Guidelines for early detection of breast cancer in Brazil. I – Development methods

Diretrizes para detecção precoce do câncer de mama no Brasil. I – Métodos de elaboração

Directrices para la detección precoz del cáncer de mama en Brasil. I – Métodos de elaboración

Arn Migowski [1,2]
Airton Tetelbom Stein [3]
Camila Belo Tavares Ferreira [1]
Daniele Masterson Tavares Pereira Ferreira [4]
Paulo Nadanovsky [5,6]

Abstract

Clinical guidelines are traditionally drafted by expert consensus. The benefits of mammographic screening have been questioned in recent years, owing to biases detected in the clinical trials that popularized its widespread use. Meanwhile, growing body of evidence on harms associated with mammographic screening also required a new approach, taking into account the uncertainties on the benefits and a balance between the gains and possible harms from screening. This article discusses the development of the new guidelines for early detection of breast cancer in Brazil, with details on the drafting methods and implications for the new recommendations. The new methodology features systematic literature reviews, assessment of the validity of the evidence, and the balance between each intervention's risks and benefits, ensuring greater transparency, reproducibility, and validity in the drafting process. The new guidelines also include recommendations for cases with suspicious signs and symptoms. The authors provide a detailed discussion of the advantages of the approach as compared to the traditional expert consensus model, as well as the methods' limitations and disadvantages. They also address the implications of various decisions, such as choices on study designs, screening effectiveness outcomes, definition of overdiagnosis, and methods for calculation.

Breast Neoplasms; Early Detection of Cancer; Mass Screening; Mammography; Practice Guidelines as Topic

Correspondence
A. Migowski
Instituto Nacional de Câncer José Alencar Gomes da Silva.
Rua Marquês de Pombal 125, Rio de Janeiro, RJ 20230-240, Brasil.
arnmigowski@yahoo.com.br

[1] *Instituto Nacional de Câncer José Alencar Gomes da Silva, Rio de Janeiro, Brasil.*
[2] *Instituto Nacional de Cardiologia, Rio de Janeiro, Brasil.*
[3] *Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brasil.*
[4] *Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.*
[5] *Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.*
[6] *Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.*

Introduction

Clinical guidelines essentially aim to assist evidence-based decision-making, both for health professionals and health system users and policymakers. Traditionally, clinical guidelines, also known as clinical practice guides or clinical protocols, are drafted by expert consensus or based on clinical protocols from what are considered excellent services. Even in guidelines with incipient incorporation of some evidence-based features, such as a certain formalization of the literature search process and the classification of levels of evidence, such evidence is often chosen by convenience in order to confirm either prevailing practice or the opinion of the group drafting the recommendations.

The Brazilian Ministry of Health has made an effort in recent years to produce clinical guidelines and replace the country's hegemonic model, based mainly on expert opinions and narrative literature reviews. This effort resulted in the creation of the so-called "Clinical Protocols and Treatment Guidelines", which have succeeded both in increasing the country's prevailing quality standard for guidelines and the publication of a wide range of guidelines on diverse themes in a short space of time, generally broad enough to cover a major portion of the line of care for each respective disease. Even so, the analysis of a random sample of these Ministry of Health protocols using the AGREE II instrument showed that there is still considerable room for improvement in the drafting process [1]. The new guidelines for early detection of breast cancer in Brazil used a pioneering development method for the country, based on systematic literature reviews and a risk-benefit analysis for each intervention according to the best available evidence [2]. This article aims to present the development process of the new guidelines for early detection of breast cancer in Brazil, with details on the methods used and their implications for the new recommendations.

History of government recommendations for early detection of breast cancer in Brazil

Following the creation of the Brazilian Unified National Health System (SUS), government recommendations for early detection of breast cancer were backed initially by the Viva Mulher Program (1996-2003), which recommended, as strategies for early detection of breast cancer in the country, monthly screening with breast self-examination and annual clinical examination. These procedures were performed by physicians or nurses in all the women, especially those 40 years or older, reserving mammography for diagnostic confirmation [3]. According to a publication from 2002 by the Brazilian National Cancer Institute (INCA), mammograms were to be used primarily for diagnostic purposes, ordered by a medical specialist in case of an abnormal physical examination or annually starting at 40 years for women at high risk of developing breast cancer. According to this same publication, all women 50 to 69 years of age should ideally undergo an annual mammogram, but according to the availability of resources, the exam could only be ordered by a medical specialist in case of an abnormal physical examination [4,5]. These recommendations reflected an institutional position at the time, but without any formal method for guidelines development.

In order to develop a more in-depth document on the issue and simultaneously involve more actors in the drafting process, the Ministry of Health (through INCA and the Technical Area on Women's Health, and with the support of medical societies) organized the "Workshop for Drafting Recommendations for the National Breast Cancer Control Program" in November 2003. The event featured participation by representatives from various areas of the Ministry of Health, state administrators, researchers, university professors, representatives of medical specialty societies, and civil society organizations. The workshop produced a consensus document that established the national guidelines for early detection of breast cancer, lasting from 2004 until September 2015 [6]. The method used to draft the recommendations was expert consensus, and their guidelines' broad scope included primary prevention, early detection, diagnosis, treatment, and palliative care. In this consensus document, mammographic screening was recommended for the first time as a public health strategy by the Federal Government [6]. The recommendation was reinforced by publication of the "Pact for Life" in 2006, whose operational guidelines included the target of expanding mammographic screening coverage to 60% of the target population [7], and later by the Plan for Chronic Noncommunicable Dis-

eases, which increased the target to 70% by 2022 [8]. By defining the target population as women 50 to 69 years of age with biennial screening, although without explicitly citing the underlying evidence for each recommendation, the guidelines from the 2004 consensus were in line with those of the World Health Organization and countries with a tradition of screening programs, especially in Europe [9,10].

