Interval carcinomas as indicators of screening programme performance
Evaluation modalities and standards

Stefano Ciatto, Carlo Naldoni, Antonio Ponti, Livia Giordano, Daniela Giorgi, Alfonso Frigerio, Marcello Vettorazzi, Lauro Bucchi, Luigi Bisanti, Marco Petrella, Eugenio Paci, Gianni Saguatti, Dolores Santini, Marco Rosselli Del Turco, Marco Zappa, Paolo Giorgi Rossi, Vittorio Corsetti, Luisella Milanesio, Priscilla Sassoli de' Bianchi.

Preface

GISMa, before and during its annual meeting, held in Peschiera del Garda on 11-12 October 2007, has promoted and completed a document on interval cancers (IC) observed in screening programmes, aimed at the systematic definition of these events, and at providing detailed guidelines for the identification and review of related mammograms. These recommendations are provided for operators involved in Italian mammography screening programmes, who must include the evaluation of IC among current quality control procedures.

GISMa is grateful to the authors and to all those who contributed to writing the document and final guidelines. The original document may be downloaded from GISMa website (www.gisma.it). We are also grateful to all those who attended the annual meeting and gave their precious contribution by participating to the open discussion.

Presentation

Interval cancer (IC) is a negative event, commonly described only within organized screening programmes. In general, IC tend to generate a negative feeling in the public opinion, and are commonly reported by mass media as major faults of the screening programme: all this may translate into a serious negative impact on the programme management and performance. IC occurrence may be ascribed to human error, as in other medical practices, but may also (in the majority of cases) depend on the intrinsic diagnostic limitations of the screening test, subject to false negativity which is unavoidable to some extent Screening operators (particularly radiologists) are afraid about IC occurrence for its negative effects on their professional image and for the associated medico-legal consequences which may occur.

The screening programme organization, in charge of careful monitoring of all phases of the screening process, should assess the magnitude and the characteristics of such events, to promote actions aimed at improving organization, technical and professional aspects, and to minimize false negatives by improving screening test quality and sensitivity.

IC must be considered under a new perspective, changing a diffuse negative attitude, as a favourable opportunity to check the quality of the programme and of involved professionals, to optimize quality, to monitor (and thus improve) performances and identify patterns of error. Monitoring IC adds value to the process pursuing the final goal of screening, that is reducing mortality.

Based on these arguments, GISMa has perceived the need for a deeper insight in the IC phenomenon, in order to provide a scientific and cultural basis of knowledge to professionals and managers involved in screening, which will help then to cope at the best level with this problem.

Description of the phenomenon

Mammography screening has been shown to be effective in reducing breast cancer (BC) mortality, and is currently recommended as a public health preventive measure. In Italy screening of women aged 50-69 years is under implementation in the whole nation [1], and coverage by the end of 2006 was 78%

Screening effectiveness is far from being absolute, and mortality reduction, as shown by several meta-analyses of randomized trials, is in the range of 20-30% [2]. These figures depend on the fact that randomized trials compare invited subjects (who do not comply totally to invitation) to non invited subjects (some of whom undergo spontaneous screening). Mortality reduction in subjects undergoing mammography as compared to those having no mammography at all, as suggested by several case-control studies, is likely to be substantially higher, in the range of 40%.

Suboptimal sensitivity is no doubt one major limitation of screening. Screening, currently adopting a biennial interval (CE guidelines), does not detect all incident cancers. From one third to one fifth of all

cancers, in fact, is not detected by screening, and is diagnosed mostly for the occurrence of subjective symptoms during the interval between a negative screen and further screening, planned after two years [3,4,5]. These cancer are thus called “interval” cancers (IC).

