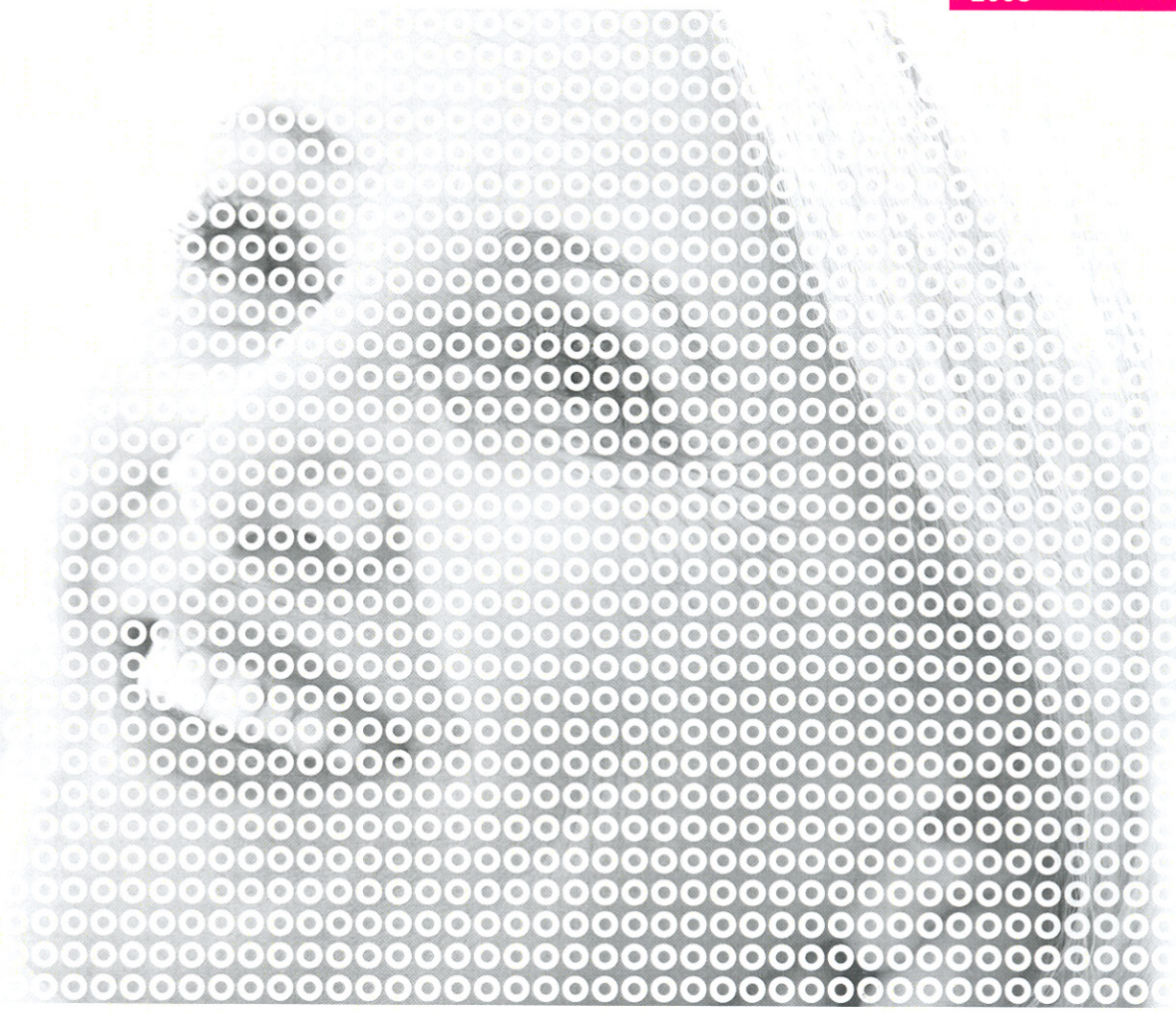
EVALUATION OF THE PERFORMANCE OF DESIGNATED SCREENING CENTERS OF THE PROGRAMME QUÉBÉCOIS DE DÉPISTAGE DU CANCER DU SEIN (PQDCS):

Description of the Methods

2008



Institut national de santé publique du Québec

Québec

EVALUATION OF THE PERFORMANCE OF DESIGNATED SCREENING CENTERS OF THE PROGRAMME QUÉBÉCOIS DE DÉPISTAGE DU CANCER DU SEIN (PQDCS):

Description of the Methods

2008

Direction des systèmes de soins et politiques publiques
Programmes de dépistage, génétique et lutte au cancer

Institut national de santé publique
Québec

AUTHORS

Éric Pelletier, M.Sc.
Direction des systèmes de soins et politiques publiques,
Institut national de santé publique du Québec

Isabelle Théberge, M.Sc.
Direction des systèmes de soins et politiques publiques,
Institut national de santé publique du Québec

Nathalie Vandal, M.Sc.
Direction des systèmes de soins et politiques publiques,
Institut national de santé publique du Québec

Jacques Brisson, M.D., D. Sc.
Direction des systèmes de soins et politiques publiques,
Institut national de santé publique du Québec,
Département de médecine sociale et préventive, Université Laval,
Unité de recherche en santé des populations (URESP) et
Centre des maladies du sein Deschênes-Fabia,
Hôpital du Saint-Sacrement du CHA de Québec

WITH THE COLLABORATION OF

Other members of the evaluation team of Programme québécois de dépistage du cancer du sein (PQDCS).

SECRETARIAT

Sylvie Muller, agente administrative
Direction des systèmes de soins et politiques publiques,
Institut national de santé publique du Québec

ACKNOWLEDGMENT

We want to thank Dr. Belkacem Abdous to having revised this document and for having allow us to benefit from his expertise through its relevant comments.

The translation of this report was made possible by the Public Health Agency of Canada.

This document is available in its entirety in electronic format (PDF) on the Institut national de santé publique du Québec Web site at: http://www.inspq.qc.ca.

Reproductions for private study or research purposes are authorized by virtue of Article 29 of the Copyright Act. Any other use must be authorized by the Government of Québec, which holds the exclusive intellectual property rights for this document. Authorization may be obtained by submitting a request to the central clearing house of the Service de la gestion des droits d'auteur of Les Publications du Québec, using the online form at http://www.droitauteur.gouv.qc.ca/en/autorisation.php or by sending an e-mail to droit.auteur@cspq.gouv.qc.ca.

