	Key secondary outcomes (disorder-related morbidity, compliance with tests and intervention, contamination, adverse effects of tests and intervention, false-positives, false-negatives; overdiagnosis, psychological impact) are defined and monitored similarly in all groups	The 2 major trials of PSA screening for prostate cancer have not documented most secondary outcomes for all individuals in all groups[41,42]
Sample size	7a	Hypotheses underlying sample size calculation are reasonable	In the Dante trial of lung cancer screening, the expected 50% reduction of mortality is not compatible with current hypotheses[48]
Statistical methods	12a	Analysis is intent to screen, and accounting for actual length of follow-up	In the ERSPC, the authors considered excluding from the analysis centers with a large proportion of contamination in the control group and performed their final analysis in a subgroup of men, therefore breaking the randomization[42]
Discussion			
Interpretation	22	Interpretation consistent with results, balancing actual benefits and all aspects of harm and feasibility issues	In screening for cervical, prostate, or breast cancer, there is a gap between evidence regarding efficacy and harm[16]

[a]CT = computed tomographic; ERSPC = European Randomized Study of Screening for Prostate Cancer; NELSON = European NEderlands-Leuvens Screening ONderzoek; NLST = National Lung Screening Trial; PSA = prostate-specific antigen.

[b]Adapted from the Consolidated Standards of Reporting Trials (CONSORT) Statement (see Schulz et al[39] for original CONSORT items and the full list of items adapted to screening trials in Supplemental Table 2).

[c]Items of the original CONSORT or proposed numbers; see the complete original list in Supplemental Table 2.

[d]See the text for a discussion of relevance and the limits of all-cause, disorder-specific, and incidence-based mortality.

both the individual and population perspectives (Figure 1), the main outcome is the most severe consequence for which it is hypothesized that screening can improve prognosis. For example, when screening for inability to drive an automobile in the elderly, the main outcome must include all occurrences of collisions with at least a severe or lethal injury to either the driver or other road users.[21] For most chronic diseases for which screening has been proposed, the main outcome is death,[20] but it is debated what is the most relevant parameter[49] (Table, items 2b and 6a).

Case fatality is relevant from the individual perspective (Figure 1A) in trials comparing early and delayed treatment, but it is not as relevant from the population perspective (Figure 1B) because it considers only people with the disorder. Furthermore, in the presence of overdiagnosis, case fatality will seem lower in screened than in unscreened populations, and the effect of lead-time bias will be increased. Disorder-specific mortality is the parameter estimated in most cancer screening trials.[49] Proponents of disorder-specific mortality argue that, first, it is the parameter most directly related to an improvement in prognosis and, second, it is unrealistic to hope for a positive effect to be reflected by a change in all-cause mortality, unless follow-up is very long[49,50]; the huge sample size needed to show an improvement in all-cause mortality, a major obstacle in trials based on all-cause mortality, is another issue discussed by proponents of disorder-specific mortality.[50,51] It can also be argued that whenever screening can reduce the incidence of the disease, for instance when the diagnostic procedure allows removal of precancerous lesions (eg, removal of polyps during colonoscopy in colonic cancer screening),[52] disorder-specific mortality can be an appropriate outcome parameter.

Proponents of all-cause mortality,[53] including the authors, argue that the follow-up needs, anyway, to be very long because screening cannot have a positive effect before a substantial number of patients in the preclinical phase are identified and the resulting improved prognosis results in a change in disorder-specific mortality.[45] Furthermore, it can be argued, from the population perspective, that only all-cause mortality can reflect the most severe consequences occurring in people with and without the disorder, resulting from disorder- and screening-related morbidity.[53] It is also relevant because a decrease in disorder-specific mortality implies that survivors remain subject to competitive causes of deaths, especially in aged populations.[54,55]

Another reason to prefer all-cause to disorder-specific mortality is that the latter is easily subject to misclassification.[53] Indeed, the validity of disorder-specific mortality assumes that cause of death is accurately determined and that screening has negligible effects on other causes of death. Not meeting these assumptions results in 2 biases.[38,53] (1) Sticky diagnosis bias can occur in a randomized trial if validation of outcomes is not adequately standardized, resulting in the screening group in some deaths from other causes being unduly attributed to the disorder; this bias underestimates the positive effect of screening. (2) Slippery linkage bias occurs because deaths due to adverse effects of screening-related procedures are not counted in disorder-related mortality; this bias overestimates the positive impact of screening. This lack of consideration of screening-related morbidity is another possible argument in favor of all-cause over disorder-specific mortality.[53]