From a biological point of view, IC seem not to differ from carcinomas occurring in the general population in absence of screening. Theoretically, one might expect IC being more aggressive, with higher growth rate, rapidly progressing from a very small size, below the threshold of radiological detection, at the time of screening, to symptomatic status, during the screening interval. Such a theory (length biased sampling) which suggests that IC sojourn time and preclinical detectable phase are shorter than average, is not confirmed by several observational findings: first of all, IC do not differ from average cancers as to grade (differentiation is associated to growth rate); moreover, survival from diagnosis, adjusted by stage, is not different from clinical cancers. Such findings tend to deny the hypothesis of IC being more aggressive and fast growing, and suggest that IC have average biological features, and are simply “missed” at screening for several reasons (see below).

Although it is evident that, due to average BC growth rate, IC occurring within two years of a negative screening mammogram was certainly “present” at that time, it was not detected a) as it was below the threshold of detection (being too small and/or masked by radiologically dense breast), or b), in spite of the fact that it was detectable, the radiologist, due to reading error (more or less severe), failed to detect it, or suspected it at screening but failed to diagnose it correctly at assessment. These possible reasons have different implications as to possible corrective actions (e.g. improving mammography detection power, improving professional skill, optimizing diagnostic assessment procedures and protocols), and as to the risk of being sued for malpractice.

Whatever may be the motivation of their occurrence, IC always account for a failure of the screening programme, which in theory is aimed at detecting all prevalent cancers, and should not allow any IC to surface. Thus the study of IC (their frequency over time, their morphological and clinical features, the review of prior negative screening mammograms) has been proposed as a fundamental aspect of screening performance evaluation [6].

Definition

We will assume as IC any cancer following a negative screening episode and occurring prior to further screening round. According to CE guidelines, IC status may apply to both invasive and in situ carcinomas. Thus it is important that also in situ carcinomas are registered, at least for a retrospective review of prior screening mammograms, to assess the presence of reading errors. Although in situ carcinomas may be indolent (overdiagnosis) or slowly progressing to invasion (thus eligible for early detection, possibly still as in situ, at further screening), a minority of these lesions (poorly differentiated types) have a high risk of rapid progression to equally aggressive invasive carcinoma, and screening efficacy may also be depending from the detection of in situ carcinomas. As most cancer registries do not consider in situ carcinomas when computing incidence, in situ carcinomas will be excluded when assessing IC proportional incidence (see below).

Most commonly IC occur after a negative screening mammogram and are detected within two years. There are some exceptions or slight variations to this pattern, which are worth being clearly defined, as they must be considered when comparing different programmes as to IC frequency.

- *IC diagnosed beyond two years*: re-screening interval may exceed two years (usually invitation delay for organization problems, or for protocol choice, as in UK). In this case, IC may occur during the third year of the interval. Two years interval is commonly used for comparing programmes as to IC incidence. When a three years interval is systematically adopted, proportional IC rate during the third year will be considered as a measure of screening sensitivity. Otherwise, IC in the third year will be considered as an indicator that the planned biennial interval was not respected, and will be only used for retrospective review of reading errors.
- *Assessment failures*: It may happen that the screening test is positive (i.e. the cancer was correctly perceived), but diagnostic assessment fails to diagnose it correctly, and IC follows. This is considered as an IC as it was not diagnosed at the screening “episode”: of course, when reviewing the determinants of IC occurrence, diagnostic failure will be ascribed to the assessment, not to the screening phase.
- *Early recall*: dubious findings at assessment may generate early recall, thus leading to some diagnostic delay. According to CE guidelines such cases are assumed as screen detected, not as IC. However, if diagnosis is done (for the onset of subjective symptoms or other causes) prior to scheduled early recall, cases must be assumed as IC. Finally, cancer diagnosed in women not showing up at planned early recall beyond scheduled early recall are not considered as IC, but it is recommended that such cases are registered and undergo retrospective review.