Information contained in the document may be cited provided that the source is mentioned.

LEGAL DEPOSIT 3rd QUARTER 2008
BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC
LIBRARY AND ARCHIVES CANADA
ISBN: 978-2-550-52893-7 (FRENCH PRINT VERSION)
ISBN: 978-2-550-52895-1 (FRENCH PDF VERSION)
ISBN: 978-2-550-53725-0 (PRINT VERSION)
ISBN: 978-2-550-53726-7 (PDF)

©Gouvernement du Québec (2008)

SUMMARY

The evaluation of the Programme québécois de dépistage du cancer du sein (PQDCS) is based primarily on the analysis of performance indicators as defined in the PQDCS guidelines. Methods were initially developed to measure performance indicators at the provincial level and then in individual regions. The goal of this report is to develop an approach for the measurement and comparison of performance indicators at the screening centre level.

The indicators selected to evaluate the performance of individual centres are: the detection rate, recall rate, and number of false positives per screen detected cancer. These indicators are estimated separately for initial and subsequent screens. The percentage of *in situ* cancers, the percentage of small invasive cancers, and the percentage of lymph node-negative invasive cancers were also selected. To evaluate a centre's performance, these indicator measurements were compared to the measurements for Quebec as a whole, as well as with the indicator targets set out in the PQDCS guidelines.

Comparing the indicators measured for a given screening centre with those measured for Quebec as a whole raises two challenges of a methodological order. At first, women characteristics differ from centre to centre. Also, certain centres perform only a limited number of mammograms per year. It was therefore necessary to develop a method for measuring performance indicators adjusted for the characteristics of the women screened, as well as a method for calculating a confidence interval around adjusted indicators. Finally, it was also necessary to develop a method for presenting results that would provide a better understanding and analysis of screening centre performance.

Standardization is an adjustment method frequently used to account for variations in the characteristics of individuals when comparing two or more sub-populations. In the literature, two primary standardization methods are used: direct standardization and indirect standardization. Direct standardization presents significant limitations in terms of the statistical stability of estimates when the number of subjects in individual sub-populations is relatively small. However, indirect standardization does not lend itself to the direct comparison of two sub-populations and only allows for the comparison of sub-populations with the total population. Still, given the large number of potential confounding factors and the relatively small sub-populations, indirect standardization was selected. Indirect standardization consists of comparing the observed value of an indicator in a given centre (taking into account its clientele) with the expected indicator to find with the same clientele if the centre's performance were equal to that observed across the PQDCS. Indirect standardization provides a means of establishing an observed rate/expected rate ratio (X/A).

Since the number of screening mammograms performed in Quebec centres is limited, several indicators need to be calculated using data from more than one year. The method presented here is based on data from three years. Also, in order to account for any random variation (statistical variation) in the number of screens performed by screening centres, a 95% confidence interval is calculated. The indicators selected are proportions (and one odds). Binomial distribution is therefore used to calculate the confidence interval of the ratio X/A.

In order to compare the performance of each individual centre with that of Quebec as a whole, each observed rate/expected rate ratio (and its confidence interval) is multiplied by the value of the indicator calculated for Quebec. For individual screening centres, this calculation produces an “adjusted” indicator value. Moreover, a means of graphically representing adjusted indicators was developed in order to facilitate the analysis and interpretation of results. The first approach is intended to demonstrate the evolution of a centre’s performance over time. The second approach is designed to simultaneously present the performance of all centres in a given period and a specific indicator.

The methods presented in these pages are now being used in Quebec to evaluate the performance of screening centres. Eventually, these measures may be used to evaluate the performance of centres in which women with abnormal screening mammograms undergo further investigation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Quality assurance in the Programme québécois de dépistage du cancer du sein (PQDCS) is largely based on the analysis of performance indicators, as defined in the PQDCS guidelines (1). Monitoring these indicators provides a means of determining whether the PQDCS is meeting its targets.

Methods for measuring performance indicators were initially developed to monitor indicators in Quebec as a whole (2). These indicators have also been analyzed at the regional level. These analyses have demonstrated that certain indicators change over time, both regionally and provincially, and that program performance sometimes shows regional variations.

Since the health ministry, as well as regional coordinators, institutions and professionals are also interested in evaluating the performance of individual centres associated with the PQDCS, it has become important to develop a suitable approach for the measurement and comparison of performance indicators in these centres. The first stage of the present project consists of developing such an approach to evaluate the performance of designated screening centres (DSCs). The indicators retained to evaluate the performance of DSCs are the detection rate, recall rate, and number of false positives per screen detected cancer, estimated separately for initial and subsequent screens (see appended table). The percentage of *in situ* cancers, the percentage of small invasive cancers, and the percentage of lymph node-negative invasive cancers were also selected. A definition of each of these indicators is provided in the appendix. To evaluate the performance of a DSC, its performance is compared with that of Quebec as a whole, as well as with the target established for the indicator in question in the PQDCS guidelines.

1. THE ISSUE

From a methodological standpoint, comparing a centre's performance indicators with those of Quebec as a whole can be problematic for two reasons (3).

First, the characteristics of women screened at a given centre may differ from those of the women with whom a comparison is being made, namely all Quebec women who participate in the PQDCS. For one thing, women's risk of developing breast cancer may vary. For example, the women screened at a given centre may be older than average or them may have a greater proportion of family history of breast cancer than is true of the larger PQDCS cohort. As a result, the detection rate for that particular centre would naturally be higher than that of the PQDCS, simply because the frequency of breast cancer is greater for that centre. Therefore, an analysis that takes into consideration the women characteristics of a centre is needed.

The second problem is linked to the fact that the number of mammograms performed every year in the DSCs is relatively small. The random error component (statistical variance) may therefore be large, resulting in a lack of precision in the measurement of indicators. Calculating a confidence interval makes it possible to interpret a performance indicator while accounting for its random error component.

2. OBJECTIVES

To develop a method for measuring PQDCS performance indicators in individual designated screening centres (DSCs), while taking into account the characteristics of the women screened.

To develop a method of calculating confidence intervals for the measurement of indicators that has been adjusted on the basis of women characteristics.

To apply the methods developed to the examination of DSC performance. That is: measure the performance of DSCs, analyze the evolution of each DSC's performance over time, and compare the performance of each DSC with the average performance of the Quebec program as a whole, as well as with the PQDCS target.

3. METHODOLOGY

The proposed methods of analysis are illustrated using the breast cancer detection rate as the performance indicator of interest.