To resolve the controversy regarding all-cause and disorder-specific mortality, some authors have suggested assessing the real-life impact of screening based on excess mortality in those with the disorder.[56] This approach consists of comparing the observed number of deaths in screening participants with the expected number derived from disease-specific mortality in the general population. However, this approach is also subject to strong overestimation of the number of expected deaths and of the benefit of screening if it is not based on an incidence-based mortality method, which standardizes for time of diagnosis and considers that screening participants do not have a date of previous diagnosis.[56] Still, because the method can be derived from data available in cancer and death registries, it probably gives a good reflection of the actual effect of a screening program in circumstances in which a trial is not feasible or not completed.[56]

Whatever the option, the population perspective implies that investigators should thoroughly document all relevant positive and negative outcomes of the disorder and

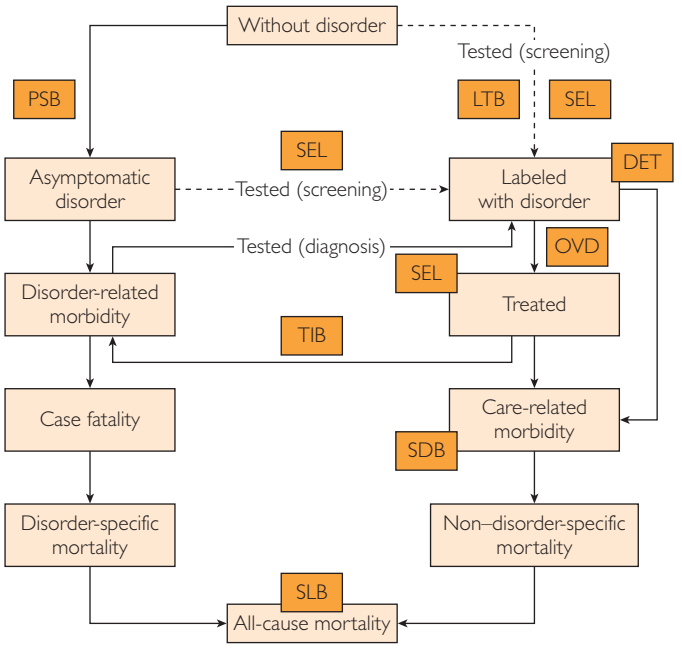

FIGURE 2. The course of a disorder and biases to consider in randomized trials, modeling, or cohort studies of the effects of screening programs. The solid lines indicate the possible transitions in the absence of screening; dashed lines, the additional possible transitions in the presence of screening. Nonspecific mortality also includes mortality related to competitive causes, which are neither disorder specific nor care related. DET = detection bias; LTB = lead-time bias; OVD = overdiagnosis; PSB = prevention success bias; SDB = sticky diagnosis bias; SEL = selection bias; SLB = slippery linkage bias; TIB = treatment improvement bias (see definitions in text).

screening. This is a huge challenge, as reflected by the fact that systematic reviews on screening usually identify limitations of published trials: focus on the initial test rather than on the program,[57] insufficient statistical power,[58] poorly compensated by attempts to mix studies or centers with different populations or screening,[42,59] insufficient follow-up,[60] lack of standardization of outcome assessment[61] and follow-up procedures,[47] unclear definition and poor documentation of secondary outcomes,[17,57,61] and major changes in the interventions or interventions becoming obsolete.[15,47] It is therefore argued that good screening trials are difficult to conduct,[38] and our proposed specifications of CONSORT (Table) make the challenge even bigger. Nevertheless, trials are important because they are the gold standard to document the effect of any intervention.

The Possibility of Simulating the Effects of Screening

Given the difficulties of conducting adequate trials and the delays before relevant results are available, some authors have suggested that simulations using, for example, Markov models[62] could be used to assess the effects of screening.[63] Such simulations can consider all steps of the natural history of the disorder and can compare the values of case fatality and of disorder-specific and all-cause mortality in the presence and absence of screening (Figure 2) for various lag times, screening intervals, starting ages, stopping ages, and ascertainment algorithms. The proposed model, however, must consider that many methodological choices are specific to a given disorder/testing procedure.

