Evaluation and Monitoring of Screening Programmes

International Agency for Research on Cancer World Heath Organisation	European Network of Cancer Registries	Europe Against Cancer European Commission

Evaluation and monitoring of screening programmes

Edited by:
R. Sankila, E. Démaret, M. Hakama, E. Lynge, L.J. Schouten, D.M. Parkin
for the **European Network of Cancer Registries**

Contents

Foreword

The European Commission considers it a priority to help in preventing death from cancer by early detection through screening programmes. Thus the third action plan to combat cancer, better known as the Europe against Cancer programme, has helped to provide evidence for European high quality cancer screening programmes. A *conditio sine qua non* to establish and to monitor the efficacy of any cancer screening programme is the European-wide availability of high quality cancer registries, which is the objective of the European Network of Cancer Registries (ENCR), another priority of the Europe against Cancer programme.

While early detection of a specific cancer by screening appears, at first sight, to be a promising method of preventing death, conclusive evidence on its efficacy at the public health level must first be established. For cancer of the cervix uteri, for example, this was possible, using time trend analyses, based originally on data from population-based cancer registries in the Nordic countries. For breast cancer screening using mammography, the evidence of benefit is largely based on fairly small screening trials, and somewhat varying results of population based time trend analyses. Overcoming these shortcomings became the prime objective of the European Breast Cancer Screening Network, which has aided in establishing European-wide guidelines for more uniform high quality breast cancer screening. In turn, this improved the comparability of the available data from different national settings, demonstrating true European added value.

It seems likely that new screening programmes will be implemented as public health policy based upon limited screening trials, usually without a mechanism for evaluation of their effectiveness. However, it is not self-evident that the positive results of screening trials will be replicated in a service setting, let alone be readily transferable to other countries with different health care systems. In this instance, one of the few tools available to evaluate the results of such new public health interventions is the population-based cancer registry.

A number of factors may affect the efficacy of a screening intervention in the long term. There are examples where the anticipated results of a new screening policy were not achieved, but where, after critical re-evaluation, the organisation of the programme was remodelled yielding satisfactory results. In the era of evidence-based medicine, this routine monitoring of the performance of <u>existing</u> screening programmes is of utmost importance. Population-based cancer registries are uniquely placed to provide the data needed for these processes.

This monograph provides a wide range of experiences from cancer registries in Europe and North America in the evaluation and monitoring of existing screening programmes. As such, I hope it will aid researchers in choosing the necessary data and applying appropriate methods. The reader should also realise that despite more than 30 years' experience, there are still very demanding aspects in the evaluation and monitoring of screening programmes. These challenges need very careful consideration in each unique setting.

David Byrne
Commissioner
Health and Consumer Protection Directorate-General
European Commission

List of contributors

Dr Freda E. Alexander
Department of Public Health Services
The University of Edinburgh
Medical School
Teviot Place
Edinburgh EH8 9AG, Scotland
United Kingdom

Dr Ahti Anttila
Mass Screening Registry
Finnish Cancer Registry
Liisankatu 21 B
00170 Helsinki
Finland

Dr Anne-Marie Benhamiche
Registre Bourguignon des Cancers
Digestifs
INSERM CRI 9505
Faculté de Médecine
7, Blvd Jeanne d'Arc
21033 Dijon Cedex
France

Dr Stefano Ciatto
Centro per lo Studio e la
Prevenzione Oncologica
Viale A. Volta 171
50131 Florence
Italy

Dr Jan Willem Coebergh
Comprehensive Cancer Centre South (IKZ)
Zernikestraat 29
P.O.Box 231
5600 AE Eindhoven
The Netherlands

Mrs Eva Démaret
Unit of Descriptive Epidemiology
International Agency for Research on
Cancer
150, cours Albert Thomas
69372 Lyon Cedex 08
France

Dr Jean Faivre
Registre Bourguignon des Cancers Digestifs
Faculté de Médecine
7, Blvd Jeanne d'Arc
21033 Dijon Cedex
France

Dr Timo Hakulinen
Finnish Cancer Registry
Institute for Statistical and Epidemiological
Cancer Research
Liisankatu 21B
00170 Helsinki 17
Finland

Dr Esa Läärä
University of Oulu
Department of Mathematical Sciences
P.O.Box 333
90571 Oulu
Finland

Dr Elsebeth Lynge
Institute of Public Health
University of Copenhagen
Blegdamsvej 3
2200 Copenhagen N
Denmark

Dr Jenny McCann
Department of Community Medicine
University of Cambridge
Strangeways Research Laboratory
Wort's Causeway
Cambridge CB1 8RN
United Kingdom

Dr Sue Moss
Cancer Screening Evaluation Unit
Institute for Cancer Research
D. Block, Cotswold Road
Sutton, Surrey SM2 5NG
United Kingdom

Dr D. Maxwell Parkin
Unit of Descriptive Epidemiology
International Agency for Research on
Cancer
150, cours Albert Thomas
69372 Lyon Cedex 08
France

Dr Risto Sankila
Unit of Descriptive Epidemiology
International Agency for Research on
Cancer
150, cours Albert Thomas
69372 Lyon Cedex 08
France

Dr Leo J. Schouten
Maastricht Cancer Registry
Comprehensive Cancer Centre Limburg
P.O.Box 2208
6201 HA Maastricht
The Netherlands

Dr Jenifer Smith
South and West Cancer Intelligence Unit
Highcroft, Romsey Road
Winchester SO22 5DH
United Kingdom

Dr Robert A. Stephenson
Department of Surgery
University of Utah School of Medicine
50 North Medical Drive
Salt Lake City, UT 84132
USA

Dr Diane Stockton
Cancer Surveillance Group
Scottish Cancer Intelligence Unit
Information & Statistics Division
Trinity Park House, South Trinity Road
Edinburgh EH5 3SQ, Scotland
United Kingdom

Dr Mohammed A. Tazi
Registre Bourguignon des Cancers Digestifs
Faculté de Médecine
7, Blvd Jeanne d'Arc
21033 Dijon
France