- *Repeat early recall:* when early recall is repeated, and diagnosis occurs more than one year after screening, assuming such cases as screen detected is questionable. When evaluating IC we suggest that cancers with repeat early recall detected more than one year after a positive screen are assumed as IC, but evaluated separately: In fact in these cases, although the diagnostic process was initiated by a positive screen, the following diagnostic assessment (repeat early recall prompting no biopsy) may be considered inadequate and early detection has been substantially reduced. If not assumed as IC, these cases should be anyhow monitored and assumed as indicative of inadequate screening process.
- *Follow-up not-attenders*: subjects with a positive screening test who do not attend the planned assessment at the screening programme (or interrupt it before completion) and are diagnosed during the interval as a) they attend assessment elsewhere or b) they become symptomatic. In the former case the diagnostic process has been initiated by screening and assessment has been completed in due time, whereas in the latter diagnosis has been delayed, though not for the screening programme responsibility. In both events cases are assumed as screen detected, but the event may suggest to review the adequacy of the assessment procedure (e.g. the modality with which assessment and its procedure, particularly invasive ones, are offered).
- *Lapsed attenders*: Cancers diagnosed more than two years after a negative screening test in women not attending to further screening are not assumed as IC, although it is recommended to monitor them separately.
- *Prevalent cancers***:** Ipsi- and controlateral metachronous cancers may be diagnosed when subjects with prior primary breast cancers are included in the invited population. Several screening programmes exclude these subjects from invitation, leaving them to general practices for follow-up surveillance, which has different modalities and frequency as compared to screening. These cancers should not be considered when evaluating IC, as these subjects are not strictly eligible for screening invitation, being eligible for other type of surveillance. However, cancers following a negative screening mammogram should be considered for radiological review.

Women excluded from screening invitation (e.g. after their 70th birthday, or emigrating outside the screening area) should be followed up for at least one screening interval, in order to identify IC occurring during this period.

Similar categories are mentioned in a document of the UK screening programme (NHSBCSP publication nr. 62, April 2006, page 7): IC are classified as a) screen IC (IC following a negative screening test/episode), b) assessment IC (IC following a positive screening test and a false negative assessment), and c) IC in follow-up not-attenders (IC following a positive screening test, in subjects not attending scheduled diagnostic assessment).

Identification

Real interval duration is measured from the date of a negative screening mammogram to the date of IC histological diagnosis (or, in alternative, the date of first recording in the cancer registry). A malignant cytology report (C5) might be also considered, due to its almost absolute positive predictive value, but a C5 report may be associated to carcinoma in situ, which should not be assumed as IC. Cancer registries are commonly based on the date of first histological diagnosis of invasive carcinoma. Histology is a more universal reference parameter (registries often ignore the date of first diagnostic suspicion, and for some IC, cancer registration is the only available evidence). Considering that substantial delay may occur between diagnosis and biopsy/surgery, such a criterion may cause shifting of a few cases from first to second or from second to third year of the interval.

Most IC are diagnosed as they become subjectively symptomatic and the woman refers for breast consultation: in a limited number of cases IC are detected for early self referral (typically after one year) to mammography. When IC is detected outside the screening programme, the diagnosing institution and/or the woman usually do not provide such information to the screening programme. This implies that only a minority of IC will be known to the screening centre at the time of diagnosis. In order to identify all IC, active search is necessary using different modalities:

- first of all IC diagnosed at the screening centre must be considered (a self-referral based breast diagnosis clinical facility is often available at the same institution running the screening programme). The proportion of IC diagnosed at the screening centre is a good indicator of a good communication between the assisted population and the screening programme, which maintains a reference role also after a negative screening test. The possibility of immediate review of previous negative mammograms for quality assessment is another advantage.

- IC not diagnosed at screening may be ideally identified through a cancer registry, the official institution responsible for identifying incident cancers, which tends to ignore only cases detected outside the NHS (usually a very limited number). Linkage with cancer registry files should not be limited only to incident invasive cancers, but also to metachronous cancers and carcinomas in situ. These two events should be registered separately and should not be used for assessing proportional incidence, as both are excluded from the denominator (underlying incidence, see below) and should not contribute to the numerator (the former type as it is prevalent, not incident, the latter for the possibility that carcinoma in situ may be overdiagnosed and/or may not progress to invasive.

