Variation in performance in low-volume mammography screening programmes: Experience from Switzerland

J.-L. Bulliard a,*, C. Ducros b, E. Dayer c, B. Arzel d, F. Levi a,e

a Unité d’épidémiologie du cancer, Institut de médecine sociale et préventive, Centre Hospitalier Universitaire Vaudois et Université de Lausanne, Lausanne, Switzerland
b Fondation pour le dépistage du cancer du sein, Lausanne, Switzerland
c Centre vaudois de dépistage du cancer du sein, Sion, Switzerland
d Fondation genevoise pour le dépistage du cancer du sein, Geneva, Switzerland
e Registres Vaudois et Neuchâtelois des Tumeurs, Centre Hospitalier Universitaire Vaudois et Université de Lausanne, Lausanne, Switzerland

1. Introduction

Most western countries have established national or regional breast cancer screening programmes since randomised controlled trials demonstrated that mammography screening was effective in reducing breast cancer mortality in females aged 50 and over [1,2]. In Switzerland, where healthcare delivery is organised at a regional (cantonal) level, the first screening programmes were implemented in 1999, following a successful pilot project [3,4]. The development of organised programmes concomitantly to pre-existing opportunistic screening was facilitated by the decision of the Federal Office of Social Insurance to reimburse from 1999 biennial mammography screening for women aged 50 and above through the Swiss compulsory health insurance coverage system, when performed within a quality-assured programme.

Wide variation in performance indicators has been documented across service screening programmes in Europe and similar variation is expected in the impact of screening on breast cancer mortality [2]. This discrepancy has broadly been ascribed to differences in screening organisation and procedures, and to individual programme size and volume of activity [5,6]. Most comparisons of short-term performance indicators have additionally been subjected to potential age confounding since screening performance is also sensitive to programme duration and the age groups screened [1,7].

The region-oriented healthcare system in Switzerland enables to compare performance indicators across programmes initiated at...
the same time, with similar screening organisation and procedures, but with substantial differences in multiple reading strategy, population size, historical screening habits and prevalence of opportunistic screening [3,4]. The aim of this study was (1) to examine the variation in quality and effectiveness of mammography screening in low-volume Swiss regional programmes after 8 years of operation (4 screening rounds), and (2) to explore determinants of this variation in order to improve mammography practice and optimise screening performance.

2. Materials and methods

2.1. Organised mammography screening in Switzerland: current state

Five mammography screening programmes are operating in Switzerland, covering six (all French-speaking) cantons and about 25% of the Swiss female resident population aged 50–69 (Fig. 1). A sixth programme will start in 2010 in German-speaking Switzerland. All programmes share the same operating software and quality control procedures, and have joint educational material and similar screening procedures that include biennial double reading with arbitration and two-view mammography. The invitation and reading processes in Swiss programmes have been described elsewhere [3]. This study focussed on the three longest standing programmes (Vaud (VD), Geneva (GE) and Valais (VS)) which all were launched in 1999 (Fig. 1). Programmes in the cantons of Fribourg (FR-2004) and Jura/Neuchâtel (JU-2005/NE-2007) were too recent to have reached their full potential. The three programmes considered took place in cantons with different population characteristics with respect to breast cancer screening, but with similar age structure of their 50–69 female populations.

2.2. The Vaud screening programme

The pioneer activities in breast cancer screening in the Vaud canton contributed to the development of service screening programmes in Switzerland [3]. This largest Swiss programme – it encompassed over 40% of all programme-based mammograms – elicited stringent criteria for second reader radiologists in order to comply with the 5000 minimum reading requirement recommended in the European guidelines (Table 1) [8]. Several accredited radiologists had prior experience with mammography reading in an organised setting from the pilot trial [3]. Exposure to organised screening from the pilot programme, but not accessibility, influenced participation in this canton [9]. In this region with a dense offer of screening facilities, overall attendance was about 50% but varied locally: the denser the population, the lower was participation [9].

2.3. The Geneva screening programme

The Geneva programme applied a similar second reader strategy to that in Vaud. Difficulties in collaborating with private radiologists temporarily hindered the development of a programme (participation rate: about 25%) set in a population with one of the highest incidence of breast cancer worldwide [10]. Opportunistic screening prevails in this highly urbanised canton and has increased since the inception of the programme. Conflicting evidence has been reported about attraction by the programme of women from predominantly lower socioeconomic classes [11,12].