Dr Laufey Tryggvadóttir
Icelandic Cancer Registry
P.O.Box 5420
125 Reyjkavik
Iceland

Dr Han J. van der Rhee
Department of Dermatology
Leyenburg Hospital
Leyweg 275
2545 CH The Haag
The Netherlands

Dr Jos A.A.M. van Dijck
Cancer Registry Department
Comprehensive Cancer Centre East
P.O.Box 1281
6501 BG Nijmegen
The Netherlands

Dr Suzanne Wait
Laboratoire de Santé Publique
Faculté de Médecine
Université Louis Pasteur
11, rue Humann
67085 Strasbourg Cedex
France

Dr Marco Zappa
Centro per lo Studio e la
Prevenzione Oncologica
Viale A. Volta 171
50131 Florence
Italy

Chapter 1. Introduction: study design and the potential of cancer registration

D.M. Parkin

A distinction should be drawn, at the outset, between studies which are concerned with the evaluation of screening, and those which aim to monitor the performance of an ongoing programme. Strictly speaking, evaluative studies should be performed before the introduction of a particular screening programme, in order to decide on the likely effectiveness, and possibly also on the resource implications (cost-effectiveness). Normally, we are concerned that screening will reduce the number of new cases of invasive cancer (if the test detects precursor lesions) or the number of deaths from the cancer (if aiming to detect early or pre-clinical invasive disease). It is only with this assurance that certain so-called 'intermediate endpoints' – such as numbers of preclinical lesions detected/treated, size/stage distribution, and survival in screened vs. unscreened populations – can be used to monitor screening programmes. By themselves, they provide no reassurance, as the lessons of screening for neuroblastoma (see Chapter 17 in this volume) clearly demonstrate.

Sometimes cancer registries have been a part of evaluative studies on effectiveness of screening. Probably more frequently, they are expected to aid the monitoring of existing programmes. The methods available depend upon the data to which the registry has access. These may be, for example, simply the incident cancer cases in a given population in which screening is ongoing. Alternatively, the registry may have access to the database of the screening programme itself, so that the screening history of individual members of the population is known. Clearly, the latter circumstance permits investigations to be better focussed, than in circumstances where the extent of screening must be inferred from population averages.

Studies using the registry database alone

Changes in incidence/mortality

Here the focus is simply upon the time trends in incidence and mortality from cancer following the introduction of screening (or changes of screening policy). This is analogous to the before/after study design in evaluative research, and like all such studies, is fraught with problems in interpretation, largely through lack of certainty of what would have happened in the absence of screening (no control group) with which to compare the outcome. The type of evidence concerning the presence of screening activity, and its intensity, is variable. It may be simply

anecdotal (e.g., 'screening was introduced in 1988'), be supported by actual data from the screening programme (rate of testing – tests per 1000 eligible subjects per year), or possibly from the registry itself (registrations of screen-detected disease). As an example of the latter, many workers have used registration rates of carcinoma *in situ* or of micro-invasive cancers as an indication of screening activity, with which to compare trends in symptomatic, invasive disease (e.g. Boyes *et al.*, 1981).

It is only for screening programmes which attempt to prevent the onset of invasive cancer that monitoring of incidence provides a direct measure of outcome. The principal example is cancer of the cervix (screening by cytology), although even here (see below) there is an additional component of benefit due to detection of early invasive disease (micro-invasive, or occult/asymptomatic invasive). For oral cancer screening, too, at least part of the potential benefit should be through detection of pre-malignant conditions such as leukoplakia, erythroplakia and submucous fibrosis. However, the appropriate treatment for such lesions is far from clear, and so far there is no objective evidence that the rate of invasive disease can be reduced by their detection.

For screening programmes designed to detect early invasive cancers, no reduction in incidence should occur. Indeed, the introduction of screening should be followed by a rise in incidence (as prevalent, asymptomatic cases are detected), followed by a fall, so that cumulative incidence is ultimately unchanged over what it would have been without screening (Walter and Day, 1983).

Cancer registry data can be used to monitor the extent of this phenomenon. The introduction of screening for breast cancer has certainly led to an increase in incidence in age groups receiving mammography in the USA (Chu *et al.*, 1996), and the UK (Quinn and Allen, 1995), and the increase in recorded 'incidence' of prostate cancer in the USA has been even more dramatic (see Chapter 13 in this volume). It remains to be seen whether these increases will be reversed.

There have been many studies of trends in incidence of cervix cancer in relation to screening policies – for example the studies of incidence in the Nordic countries with respect to the date of introduction and intensity of organised screening in different countries, and the trends by age group with respect to those targeted for screening (Hakama, 1982; Hakama *et al.*, 1991).

Trends in size or stage of tumours

Many cancer registries attempt to record stage of registered tumours, and sometimes also their size. One good reason for doing so is to study the effects of screening, since cancers detected by screening programmes must have a more

favourable stage distribution, and be of smaller size than those detected clinically (via symptoms), if the programme is to have any benefit.

Using registry data alone, it has been common to examine the proportionate distribution of diagnosed cancer cases by stage at diagnosis, in relation to the presence of screening (or measures of screening intensity, as above). More meaningful is calculation of rates of disease, by stage, in relation to screening activity. In principle, screening should lead to a reduction in the incidence of advanced cancers. This is usually in the context of trends over time (see, for example, for cervix cancer: Christopherson et al., 1976; Johannesson et al., 1978; for stomach cancer: Hisamichi et al., 1991; for breast cancer: Chu et al., 1996; for colon cancer: Chu et al., 1994).

An inevitable problem of studying temporal trends in stage (without a true comparison group) is the phenomenon of stage migration. This occurs because improvements of investigative methodology over time result in tumours being classified to higher stages than they would have been in an earlier period (by detection of small metastases, for example).

Comparisons between sub-populations receiving screening or not (or different intensities of screening) (e.g., Taplin et al., 1997) largely avoids problems of stage migration, but does raise other questions concerning the comparability of the populations being compared.