3.1. ADJUSTMENT FOR THE CHARACTERISTICS OF THE WOMEN SCREENED

Standardization is an adjustment method frequently used to reduce to the greatest extent possible the confounding effect caused by variations in cohort characteristics, when comparing two or more sub-populations (for example, when comparing the performance of a DSC with that of the PQDCS as a whole, or when comparing the performance of one DSC with that of another DSC). Two primary standardization methods are used: direct standardization and indirect standardization. Direct standardization consists of applying the category distribution of the adjustment variable(s) for the reference population to the rates by category of the adjustment variable(s) for the study populations. This method has significant limitations in terms of estimate stability when calculations are based on small numbers (4, 5). Indirect standardization consists of applying the rates by category of the adjustment variable(s) for the reference population to the distribution by category of the adjustment variable(s) for the study populations. Indirect standardization also has a significant drawback, since it lends itself solely to the comparison of study populations with a reference population. For example, indirect standardization allows us to compare the performance of a DSC with that of the PQDCS as a whole (the reference population), but it does not allow us to make comparisons between two DSCs. Still, given the large number of women characteristics that can differ from one DSC to another, while also having an effect on the various performance indicators being studied (i.e., given the large number of potential confounding factors), as well as the relatively small cohorts (for example, the number of breast cancer cases detected in a DSC on a yearly basis), indirect standardization was retained as the adjustment method.

The PQDCS performance indicators will therefore be adjusted to reflect the characteristics of the women screened, using indirect standardization. Indirect standardization is applied by calculating a rate ratio, namely the observed rate divided by the expected rate. Calculation of standardized mortality ratios (SMR) is a common application of indirect standardization. This method allows for adjustment according to several women characteristics. Indirect standardization has been used in other countries to obtain clinical indicators that are adjusted according to the characteristics of women at the hospital and/or physician level (6-8).

3.1.1. Observed rate

The observed rate corresponds to the crude measurement of a performance indicator in a sub-population of interest, which is to say a measurement that has not been adjusted in any way.

For example, the observed detection rate ($DR_{observed}$) for a given DSC corresponds to the proportion of women who were diagnosed with breast cancer following an abnormal screening mammogram, in relation to the total number of screening mammograms performed by the DSC.

$$DR_{observed} = \frac{\text{number of diagnosed cancers}}{\text{number of screening mammograms}} \text{ X 1,000 women}$$

For example, in DSC_1, 2,000 women had a screening mammogram through the PQDCS. Of these women, 200 had abnormal mammograms and 12 out of the 200 were diagnosed with breast cancer after further investigation. Therefore, the observed detection rate for DSC_1 is 6 cancers/1,000 women:

$$\frac{\text{12 diagnosed cancers}}{\text{2,000 mammograms}} = \text{6 cancers/1,000 women}$$

3.1.2. Expected rate

The expected rate is obtained using a logistic regression model (9, 10). In our case, the reference population is Quebec as a whole and the study sub-populations are the various DSCs of the PQDCS. Since indirect standardization calculations are based on the rate of the reference population, the logistic regression model uses the data for all participants in the PQDCS. The logistic regression model makes it possible to obtain specific rates for different combinations of characteristics of women in the study. At first, a selection of characteristics is made: only characteristics considered to be important in terms of predicting the performance indicator are retained. Specific rates can then be obtained for each combination of women characteristics retained in the final model. Finally, these specific rates are used to calculate the expected rate for the DSC, which is done by calculating the average of the specific rates for each woman in the DSC.

Three steps are needed to calculate the expected rate for a specific DSC: (1) selecting the women characteristics to use in the logistic regression model; (2) obtaining specific rates by logistic regression; and (3) calculating the expected rate for the DSC.

1. Selecting the women characteristics to use in the logistic regression model

Let us continue with our example, namely the detection rate for DSC_1. The dependent variable of the logistic regression model indicates the presence or absence of a diagnosis of screening-detected breast cancer. The dependent variables correspond to the characteristics of the women who have had a screening mammogram through the PQDCS. The characteristics of these women are collected in the PQDCS information system (SI-PQDCS) and are also listed in Table 1.

Table 1. Characteristics of women available in the SI-PQDCS

- Age
- History of breast examination(s): mammography, aspiration/biopsy, breast reduction, mastectomy, other procedure
- Physical breast exam during the past year
- Presence of symptoms at time of screening mammogram
- Breast prosthesis
- Family history of breast cancer
- Hormone-replacement therapy
- Menopausal status and age at onset of menopause
- Parity, age at first childbirth, number of children
- Body mass index
- Breast density

Only characteristics considered to be important in the prediction of the detection rate are included in the model. The importance of a given characteristic in predicting the expected detection rate is determined with the aid of discriminant value *c*, obtained using the logistic regression model. Value *c* corresponds to the area below the Receiver Operating Characteristic (ROC) curve. The ROC curve corresponds to a graph that relates the proportion of true positives (sensitivity) to the proportion of false positives (1 – specificity) of a test for different decision thresholds. The area below the ROC curve, and therefore discriminant value *c,* can be interpreted as being the proportion of pairs, one with breast cancer, the other without breast cancer, correctly predicted (11, 12). A *c* value is equal to 0.5 when the model does not discriminate (i.e., when the model has correctly predicted women's outcomes for 50% of pairs) and is equal to 1.0 when the model discriminates perfectly between cases and non-cases (i.e., when the model has correctly predicted women's outcomes for 100% of pairs).

To obtain the desired prediction model, discriminant value *c* from the complete model (with all available women characteristics) is calculated first. Then, the effect on discriminant value *c* of withdrawing each characteristic in turn is evaluated. When the withdrawal of a given characteristic has little effect in terms of altering discriminant value *c* (a change of less than 0.010), the characteristic is considered to be unimportant in the final predictive model.

The PQDCS information system lacks a few characteristics for certain women. However, the proportion of women with missing values is very small (< 2%). During the characteristic selection process, women with missing values are grouped together and incorporated into the model with the aid of an indicator variable. Once the characteristics for the final model have been selected, the missing values for a given variable are included in the variable category containing the largest proportion of women. This approach to missing values is designed to avoid significant instability in the coefficient corresponding to the missing values and produces little or no change in the coefficient of the category with the largest proportion of women.

Table 2 illustrates the selection of characteristics for the logistic regression model. Only three characteristics are considered in this example, namely age, body mass index and a history of breast aspiration/biopsy.