Although the 2004 consensus did not recommend teaching breast self-examination, it maintained the traditional recommendation of annual screening with clinical breast examination in women 40 years or older [6]. Although the evidence for this recommendation is very weak [11], other similar recommendations are found in the guidelines of developing countries in Latin America, Africa, and Asia [3,10,12,13], generally including women under 50 years in the target population. Indirect criteria, such as a younger population age structure than in Europe and North America, less access to mammography, and lower accuracy of this exam in young women, as well as habitually later tumor detection in these countries, are the justifications usually presented for this recommendation of annual screening with clinical examination.

Evolution of evidence on early detection of breast cancer and the need for a new conceptual and methodological approach

Scientific acceptance of mammographic screening reached its peak in the 1990s after important national screening programs were implemented in various European countries in the 1980s and a meta-analysis of Swedish clinical trials, published in 1993, showed a 29% relative reduction in breast cancer mortality [14]. Still, early in the last decade a systematic Cochrane Collaboration review identified several biases in most of the mammographic screening trials, which may have overestimated the effect sizes in the reduction of breast cancer mortality [15], thus triggering a long period of controversies on screening which has lasted to this day. Some biases involved the randomization process, including random sequence generation, allocation concealment, and evidence of imbalance in the comparison groups at baseline, thereby compromising comparability between groups in some trials [16]. Most trials may also have been affected by measurement bias of the breast cancer mortality outcome, due to lack of blinding of the persons responsible for assessing cause of death in relation to allocation of the intervention (screening). The presence of biases in some studies is also suggested by the fact that older clinical trials with important contamination presented larger effects in the reduction of mortality, which may have been overestimated [16].

Bias in the estimation of screening efficacy would be even greater if observational studies were considered, since they potentially introduce other biases, including healthy screenee bias, since individuals that agree to be screened tend to be healthier, more health-conscious, and more adherent to medical recommendations. Evidence suggests that women who agree to mammographic screening have less risk of dying from other causes unrelated to breast cancer or to screening [17]. Therefore, two main points for the evaluation of screening efficacy in new guidelines would be to include only systematic reviews of clinical trials on the efficacy of mammographic screening and quality evaluation of the selected studies.

In addition to these questions on screening efficacy, recent years have witnessed growing evidence on the harms of mammographic screening. The most serious and important harms are overdiagnosis and overtreatment [18]. Overdiagnosis means the diagnosis of breast cancer cases that would never manifest clinically if they had not been detected by routine screening of asymptomatic women. They are not false-positives, since they meet the histopathologic criteria for breast cancer. That is, they were first detected on mammography and subsequently confirmed through biopsy. This is a limitation to the state-of-the-art in the determination of breast cancer prognosis. Current research indicates that overdiagnosis involves cases of both in situ and invasive breast cancer [19]. An observational study with data from Surveillance, Epidemiology, and End Results (SEER) estimated that 31% of all cancer cases diagnosed in the United States in women 40 years or older corresponded to overdiagnosis [20]. This proportion would probably be higher than that found in Canadian clinical trials if the researchers had only considered the cancers diagnosed by screening. The tumor's own biological characteristics, many of which still unknown to science, are manifested as this non-progressive or scarcely aggressive behavior. At the individual level, it is impossible to know whether a case of breast cancer discovered

by screening is overdiagnosis or not, generating overtreatment in most of these cases. Thus, unnecessary treatments are performed with no benefit whatsoever for the women, and potentially producing health harms due to the inherent risks of the existing treatments.

The inclusion of harms associated with screening is another innovative characteristic of the new guidelines, especially in the Brazilian context, where this kind of outcome is rarely addressed in clinical guidelines. A recent systematic review found that 69% of guidelines that were identified for cancer prevention or early detection either failed to quantify the harms and benefits or presented them asymmetrically [21]. Thus, although the inclusion of harm outcomes is recommended by GRADE (Grading of Recommendations, Assessment, Development, and Evaluation), its implementation in guidelines for early detection of cancer is still incipient, even in the international context. One possible explanation is that historically, harms resulting from screening have not been investigated adequately, even in clinical trials focusing specifically on the subject. In a review that assessed 57 clinical screening trials, even the most important harms such as overdiagnosis and false-positive results were only quantified in 7% and 4% of the studies, respectively [22].

The new guidelines also include the evaluation of alternative screening methods widely used in clinical practice, like clinical breast examination, teaching breast self-examination, and ultrasonography, which also required a more rigorous assessment of their efficacy and risks. This also applies to emerging methods or that could potentially be used in breast cancer screening, like magnetic resonance imaging, breast tomosynthesis, and thermography.