Trends in survival

Cancer registries are increasingly concerned with follow-up of registered cases in order to perform survival analyses. Survival is influenced by many factors (Sankar and Black, 1999), but size and stage of tumour are powerful predictors of outcome. Screening programmes might well expect to be associated with improvements in overall survival, therefore. Survival trends, by age, have been studied in relation to early diagnosis of breast cancer due to screening (Chu et al., 1996) or to improving breast-awareness (Stockton et al., 1997) and to early detection of colon cancer (Chu et al., 1994). In both cases, the favourable trend in survival was due to a shift to earlier stage at diagnosis as well as better survival within stage. This latter effect is probably the results of improved therapy, although the possibility that some of it is due to stage migration should also be considered (Feinstein et al., 1985).

Ecological (correlation studies)

The idea is to compare rates of disease in different populations, with some measure of screening intensity. As far as cancer registries are concerned, the relevant rates are incidence rates, or, more usefully, the change in incidence in relation to the screening input. This means that studies are confined to cervix cancer screening, with the 'exposure' being expressed as the rate of screening (in relation to the

eligible population), or as registrations of *in situ* cancers (Fouquet and Gage, 1996; Bergström *et al.*, 1993; Lynge, 1983).

Studies with linkage to data on exposure to screening

Generally, such studies will involve linkage to the database of a screening programme, in order to allocate individuals to different categories of screening exposure.

The simplest study design is the cohort study, in which incidence of disease is compared in screened vs. unscreened individuals with, for cervix cancer, the objective of deciding the decrease in risk which might plausibly be ascribed to screening. Thus, in British Colombia the relative risk of invasive cervix cancer, adjusted for age, was 6.8 times higher in unscreened women, compared with those who had had one or more tests (Fidler, 1968).

Pettersson *et al.* (1986) carried out linkage between the screening register and the national Cancer Registry of Sweden. They were thus able to calculate incidence rates of cervix cancer according to screening history (number of tests, time since previous test, result of test) as well as by age. These rates were compared with those recorded prior to screening (taken to represent 'expected' values in the absence of screening).

The problem with these studies, as in all non-randomised comparisons, is the difficulty in being certain that the groups being compared are similar except for their experience of screening. Women who do not take part in screening programmes (or who attend few times, in contrast to regularly) are probably at higher risk of cervix cancer than those who do not (Hakama and Räsänen-Virtanen, 1976). In a more recent linkage study of the screening programme in Sweden with the cancer registry, Sparén *et al.* (1996) found that incidence in unscreened women was similar to that pre-screening (Figure 1). The authors suggest that this implies that bias due to self-selection might be small, although the curves pertain to different time periods (1958-67 and 1968-92), and hence very different birth cohorts, so that it is difficult to be certain.

The tactic usually employed by epidemiologists in analysing cohort studies involving self selection into exposed vs. non-exposed groups is to adjust for potential confounders (in the screening context, this would be factors related to both screening uptake and risk of cervix cancer). This has rarely been done in the cohort-type of analysis mentioned above. Probably this is due to the relatively sparse data on individuals which can be obtained from the records of screening programmes – which do not usually include risk factors for the disease, or for death from the disease. In general, this is not a problem for evaluating cervix cancer

4

screening programmes, where the reduction in risk in screened vs. unscreened women is simply too big to be ascribed to confounding. But the level of mortality reduction expected in breast cancer screening programmes is small, and could easily be obscured by the inclusion of women at higher than average risk of death
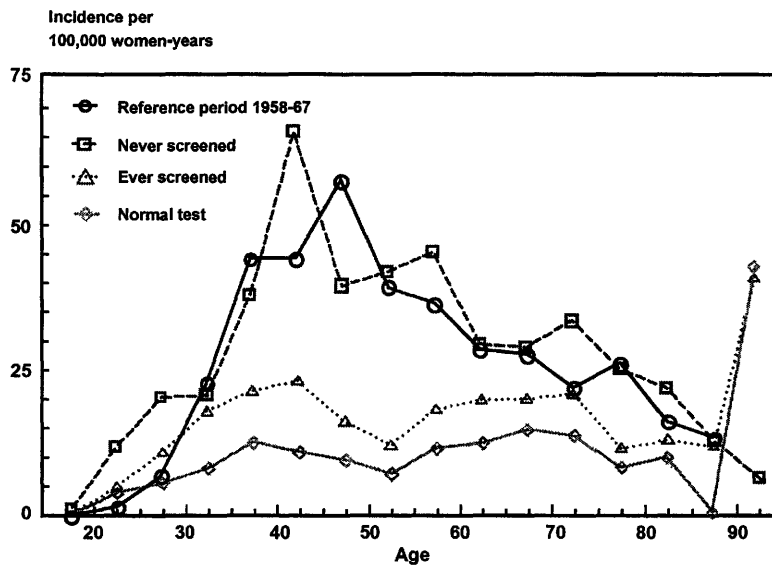
Figure 1. Age-specific incidence rates of cervix cancer in a cohort of women in Sweden according to screening status, and for the same geographic area pre-screening (1958-67) (Sparén, 1996, reproduced by permission of the author)

from breast cancer in the screened population. In fact, it seems clear that women who choose to undertake regular breast self-examination (Gastrin *et al.*, 1994) or to attend mammographic screening programmes (Moss *et al.*, 1992; Taplin *et al.*, 1989) do have a higher than average risk (incidence) of breast cancer, by virtue of their risk profile (educational level, height, weight, fertility, family history, etc.). Thompson *et al.* (1994) used a case-cohort study design, in which detailed information on screening and on risk factors was obtained from a sample of the entire cohort of women (subjects enrolled in a health maintenance organisation) as well as for the breast cancer deaths occurring. Adjusting for confounders (related to death from breast cancer) reduced the risk of death from breast cancer among women aged 50 or more and screened within the previous year from 1.08 to 0.61.