One strength of models is that they can be fed by data from many sources (randomized trials; observational studies such as cohort, accuracy, and repeatability studies; and expert judgments using formal consensus techniques) and can simulate the effect of uncertainties in estimates of relevant parameters.[64] For example, a Markov model could be used to simulate the impact of a hypothetical program screening for pediatric abusive head trauma despite an extremely scarce literature on the performance of possible tests and the definition and length of the preclinical phase.[65] Similar models have been developed, for example, for screening for psychosis,[66] age-related macular degeneration,[67] and prostate cancer.[68] Models are also potentially appropriate to consider the potential effects of issues such as selection, detection, lead time, overdiagnosis, sticky diagnosis, and slippery linkage.[69]

Another advantage is that models can complement trials by simulating more complex questions than can be addressed in trials.[70] The typical trial compares one screening strategy with one control group; there are a few instances of trials comparing several screening strategies with a control group. For example, the Multicenter Italian Lung Detection trial compared 2 rhythms of screening for lung cancer by low-dose computed tomography, annually or every other year, with the absence of screening.[71] Such trials are even more difficult to conduct than 2-group trials and are

Mayo Clin Proc. ■ November 2016;91(11):1594-1605 ■ http://dx.doi.org/10.1016/j.mayocp.2016.07.017
www.mayoclinicproceedings.org

less likely to meet all the expectations of CONSORT.[39] It is, therefore, difficult to imagine larger trials comparing more screening strategies or screening with other interventions (prevention or treatment).

The Need to Compare Screening With Other Options

Prevention and treatment are interventions that should be considered alternatives to screening. Failure to consider these alternatives when assessing screening can result in 2 issues, similar to the classic history bias in evaluation research[72]: treatment improvement bias (obstacle 9) and prevention success bias (obstacle 10). Treatment improvement bias occurs when progress has occurred between the start and the final analysis of a trial. For example, a dramatic improvement has been observed in the prognosis of breast cancer[73] owing to the availability of targeted treatments.[74] Were trials of mammography screening to be started in the present decade, rather than 40 years ago, it is likely that the observed positive effect of screening, if any, would be small.[15] Treatment improvement bias decreases the disorder-specific mortality in the usual care group; it also dramatically decreases the power of trials, making it even less likely to demonstrate a statistically and clinically significant difference between screening and usual care.[45]

Prevention success bias is a consequence of a decrease in the incidence of the disorder owing to effective preventive interventions. For example, the positive effect of screening for driving under the influence of alcohol is lowered by the effectiveness of laws.[75] In a trial of prevention vs screening, the result will favor the prevention arm. If a trial compares screening with no intervention in a population in which prevention has been disseminated, then the trial will lack power. Treatment improvement bias and prevention success bias are more likely to occur when the length of follow-up needed to expect a positive effect of screening is long.[72]

Another effect of the decreased incidence of the disorder is to worsen detection and selection issues: fewer people with the disorder increases the likelihood of false-positive results and their consequences.[34] For example, the model comparing prevention, screening, and the combination of prevention and screening with no intervention to deal with pediatric abusive head trauma showed that prevention would be effective in decreasing mortality but that both screening and the combination of screening and prevention could actually increase mortality, mostly because stigmatization of parents—by increasing parental poor self-esteem, depression, and social isolation—might increase the risk of abuse.[65]

RECONCILING PERSPECTIVES AND MEETING ETHICAL CHALLENGES

Reconciling individual and population health perspectives is possible by considering that verification of hypotheses underlying the former is a prerequisite for the latter (Figure 3). The only subtlety is that a screening program or policy is justified only if the disorder is a public health issue (Figure 3, box 1), that is, that its consequences are important in terms of mortality, severe morbidity, or costs to society.[1,4,76] Mass screening is a public health option, but issues common to the individual and population perspectives (a well-defined and long preclinical phase, accurate and reliable tests in the preclinical phase, and early treatment demonstrated to be better than later treatment) have to be addressed whether screening is systematic, targeted, or case finding.[63] Overdiagnosis, acceptability, selection, and detection issues are also relevant to both perspectives.