Considering this body of evidence, the new guidelines are also expected to address the balance between these risks and the possible benefits of each screening proposal. Another important innovation is that the recommendations should be accompanied by an estimate of the level of certainty associated with each of them. The GRADE system was chosen by the guidelines steering committee for conducting the synthesis and grading the quality of evidence and strength of the recommendations [23]. Some of the advantages of the GRADE approach over other methods for drafting recommendations are the definition of the quality of evidence for each outcome and the fact that this evaluation is not only related to the study design. Another great advantage is that in GRADE, the recommendations do not depend only on the quality of evidence, but also include the balance between harms and possible benefits. With this system, even evidence from randomized clinical trials can have its level of evidence reduced if the following limitations are identified: risk of bias, imprecision in effect size measurement, inconsistency (or heterogeneity), indirectness (such as proxy outcomes or differences between the study population and the consensus population), or publication bias [24].

Another innovation in the new guidelines was the division of the early detection strategies into two distinct fields: screening and early diagnosis. Screening is the application of tests in asymptomatic individuals, while early diagnosis refers to the strategies for women with signs and symptoms suggestive of breast cancer [10]. Evidence has shown that delays of more than three months between the onset of symptoms and initiation of breast cancer treatment results in a 5% mean decrease in patient survival [25]. The overemphasis on screening in some guidelines is based on the false premise that with wide coverage of mammographic screening, the symptomatic cases would practically disappear, which has not proven to be the case, even in countries with well-consolidated national screening programs [25].

Early diagnosis strategies can take various forms, but they should be based on the following triad: (1) population awareness-raising on cancer signs and symptoms, together with adequate access to health services for symptomatic cases; (2) clinical evaluation with high-quality and timely diagnostic confirmation; and (3) quality and timely access to adequate treatment for confirmed cancer cases [10]. The first two dimensions of these were included in the scope of the new guidelines and translated into three different strategies. The first was the so-called "breast awareness" strategy, based on the promotion of women's own knowledge of their breasts in different life phases, acknowledging what is normal and habitual for each woman and the suspicious findings for breast cancer, aimed at streamlining and upgrading access to health services. The second evaluation strategy was the identification of suspicious signs and symptoms in primary care and priority referral for diagnostic confirmation, aimed at a referral flow to secondary care that avoids repetitive consultations in cases with strong clinical suspicion of breast cancer. The third strategy was diagnostic confirmation in a single service (or one stop clinic), aimed at decreasing the time between the various stages of diagnostic confirmation in symptomatic cases until final determination of the diagnosis, including clinical, histologic, and imaging assessment.

Stages and methods in the drafting process for new guidelines

The guideline development stages include formulation of the research question, search, selection, evaluation of the quality and synthesis of the evidence, drafting of the recommendations, and production of the final text. Still, before beginning the drafting itself, the first step should be the creation of the steering committee. In the Brazilian case, there was the need for first paradigm shift in relation to the traditional model, from a team that merely monitors the experts' work administratively to a steering committee capable of defining methods to be used in each development stage, in order to innovate and overcome the prevailing development standards. A steering committee was thus formed, consisting of members from various areas of the Ministry of Health and two outside experts from the academic community, with the aim of forming a group with expertise in systematic reviews and evidence-based medicine, capable of defining the scope and methods for drafting guidelines. Next, a development group was formed to add expertise on the theme of "early detection of breast cancer" and the method to be used, that is, expertise for conducting systematic literature reviews and critical assessment of the evidence. Some members of the steering committee (50% of the total) also participated in the development group, and two of these members also had a role in coordinating the drafting process.

In the absence of uniform methods for guidelines development in Brazil, the option to standardize the process and even the development group's expertise was to create a manual of methods. As a pioneering initiative in the country, this manual also served as the basis for producing a manual of methodological guidelines for drafting clinical guidelines, under the Ministry of Health [24].

As for editorial independence, one strategy was the inclusion of external experts (non-Ministry of Health) in both the steering committee and the development group. No recommendation proposed by the development group was changed by the steering committee, and there was no outside interference in the drafting process at any time. The only external interference in the drafting process was at the beginning, with a request to broaden the scope and to establish a short drafting deadline. These two issues were clearly related to the expectation raised by the traditional drafting model for clinical guidelines, which had allowed a very wide scope and a very short drafting timeline, as with the consensus in 2004.

Thus, the main problem became the lack of direct involvement by other important actors, like groups from organized civil society and medical specialty societies (the latter had just published their own consensus, based on expert opinion). The solution was to use a public consultation in which the contributions by these actors would be assessed with the same methods and rigor as any other evidence identified during the drafting process.

Three main steps were taken to manage conflicts of interest. The first was to adopt the method for selection of evidence based on blind peer review, with discordant cases assessed by a third independent reviewer, in the same way as with traditional systematic reviews. This procedure also aimed to decrease the likelihood of errors in the selection process. The other procedure was to keep the steering committee and development group from including any specialists with economic interests in the screening procedures, which would inevitably create a conflict of interest. This was considered an important procedure, since the development group would have to be free to recommend abandoning mammographic screening if necessary. There is evidence that the inclusion of this type of specialists in drafting breast cancer screening guidelines is associated with higher likelihood of favorable recommendations for mammographic screening [26]. This issue can be challenging for other guidelines in which it is not possible, in terms of expertise, to form a development group without including this type of professional. For these cases, a rule was elaborated for managing conflicts of interest, published elsewhere [24]. The third procedure was the recording and subsequent disclosure of potential conflicts of interest by all the participants, along with the guidelines, as well as a detailed description of each member's participation [2].