- in areas uncovered by a cancer registry (still a substantial proportion of the country), an alternative may be the creation of a pathology registry (by monitoring local pathology departments), considering that breast cancers lacking histological confirmation are quite rare. Such an alternative, however, tends to ignore cases diagnosed and treated outside the screening area.

- checking Hospital Discharge Records may be very useful, as these records cover the whole regional territory and, with some delay, the whole nation. Such a method tends to ignore cases treated outside the NHS structures and/or having no hospitalization, which is quite unlikely as BC, in most cases, is referred to hospital for surgical treatment [7,8].

The most accurate identification of IC according to above mentioned methodologies is one of the duties of the screening programme, with the aim of verifying the quality of its performance. Of all measures put in action to monitor a screening programme performance, the analysis of IC has been the most neglected thus far, probably also for its intrinsic difficulties, although this is one of the most reliable indicators.

Indicators

The frequency of IC occurrence over time is commonly measured as the fraction (proportional incidence) of IC observed as compared to “expected” BC in absence of screening (baseline, or underlying incidence). Expected BC are known only in areas covered by a population cancer or pathology registry, but incidence estimates, based on mortality comparisons and adjacent areas incidence, are available almost for all areas in Italy. Expected BC are calculated by attributing the age specific incidence rates to age-specific person-years of subjects with a negative screening test.

The ideal screening programme allows no IC to surface, and proportional incidence is null. On the opposite, the worst screening (or no screening) implies that IC equal expected BC, with a 100% proportional incidence. Reality is half way. The more sensitive the screening, the lower IC rate will be. Proportional incidence is complementary to sensitivity (1-sensitivity): e.g. a 30% proportional incidence (30 IC observed out of 100 expected BC) during the first year of the interval corresponds to a 70% (100 – 30) one-year sensitivity.

CE guidelines [6] suggest that sensitivity may be also calculated according to the formula

screen detected cancers / screen detected cancers + IC

Such a formula is questionable: in fact it compares screen detected cancers (that is those bound to surface clinically also beyond two years, as mammography allows for a larger diagnostic anticipation) with IC (which by definition occurs during the two years interval). Such a method thus implies some degree of overestimation of sensitivity, particularly when determined at the first prevalence screening. The method allows the inclusion of carcinomas in situ both at the numerator (screen detected) and denominator (screen detected + IC). Were this the case, this must be specified, as diagnostic anticipation for screen detected carcinomas in situ is even greater, and thus overestimation of sensitivity may be even greater.

However, also the proportional incidence method has its limitations, mainly for the difficultly of estimating underlying incidence in absence of screening

a) when a cancer registry is lacking,

b) when screening has been ongoing since a long time (incidence in a pre-screening scenario refers to many years before, when cancer registration was not available, or when underlying incidence was different for a different prevalence of risk factors)

c) for the impact of opportunistic screening on underlying incidence,

d) for the limited reliability of incidence rate estimates based on mortality and geographical comparison and

e) for the necessary exclusion of in situ carcinomas from calculation.

Both methods to estimate sensitivity may be used, but it is essential that comparison of different scenarios are based on the same method.

CE Guidelines suggest the rate of IC / 10,000 screened subjects as another possible indicator. Such a figure, as well as cancer detection rate at screening, is an imperfect indicator, as it is strongly dependent on underlying incidence, which may vary from one to another setting. If such an indicator is used to compare different scenarios, it should be adjusted according to local underlying incidence. Such an indicator needs no adjustment when used for comparisons within the same programme (e.g. between operators or units).

Moreover, in order to evaluate the impact of IC on screening efficacy, the rate of IC with advanced stage (T2+or stage II+) could be considered, which is not influenced by early detected IC due to spontaneous screening (e.g. early annual re-screening).