The introduction of breast cancer screening in Finland was very systematic, involving certain municipalities, and, within these, invitations to women in different (single year) birth cohorts over a five-year period. This population could be followed up by the cancer registry and, for a short period after screening (3-4

years), the mortality rates were compared in screened, invited but unscreened, and control women (Hakama *et al.*, 1997). Although technically a cohort study, women were <u>allocated</u> to screened and unscreened categories by adjacent birth cohorts, so there could have been little difference between the groups in terms of confounders. A small but significant reduction in mortality was present in women invited for screening (vs. unscreened controls) at 3-4 years post screening.

Cancer registration and the randomised controlled screening trial

Randomised trials remain the established method of evaluating the effectiveness of screening (they have no role in monitoring ongoing programmes). The fact that they are difficult to perform, of long duration, and expensive, means that few have been completed (for the evaluation of screening for breast cancer, and for colon cancer). In theory, a randomised trial established in the area which is served by a population-based cancer registry, would have the enormous advantage of automatic surveillance of the study groups for the onset of new cases in intervention and control populations. The issue of loss to follow up from out-migration would be a problem for regional registries, but it would probably be non-differential if randomisation was thorough, and lead to a loss of power, rather than any bias.

The role of the registry in this context would be to follow up the screened and intervention groups in order to monitor

a)   overall (cumulative) incidence in screened/control groups (as a check on randomisation and/or the presence of any over-diagnosis);

b)   stage/size of tumours, and survival, in screened and control groups (intermediate endpoints, but change in these must be present if mortality reduction is to be achieved);

c)   deaths from cancer in screened/unscreened (the outcome measure, collected by registries undertaking follow-up of registered cases of all cancers, or the cancer under study).

In fact, most randomised trials of screening in Europe and the United States have established their own follow-up of the study populations, rather than rely upon cancer registration (although registries may comprise part of the case-finding network, e.g. UK Trial of Early Detection of Breast Cancer (1993)). Perhaps this reflects the investigators lack of knowledge of registry work (or potential), as well as study funding sufficiently lavish to dispense with economies at this stage of the investigation. In our own randomised studies of screening for breast cancer by physical examination in Manila, Philippines (Parkin *et al.*, 1997), or for oral cancer by visual inspection in Trivandrum, India, we depend on linkage between the study database which contains details of each individual's screening history, and the cancer registry database.

6

Monitoring screening programmes

Monitoring of incidence rates of cancers for which screening aims at early detection (and reduction of deaths rather than of incidence) has several purposes. The interpretation of overall incidence in screened vs. unscreened populations is not straightforward. On the one hand, any observed differences might be interpreted as reflecting self-selection for screening as noted. On the other hand, early detection programmes clearly influence incidence of disease, either by simply advancing date of diagnosis (with a short term increase in incidence) or by detecting cancers that would never have otherwise surfaced during life (over-diagnosis). The latter effect seems particularly strong for prostate cancer, so that incidence rates in a group of individuals that has been screened will be permanently higher than the unscreened.

Day *et al.* (1989) have described how information on incident cancers in screened/unscreened individuals may be used to monitor breast cancer screening programmes (Table 1). Cancer registry data are required in order to estimate the 'expected' distribution of cancers by stage, or the 'expected' incidence rates (overall, by age group, and/or by stage), against which the results in the screened population can be evaluated. Linkage of the cancer registry and the screening programme is required in order to calculate incidence rates of interval cancers (detected between screenings) and incidence of advanced cancers. The methods concerned are reviewed by Wait (see Chapter 9 in this volume). An example of this type of analysis is provided by McCann *et al.* (1998), who studied the incidence of breast cancer, by stage, according to mode of detection, following the introduction of screening for breast cancer in East Anglia. By linkage of the cancer registry with the data base of the screening programme, cancers could be separated into screen detected, interval cancers, cancers in non-attenders, etc. Schouten *et al.* (1998) have performed a similar analysis for the population of Limburg in the Netherlands.

The case-control study has a rather long history of use in the evaluation of the effectiveness of cancer screening programmes (Morrison, 1985). These studies are of course much more simply and rapidly applicable than the cohort designs described earlier. They also have the advantage that screening histories need only be obtained for a limited number of subjects. This may even be done by interviewing the subjects (if deaths from cancer do not comprise the case-group), without recourse to screening records, although several studies have suggested that interview data may not always reflect that from more objective sources (Gordon *et al.*, 1993). Although traditionally the case-control study has been used to investigate the effectiveness of screening, there is no conceptual difference to the 'audit' of ongoing programmes with the same techniques (Sasieni *et al.*, 1996). For cervix cancer, registries should be particularly well placed to provide a list of all cases of invasive cancer as the potential case group. As in the cohort studies

described earlier, the main problem in interpreting the results of case-control studies is selection bias (Moss, 1991; Friedman and Dubin, 1991).

Table 1. Monitoring measures for breast cancer screening programmes (Adapted from Day *et al.*, 1989, Br J Cancer 59:954-958, by permission of the publisher Churchill Livingstone)

Measure	Comment	Additional information	Evaluation provided
Prevalence at the first screen	As a multiple of 'expected' incidence in screened women	'Expected' estimated from incidence in non-compliers (un-screened) and a comparable population with no screening programme (e.g., historical rates)	Sensitivity, lead time, sojourn time, predictive value
Rate of interval cancers	As a proportion of 'expected' incidence in screened women, by time since last screen	Identify all interval cancers. 'Expected' rate calculated as above	Sensitivity, lead time, sojourn time, predictive value
Stage/size distribution of screen-detected cancers (1) initial screen (2) subsequent screen	Compare to 'expected' stage distribution in the absence of screening	Stage/size distribution in non-compliers (unscreened) and in a comparable population with no screening programme (e.g., historical data)	Indicates potential for reduction in absolute rate of advanced cancer
Incidence of advanced cancers	'Advanced' must be defined, and majority of cases classified as early/advanced. Compare to 'expected' incidence	Stage/size distribution in non-compliers (unscreened) and in a comparable population with no screening programme (e.g., historical data)	Early surrogate of mortality

References

Bergström R, Adami HO, Gustafsson L, Pontén J & Sparén P (1993) Detection of preinvasive cancer of the cervix and the subsequent reduction in invasive cancer. *J Natl Cancer Inst* **85**:1050-1057