Following the formation of the steering committee, the next step was definition of the scope, a key stage in the drafting process, since an excessively broad scope can hinder the drafting of evidence-based guidelines and compromise the quality, due to the workload. The scope excluded topics like primary prevention, evaluation of the risk of developing cancer, approaches to the high-risk population, diagnostic confirmation, prognosis, staging, treatment, and palliative care. Cost issues were

also excluded. Although the cost dimension is one criterion in drafting recommendations under the GRADE system, the steering committee opted not to include it, in order to make clear that the only criteria used in the recommendations would be the scientific quality of the evidence and the balance between risks and possible benefits for the population's health, associated with each intervention. In other words, the focus was health and not financial cost, even though the latter is also a relevant dimension from the health system's perspective, so that this choice can be considered a limitation to these guidelines.

Based on the scope of the guidelines, 13 structured research questions were formulated, containing the following eligibility criteria: target population, intervention, comparison, outcome, and study design (PICOS). The information sources were: MEDLINE (via PubMed), LILACS (via BVS Prevention and Cancer Control), Embase, and Cochrane Library (including at least Systematic Cochrane Reviews, DARE, and Cochrane Central Register of Controlled Trials – CCTR). Next, search strategies were elaborated, based on these criteria for each question or for a set of interventions grouped by the same type of intervention (mammography and other imaging tests). Details on the research questions, search strategies, and PICOS eligibility criteria are available in the Supplementary Material (see http://cadernos.ensp.fiocruz.br/csp/public_site/arquivo/material-suplementar-ingles_2381.pdf). Unlike a classical systematic review, the process prioritized the selection of syntheses from the literature, in systematic review format. Primary studies were only included in the absence of systematic reviews or in case these reviews were outdated. This strategy was particularly important in the case of questions for which there was little published research, as in the case of questions on early diagnosis. The search for evidence was performed jointly with two librarians specialized in search strategies and references, in order to guarantee the sources' comprehensiveness, balance in the article retrieval, and the retrieved records' precision, in order to respond to the specificity of the questions [27].

Systematization of the search for evidence considered the application of validated filters according to study design, management of located references, and documentation of the entire process to guarantee transparency, reproducibility, and use of the guidelines. Participation by these professionals in the methodological development of guidelines is also new to the health information field in Brazil and is associated with quality improvement in the search strategies used in systematic reviews in the international literature [28]. In the screening questions, outcomes were not used to comprise the search strategies, in order to increase the strategies' sensitivity. A conceptual analysis was elaborated for the representation and translation of the main terms in each question's variables. These conceptual blocks included terms extracted from the controlled vocabularies of the reference bases, in association with free terms in the "title" and/or "abstract" fields. The use of free terms with synonyms from the controlled vocabularies or terms not otherwise covered aimed to increase the search strategies' sensitivity. The combination of free terms and MeSH terms is essential for retrieving recently inserted new articles and updated articles, as well as for those in which there is no indexation in the PubMed records [29]. The study designs (randomized controlled trials and systematic reviews) were represented in the search strategies by means of validated filters for the respective types of design [30].

Selection of the 3,488 references retrieved in the searches was performed by the drafting through evaluation of the articles' abstracts and titles, besides evaluation of duplication between databases. Selection of the titles and abstracts was done in pairs to guarantee that each reference was evaluated by two reviewers independently and blindly. In this stage, the titles and abstracts were classified as eliminated or not eliminated. Articles classified as not eliminated were retrieved as full text for a more detailed evaluation and their subsequent inclusion or exclusion as evidence in the guidelines. In case of disagreement between the experts, a third member of the team was asked to classify the article.

The previously defined inclusion and exclusion criteria were used in the selection of articles related to the defined clinical questions. These criteria were applied twice: first in the title and the analysis of the abstracts and later in the evaluation of the complete article. This two-stage process is similar to the one used to draft systematic reviews and was planned to minimize errors and to be efficient, transparent, and reproducible. The selection of each complete article followed the previously defined inclusion/exclusion criteria for articles, according to the review protocol, based on the questions' definition in the PICOS format.

At the end of the selection process, the remaining articles had their quality critically evaluated using the criteria set by the steering committee for each study design, as shown in Box 1. The use of

Box 1

Quality assessment criteria for systematic reviews.

DOMAIN	EVALUATION QUESTION	COMPLEMENT
Design	Is the design in the primary studies included in the review the same as that of the PICOS question?	
Search	Does the methods section describe how all the relevant studies were located and selected?	
	Were searches done in all the relevant databases?	
	Was a search done in the "grey literature"?	
	Was a manual search of the journals performed?	
	Did the authors check the references of all the retrieved articles?	
	Were articles in all languages included?	
	Did the reviewers contact authors to obtain access to unpublished studies?	
Selection	Are the selection criteria for studies described?	
	Was the selection done by at least two independent (blinded) reviewers?	
Validity of primary studies	Does the methods section describe how each study's validity was evaluated (potential biases)?	
	Did the majority of the selected studies present...	high risk of bias?
		moderate risk of bias?
		low risk of bias?
Publication bias	Were fewer than 10 studies selected?	
	In case more than 9 studies were selected, was the presence of publication bias investigated?	
	Was the presence of publication investigated with...	funnel graph?
		Egger regression?
		trim and fill method?
Heterogeneity	Were the studies' results all in the same direction?	
	Was there a statistical test to verify whether the studies' results were heterogeneous?	
	Was there statistically significant heterogeneity?	
	In case of statistically significant heterogeneity, was it discussed and explored by the authors via subgroup analysis or meta-regression?	
	Describe the I^2 value.	

these instruments supported the assessment of risk of bias according to GRADE. After this stage, the body of evidence for each outcome had its evidence assessed according to the GRADE system criteria, as described previously, and provided a basis for drafting the recommendations, along with each intervention's risk-benefit balance. Articles not retrieved in the searches but known previously to the guest experts were treated the same way as the articles retrieved in the previously described searches and could either be included in or excluded from the body of evidence for a given clinical question. Finally, the recommendations were drafted according to the GRADE system, with classification of the quality of evidence and strength of the recommendations for the clinical guidelines [23], considering not only the quality of the body of evidence for each outcome, but also the respective intervention's risk-benefit balance.