Proportional incidence is commonly measured separately for the first and second year of the interval, or for the whole biennial interval. CE guidelines suggest as a standard a proportional incidence <30% for the first year, <50% for the second year, or <40% for the whole biennial interval.

All indicators concerning IC should be also stratified or adjusted by age (5- or 10-year classes). On the contrary, no distinction is recommended between first or repeat screening round.

Stage at diagnosis

CE guidelines [6] recommend that stage at diagnosis should be recorded for IC, and compared to that of screen detected cancers and of cancers occurring in not-attenders. Such a comparison is important, as it allows for the indirect estimate of length biased sampling and of the possible negative impact of IC on screening efficacy. Such information, however, requires a reliable and detailed source of information, such as a cancer registry, which unfortunately is not always available.

Radiological review

Review of negative screening mammograms followed by an IC is commonly associated to IC identification, being one of the most effective instruments to improve the quality of mammography interpretation. Such a procedure is applicable also when identification of IC is not complete, due to inadequate information, or to missing previous screening mammograms: in fact such conditions are not likely to be associated with a specific reading error type. However, diagnostic mammograms are usually needed, to confirm exact cancer side and site, but these are not often available at the screening centre.

Review should be ideally performed with a blind modality: negative mammograms prior to IC should be randomly mixed with negative controls, with a 1:4-1:5 ratio. In this way the review is more close to the original screening setting, when an interpretation error might have occurred. Such a modality is more correct (and respectful of radiologist's rights) that a review "partially" (only IC are reviewed), or "totally" informed (IC are reviewed together with diagnostic mammograms). Unfortunately informed review ("totally" informed in particular) is most commonly employed, also within a malpractice legal procedure, and tends to overestimate the degree of presumed diagnostic error, as minimal radiological abnormalities, which may be reasonably ignored in the current screening practice, are much easier to identify when information on their side, site and type is available.

Blind review is the most reliable and should be preferred to partially or totally informed review. Blind review, however, is complex, and might be difficult to be applied when reviewing regional or national databases with a large number of IC. In any case the review modality used must be declared and comparisons between different programmes are reliable only if the same review modality was used.

At review mammograms may be classified as negative (occult, true interval), and may be radiologically occult also at diagnosis: in this case no error is ascribed to radiologists. In alternative, one or more suspicious abnormalities must be marked with precision by the reviewer on the film: only after that phase the exact site of the IC will be disclosed, and the correspondence with review marks will be checked. When marking an abnormality the reviewer must specify a) if the lesion is definitely suspicious and deserves further diagnostic assessment (screening error, false negative), thus implying a precise responsibility of the radiologist not reporting it, or b) if the abnormality is minor, subtle (minimal sign), and does not imply a major fault of the radiologist as it does not necessarily require diagnostic assessment: the reviewer marks it as he has perceived it, but he is also aware of the judgment distortion due to the review scenario. When performing a review, even if with a blind modality, the reviewer is aware he is looking at a limited series of cases enriched with IC. Due to this, as usually happens in an artificial "test" setting, the reviewer is alerted and may have the tendency to lower his threshold for suspicion, over-reporting minor abnormalities which he would consider as benign morphological variables in the everyday screening practice. Reporting minimal signs is mainly relevant to training, as it identifies subtle morphological changes which may correspond to a cancer

the knowledge of such changes helps radiologists to refine their diagnostic categories and possibly improve their judgment.

After blind review, IC are disclosed and their correspondence to lesions marked by the reviewer is checked. In this way review sensitivity (proportion of IC marked as screening error) and specificity (proportion of marked negative controls) are determined. It is evident that a given sensitivity has a different meaning when associated to high or low specificity.

Reproducibility studies indicate that review is largely subjective. For this reason, although such a procedure may turn out to be rather complicate, review (especially for study purpose) should involve more than one reviewer, allowing for a majority report to be defined.