Furthermore, we argue that consideration of acceptability (of tests, early treatment, and errors) is a foundation of evidence-based medicine and evidence-based policy.[77-79] The challenge of evidence-based medicine is to integrate scientific evidence, the values and perspectives of the patient/person on his or her health and the intervention, the experience of the physician, and the availability of health care resources.[78] Addressing the patient's values and perspectives is central to acceptability (Figure 3, box 6); individuals with different expectations from life and different coping may, indeed, give different values to benefits and harms. Nevertheless, the explanation of options (to screen or not to screen) and their consequences requires a previous demonstration that the advantages of screening the target population are larger than the disadvantages (Figure 3, box 5). Also, it should be the experience of physicians that

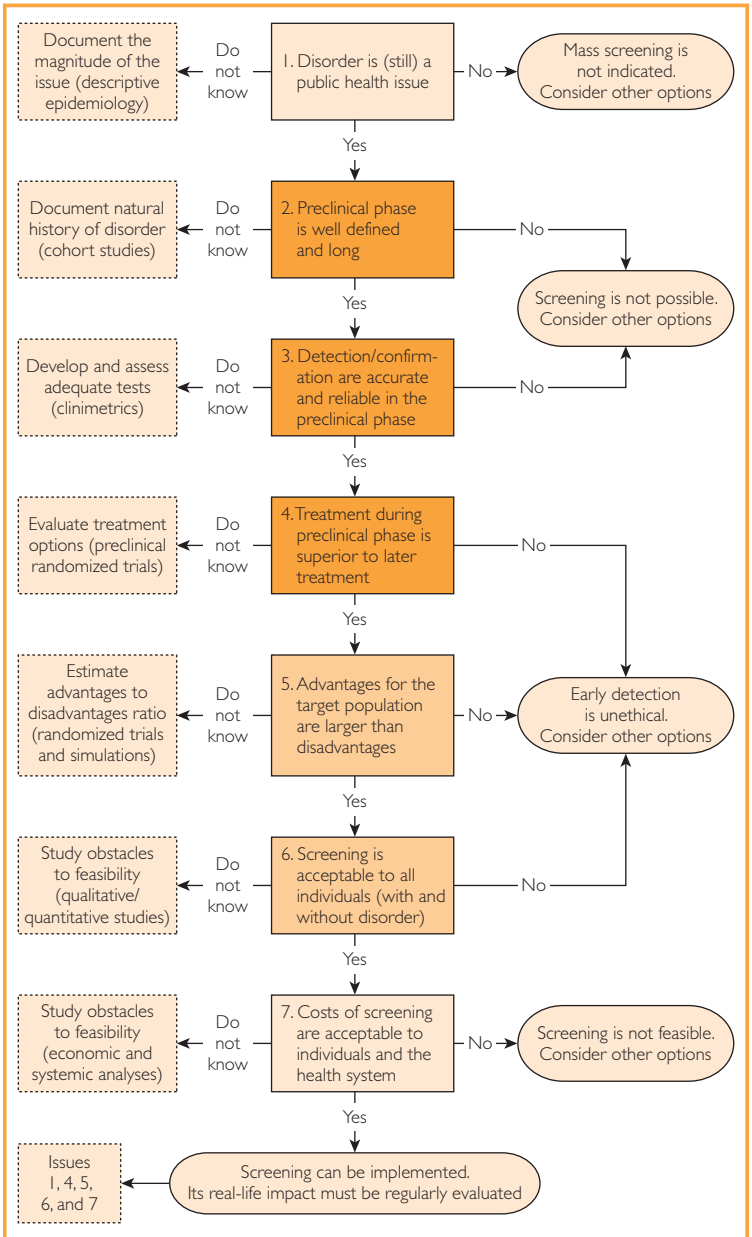

FIGURE 3. Seven key issues reconciling the individual and population perspectives on screening and their consequences for decisions (ovals) and implications for research (dotted-line boxes). Dark orange boxes indicate the only issues traditionally considered in the individual model; light orange boxes are suggested to be part of an evidence-based model (see the text for details).

most tests are negatives, most positive tests are false-positives, and screening-related morbidity can be frequent for some disorders.[63] Thus, issues of effectiveness and acceptability belong to both perspectives. Finally, screening is feasible only if costs are acceptable for the individual and the health system (Figure 3, box 7). However, any evaluation of costs should not be limited to the cost of testing but should include costs directly or indirectly related to treatment and errors[6]; costs should also be reevaluated during the program.