Implications for the chosen methods

Out of the entire complex drafting process, perhaps the stage that most influenced the recommendations was the formulation of questions with the respective eligibility criteria, precisely a stage that is not usually present in guidelines using traditional methods. An example was the decision to limit the study design to randomized trials in the questions on the efficacy of screening strategies. This restriction was essential for controlling biases that are likely to be present in observational studies, especially selection bias, as well as confounding factors, whether known or unknown.

Another critical aspect addressed in this stage was the choice of outcomes for each research question. In the evaluation of breast cancer screening efficacy, traditional and clinically relevant outcomes in oncology, like survival time and staging distribution, are not valid. This is because they result in spurious inferences concerning screening efficacy, since they are susceptible to overdiagnosis and length-time and lead-time bias [31]. The lack of validity (bias risk) in these outcomes occurs even when they are used in high-quality randomized and controlled clinical trials, since these biases are inherent to screening. Breast cancer is a heterogeneous disease and can present in various forms, clinically more aggressive or indolent, depending on the tumor's various biological characteristics. The less aggressive forms have a long asymptomatic period and are thus more likely to be identified by screening. When comparing women who had breast cancer identified by screening and those whose cancer was identified by signs and symptoms, the tumors in the latter group tend to be more aggressive. Length-time bias occurs when evaluating outcomes like survival in these two groups, and it is believed that the difference in outcome is due to the screening and treatment of diagnosed cases, when in fact the former group's prognosis is better even in the absence of these interventions (i.e., spurious causal inference). Screening necessarily introduces lead time in the date of the cancer diagnosis. Therefore, when comparing women whose breast cancer was identified by screening to those whose cancer was identified by signs and symptoms, the screened group will have longer survival due to the lead time, even if there was no effect from screening on the women's real survival. In such cases, in fact, screening does not give additional life, but rather lead time living with the breast cancer diagnosis. The use of survival time as an outcome in screening studies introduces lead-time and length-time biases, and conclusions on the screening method's efficacy are spurious.

Even if the selection of mortality as outcome would control these biases, it would still be necessary to select which outcome is considered "critical" according to GRADE: all-cause mortality or breast cancer-specific mortality. According to GRADE, critical outcomes are highly influential in determining the overall level of evidence for each research question. The reduction of breast cancer mortality may not be translated as a real experience of prolonging life if screening increases the risk of dying from other causes, in addition to being more susceptible to biases. Furthermore, since deaths from other causes are much more frequent, a possible reduction in breast cancer-specific mortality becomes "diluted" to the point of making the studies' statistical power insufficient for detecting a significant difference in all-cause mortality, despite the high number of screened participants. The methodological option here was to consider breast cancer-specific mortality as the critical outcome and penalize the quality of evidence, given the possibility of biases. This penalization and the borderline balance between the risks and benefits of mammographic screening [11] were the two factors that resulted in the weakly favorable recommendations for screening, even in the 50 to 69-year target population. For women in other brackets, the imprecision of the effect estimates (wide confidence intervals) in the meta-analyses resulted in further penalizing the quality of the evidence. If all-cause mortality had been considered the only critical outcome, the conclusion would have been lack of evidence of efficacy in mammographic screening, resulting in a recommendation against screening in any age bracket, since it would refute the evidence of benefits and that screening is associated with various harms.

Another important methodological definition was the non-incorporation of long-term follow-up results after the conclusion of mammographic screening clinical trials. Thus, the differences in the dates of the selected systematic reviews on this theme were not considered a relevant problem, since the mammographic screening clinical trials are old and their original results were published some time ago (the most recent one, the UK Age Trial, had its findings published in 2006 and only referred to women in the 40 to 49-year age bracket). Therefore, it is not expected to find great variability in the results of the selected systematic reviews, although the dates differ. The inclusion of more recent

results based on follow-up after completion of the study (often decades later) increases the problem of contamination of the control group by screening and tends to dilute its effect, even though this decrease is small [32]. The same applies to estimates of overdiagnosis, i.e., contamination of the control group tends to dilute its magnitude [33]. Currently there is already evidence that the lead time with screening is roughly less than four years (generally one year) and that five years after completion of the clinical trials it is already possible to have reliable estimates of the overdiagnosis rate [34,35]. What actually creates important discrepancies in the calculation of overdiagnosis is the denominator used [34]. In the current guidelines, we opted to use total cancers detected by mammographic screening as the denominator, since larger denominators like total cancers detected in the experimental group in long follow-up times after completion of the trials greatly dilutes the estimates of overdiagnosis [34].

Advantages of the new methodological approach

The main advantages of evidence-based guidelines compared to the traditional drafting model based on expert consensus are greater transparency, reproducibility, clarity of presentation, and control of risk of bias [36]. These qualities allow readers to identify how the evidence was searched, selected, and used to generate recommendations. Box 2 summarizes the principal methodological differences between the approach taken here and other national guidelines for early detection of breast cancer.