First, it is necessary to adjust the complete model, which comprises all the variables (Step A, Table 2). Then, each variable in the complete model is excluded in turn in order to calculate its contribution to discriminant value *c.* Three new models are therefore created. The discriminant value *c* for each model is then compared with that of the complete model. If the difference is less than 0.01, the variable is excluded from the model, since its contribution is considered to be insufficient. It is important to exclude only one variable at each step. In our example, the "history of breast aspiration/biopsy" variable is excluded from the model, since its contribution to the complete model is only 0.003.

Second (Step B, Table 2), the remaining variables in the model are removed in turn in order to calculate the new contribution of each variable to discriminant value *c*. In our example, the difference between the discriminant value *c* of the two remaining variables and that of the complete model is greater than 0.01. Therefore, the characteristics retained for the final predictive model are women's age and body mass index.

Table 2. Selection of characteristics for the final model

	Discriminant value *c*	Difference between discriminant value *c* and the complete model	Retention or exclusion of the variable in the final model
Step A			
Complete model with all participant characteristics	0.573	---	
Model without the woman's age	0.534	0.039	Retain
Model without body mass index	0.560	0.013	Retain
Model without history of breast aspiration/biopsy	0.570	0.003	Exclude
Step B			
Model without history of breast aspiration/biopsy and without age	0.521	0.052	Retain
Model without history of breast aspiration/biopsy and without body mass index	0.558	0.015	Retain

The analysis described above was carried out for each women characteristic available in the SI-PQDCS (Table 1). The characteristics retained for the purposes of adjusting each performance indicator are shown in Table 3. The choice of variables was made on the basis of the cohort of women who had a screening mammogram between January 2001 and December 2003.

Table 3. **Adjustment variables for the analysis of performance indicators**

	RR	DR	FP/C	IS	SI	NN
Body mass index	√	√	√	√	√	√
Mass		√	√	√	√	√
Breast density	√	√		√	√	√
History of mammography	√	√		√	√	√
Age		√	√	√	√	
Menopause and age at onset of menopause		√	√			√
Family history of breast cancer		√	√			
Parity and age at first childbirth		√	√			
History of breast aspiration or biopsy	√					
Symptoms (pain, inversion, discharge)						
Breast reduction or prosthesis						

√: indicates that the variable has been retained in the final model.

RR: Recall rate, DR: Detection rate, FP/C: False positive per cancer, IS: *In situ* cancers, SI: Small invasive cancers, NN: Lymph node-negative invasive cancers.

2. Obtaining specific rates by logistic regression

Using all PQDCS screening mammograms performed in a given period, the logistic regression model allows us to attribute a specific detection rate to each woman. This specific detection rate is based on the characteristics of the women and can be interpreted as the probability of breast cancer detection for each woman in the reference population, based on her personal characteristics.

For a group of possible characteristics, X_1, …, X_p, the logistic regression model equation is equal to:

$$\log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$

therefore, the specific rate is estimated as follows:

$$\pi(x) = \frac{e^{\alpha+\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p}}{1+e^{\alpha+\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p}}$$

In the above example, two characteristics were retained in the logistic regression model. These two characteristics are treated as categories with the aid of indicator variables (Table 4).

Table 4. Characteristics retained in the model and their indicator variables

Characteristics	Indicator variables
Age (years)	
50-54	--- (referral)
55-59	ag_2
60-64	ag_3
65-69	ag_4
Body mass index (kg/m^2)	
<20	--- (referral)
20-24	bmi_2
25-29	bmi_3
30-34	bmi_4
≥ 35	bmi_5

The logistic regression model equation therefore becomes:

$$\log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \alpha + \beta_1 ag_2 + \beta_2 ag_3 + \beta_3 ag_4 + \beta_4 bmi_2 + \beta_5 bmi_3 + \beta_6 bmi_4 + \beta_7 bmi_5$$

therefore,

$$\pi(x) = \frac{e^{\alpha+\beta_1 ag_2 + \beta_2 ag_3 + \beta_3 ag_4 + \beta_4 bmi_2 + \beta_5 bmi_3 + \beta_6 bmi_4 + \beta_7 bmi_5}}{1+e^{\alpha+\beta_1 ag_2 + \beta_2 ag_3 + \beta_3 ag_4 + \beta_4 bmi_2 + \beta_5 bmi_3 + \beta_6 bmi_4 + \beta_7 bmi_5}}$$

where $\pi(x)$ = the specific detection rate

The logistic regression model for the total population of screened women enables us to determine the value of coefficients α and β in this equation. With the aid of these coefficient values, specific rates based on the characteristics of the screened women can be calculated.

$$\pi(x) = \frac{e^{-5.75+0.33ag_2+0.40ag_3+0.57ag_4+0.44bmi_2+0.49bmi_3+0.63bmi_4+0.78bmi_5}}{1+e^{-5.75+0.33ag_2+0.40ag_3+0.57ag_4+0.44bmi_2+0.49bmi_3+0.63bmi_4+0.78bmi_5}}$$

For example, a woman aged 60 with a body mass index of 23 kg/m^2 will have a specific detection rate of:

$$\pi(x) = \frac{e^{-5.75+0.40\,ag_3+0.44\,bmi_2}}{1+e^{-5.75+0.40\,ag_3+0.44\,bmi_2}} = \frac{e^{-4.91}}{1+e^{-4.91}} = 0.00732$$

3. Calculating the expected rate of individual DSCs

Once a specific rate has been obtained for each woman, the expected rate for a DSC can be estimated. The expected rate corresponds to the average of the specific rates of the women screened at the DSC:

$$\text{rate}_{\text{expected}} = \frac{\sum \text{specific rates of all women having had a mammogram at the CDD}}{\text{Number of screening mammograms performed at the CDD}}$$

In our example, 2,000 screening mammograms were done at DSC_1. A specific rate is assigned to each woman attached to that DSC, based on her characteristics (Table 5).

Table 5. Specific detection rate based on the characteristics of the women screened at DSC_1.

Characteristics of women	Specific detection rate*
1st woman: age 60, bmi = 23 kg/m^2	0.00732
2nd woman: age 52, bmi = 28 kg/m^2	0.00517
…	…
2,000th woman: age 67, bmi = 19 kg/m^2	0.00560

* The specific detection rate is derived from the logistic regression model.