2.4. The Valais screening programme

In Valais, the second reading approach favoured volunteering, self-motivation and self-responsibility of radiologists. This led to...
more second readers with lesser annual throughput. Participation has been the highest among Swiss service screening programmes, almost reaching the European recommendation (66% vs 70%) [13]. Participation however differed within this alpine canton, being higher in the French than the German-speaking community [13].

2.5. Opportunistic screening

Self-reported biennial mammography (screening and diagnostic) coverage was in 2002 about 90% in the 3 cantons, above the Swiss average for 50–69-year-olds females (71.5%) [12]. The wide difference in participation across programmes (GE: 26%, VD: 49%, VS: 66%) [14] reflected different prevalences of opportunistic screening [4].

2.6. Analyses

Analyses included 197,608 screens performed during the first 8 years (1999–2006) of operation among 50–69-year-old women residing in these 3 cantons. Performance and prognostic indicators as defined in the European Guidelines were calculated for prevalent (initial) and incident (subsequent) screening rounds for each programme [8]. Interval cancers were identified by linkage between screening and cancer registries records, performed separately for the Vaud, Geneva and Valais Cancer Registries. To estimate the proportional incidence of interval cancers (the ratio of observed to expected incidence in the absence of screening), age-specific breast cancer incidence rates for females aged 50–69 in the 5-year period preceding the screening programmes were used.

3. Results

The volume of activity for each regional programme is described in Table 1. The average throughput ranged between circa 5000 and 13,600 yearly readings between 1999 and 2006. Both the overall and individual volume increased over time as participation raised (data not shown). The number of accredited second readers was inversely related to the population size of the programme. The combined effect of different participation rates and strategies for second readers led to a four-fold variation across programmes in the annual volume interpreted by second readers.

Performance indicators for the 3 programmes, ranked by increasing size, are presented by type of screening round in Table 2 and the relationship between some key indicators is illustrated in Fig. 2. The proportion of subsequent screens ranged from 44% in GE to 57% in VD, and was 50% in VS (data not shown). In terms of quality, the biopsy yield and the referral rate met European standards in subsequent screens and, for one programme only, in initial screens. Overall, screening performance improved between prevalent and incident rounds (higher PPV, lower false-positive and referral rates, Fig. 2). This improvement was accompanied by less heterogeneity in performance across programmes in subsequent screens (Fig. 2).

The fraction of women reinvited within 2 years was too low for the 3 programmes (65–92% vs 95% recommended, Table 2), although the reinvitation rate at 30-month substantially improved and reached the European norm for one programme. The median time between two screening invitations was 719 days (VD), 739 days (VS) and 742 days (GE), respectively, and these delays decreased over screening rounds (data not shown).

Most indicators of screening effectiveness fulfilled European requirements for all programmes (Table 2). The proportion of advanced cancers was consistently high, irrespective of screening round. Substantial regional variations were observed for (1) in situ cancers whose rates were 50% lower in VS than elsewhere, (2) small cancers (<1 cm) in subsequent screens whose proportion was in VS half that in VD and GE, and (3) interval cancers for which