The evidence-based guidelines method appeals to health professionals (at least theoretically), due to its greater reliability. However, the term "evidence" is worn out and its meaning is still not totally clear to the majority of these professionals. In fact, expert opinion is a source of evidence, as is the result of a study chosen by convenience. The fact that the term "evidence-based" is now timeworn has led to difficulty in communicating it to users of guidelines (health professionals and managers and the general population), due to discordant recommendations between diverse actors with legitimacy vis-à-vis public opinion. The current proposal's main difference is that the recommendations have to be based on the best available evidence. The proposal thus takes into account: systematic search, selection based on predefined eligibility criteria, and quality assessment of the studies. Although it was not used in the

Box 2

Comparison of methods between Brazilian guidelines for breast cancer early detection.

Criteria pertaining to the methodological rigor of development, transparency and reproducibility of the process	Breast cancer 2004 consensus *	New guidelines of three Brazilian medical societies **	New national guidelines ***
Was there a systematic literature search?	No	No	Yes
Were the eligibility criteria for article selection specified?	No	No	Yes
Was there double-blind article selection?	No	No	Yes
Were the main harms from the target interventions specifically considered in the recommendations?	No	No	Yes
Were the effect sizes of the benefits and harms considered in the recommendations?	No	No	Yes
Were the existing uncertainties (especially statistical significance and risk of bias) considered in the recommendations?	No	No	Yes
Were the conflicts of interest reported?	No	No	Yes

* Instituto Nacional de Câncer [6];

** Urban et al. [41];

*** Migowski & Dias [2].

current guidelines, the classification of "levels of evidence" of the Oxford Centre for Evidence Based Medicine (CEBM) is a good example of this difference. In that classification, systematic reviews with homogeneity of meta-analyses of randomized clinical trials are considered the highest level of evidence for intervention studies, while expert opinion appears as the lowest level of available evidence.

A good example of the qualitative leap in support for clinical decision-making with the new methods is the "6 S model" [37]. In this model, evidence-based guidelines are close to the top of the symbolic pyramid representing the hierarchy of sources of evidence for clinical decision-making. Such clinical guidelines are classified as a "summary", since they are capable of synthetizing the evidence from systematic reviews and primary studies comprising the pyramid's base [31], clearly distinguishing them from guidelines that simply cite primary studies to base their recommendations.

In screening, the expert's opinion undergoes other spurious inferences stemming from personal clinical experience. This occurs because length-time and lead-time biases give the impression of better prognosis in screened women, even in the absence of real efficacy. Traditionally, screening studies and guidelines tend not to present information on harms, while inducing an overestimated interpretation of their benefit due to their use of relative measures of comparison rather than absolute differences in the risks between screened and unscreened individuals, which would be more recommendable.

Limitations of the approach

Greater complexity and longer drafting time are disadvantages of the approach taken here when compared to the traditional expert consensus model. The tension between the demands' urgency and scope and the need for greater rigor in the drafting methods will be determinant factors for the feasibility of consolidation of the proposed new model for drafting clinical guidelines in Brazil.

Non-inclusion of patients in the drafting process is also a limitation. This issue was discussed by the guidelines steering committee, and the decision against this procedure was based on evidence of a tendency to overestimate the risk of death from breast cancer and the effect of mammographic screening [38]. This is reinforced by equivocal technical messages by the health professionals themselves [39]. The risks are generally unknown and difficult to understand even for health professionals, such as: false-positive results, overdiagnosis, overtreatment, and cancers induced by ionizing radiation from tests. A possible improvement in future versions would be to succeed in transforming these more direct outcomes, such as morbidity and mortality caused by screening, which would allow a more objective judgment on values and preferences.

Another limitation was the synthesis of evidence from systematic reviews. It was a qualitative synthesis that presented the results of each systematic review in the summary of findings. This limitation was not considered important, since there was no discrepancy in the efficacy results. As for overdiagnosis, the existing discrepancies refer mainly to the denominator used in the calculation, as discussed above.

Another limitation of the current guidelines, particularly in relation to evaluation of the effectiveness of mammographic screening, is that recent decades have seen a decrease in case-fatality in locally advanced cases and palpable tumors in general, due to improvement in adjuvant therapy [40]. Thus, the difference in prognosis decreased between non-palpable tumors detected by screening mammography and those detected clinically, which very probably also reduces the effectiveness of screening in the reduction of breast cancer mortality in more recent cohorts. Since the mammographic screening clinical trials are old, they generally fail to reflect this change. Therefore, for the current reality this external validity problem was assessed as indirect evidence of effectiveness, weakening the favorable recommendation for screening.

It is important to recall that none of the mammographic screening trials was conducted in Brazil, and that the current guidelines did not quantitatively estimate the benefits and harms in the country. In order to attempt to address this problem, the recommendations were penalized as indirect evidence, especially in the North of the country, where breast cancer incidence and mortality are lower.

As mentioned above, for quality assessment of the selected studies for backing the evaluation of bias risk by GRADE, criteria were created, based on preexisting instruments in the literature. Comparison of the criteria used for quality evaluation of systematic reviews (Box 1) and the AMSTAR

instrument [24] shows that the criteria used here contemplate all the dimensions evaluated by this instrument. The only issue not addressed in any way by the adopted criteria is the existence of a protocol. However, we do not see this as an important limitation, since it is standard practice in the main selected systematic reviews, such as Cochrane Collaboration reviews and the Canadian and U.S. task forces (CTFPHC and USPSTF).