Therefore,

$$DR_{\text{expected}} = \frac{0.00732 + 0.00517 + \ldots + 0.00560}{2{,}000} = 5.9 \text{ cancers/1,000 women}$$

This expected detection rate is interpreted as the detection rate that would have been observed at this DSC had its performance been similar to that of the PQDCS as a whole.

3.1.3. Adjusted performance indicator

Dividing a DSC's observed rate by its expected rate provides a rate ratio generally referred to as the SMR. This ratio compares the performance of the DSC with that of Quebec as a whole, after an adjustment is made to reflect the composition of the DSC's clientele. When the observed rate is lower than the expected rate, the ratio is less than 1, indicating that the performance of the DSC is lower than that measured for the PQDCS as a whole, taking into consideration differences in women characteristics. When the observed rate is equal to the expected rate, the ratio equals 1, and when the observed rate is higher than the expected rate, the ratio is greater than 1 indicating, respectively, that the DSC's performance is equal to or greater than that of Quebec as a whole. However, interpreting a series of ratios can sometimes generate confusion. Therefore, we have chosen to multiply the ratio by the value of the performance indicator for Quebec as a whole. The result of this multiplication can be interpreted as a measurement of the performance indicator "adjusted" for the characteristics of the women screened at the DSC.

Pursuing the above example, the observed detection rate (6.0 cancers/1,000 women) and the expected detection rate (5.9 cancers/1,000 women) of DSC_1 produce a observed rate/expected rate ratio of 1.017. Since this ratio is very close to a value of 1, the performance of DSC_1 with respect to detection rate is not very different from that of the Quebec program as a whole. Given that the Quebec detection rate for the same period is 6.4 cancers/1,000 women, the "adjusted" detection rate of DSC_1 will be 6.5/1,000 women:

$$DR_{adjusted} = \frac{DR_{observed}}{DR_{expected}} \times DR_{provincial}$$

$$DR_{adjusted} = \frac{6.0 \text{ cancers/1,000 women}}{5.9 \text{ cancers/1,000 women}} \times 6.4 \text{ cancers/1,000 women}$$

$$_{adjusted} = 6.5 \text{ cancers/1,000 women}$$

It can therefore be said that the adjusted detection rate for DSC_1 (6.5 cancers/1,000 women) is almost identical to the Quebec detection rate (6.4 cancers/1,000 women), taking into consideration the composition of the clientele of DSC_1.

3.2. ESTIMATION OF CONFIDENCE INTERVALS

Since the number of mammograms performed annually is relatively low in a small number of DSCs, performance indicator calculations are based on data covering a three-year period. Furthermore, in order to account for random variability (statistical variability) in the measurement of performance indicators, a 95% confidence interval is estimated.

The confidence interval calculation for an adjusted performance indicator is derived directly from the confidence interval of the observed rate/expected rate ratio, with the aid of logistic regression (13). To ensure that the distribution of ratios respects the postulate of normality, the normal approximation method, based on a logarithmic transformation of the observed

rate/expected rate ratio, is used. The 95% confidence intervals are calculated as follows using the approximation method:

Rate ratio (R) = $\frac{observed\ rate}{expected\ rate} = \frac{X/N}{A/N} = \frac{X}{A}$

Where X = Number of cancer cases observed
A = Number of cancer cases expected
N = Number of screening mammograms performed

The variance of the log ratio is therefore equal to:

$$Var(\log R) = Var(\log(\frac{X}{A})) = Var(\log X - \log A) = Var(\log X) + Var(\log A) - 2Cov(\log X, \log A)$$

Using Taylor linearization and assuming that the covariance between log X and log A is negligible, we obtain:

$$Var(\log R) \cong \frac{Var(X)}{X^2} + \frac{Var(A)}{A^2}$$

When the indicator in question is a proportion (detection rate, recall rate, percentage of *in situ* cancers, percentage of small invasive cancers and percentage lymph node-negative invasive cancers), X follows the binomial probability law (Bin (n,p)). Therefore,

$$E[X] = np \quad and \quad Var(X) = np(1\text{-}p)$$

If p is estimated by x/n, we obtain:

$$\hat{Var}(x) \cong x\,(1-\frac{x}{n}) = \frac{x(n-x)}{n}$$

Therefore,

$$\hat{Var}(\log(R)) \cong \frac{(n-x)}{nx} + \frac{\hat{Var}(A)}{\hat{A}^2}$$

Which give us:

$$\log(R) \pm z_{\alpha/2} \sqrt{\frac{(n-x)}{nx} + \frac{\hat{Var}(A)}{\hat{A}^2}}$$

Through inverse transformation, we obtain:

$$95\%\,C.I.\,of\ the\ ratio = \frac{observed\,rate}{expected\,rate} * e^{\left(\pm Z_{\alpha/2}\sqrt{\frac{n-x}{nx}+\frac{\hat{Var}(A)}{\hat{A}^2}}\right)}$$

Variance with respect to the expected rate is calculated on the basis of the variance-covariance matrix derived from the logistic regression used to estimate the expected rates. The formula is described in the article by Hosmer and Lemeshow (13). Variance in the expected rate is minimal compared to that of the observed rate. For most centres in our study, the addition of this variability does not alter the confidence interval.

The confidence interval for the adjusted performance indicator is then obtained by multiplying the upper and lower limits of the ratio's confidence interval by the average value of the performance indicator for the PQDCS as a whole. The use of logarithmic transformation results in asymmetric confidence intervals around the ratio. If the confidence interval of the performance indicator for a given DSC does not cover the value of the performance indicator for all of Quebec, the performance of that DSC is statistically different from the overall performance of DSCs participating in the PQDCS.

If, when calculating the performance indicator for individual DSCs, cohort numbers are too small (when the number of cases is equal to or less than 5), the confidence interval is calculated with the aid of an exact method. The upper and lower limits of the confidence interval are determined by solving equations by iteration (14).