Review of IC mammograms is mainly performed with the aim of improving the quality of reading, by comparing screening and diagnostic mammograms, but it is also used for medico-legal purpose, when reading error is suspected. According to what has been previously discussed it is quite evident that in the latter case blind review is the only valid modality.

CE guidelines [6] suggest as a standard that IC reviewed as screening error should not exceed 20% of reviewed IC. EC guidelines indicate the reporting categories to be used at review and the above mentioned standards, but do not go into details as to review modality to be used (blind, partially or totally informed, single or multiple reviewers, with consensus of discrepancies or with majority report). In absence of a precise definition of these parameters, the variability of results may be large, and comparison between programmes may look quite unreliable.

If possible, review might be extended to negative screening mammograms followed by cancer detected at further screening. This review might be limited to screen detected cancers with advanced stage, and might help in identifying systematic interpretation errors, and thus prompt adequate corrective actions.

Reccomendation

All screening programmes should include among performance indicators to be checked periodically:

- absolute and proportional IC incidence, providing information as to the adopted modalities to define expected (underlying) incidence and to identify observed IC. Where a cancer registry is lacking, IC may be identified through Hospital Discharge Records, indicating linkage modalities until a standard protocol will be defined on a national basis
- the proportion of IC reviewed as screening error, providing information as to review modalities adopted (blind or informed, number of reviewers, definition of review discrepancies)
- stage distribution of IC, compared with screen detected cancers and with cancer occurring in not-attenders

As the definition of these indicators is rather complicate, they may be determined with less than yearly frequency. On the contrary radiological review (not necessarily to be done on all IC) should be a routine procedure with annual frequency.

GISMa defines two initiatives that are urgently needed to allow for the implementation on a national basis of the above discussed monitoring and quality assurance procedures for mammography screening:

- a working group must be created, involving the Italian Association of Cancer Registries, to provide incidence estimates for all Italian areas, to be used for the estimate of proportional IC incidence
- a working group must be created, aimed at defining standardized modalities for the use of Hospital Discharge Record to identify IC in areas lacking cancer or pathology registries.

References

1. Ministero della Sanità. Linee guida elaborate dalla Commissione oncologica nazionale, in applicazione di quanto previsto dal Piano Sanitario Nazionale per il triennio 1994-1996, relativo all'azione programmatica "Prevenzione e cura delle malattie oncologiche", concernenti l'organizzazione della prevenzione e dell'assistenza oncologica. Gazzetta Ufficiale 1996;127:25-38.
2. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. JAMA. 1995 Jan 11;273(2):149-54.
3. Paci E, Ciatto S, Buiatti E, et al. Early indicators of efficacy of breast screening programmes. Results of the Florence District Programme. Int. J. Cancer 1990;46:198-202.
4. Ciatto S, Rosselli Del Turco M, Zappa M. The detectability of breast cancer by screening mammography. Br. J. Cancer 1995;71:337-339.
5. Peeters PHM, Verbeek ALM, Hendricks JHCL, et al. The occurrence of interval cancer in the Nijmegen screening programme. Br. J. Cancer 1989;59:929-932.

6. Perry NM, Broeders M, de Wolf C, et al. European guidelines for quality assurance in breast cancer screening and diagnosis. European Commission. Fourth editon, Luxembourg, 2006.
7. Vettorazzi M, Stocco C, Chirico A, et al. Quality control of mammography screening in the Veneto Region. Evaluation of four programmes at a Local Health Unit level – Analysis of the frequency and diagnostic pattern of interval cancers. Tumori, 2006;92:1-5
8. Burani R, Caimi F, Maggioni C, et al. Quality assessment of the mammographic screening programme in the Azienda Sanitaria locale Provincia Milano 1 - Analysis of interval cancers and discussion of possible causes of diagnostic error. Radiol. Med. (Torino). 2005;109:260-267.