Conclusion

The drafting methods used here produced a paradigm shift for drafting guidelines in Brazil. The new approach also raises challenges, like the need for more drafting time and the addition of new actors with knowledge in systematic literature reviews and clinical epidemiology. The method's main advantages are greater transparency, reproducibility, and validity in the drafting process. For this, it is essential that the clinical guidelines explicitly consider in each recommendation the uncertainties involved in the decision-making process and the magnitude of each intervention's benefits, as well as a comparison to the associated risks. This is particularly relevant in cancer screening due to the various biases involved in the evaluation of its efficacy and the borderline risk-benefit ratio.

Contributors

A. Migowski undertook the literature review and conceived, wrote and revised the paper. A. T. Stein, C. B. T. Ferreira, D. M. T. P. Ferreira, and P. Nadanovsky collaborated to the writing and review of the paper.

References

1. Ronsoni RDM, Pereira CCA, Stein AT, Osanai MH, Machado CJ. Avaliação de oito Protocolos Clínicos e Diretrizes Terapêuticas (PCDT) do Ministério da Saúde por meio do instrumento AGREE II: um estudo piloto. Cad Saúde Pública 2015; 31:1157-62.
2. Migowski A, Dias MBK, organizadores. Diretrizes para a detecção precoce do câncer de mama no Brasil. Rio de Janeiro: Instituto Nacional de Câncer José Alencar Gomes da Silva; 2015.
3. Adib SM, El Saghir NS, Ammar W. Guidelines for breast cancer screening in Lebanon Public Health Communication. J Med Liban 2009; 57:72-4.
4. Instituto Nacional de Câncer. Falando sobre as doenças da mama. Rio de Janeiro: Instituto Nacional de Câncer; 1996.
5. Coordenação de Prevenção e Vigilância, Instituto Nacional de Câncer. Falando sobre câncer de mama. Rio de Janeiro: Instituto Nacional de Câncer; 2002.
6. Instituto Nacional de Câncer. Controle do câncer de mama: documento de consenso. Rio de Janeiro: Instituto Nacional de Câncer; 2004.
7. Ministério da Saúde. Diretrizes operacionais dos Pactos pela Vida, em Defesa do SUS e de Gestão. Brasília: Ministério da Saúde; 2006.
8. Ministério da Saúde. Plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil 2011-2022. Brasília: Ministério da Saúde; 2011.
9. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L, editors. European guidelines for quality assurance in breast cancer screening and diagnosis. Luxembourg: Office for Official Publications of the European Communities; 2006.
10. World Health Organization. Cancer control: early detection. WHO guide for effective programmes. http://www.who.int/cancer/publications/cancer_control_detection/en/ (accessed on 01/Feb/2017).
11. Migowski A. A interpretação das novas diretrizes para a detecção precoce do câncer de mama no Brasil. Cad Saúde Pública 2016; 32:e00111516.
12. Viniegra M. Cáncer de mama en Argentina: organización, cobertura y calidad de las acciones de prevención y control. Informe final julio 2010: diagnóstico de situación del Programa Nacional y Programas Provinciales. Buenos Aires: Organización Panamericana de la Salud; 2010.

13. Instituto Nacional de Cancerología. Recomendaciones para la tamización y la detección temprana del cáncer de mama en Colombia. Bogotá: Instituto Nacional de Cancerología; 2006.

14. Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet 2000; 355:129-34.

15. Olsen O, Gøtzsche PC. Cochrane review on screening for breast cancer with mammography. Lancet 2001; 358:1340-2.

16. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2011; (1):CD001877.

17. Gordis L. Epidemiology. Rio de Janeiro: Elsevier; 2013.

18. Harding C, Pompei F, Burmistrov D, Welch HG, Abebe R, Wilson R. Breast cancer screening, incidence, and mortality across US counties. JAMA Intern Med 2015; 175:1483-9.

19. Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. BMJ 2014; 348:g366.

20. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med 2012; 367:1998-2005.

21. Caverly TJ, Hayward RA, Reamer E, Zikmund-Fisher BJ, Connochie D, Heisler M, et al. Presentation of benefits and harms in US Cancer Screening and Prevention Guidelines: systematic review. J Natl Cancer Inst 2016; 108:djv436.

22. Heleno B, Thomsen MF, Rodrigues DS, Jørgensen KJ, Brodersen J. Quantification of harms in cancer screening trials: literature review. BMJ 2013; 347:f5334.

23. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol 2013; 66:719-25.

24. Migowski A, Fernandes MM, organizadores. Diretrizes metodológicas: elaboração de diretrizes clínicas. Brasília: Ministério da Saúde; 2016.

25. Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. Influence of delay on survival in patients with breast cancer: a systematic review. Lancet 1999; 353:1119-26.

26. Norris SL, Burda BU, Holmer HK, Ogden LA, Fu R, Bero L, et al. Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. J Clin Epidemiol 2012; 65:725-33.

27. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. J Clin Epidemiol 2016; 75:40-6.

28. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. J Clin Epidemiol 2015; 68:617-26.

29. Lefebvre C, Manheimer E, Glanville J. Searching for studies. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011. http://handbook.cochrane.org.