Let us return to our example concerning the detection rate of DSC_1. This DSC performed 2,000 mammograms and 12 women received a diagnosis of breast cancer following an abnormal mammogram. The centre's observed detection rate is 6.0 cancers/1,000 women and its expected detection rate is 5.9 cancers/1,000 women, which produces a rate ratio of 1.017 and an adjusted detection rate of 6.5 cancers/1,000 women. In this example, the provincial detection rate is 6.4 cancers/1,000 women. Since the number of diagnosed cancer cases in women who have had a screening mammogram at the DSC is greater than 5 (namely 12), the approximation method can be used to calculate the confidence intervals. The 95% confidence interval for the DSC's adjusted detection rate will be:

$$95\%\,C.I.\,of\ the\ ratio = \frac{6.0\ cancers/1{,}000\ women}{5.9\ cancers/1{,}000\ women} * e^{\left(\pm 1.96\sqrt{\frac{2{,}000-12}{2{,}000*12}+\frac{0.38}{(11.8)^2}}\right)}$$

$$95\%\ C.I.of\ the\ ratio = (0.58 - 1.79)$$

The 95% confidence interval for the ratio ($DR_{observed}$ $DR_{expected}$) is therefore (0.58 – 1.79). In order to obtain the 95% confidence interval for the adjusted detection rate of DSC_1, both limits of the confidence interval must be multiplied by the Quebec detection rate (namely 6.4 cancers/1,000 women). The 95% confidence interval with respect to the adjusted detection rate for DSC_1 is therefore (3.7‰ – 11.5‰). Consequently, the adjusted detection rate of DSC_1 is 6.5 cancers/1,000 women and its 95% confidence interval ranges between 3.7 to 11.5 cancers/1,000 women. Therefore, the adjusted detection rate of DSC_1 is not statistically different from the Quebec (PQDCS) detection rate, since its confidence interval covers that of Quebec. Consequently, the performance of DSC_1, in terms of its breast cancer detection rate, is comparable to the average performance of all DSCs in Quebec.

4. RESULTS

The method used to present results must lend itself to an examination of the performance of the various centres. In this section, two approaches are proposed. The first consists of presenting the evolution of an individual DSC's performance over time, while the second seeks to simultaneously present the performance of all DSCs for a given period.

4.1. EVOLUTION OF A DSC'S PERFORMANCE OVER TIME

This way of presenting results can be used to compare the performance of a DSC with that of the entire PQDCS, as well as with established program targets.

Figure 1 shows the nine retained performance indicators on a single page and thus provides an overview of the DSC's performance. The upper three graphs deal with the DSC's performance with respect to initial screening mammograms (recall rate, detection rate and number of false positives per screen detected cancer). The middle three graphs relate to the centre's performance with respect to subsequent screening mammograms (recall rate, detection rate, and number of false positives per screen detected cancer). Finally, the three graphs on the bottom are concerned with the stage of cancers detected through initial or subsequent mammograms (namely the percentage of *in situ* cancers, the percentage of small invasive cancers, and the percentage of lymph node-negative invasive cancers).

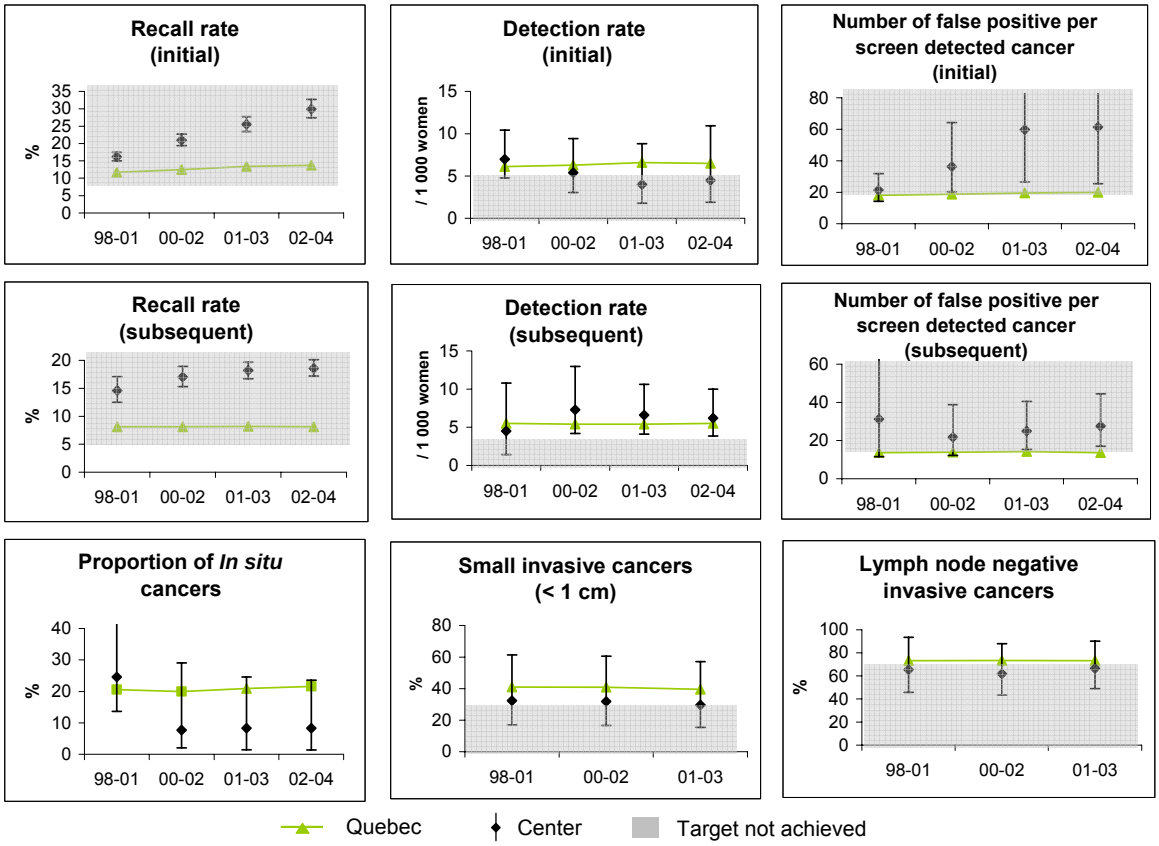
Each point on these graphs represents the performance of the DSC over a 3-year period. For example, the points situated above category "02-04" indicate the results for the years 2002, 2003 and 2004 combined. Category "98-01" is the only category that encompasses more than three years, an exception that is due to the fact that the PQDCS did not begin until May 1998. The reader will also note that the years overlap. This method, referred to as smoothing, makes it possible to reduce random variability (statistical variability) between points on a graph.

Each graph depicts the DSC's adjusted performance (black diamonds) with a 95% confidence interval (CI) (vertical lines). The DSC's performance can be compared with the Quebec average (grey triangle). The 95% confidence interval represents the random variation around the DSC result. A DSC can therefore view its performance as statistically comparable to that of the PQDCS as a whole when its confidence interval covers the value of the indicator for all of Quebec. Moreover, if the DSC's result falls within the grey zone shown on the graph, this indicates that the DSC did not attain the PQDCS target. It should be noted that there is no target for the percentage of *in situ* cancers.