Boyes DA, Worth AJ & Anderson GH (1981). Experience with cervical screening in British Columbia. *Gynaecol Oncol* **12**:S143-155

Christopherson WM, Lundin FE Jr, Mendez WM & Parker JE (1976). Cervical cancer control: a study of morbidity and mortality trends over a twenty-one-year period. *Cancer* **38**:1357-1366

Chu KC, Tarone RE, Chow WH, Hankey BF & Ries LA (1994). Temporal patterns in colorectal cancer incidence, survival, and mortality from 1950 through 1990. *J Natl Cancer Inst* **86**: 997-1006

Chu KC, Tarone RE, Kessler LG, Ries LA, Hankey BF, Miller BA & Edwards BK (1996). Recent trends in U.S. breast cancer incidence, survival, and mortality rates. *J Natl Cancer Inst* **88**:1571-1579

Day NE, Williams DR & Khaw KT (1989). Breast cancer screening programmes: the development of a monitoring and evaluation system. *Br J Cancer* **59**:954-958

Feinstein AR, Sosin DM & Wells CK (1985). The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med* **312**:1604-1608

Fidler HK, Boyes DA & Worth AJ (1968). Cervical cancer detection in British Columbia. A progress report. *J Obstet Gynaecol Br Commonw* **75**:392-404

Fouquet R & Gage H (1996). Role of screening in reducing invasive cervical cancer registrations in England. *J Med Screen* **3**:90-96

Friedman DR & Dubin N (1991) Case-control evaluation of breast cancer screening efficacy. *Am J Epidemiol* **133**:974-984

Gastrin G, Miller AB, To T, Aronson KJ, Wall C, Hakama M, Louhivuori K & Pukkala E (1994). Incidence and mortality from breast cancer in the Mama Program for Breast Screening in Finland, 1973-1986. *Cancer* **73**:2168-2174

Gordon NP, Hiatt RA & Lampert DI (1993). Concordance of self-reported data and medical record audit for six cancer screening procedures. *J Natl Cancer Inst* **85**:566-570

Hakama M (1982). Trends in the incidence of cervical cancer in the Nordic countries. In: Magnus K (ed) *Trends in Cancer Incidence: Causes and Practical Implications.* Washington, Hemisphere, pp. 279-292

Hakama M & Räsänen-Virtanen U (1976). Effect of a mass screening program on the risk of cervical cancer. *Am J Epidemiol* **103**:512-517

Hakama M, Magnus K, Pettersson F, Storm H & Tulinius H (1991). Effect of organized screening on the risk of cervical cancer in the Nordic countries. In: Miller A, Chamberlain J, Day N, Hakama M & Prorok P (eds) *Cancer Screening.* Cambridge University Press, pp. 153-162

Hakama M, Pukkala E, Heikkila M & Kallio M (1997). Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *Br Med J* **314**:864-867

Hisamichi S, Fukao A, Sugawara N, Nishikouri M, Komatsu S, Tsuji I, Tsubono Y & Takano A. (1991). Evaluation of Mass Screening Programme for Stomach Cancer in Japan. In: Miller A, Chamberlain J, Day N, Hakama M & Prorok P (eds) *Cancer Screening..* Cambridge University Press

Johannesson G, Geirsson G & Day N (1978). The effect of mass screening in Iceland, 1965-74, on the incidendence and mortality of cervical carcinoma. *Int J Cancer* **21**:418-425

Lynge E (1983). Regional trends in incidence of cervical cancer in Denmark in relation to local smear-taking activity. *Int J Epidemiol* **12**:405-413

McCann J, Stockton D & Day N (1998). Breast cancer in East Anglia: the impact of the breast screening programme on stage at diagnosis. *J Med Screen* **5**:42-48

Morrison AS (1985). Screening in chronic disease. *Oxford University Press, New York*, USA

Moss SM (1991) Case-control studies of screening. *Int J Epidemiol* **20**:1-6

Moss SM, Summerley ME, Thomas BT, Ellman R & Chamberlain JO (1992). A case-control evaluation of the effect of breast cancer screening in the United Kingdom trial of early detection of breast cancer. *J Epidemiol Community Health* **46**:362-364

10

Parkin DM, Ngelangel C, Pisani P, Gibson, L & Esteban D (1997). *Proceedings of the Department of Defense Breast Cancer Research Program Meeting 'Era of Hope'*, Washington, October 31 - November 4

Pettersson F, Näslund I & Malker B (1986). Evaluation of the effect of Papanicolaou screening in Sweden: Record linkage between a control screening registry and the national cancer registry. In: Hakama M, Miller AB & Day NE (eds) *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, pp. 91-105

Quinn M & Allen E (1995). Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. United Kingdom Association of Cancer Registries. *Br Med J* 311:1391-1395

Sasieni PD, Cuzick J & Lynch FE (1996). Estimating the efficacy of screening by auditing smear histories of women with and without cervical cancer. The National Co-ordinating Network for Cervical Screening Working Group. *Br J Cancer* 73:1001-1005

Schouten LJ, de Rijke JM. Schlangen JT & Verbeek AL (1998). Evaluation of the effect of breast cancer screening by record linkage with the cancer registry, The Netherlands. *J Med Screen* 5:37-41

Sparén P, Gustafsson L, Hsich EC, Yven J & Adami HO (1996). Pap smear screening and the risk of squamous cell cancer of the cervix: a prospective study in Sweden. In: Sparén P (ed) *Early detection and screening for cancer of the cervix in Sweden during the 20th Century.* Comprehensive summaries of Uppsala dissertations from the Faculty of Medicine. 599. Uppsala, Acta Universitatis Upsaliensis

Stockton D, Davies T, Day N & McCann J (1997). Retrospective study of reasons for improved survival in patients with breast cancer in east Anglia: earlier diagnosis or better treatment. *Br Med J* 314:472-475

Taplin S, Anderman C & Grothaus L (1989). Breast cancer risk and participation in mammographic screening. *Am J Public Health* 79:1494-1498