30. McKibbon KA, Wilczynski NL, Haynes RB. Retrieving randomized controlled trials from medline: a comparison of 38 published search filters. Health Info Libr J 2009; 26:187-202.

31. Migowski A. A detecção precoce do câncer de mama e a interpretação dos resultados de estudos de sobrevida. Ciênc Saúde Coletiva 2015; 20:1309.

32. Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med 2016; 164:244-55.

33. Nelson HD, Cantor A, Humphrey L, Fu R, Pappas M, Daeges M, et al. Screening for breast cancer: a systematic review to update the 2009 U.S. Preventive Services Task Force Recommendation. Rockville: Agency for Healthcare Research and Quality; 2016. (Report, 14-05201-EF-1).

34. Zahl PH, Jørgensen KJ, Gøtzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. Br J Cancer 2013; 109:2014-9.

35. Baines CJ, To T, Miller AB. Revised estimates of overdiagnosis from the Canadian National Breast Screening Study. Prev Med 2016; 90:66-71.

36. Khan GSC, Stein AT. Adaptação transcultural do instrumento Appraisal of Guidelines for Research & Evaluation II (AGREE II) para avaliação de diretrizes clínicas. Cad Saúde Pública 2014; 30:1111-4.

37. DiCenso A, Bayley L, Haynes RB. ACP Journal Club. Editorial: accessing preappraised evidence: fine-tuning the 5S model into a 6S model. Ann Intern Med 2009; 151:JC3-2, JC3-3.

38. Biller-Andorno N, Jüni P. Abolishing mammography screening programs? A view from the Swiss Medical Board. N Engl J Med 2014; 370:1965-7.

39. Gigerenzer G. Full disclosure about cancer screening. BMJ 2016; 352:h6967.

40. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. N Engl J Med 2016; 375:1438-47.

41. Urban LABD, Chala LF, Bauab SDP, Schaefer MB, Santos RP, Maranhão NMA, et al. Breast cancer screening: updated recommendations of the Brazilian College of Radiology and Diagnostic Imaging, Brazilian Breast Disease Society, and Brazilian Federation of Gynecological and Obstetrical Associations. Rev Bras Ginecol Obstet 2017; 39:569-75.

Resumo

Tradicionalmente, diretrizes clínicas são elaboradas a partir do consenso de opiniões de especialistas. Nos últimos anos, a magnitude dos benefícios do rastreamento mamográfico vem sendo questionada em função dos vieses detectados nos ensaios clínicos que popularizaram a disseminação dessa prática. Paralelamente, o crescente corpo de evidências sobre danos associados ao rastreamento mamográfico também demandava uma nova abordagem que considerasse as incertezas sobre os benefícios e um balanço entre ganhos e possíveis danos. O presente artigo tem por objetivo apresentar o processo de elaboração das novas diretrizes para detecção precoce do câncer de mama no Brasil, detalhando os métodos utilizados, bem como suas implicações para as novas recomendações. A nova abordagem metodológica apresenta como pilares a realização de revisões sistemáticas da literatura, a avaliação da validade das evidências e o balanço entre riscos e benefícios de cada intervenção, garantindo maior transparência, reprodutibilidade e validade no processo de elaboração. Outra inovação das novas diretrizes é a presença de recomendações dirigidas a casos com sinais e sintomas suspeitos. As vantagens da abordagem adotada frente ao modelo tradicional de consenso de especialistas são discutidas com detalhes, bem como os limites e desvantagens dos métodos utilizados. Também são discutidas as implicações de diversas decisões, como escolhas sobre desenhos de estudo, desfechos sobre efetividade do rastreamento, além da definição de sobrediagnóstico e forma de cálculo.

Neoplasias da Mama; Detecção Precoce de Câncer; Programas de Rastreamento; Mamografia; Guias de Prática Clínica como Assunto

Resumen

Tradicionalmente, las directrices clínicas se elaboran a partir del consenso de opiniones de especialistas. En los últimos años, la magnitud de los beneficios del rastreo mamográfico ha sido cuestionada, debido a los sesgos detectados en los ensayos clínicos que popularizaron la propagación de esta práctica. Paralelamente, el creciente cuerpo de evidencias sobre daños asociados al rastreo mamográfico también demandaba un nuevo enfoque que considerase las incertidumbres sobre los beneficios y un balance entre ventajas y posibles daños. Este artículo tiene como objetivo presentar el proceso de elaboración de las nuevas directrices para la detección precoz del cáncer de mama en Brasil, detallando los métodos utilizados, así como sus implicaciones para las nuevas recomendaciones. El nuevo enfoque metodológico presenta como pilares la realización de revisiones sistemáticas de la literatura, la evaluación de la validez de las evidencias y el balance entre riesgos y beneficios de cada intervención, garantizando una mayor transparencia, reproductibilidad y validez en el proceso de elaboración. Otra innovación de las nuevas diretrices es la presencia de recomendaciones dirigidas a casos con signos y síntomas sospechosos. Las ventajas del enfoque adoptado, frente al modelo tradicional de consenso por parte de los especialistas, se discute en detalle, así como los límites y desventajas de los métodos utilizados. Asimismo, se discuten las implicaciones de diversas decisiones, como las decisiones en relación con diseños de estudio, resultados sobre efectividad del rastreo, así como la definición de sobrediagnóstico y forma de cálculo.

Neoplasias de la Mama; Detección Precóz del Cáncer; Tamizaje Masivo; Mamografia; Guías de Práctica Clínica como Asunto