In order to analyze these graphs correctly, it is crucial that the DSC's performance over time be examined. The evolution of a DSC's performance over time constitutes the best means of determining its overall performance. For example, if the DSC's confidence intervals coincide with those of the entire PQDCS and if the DSC's performance is systematically above the provincial average, one can be reasonably confident that the centre's performance is indeed superior to that of the PQDCS. This approach to the interpretation of results is essential primarily for small DSCs that detect only a few cases of cancer every year and have confidence intervals that are very wide for each point on the graph. However, care must be

taken in interpreting results, since indicators are adjusted with respect to a specific period. Interpreting a DSC's performance between different periods could be problematic in cases where the characteristics of the women screened vary markedly over time.

Figure 1. Evolution of performance* over time, by mammogram type (initial or subsequent) for DSC_2



*: Adjusted indicators for characteristics of women

4.2. PERFORMANCE OF ALL DSCS OVER A GIVEN PERIOD

In this section, we propose a means of simultaneously presenting the performance of all DSCs with respect to a specific indicator and a specific period. The figure we propose to use makes it possible to compare the performance of each DSC over a three-year period with the Quebec average, as well as with the target set under the PQDCS. And, since all the DSCs are represented on the same page, the figure also provides some insight into inter-centre performance variations. Again, care should be taken in analyzing this graph; because the results were calculated using the indirect standardization method, the focus should be on comparing each DSC's results with the provincial average, rather than on inter-centre comparisons.

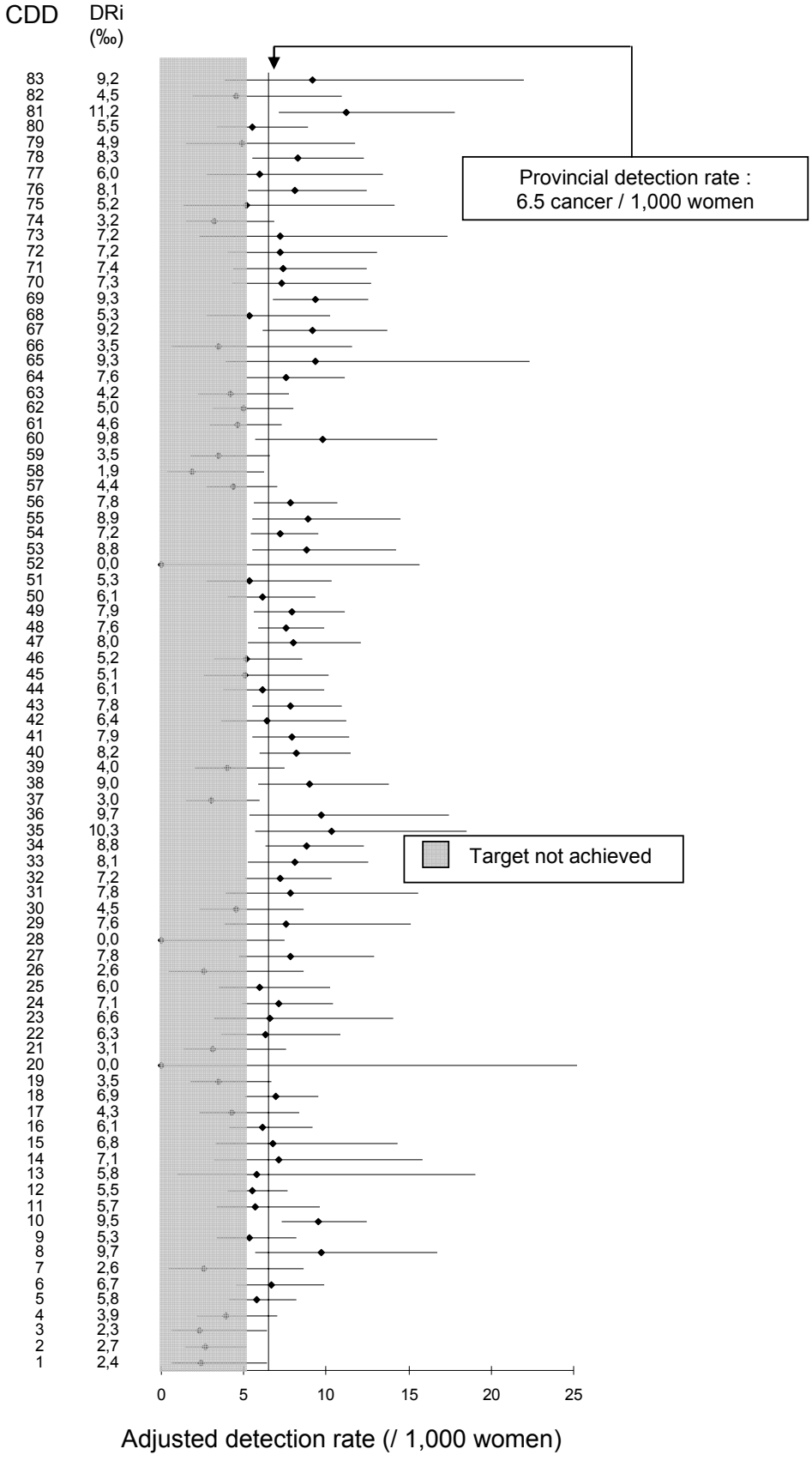
The detection rate for initial mammograms is used in Figure 2 to illustrate the proposed approach. The first column in the figure lists the identification numbers of the various DSCs. These identification numbers are designed to safeguard the confidentiality of DSC results. The second column indicates the value of the adjusted performance indicator. The indicator used in Figure 2 is the adjusted detection rate in women who have had an initial screening mammogram.

The diamonds in the graph provide a visual representation of the adjusted indicator values. A 95% confidence interval has been added (horizontal line).

The vertical line that appears along the entire length of the graph represents the value of the performance indicator for Quebec as a whole. This value is always indicated in a rectangle at the top of the figure. A DSC may assume that its performance is statistically comparable to that of Quebec as a whole when its confidence interval covers the Quebec value.

The grey rectangle represents the zone in which the PQDCS target is not met. The value of this target is also indicated in a rectangle.