Taplin SH, Mandelson MT, Anderman C, White E, Thompson RS, Timlin D & Wagner EH (1997). Mammography diffusion and trends in late stage breast cancer: evaluating outcomes in a population. *Cancer Epidemiol Biomarkers Prev* 6:625-631

Thompson RS, Barlow WE, Taplin SH, Grothaus L, Immanuel V, Salazar A & Wagner EH (1994). A population-based case-cohort evaluation of the efficacy of mammographic screening for breast cancer. *Am J Epidemiol* **140**:889-901

UK Trial of Early Detection of Breast Cancer Group (1993). Breast cancer mortality after 10 years in the UK Trial of early detection of breast cancer. *The Breast* **2**:13-20

Walter SD & Day NE (1983). Estimation of the duration of a pre-clinical disease state using screening data. *Am J Epidemiol* **118**:865-886

Chapter 2. Planning and designing of screening programmes

M. Hakama

In the evaluation and monitoring of screening for cancer, the design of a programme can not be separated from the analysis. The programme should be designed in such a way that it can be evaluated. This simple principle will be illuminated by the experience from Finland. The programmes considered are for cervical cancer and for breast cancer. These are the modes of screening which can as a rule be run as public health policies. Much which follows below applies to screening for other cancers. However, these other programmes are still in experimental phases.

Indicators of effect

The purpose of screening is to reduce the mortality from the cancer subjected to screening. Therefore, the primary indicator of effect is the observed mortality compared with the expected, assuming no screening. For cervical cancer, the preinvasive disease is detected by screening and therefore reduction in incidence of invasive cancer is also a valid indicator of effectiveness.

Process or intermediate indicators are also used in the evaluation of screening. They are applicable if there is proof of effect in terms of reduction in mortality, and evidence of a relationship between the intermediate indicators and the outcome indicator. The best of these intermediate indicators is a change in the incidence (not proportion) of advanced disease, due to screening. Most process indicators are the results of the screening examinations, such as the numbers or proportions of early or preinvasive cancers detected at screening or the proportion which such cases comprise of all cancers. Short term follow-up may permit estimates of the validity of the screening test.

The basic measures of validity of a screening test are sensitivity and specificity. Sensitivity is an indicator of the yield, and specificity that of the cost of the screening. For a successful screening programme the performance of the test alone is clearly not sufficient. There are other problems, e.g. in participation in the programme, and in the referral system (for individuals with a positive screening test).

It is, therefore, useful to consider also the programme sensitivity and programme specificity (Hakama, 1984). The programme sensitivity is the proportion of the

persons diagnosed as having the disease as a result of participating in the screening programme out of the total number of persons with the disease in the target population covered by the programme. The programme specificity is the proportion of persons not diagnosed as having the disease in the disease-free part of the target population. Not diagnosing the cancer will be either due to a negative test, or to the test not being applied, i.e. the person was not screened at all, or that a person with a positive test did not have diagnostic confirmation of the test result. A valid test is a prerequisite for a successful programme, but the programme may fail even if the test is valid. There may be substantial differences in test sensitivity and programme sensitivity, on the one hand, and in test specificity and programme specificity on the other.

The cancer registry collects data on the incident cases, and provides indicators of programme validity through the follow-up of the target population and the linkage of cancer registry and screening information.

Mortality data can be extracted from death certificates. There are reasons, however, to prefer information on cancer deaths from the cancer registry. The cancer registry diagnosis is more accurate because more is known of the individual at the registry than by the one who signs the death certificate. This is especially true for cervical cancer because in the death certificate the diagnosis is often given as uterus unspecified. Such diagnoses were common in the early years of cervical cancer screening. Furthermore, incidence of invasive disease for cervical cancer and sometimes incidence of advanced disease for breast cancer is available at the cancer registry before the deaths occur, so that the process of evaluation can be shortened.

Finally, only deaths occurring in patients for whom the disease was diagnosed after the introduction of screening can be prevented by screening. The time of diagnosis is reliably available only at the cancer registry. In the following, 'refined mortality' is used if cases diagnosed before the first screening round were excluded.

Design of a screening programme

Two main issues must be decided when designing a screening programme. First, whether to launch a special programme or to implement the screening spontaneously in the target population. Second, whether to advocate general population-based screening or selective screening targeted of high-risk groups only.

14

Organised programmes or spontaneous screening

Screening can be spontaneously implemented or it can be implemented through active planning. There are several reasons why active planning resulting in an organised programme is better than spontaneous screening. First, only an organised programme can be evaluated reliably and in detail. Second, when comparisons between effects of organised and spontaneous programmes have been feasible, organised programmes have been shown to result in larger effects than spontaneous ones, as will be described in the next chapters. Third, without active planning, there is the danger of an unpredicTable increase in cost. Spontaneous screening has no inbuilt mechanism to prevent unnecessarily frequent screens which result in a high marginal cost. Spontaneous screening is likely to pay the greatest attention to high sensitivity without regard for the effect of high sensitivity on reducing specificity. For a high-technology screening programme, such as mammography for breast cancer, low specificity results in high costs and uncontrolled adverse effects.

Ages for and frequencies of screening

Major organisational considerations of any screening programme are the age at which to start the programme, the interval at which the test is applied and the age at which to stop screening. For example, recommended practice for cervical cancer screening in western populations, which share about the same risk of disease and resources available, ranges from an annual smear from the start of sexual activity to a smear every five years from age 30 to age 55. Hence the difference in number of tests is tenfold.

The age range to be screened depends on variation in the risk of disease with age, the protective effect of screening by age, and the person-years saved by screening. Many cancers are rare at young ages and their incidence increases rapidly with age. The cost-effectiveness will be low if screening attempts to prevent rare occurrences at young ages. The protective effect of a screening test varies with age. Such variation is due to the test itself, to the disease or to the subject. For example, taking an adequate cervical smear for cancer becomes more difficult after the menopause, and the sensitivity of the test will decrease. Also, the results of treatment may be poorer at old ages. For clinical cases, survival is better for young cervical cancer patients. In many screening programmes the attendance – often the most important single determinant of success of a screening programme – goes down with age. Therefore, it is likely that the effectiveness of screening is inversely related to age. Screening of older age groups results in fewer person-years saved because of the competing risk of death from other causes. Again, cost-effectiveness will be reduced if older populations are screened.