We further suggest that recommending screening or prescribing a test is unethical when all issues have not been properly addressed.[18,80-82] Addressing all issues by adequate research, however, is a huge challenge.[83] Nevertheless, we submit that a positive benefit-to-harm balance should be better documented, whether through better-conducted randomized controlled trials, modeling, or both, before any recommendation of screening. One has to accept that although screening is the responsibility of health care professionals, relevant research documenting the benefits and harms of screening is not limited to the clinical setting; integrating epidemiologic, clinical, and behavioral and social sciences, mathematical modeling and economics can properly be addressed by multidisciplinary research groups familiar with the evaluation of complex public health interventions.[6]

In that context, one should question whether it is ethical to continue trials that will never provide the information needed to guide recommendations.[38] We believe that modeling is a promising approach because models can be used to estimate the potential effects of proposed screening and, using data from cohorts of populations where screening has been implemented, estimate the real-life effects of existing programs.[70] Nevertheless, we suggest that trials, if better designed and monitored to appropriately consider key issues discussed herein, remain important sources of evidence to inform benefit and harm. Modeling and trials are not in opposition; both should be used whenever screening is considered. Finally, we suggest that better formalization of reflections on screening, as proposed in this article, will make it easier for clinicians and other decision makers, as well as researchers, to define best-evidence practices and policies. One has to accept, however, that not prescribing screening is often ethically the best option and that research, prevention, and improved treatment should always be considered as alternatives to screening.

Mayo Clin Proc. ■ November 2016;91(11):1594-1605 ■ http://dx.doi.org/10.1016/j.mayocp.2016.07.017
www.mayoclinicproceedings.org

CONCLUSION

Screening—the early detection of a latent disorder to improve prognosis—is an important public health intervention for which the individual and population perspectives can be reconciled. Many obstacles and biases are well documented, and most apply equally to the individual and population perspectives and can be addressed in randomized controlled trials or modeling of the effects of screening, provided these studies are adapted to the specific disorder and testing procedures. The individual and population perspectives can both benefit from stronger emphasis on an evidence-based model of screening, in which the interests and values of all individuals, whether they have or do not have the targeted disorder, are considered.

SUPPLEMENTAL ONLINE MATERIAL

Supplemental material can be found online at: http://www.mayoclinicproceedings.org. Supplemental material attached to journal articles has not been edited, and the authors take responsibility for the accuracy of all data.

Abbreviations and Acronynms: **CONSORT** = Consolidated Standards of Reporting Trials; **CT** = computed tomographic; **DET** = detection bias; **ERSPC** = European Randomized Study of Screening for Prostate Cancer; **FN** = false-negative; **FP** = false-positive; **LTB** = lead-time bias; **NELSON** = European NEderlands-Leuvens Screening ONderzoek; **NLST** = National Lung Screening Trial; **OVD** = overdiagnosis; **PSA** = prostate-specific antigen; **PSB** = prevention success bias; **SDB** = sticky diagnosis bias; **SEL** = selection bias; **SLB** = slippery linkage bias; **TIB** = treatment improvement bias; **TN** = true-negative; **TP** = true-positive

Potential Competing Interests: Drs Salmi, Coureau, and Mathoulin-Pélissier have performed expert consulting for the French Haute Autorité de Santé and have received honoraria from that public institution for work on screening for prostate cancer (Dr Salmi) and screening for lung cancer (Drs Salmi, Coureau, and Mathoulin-Pélissier).

Correspondence: Address to L. Rachid Salmi, MD, PhD, Institut de Santé Publique, d'Épidémiologie et de Développement, Université de Bordeaux, 146 rue Léo-Saignat, F-33076 Bordeaux cedex, France (rachid.salmi@isped.u-bordeaux2.fr).