<table>
<thead>
<tr>
<th>Performance indicator</th>
<th>GE n = 40,553</th>
<th>VS n = 52,889</th>
<th>VD n = 104,166</th>
<th>European norms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referral rate (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial screen</td>
<td>8.0</td>
<td>7.1</td>
<td>6.4</td>
<td>&lt;7</td>
</tr>
<tr>
<td>Subsequent screen</td>
<td>4.7</td>
<td>3.3</td>
<td>3.9</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Benign to malignant biopsy ratio</td>
<td>n.a.</td>
<td>0.55</td>
<td>0.47</td>
<td>&lt;0.5</td>
</tr>
<tr>
<td>Subsequent screen</td>
<td>n.a.</td>
<td>0.24</td>
<td>0.28</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td>Timeliness (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Re-invitation within 24 months</td>
<td>65.0</td>
<td>83.2</td>
<td>91.8</td>
<td>&gt;95</td>
</tr>
<tr>
<td>Re-invitation within 30 months</td>
<td>92.0</td>
<td>94.6</td>
<td>98.0</td>
<td>&gt;98</td>
</tr>
<tr>
<td>Detection rate (/1000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial screen</td>
<td>6.5</td>
<td>6.2</td>
<td>7.6</td>
<td>(3 x I.R.)</td>
</tr>
<tr>
<td>Subsequent screen</td>
<td>6.2</td>
<td>4.8</td>
<td>5.9</td>
<td>(1.5 x I.R.)</td>
</tr>
<tr>
<td>In situ cancers (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial screen</td>
<td>18.9</td>
<td>12.2</td>
<td>18.2</td>
<td>10–20</td>
</tr>
<tr>
<td>Subsequent screen</td>
<td>18.9</td>
<td>12.5</td>
<td>17.1</td>
<td>10–20</td>
</tr>
<tr>
<td>Invasive cancers ≤ 10 mm (%)</td>
<td>27.5</td>
<td>25.2</td>
<td>39.6</td>
<td>&gt;25</td>
</tr>
<tr>
<td>Stage II+ (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial screen</td>
<td>31.5</td>
<td>35.4</td>
<td>29.7</td>
<td>&lt;30</td>
</tr>
<tr>
<td>Subsequent screen</td>
<td>29.6</td>
<td>39.3</td>
<td>28.9</td>
<td>&lt;25</td>
</tr>
<tr>
<td>Node-negative cancers (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial screen</td>
<td>75.4</td>
<td>74.3</td>
<td>72.3</td>
<td>&gt;70</td>
</tr>
<tr>
<td>Subsequent screen</td>
<td>74.7</td>
<td>75.9</td>
<td>73.8</td>
<td>&gt;75</td>
</tr>
<tr>
<td>Interval cancers (initial screen, proportional incidence)</td>
<td>n = 56</td>
<td>n = 59</td>
<td>n = 77</td>
<td></td>
</tr>
<tr>
<td>0–11 months after screening</td>
<td>15.0</td>
<td>27.3</td>
<td>24.0</td>
<td>≤30</td>
</tr>
<tr>
<td>12–23 months after screening</td>
<td>67.0</td>
<td>53.2</td>
<td>41.9</td>
<td>≤50</td>
</tr>
</tbody>
</table>

GE: Geneva; VS: Valais; VD: Vaud; n.a.: not available; I.R.: incidence rate.
in The Netherlands [16], or a high referral rate and low interval and high interval cancer rate (high specificity/low sensitivity), like tive results, which are costly and associated with psychological avoiding unnecessary diagnostic evaluation following false-posi-
cancer detection rate while minimising interval cancer and physical/technical quality control of radiological equipment.

biennial mammography, same length of operation) and common same screening regimen (double reading with arbitration, two-view
Fig. 2. Performance of mammography screening programmes in Switzerland, (a) detection rate vs false-positive rate, (b) positive predictive value vs referral rate, by type of screening round.

the second-year proportional incidence was 50% greater in GE (67%) than in VD (42%).

4. Discussion

Swiss mammography screening programmes met most interna-
tional standards of quality and effectiveness [8]. Performance was in line with other European service screening programmes [2,15]. However, up to a two-fold variation was observed in interim indicators of performance across these programmes despite the same screening regimen (double reading with arbitration, two-view biennial mammography, same length of operation) and common physical/technical quality control of radiological equipment.

The goal of any cancer screening initiative is to obtain a high cancer detection rate while minimising interval cancer and avoiding unnecessary diagnostic evaluation following false-positi-
tive results, which are costly and associated with psychological morbidity [1]. Consequently, a programme with a low referral rate and high interval cancer rate (high specificity/low sensitivity), like in The Netherlands [16], or a high referral rate and low interval cancer rate may reflect two screening policies or strategies with similar overall performance.

Screening policy and criteria are uniform in Swiss programmes [4] and indicators of programmes’ sensitivity and specificity were weakly related (Fig. 2). For instance, one programme (VD) had in the initial round the lowest referral and false-positive rates, the lowest 2-year interval cancer rate (1.9/1000 vs 2.3/1000 (VS) and 2.5/1000 (GE)) but the highest detection rate. Regional variations in cancer detection appear unlikely to be due to differential background risk since pre-screening levels of incidence differ by less than 5% in these regions (ranging from 290 to 303 per 100,000 for women aged 50–69) and self-selection of participants seems to occur on common factors across Swiss programmes (Swiss nationality, married and younger women) [17,18].

Although random fluctuations cannot be discarded, our results are compatible with a genuine difference in performance across Swiss programmes, at least in initial round. Higher performance was observed in the region (VD) that had both a prior pilot trial [3] and the highest annual volume per second reader. Our results show that expertise has successfully been transferred from an experi-
mental setting to a service programme and confirm that quality improved in service programmes with duration of activity [7].