15

The variation in the protective effect of screening resulting from changing to the interval between the screening rounds depends on the length of the pre-clinical phase of the disease and on the distribution of sojourn times, and not simply their average length.

For cervical cancer, data on the protective effect have been obtained from several screening programmes (IARC Working Group on Evaluation of Cervical Cancer Screening Programmes, 1986). It seems that the protective effect after a negative smear is high (more than 90 per cent) and is only marginally dependent on the interval between screenings of up to three years (Table 1). Even ten year intervals yield a two-thirds reduction in the risk. This is in sharp contrast with breast cancer, for which the sojourn times seem to be considerably shorter (Day *et al.*, 1988). Invasive cervical cancer is rare below the age of 30 and starting the screening before that age results in few prevented cases, whereas many preinvasive lesions will be detected and require treatment. Because most of the latter would regress, or remain preinvasive until age 30, the benefit is small, but the harm (in terms of adverse effects and cost) is large.

Table 1. Protective effect of screening for cervical cancer after a negative smear, ages 35–64 (reproduced by permission of the BMJ Publishing Group from IARC Working Group on Evaluation of Cervical Screening Programmes, Br Med J 293:659-664,1986)

Interval between screenings (years)	Reduction in cumulative incidence (%)
1	93
2	92
3	91
5	84
10	64

Selective screening

Selective screening means applying the screening test to only a proportion of the target population, and is based on selecting groups at particularly high risk. The purpose of confining screening to high-risk groups is to reduce the resources required for the programme, or sometimes to reduce the rate of adverse effects of the screening test among the target population. A selective screening programme should detect a substantial proportion of the disease in the entire target population

(i.e., only a few cases of the disease are assumed to originate among the low-risk groups that are not subjected to screening).

Selective screening has an effect only on the programme validity but not on the test validity. Because the purpose of selective screening is to reduce the cost, the programme specificity is increased. Because of the inverse relationship between specificity and sensitivity, the programme sensitivity of selective screening will be less than that for non-selective screening.

To decide whether screening of selected high-risk groups may be preferable to general population screening, the sensitivity and specificity (i.e., costs and yield) of the programmes must be compared. In Table 2, the reduction in cost is given in terms of the high-risk group as a percentage of the target population. The yield is given as the proportion of all cases of disease in the target population that fall in the high-risk group. The cost percentage is an upper limit of the programme specificity and the yield percentage is an upper limit of programme sensitivity. The yield depends strongly on the risk of disease in the high-risk group divided by the risk in the low-risk group (relative risk). For example, in order to have 90 per cent of the cases in the high-risk group comprising 10 per cent of the target population, the relative risk must be almost 100. Very few such strong risk factors are known for any disease. Instead, combination of several risk factors at the same time has been attempted.

Table 2. Proportion of cases belonging to the high-risk group out of all cases in the target population, by percentage of the high-risk group out of total target population and by relative risk (high vs. low risk)

Size of high-risk group (%)	Relative risk		
	2	10	100
1	0.02	0.09	0.50
10	0.18	0.53	0.92
50	0.67	0.91	0.99

Combination of risk factors can be defined in several ways. Such a combination should identify persons exposed to several risk indicators at the same time. There

is good theoretical evidence that the combined effect is usually multiplicative (Day and Brown, 1980; Peto, 1977). If the risk factors are independent, then the size of the population exposed to several risk indicators decreases rapidly as the number of risk factors increases. Table 3 gives an example of this inverse relationship, which implies that the proportion of the total cases found in each high-risk group defined by a simultaneous occurrence of several risk indicators is less than that found by means of a single risk factor.

Table 3. Combining independent risk factors; effect on size of high-risk group, on risk and on yield

	Risk factor					
	A	B	C	A and B	B and C	A and B and C
Size of high-risk group (%)	50	20	10	10	2	1
Relative risk	2	5	10	10	50	100
Proportion (%) of cases in high-risk group of all cases	67	56	53	53	51	50

On the other hand, if the high-risk group is identified by a combination of risk factors, assuming at least one risk factor only to characterise the individual belonging to the high-risk group, then the size of the population to be screened will increase. In the hypothetical situation (assuming independent occurrence of A, B and C) of Table 3 the high-risk group identified by either A or B or C is 64 per cent of the total population. The final reduction in cost will be less than 36 (100–64) per cent because only a part of the total cost of the programme is directly proportional to the number of screening tests. Some basic expenses are relatively fixed for any programme.

The best combination of risk indicators can be defined also in other ways. Multiple discriminant analysis probably has the widest applicability. However, most applications of the method show that, if the high-risk group is small enough to result in a substantial reduction in cost, then the cases belonging to the low-risk group are too numerous, i.e. the programme sensitivity is unacceptably low. This

seems to be the experience for breast cancer (Farewell, 1977; Soini and Hakama, 1978) and cervical cancer (Hakama *et al.*, 1979) and in general the finding is rather independent of the disease or defect.

There are some important exceptions. The definition of high-risk groups does not have to be in the terms of known aetiological factors, it is sufficient that the high-risk characteristics are correlates of the risk, yielding high programme sensitivity. Age and sex are used in almost all screening programmes as indicators of high-risk groups. Symptoms and diagnostic results based on previous screening may also be used to define a high-risk group, although these generally modify the time interval between screenings, rather than define an a priori high-risk group. Thus, women with positive mammography or Pap smear followed by a normal histology are at high-risk of breast cancer and cervical cancer respectively and re-screening at shorter intervals may be warranted.

So far, selective screening based on high-risk populations defined by aetiological risk factors has not been useful. The programme sensitivity has been low and a substantial proportion of the disease in the total target population originates in the low-risk group that is not subject to screening. In a developing country with very limited resources the approach may be feasible, however.