REFERENCES

1. Wilson JMG, Jungner G. *The Principles and Practice of Screening for Disease*. Geneva, Switzerland: World Health Organization; 1968.
2. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2009; 151(10):727-737.
3. Salunkhe V, van der Meer PF, de Korte D, Seghatchian J, Gutiérrez L. Development of blood transfusion product pathogen reduction treatments: a review of methods, current applications and demands. *Transfus Apher Sci*. 2015;52(1):19-34.
4. Morrison AS. *Screening in Chronic Disease*. New York, NY: Oxford University Press; 1985.
5. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology: The Essentials*. Baltimore, MD: Williams & Wilkins; 1982.
6. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ*. 2000;321(7262):694-696.
7. Rosenberg J, Chia YL, Plevritis S. The effect of age, race, tumor size, tumor grade, and disease stage on invasive ductal breast cancer survival in the U.S. SEER database. *Breast Cancer Res Treat*. 2005;89(1):47-54.
8. Lyratzopoulos G, Vedsted P, Singh H. Understanding missed opportunities for more timely diagnosis of cancer in symptomatic patients after presentation. *Br J Cancer*. 2015;112(suppl 1): S84-S91.
9. Tursz A, Crost M, Gerbouin-Rérolle P, Cook J. Underascertainment of child abuse fatalities in France: retrospective analysis of judicial data to assess underreporting of infant homicides in mortality statistics. *Child Abuse Negl*. 2010;34(7):534-544.
10. Thorpe EL, Zuckerbraun NS, Wolford JE, Berger RP. Missed opportunities to diagnose child physical abuse. *Pediatr Emerg Care*. 2014;30(11):771-776.
11. Perez P, Salmi LR, Folléa G, et al. Determinants of transfusion-associated bacterial contamination: results of the French BACTHEM case-control study. *Transfusion*. 2001;41(7):862-872.
12. Lauterbach D, Armour C. Symptom trajectories among child survivors of maltreatment: findings from the Longitudinal Studies of Child Abuse and Neglect (LONGSCAN). *J Abnorm Child Psychol*. 2015;44(2):369-379.
13. Wao H, Mhaskar R, Kumar A, Miladinovic B, Djulbegovic B. Survival of patients with non-small cell lung cancer without treatment: a systematic review and meta-analysis. *Syst Rev*. 2013;2:10.
14. Schwartz M. Estimates of lead time and length bias in a breast cancer screening program. *Cancer*. 1980;46(4):844-851.
15. Bell KJ, Bossuyt P, Glasziou P, Irwig L. Assessment of changes to screening programmes: why randomisation is important. *BMJ*. 2015;350:h1566.
16. Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ*. 2015;350:g7773.
17. Carter SM, Rogers W, Heath I, Degeling C, Doust J, Barratt A. The challenge of overdiagnosis begins with its definition. *BMJ*. 2015;350:h869.
18. Carter SM, Williams J, Parker L, et al. Screening for cervical, prostate, and breast cancer: interpreting the evidence. *Am J Prev Med*. 2015;49(2):274-285.
19. Morgan DJ, Brownlee S, Leppin AL, et al. Setting a research agenda for medical overuse. *BMJ*. 2015;351:h4534.
20. Pinsky PF. Principles of cancer screening. *Surg Clin North Am*. 2015;95(5):953-966.
21. Leproust S, Lagarde E, Salmi LR. Systematic screening for unsafe driving due to medical conditions: still debatable. *BMC Public Health*. 2008;8:27.
22. Bond M, Pavey T, Welch K, et al. Systematic review of the psychological consequences of false-positive screening mammograms. *Health Technol Assess*. 2013;17(11):1-170.
23. Petticrew M, Sowden A, Lister-Sharp D. False-negative results in screening programs: medical, psychological, and other implications. *Int J Technol Assess Health Care*. 2001;17(2): 164-170.
24. Duffy SW, Nagtegaal ID, Wallis M, et al. Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. *Am J Epidemiol*. 2008;168(1):98-104.