Performance indicators were closer for the two programmes (VD and GE) that elicited a restrictive second reading strategy. In Switzerland, accredited radiologists interpret screening mammo-
grams both within and outside the setting of a service programme. Although the total annual throughput of radiologists was unknown, those reading less mammograms within a programme were likely to interpret more mammograms outside a programme as overall screening coverage and number of radiologists per capita were roughly comparable in the three regions [12]. These results confirm that performance per se is independent from the level of attendance [19]. In presence of substantial, concurrent opportunistic screening, the relationship between volume and accuracy remains difficult to measure in screening programmes.

Comparisons are generally more reliable for subsequent rounds since differences in women’s prior screening practices (i.e. opportunistic screening) are largely attenuated after an initial round. Performance appears to vary mostly in subsequent rounds between the regional programme (VS) that favoured self-motiva-
tion and self-responsibility of radiologists to perform second readings and those programmes (VD and GE) that elicited a selective strategy for accrediting 2nd readers. The VS programme showed the highest participation but the lowest detection rate, with both fewer small cancers (<1 cm or in situ) and more advanced lesions being detected for rather similar proportional incidence of interval cancers as GE and VD. A restrictive accreditation of second readers led to a median throughput four times higher for second than first readers. Despite this measure, no Swiss programme reached the reading threshold of 5000 cases per year for second readers. Such a threshold has rarely been met in low-volume screening programmes run in decentralised health-care systems [20]. To our knowledge, there is no specific recommendation aimed at optimising accuracy of mammography interpretation for low-volume screening programmes. Strong restriction on second readers’ accreditation and/or a centralisation of breast assessment in liberal healthcare system can improve performance in low-volume programmes [21].

Even though indicators of early effectiveness (% of cancers in situ, ≤1 cm, NO) met the criteria of the European community, the proportion of screen-detected cancers of stage II+ was consistently higher than the reference value [8]. Similar proportions of advanced cancers have been observed in two high-quality Scandinavian service programmes, and a revision of the European criterion for the stage distribution has been proposed [22,23]. A trend analysis of advanced breast cancer rates by detection mode in these Swiss populations is warranted prior to any speculative interpretation.

The relative incidence of first-year but not second-year interval cancer met the levels set by the European guidelines [8]. Similar findings were reported in a pooled analysis of interval cancers in six European countries, as well as in Norway [15,24]. Apart from random variability due to few numbers in our series, an update of the current guidelines (based on the Swedish Two-County trial) has been suggested since they do not consider changes in
background incidence unrelated to the screening programme [22]. A high completeness of cancer registration, an accurate linkage procedure to identify interval cancers and a substantial prevalence of concomitant opportunistic screening have been evidenced to contribute to a greater frequency of interval cancer, regardless of screening quality [25]. All these factors apply to Switzerland.

A temporal bias has been advanced to explain the sharp increase between first and second-year proportional incidence [26]. The date of incidence mostly recorded by cancer registries is the date of the first positive histology report or hospital admission, and not that of the onset of symptoms. This shift in time of diagnosis led some first-year interval cancers to appear as second-year cancers in non-research setting.

To our knowledge, evaluation of timeliness in screening programmes has not been documented in the scientific literature. No Swiss programme invited 95% of screen-eligible women within 24 months, as recommended in the European guidelines [8]. Invitation letters do not specify a time or place in Switzerland and subsequent invitations are automatically sent 22-month after the prior screening test. Waiting time for a screening appointment in a radiology centre is a major contributor to this delay [9]. Delays do not affect screening performance per se but, if too lengthy, may incite women to be screened outside the programme and decrease attendance.

In conclusion, variation of performance across Swiss service programmes was related to difference in multiple reading strategies and a prior experience in service screening, but could not be explained by levels of opportunistic screening and programme attendance. The lesser second readers, the higher the screening performance. Performance improved over time in all programmes but variability in process indicators subsisted between programmes with and without a selective 2nd readers strategy. Centralisation of second readings to fewer radiologists should further improve mammography practice and screening performance in low-volume programmes.

Conflict of interest

None.

Acknowledgements

The authors are grateful to Drs. J.-P. de Landtsheer and C. Jemelin (respectively former Directors of the Vaud and Valais breast cancer screening programmes) for collecting and providing data from their programmes. The valuable assistance of Mr. L. Randimbison (Vaud Cancer Registry), Dr. D. de Weck (Valais Cancer Registry) and Mr. G. Fioretta (Geneva Cancer Registry) for matching screening records with cancer registries databases is also acknowledged.

References