Figure 2. Adjusted detection rates and 95% confidence intervals for DSCs in Quebec, initial mammogram, PQDCS 2002-2004



5. DISCUSSION

In our study, indirect standardization of performance indicators was used to estimate adjusted rates. An INSPQ report (5) on the use of direct and indirect standardization concludes that the two methods produce very similar results but that direct standardization should always be given preference. However, the authors also point out that when case numbers are low (fewer than 10) for a given centre, this method produces highly unstable estimates and may even necessitate the exclusion of data from that centre, which is something we wish to avoid as much as possible.

The adjusted rates obtained using the indirect method enable us to compare the performance of a DSC with that of all DSCs in Quebec as a group. In theory, however, this type of standardization is not appropriate for direct comparisons between centres. The comparison of adjusted performance indicators obtained through indirect standardization can be problematic since the adjusted rates are based on the distribution of women characteristics at each centre (15, 16). The adjusted rates are not standardized with the aid of a single weighting system derived from a reference population (such as data for the province), as is the case with direct standardization. Consequently, emphasis should be placed on comparing the adjusted performance indicator of individual DSCs with the performance indicator value for the entire province, rather than on inter-centre comparisons.

The adjustment of performance indicators based on women characteristics depends on the data available in the SI-PQDCS, as well as on the completeness and validity of this information. Most breast cancer risk factors, such as age, family history of breast cancer, and breast density, are collected in the SI-PQDCS. In our case, many characteristics retained in our model had no missing data, while other characteristics had a very low percentage of missing data (less than 2%).

We considered adjusting the performance indicators based on the time elapsed between the initial screening mammogram and the subsequent screening mammogram. This variable can only be measured in women who have had a subsequent screening mammogram. The literature has shown that detection rates, recall rates and sensitivity increase and that specificity decreases as the time elapsed between mammograms increases (17). However, this variable does not represent a women characteristic but rather a program characteristic. It can be influenced by decisions that are external to the women themselves. We did, however, verify its impact on the results obtained. The “time between mammograms” variable was retained for one performance indicator in our study, namely the recall rate. However, even in the case of this indicator, removal of the “time between mammograms” variable from the logistic regression model produced only a minimal variation in discriminant value *c* just above the established threshold of 0.009.

The percentage of invasive cancers detected following a normal screening could also have been included in the list of indicators to analyze. This indicator is complementary to the detection rate. In Quebec, however, the average number of cancer cases detected less than a year after a normal screening mammogram is less than 120. Even over a three-year period, many DSCs will have no “post-normal screen cancers.” Moreover, this indicator is not even calculated in regional analyses because the cohorts are too small. For this reason, the percentage of invasive cancers detected after a normal screen is not included in the list of DSC performance indicators to analyze.

REFERENCES

1. Programme québécois de dépistage du cancer du sein. Cadre de référence. Ministère de la Santé et des Services sociaux. 5-15. 1996.

2. Équipe d'évaluation du PQDCS. Tableau de bord de l'Équipe d'évaluation du PQDCS. INSPQ . 2007.

3. Zaslavsky AM. Statistical issues in reporting quality data: small samples and casemix variation. Int.J Qual.Health Care 2001;13:481-8.

4. MacLeod M, Yeung S, Sutton M. The stability of emergency admission rates for chromic diseases as a clinical indicator. Clinial Indicators Support Team. 1-14. 2004.

5. Muecke C et al. Doit-on utiliser la standardisation directe ou indirecte dans l'analyse de la mortalité à l'échelle des petites unités géographiques? 2005. Institut national de santé publique du Québec, 14 pages.

6. Kendrick S, MacLeod M. Adjusting outcomes for case mix: indirect standardisation and logistic regression. Clinical Indicators Support Team Working Paper (No 3). 1-9. 2001.

7. MacLeod M, Kendrick S. Effect of case-mix on outcome: how are case mix effects at individual level reflected in case mix effects at hospital level. Clinical Indicators Support Team Working Paper (No 4). 1-9. 2001.

8. New York State Department of Health. Coronary artery bypass surgery in New York State 1996-1998. 2001.

9. Glance LG, Osler TM. Comparing outcomes of coronary artery bypass surgery: Is the New York Cardiac Surgery Reporting System model sensitive to changes in case mix? Critical care medicine 2001;29:2090-6.

10. Glance LG et al. Impact of changing the statistical methodology on hospital and surgeon ranking: the case of the New York State cardiac surgery report card. Medical care 2006;44:311-9.

11. Poses RM, Cebul RD, Centor RM. Evaluating physicians' probabilistic judgments. Med Decis.Making 1988;8:233-40.

12. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med 1999;130:515-24.

13. Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance based on logistic regression models. Stat.Med 1995;14:2161-72.

14. Bernard PM. Analyse des tableaux de contingence en épidémiologie. Presses de l'Université du Québec, 2004.

15. Breslow NE, Day NE. Statistical methods in cancer research. Volume II--The design and analysis of cohort studies. IARC Sci.Publ. 1987;1-406.

16. Bernard PM, Lapointe C. Mesures statistiques en épidémiologie. Presses de l'Université du Québec, 1998.

17. Yankaskas BC et al. Association between mammography timing and measures of screening performance in the United States. Radiology 2005;234:363-73.

APPENDIX

The table below lists and defines the indicators used to evaluate the performance of the DSCs and for which calculations are to be made on a regular basis. Please note that the recall rate, detection rate, and number of false positives per screen detected cancer are calculated separately for women who have had an initial screen and those who have had a subsequent screen. The other indicators are calculated based on the totality of all initial and subsequent screening mammograms performed. The number of false positives per screen detected cancer (FP/C) measures the same information as the positive predictive value (PPV). It represents the average number of false positives recorded before a true positive (detected breast cancer) is obtained. The exact relationship between FP/C and PPV corresponds to:

$$FP/C = (1 - PPV) / PPV.$$

Table. List of performance indicators used to evaluate the performance of PQDCS designated screening centres (DSCs).

Indicators	Definitions
Recall rate	Number of abnormal mammograms / Total number of mammograms X 100
Detection rate	Number of cancers detected / Total number of mammograms X 1,000
Number of false positives per screen detected cancer	Number of false positives / Number of cancers detected
In situ cancers detected	Number of *in situ* cancers / Number of cancers detected X 100
Small invasive cancers (≤ 1 cm)	Number of small cancers (≤ 1 cm) / Number of invasive cancers of known size X 100
Lymph node-negative invasive cancers	Number of lymph node-negative invasive cancers / Number of invasive cancers with dissection X 100

N° de publication : 811