25. Demirci I, Herold S, Kopp A, Flaßhove M, Klosterhalfen B, Janßen H. Overdiagnosis of a typical carcinoid tumor as an adenocarcinoma of the lung: a case report and review of the literature. *World J Surg Oncol*. 2012;10:19.
26. Wools A, Dapper EA, Leeuw JR. Colorectal cancer screening participation: a systematic review. *Eur J Public Health*. 2015; 26(1):158-168.
27. Aberle DR, Adams AM, Berg CD, et al. Baseline characteristics of participants in the randomized National Lung Screening Trial. *J Natl Cancer Inst*. 2010;102(23):1771-1779.
28. Hestbech MS, Siersma V, Dirksen A, Pedersen JH, Brodersen J. Participation bias in a randomised trial of screening for lung cancer. *Lung Cancer*. 2011;73(3):325-331.
29. McRonald FE, Yadegarfar G, Baldwin DR, et al. The UK Lung Screen (UKLS): Demographic profile of first 88,897 approaches provides recommendations for population screening. *Cancer Prev Res*. 2014;7(3):362-371.
30. Lagerlund M, Sontrop JM, Zackrisson S. Psychosocial factors and attendance at a population-based mammography screening program in a cohort of Swedish women. *BMC Womens Health*. 2014;14(1):33.
31. Hobbs FD, Cherry RC, Fielding JW, Pike L, Holder R. Acceptability of opportunistic screening for occult gastrointestinal blood loss. *BMJ*. 1992;304(6825):483-486.
32. Irwin KL, Valdiserri RO, Holmberg SD. The acceptability of voluntary HIV antibody testing in the United States: a decade of lessons learned. *AIDS*. 1996;10(14):1707-1717.
33. Streiner DL, Norman GR. *Health Measurement Scales. A Practical Guide to Their Development and Use*. 3rd ed. Oxford, England: Oxford University Press; 2003.
34. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ*. 2013;185(11):E537-E544.
35. Soh BP, Lee W, Kench PL, et al. Assessing reader performance in radiology, an imperfect science: lessons from breast screening. *Clin Radiol*. 2012;67(7):623-628.
36. Elfström KM, Arnheim-Dahlström L, von Karsa L, Dillner J. Cervical cancer screening in Europe: quality assurance and organisation of programmes. *Eur J Cancer*. 2015;51(8):950-968.
37. Montori VM, Wyer P, Newman TB, Keitz S, Guyatt G; for the Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine, 5: the effect of spectrum of disease on the performance of diagnostic tests. *CMAJ*. 2005;173(4):385-390.
38. Juffs HG, Tannock IF. Screening trials are even more difficult than we thought they were. *J Natl Cancer Inst*. 2002;94(3): 156-157.
39. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
40. Cals JW, van Amelsvoort LG, Kotz D, Spigt MG. CONSORT 2010 Statement: unfinished update? *J Clin Epidemiol*. 2011; 64(6):579-582.
41. Andriole GL, Crawford ED, Grubb RL, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med*. 2009;360(13):1310-1319.
42. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med*. 2009;360(13):1320-1328.
43. van Iersel CA, de Koning HJ, Draisma G, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). *Int J Cancer*. 2007;120(4):868-874.
44. Crawford MJ, Thana L, Methuen C, et al. Impact of screening for risk of suicide: randomised controlled trial. *Br J Psychiatry*. 2011;198(5):379-384.
45. Hanley JA, McGregor M, Liu Z, Strumpf EC, Dendukuri N. Measuring the mortality impact of breast cancer screening. *Can J Public Health*. 2013;104(7):e437-e442.
46. Cuzick J, Cafferty FH, Edwards R, Møller H, Duffy SW. Surrogate endpoints for cancer screening trials: general principles and an illustration using the UK Flexible Sigmoidoscopy Screening Trial. *J Med Screen*. 2007;14(4):178-185.
47. Jonas DE, Feltner C, Amick HR, et al. Screening for asymptomatic carotid artery stenosis: a systematic review and meta-analysis for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;161(5):336-346.
48. Infante M, Lutman FR, Cavuto S, et al. Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer*. 2008;59(3):355-363.
49. Hanley JA. Measuring mortality reductions in cancer screening trials. *Epidemiol Rev*. 2011;33:36-45.
50. Tabar L, Duffy SW, Yen MF, et al. All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. *J Med Screen*. 2002;9(4): 159-162.
51. Steele R, Brewster D. Should we use total mortality rather than cancer specific mortality to judge cancer screening programmes? no. *BMJ*. 2011;343:d6397.
52. Schoen RE, Pinsky PF, Weissfeld JL, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *N Engl J Med*. 2012;366(25):2345-2357.
53. Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst*. 2002;94(3):167-173.
54. Seidman H, Silverberg E, Bodden A. Probabilities of eventually developing and of dying of cancer (risk among persons previously undiagnosed with the cancer). *Cancer*. 1978; 28(1):33-46.
55. Jemal A, Simard EP, Dorell C, et al. Annual Report to the Nation on the Status of Cancer, 1975—2009, featuring the burden and trends in human papillomavirus (HPV)—associated cancers and HPV vaccination coverage levels. *J Natl Cancer Inst*. 2013;105(3):175-201.
56. Brenner H, Hoffmeister M, Jansen L. Comparisons of colorectal cancer mortality between screening participants and the general population are strongly biased unless an incidence-based mortality approach is used. *J Clin Epidemiol*. 2014;67(2): 184-189.
57. Zakher B, Cantor AG, Pappas M, Daeges M, Nelson HD. Screening for gonorrhea and chlamydia: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;161(12):884-894.
58. Cantor AG, Bougatsos C, Dana T, Blazina I, McDonagh M. Routine iron supplementation and screening for iron deficiency anemia in pregnancy: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2015;162(8):566-576.
59. Oudkerk M, Heuvelmans MA. Screening for lung cancer by imaging: the Nelson study. *JBR-BTR*. 2013;96(3):163-166.
60. LeBlanc ES, Zakher B, Daeges M, Pappas M, Chou R. Screening for vitamin D deficiency: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2015;162(2): 109-122.
61. Lin JS, O'Connor E, Rossom RC, Perdue LA, Eckstrom E. Screening for cognitive impairment in older adults: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2013;159(9):601-612.
62. Chiu SY, Malila N, Yen AM, Anttila A, Hakama M, Chen HH. Analytical decision model for sample size and effectiveness projections for use in planning a population-based randomized controlled trial of colorectal cancer screening. *J Eval Clin Pract*. 2011;17(1):123-129.
63. Etzioni R, Gulati R, Cooperberg MR, Penson DM, Weiss NS, Thompson IM. Limitations of basing screening policies on screening trials: the US Preventive Services Task Force and prostate cancer screening. *Med Care*. 2013;51(4):295-300.
64. Hamra G, MacLehose R, Richardson D. Markov Chain Monte Carlo: an introduction for epidemiologists. *Int J Epidemiol*. 2013;42(2):627-634.

65. Bailhache M, Bénard A, Salmi LR. Simulation of the impact of programs for prevention and screening of Pediatric Abusive Head Trauma. *J Neurotrauma*. 2016;33(14):1397-1403.
66. Cougnard A, Salmi LR, Salamon R, Verdoux H. A decision analysis model to assess the feasibility of the early detection of psychosis in the general population. *Schizophr Res*. 2005; 74(1):27-36.
67. Tamura H, Goto R, Akune Y, Hiratsuka Y, Hiragi S, Yamada M. The clinical effectiveness and cost-effectiveness of screening for age-related macular degeneration in Japan: a Markov modeling study. *PLoS One*. 2015;10(7):e133628.
68. Mühlberger N, Kurzthaler C, Iskandar R, et al. The ONCOTYROL prostate cancer outcome and policy model: effect of prevalence assumptions on the benefit-harm balance of screening. *Med Decis Making*. 2015;14(6):758-772.
69. Seigneurin A, Labarère J, Duffy SW, Colonna M. Overdiagnosis associated with breast cancer screening: a simulation study to compare lead-time adjustment methods. *Cancer Epidemiol*. 2015;39(6):1128-1135.
70. Karnon J, Brennan A, Akehurst R. Decision modeling to inform decision making: seeing the wood for the trees. *Med Decis Making*. 2010;10(3):E20-E22.
71. Pastorino U, Rossi M, Rosato V, et al. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. *Eur J Cancer Prev*. 2012;21(3):308-315.
72. Campbell DT, Stanley JC. *Experimental and Quasi-experimental Designs for Research*. Boston, MA: Houghton Mifflin Co; 1963.
73. Gøtzsche PC, Jørgensen KJ. The Breast Screening Programme and misinforming the public. *J R Soc Med*. 2011;104(9):361-369.
74. Vera-Badillo FE, Templeton AJ, de Gouveia P, et al. Androgen receptor expression and outcomes in early breast cancer: a systematic review and meta-analysis. *J Natl Cancer Inst*. 2014; 106(1):djt319.
75. Voas RB, Fell JC. Preventing impaired driving opportunities and problems. *Alcohol Res Health*. 2011;34(2):225-235.
76. Ahern J, Hubbard A, Galea S. Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. *Am J Epidemiol*. 2009;169(9):1140-1147.
77. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71-72.
78. Charles C, Gafni A, Freeman E. The evidence-based medicine model of clinical practice: scientific teaching or belief-based preaching? *J Eval Clin Pract*. 2011;17(4):597-605.
79. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*. 1992;268(17):2420-2425.
80. Scott PA. Unsupervised self-testing as part public health screening for HIV in resource-poor environments: some ethical considerations. *AIDS Behav*. 2014;18(suppl 4):S428-S444.
81. van Dam L, Bretthauer M. Ethical issues in colorectal cancer screening. *Best Pract Res Clin Gastroenterol*. 2014;28(2): 315-326.
82. Petrova D, Garcia-Retamero R, Cokely ET. Understanding the harms and benefits of cancer screening: a model of factors that shape informed decision making. *Med Decis Making*. 2015; 35(7):847-858.
83. Rossi PG, Ferroni E, Barca A, et al. Quality appraisal of documents producing recommendations for breast, colorectal and cervical cancer screening. *Epidemiol Biostat Public Health*. 2014;11(3):e9062.