International comparisons

If the effect of screening is large, it may be evident in population rates of disease. Perhaps the most convincing evidence for the effectiveness of organised screening programmes (Hakama, 1982; Läärä *et al.*, 1987) stems from the comparison of trends and differentials in incidence with the screening activities in the Nordic countries (Hakulinen *et al.*, 1986; Engeland *et al.*, 1993). Only Norway had not started an organised screening programme by the 1990's and the reduction in incidence there was much smaller than in the other countries (Table 4). Denmark was partly covered by an organised programme and the reduction in incidence was largest in areas with organised programme within Denmark (Lynge *et al.*, 1989). Spontaneous Pap smears were common in all the Nordic countries (Hakama, 1982) and smears taken within the organised programmes were, in fact, fewer than the spontaneous ones except in Iceland. However, the decrease in the incidence of cervical cancer seems to have been proportional to the intensity of the organised screening programme. Such comparisons are too crude to reveal the effects of different intervals between the screening rounds or of the age to start and stop screening (Hakama, 1982).

Table 4. Observed and predicted annual age-adjusted incidence rates (per 100 000) of invasive cervical cancer in the Nordic countries in selected time periods. Data from the Nordic Cancer Registries (reproduced by permission of APMIS from Hakulinen *et al.*, 1986, and Engeland *et al.*, 1993)

Country	Observed				Predicted	
	1956 –1960	1966 –1970	1976 –1980	1983 –1987	1998 –2002	2008 –2012
Denmark	30	30	19	16	11	11
Finland	14	14	6	4	2	2
Iceland	16	26	9	13	10	9
Norway	15	17	17	13	9	8
Sweden	18	18	10	9	7	7

Individual level comparison

The most important characteristics of an organised screening programme is the personal invitation. From the point of view of research, i.e. evaluation of the effectiveness of screening, this invitation defines the population to be screened. The invitees can then be divided into screenees and non-responders, but the reduction in risk should be evaluated in the total target population of invitees, and compared with the risk in the invitees before screening, or with an independent, apparently similar, unscreened population.

Screening for cervical cancer in Finland

Every woman in Finland, aged 30–55, receives a personal invitation at regular intervals (every 5 years) to attend the organised screening programme for cervical cancer. In the invitation letter she is given an appointment place and time. The result of the Pap test is also given by mail, independent of whether normal, suspicious or malignant. Approximately 400 000 women with 1 400 000 woman-years, with information on the actual (participants) or potential (non-respondents) participation in the first screening, were followed up and analysed by a cohort design (Hakama and Räsänen-Virtanen, 1976). Among these women were the first ones, under the national policy, to reach the first rescreening after the 5-year

interval. The protective effect, in terms of reduction in the incidence of invasive disease among the screenees, was about 80% compared to the national rates before the screening started (Table 5). This is not simply due to the responders being a selected (low-risk) subset of the target population. If this were so, risk of cervical cancer among the total target population would have remained unchanged. To eliminate bias due to selection, incidence in the total target population should be compared to the expected one. The incidence of invasive cervical cancer in the target population in our study, responders and non-responders combined, was 32% of that among the controls showing, therefore, a 68% protective effect due to screening (Table 5).

Although selection could not explain the protective effect of screening, there still was the problem of unbiased choice of controls for the target population. We used the incidence for the whole female population of Finland shortly before the start of the national programme to define the expected risk. It could be argued that a decreasing trend in the overall incidence was already taking place before the start of the programme, and that the estimate of the protective effect was exaggerated, due to biased expected rates. However, the lag between the control rates and screening rates was short and, if anything, there was an increasing trend in the overall incidence of cervical cancer in all the Nordic countries, including Finland, before the start of screening (Hristova and Hakama, 1997)

Table 5. Annual age-specific incidence rates (per 100 000 population) of invasive carcinoma of cervix uteri among the Finnish population before intense screening period in 1962–1966 and after a written invitation to participate in the programme in 1963–1971 (reproduced by permission of Oxford University Press from Hakama and Räsänen-Virtanen, 1976, Am J Epidemiol 103:512-517)

	Age						Cumulative risk
	30–34	35–39	40–44	45–49	50–54	55–59	
Before screening	7	20	35	46	47	46	0.010
Intention to screen							
Screenees	2	2	5	10	12	12	0.002
Non-responders	19	34	37	82	68	26	0.016
Total	6	5	7	15	17	13	0.003

The effect of screening on the risk of invasive cervical cancer is large. Many routine health services activities have only small effects at a public health level and one cannot presume that these effects will be demonstrable by a crude design or analysis. Instead, public health policy should be designed in such a way that small effects can be identified with reasonable accuracy. Screening for breast cancer will, at best, have a much smaller effect on disease mortality than screening for cervix cancer.

In Finland, nationwide population-based screening for breast cancer was started in 1987 (Hakama *et al.*, 1997). Women in birth cohorts recommended by the National Board of Health were individually identified and invited for screening. The programme started with women born in 1928, 1932 and in 1936 and it covers the ages of 50 to 59 years and can be continued up to 64. Further birth cohorts were invited in subsequent years and the women are rescreened every 2 years (Figure 1). Individual municipalities decide on organisation of the activity with state subsidising and most of them have a special agreement with the Cancer Society of Finland or its member societies to run the screening.

Year of screening	Year of birth											
	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939
1987	■	□	□	□	■	□	□	□	■	□	□	□
1988	□	□	■	□	□	□	■	□	□	□	■	□
1989	■	□	□	■	□	□	□	□	■	■	□	□
1990	□	□	■	□	□	■	■	□	□	■	■	□
1991	■	■	□	■	□	■	■	□	■	□	□	□
1992	□	□	■	□	□	■	■	□	■	□	■	□
1993	□	■	□	■	□	■	■	■	■	□	■	□
1994	□	□	■	□	□	■	■	□	■	□	■	□
1995	□	□	□	■	■	■	■	■	■	□	■	□
1996	□	□	□	□	□	■	■	□	■	□	■	□

Figure 1. Finnish National Board of Health's recommendation of screening rounds in an organised screening programme for breast cancer, by birth cohort and calendar year (reproduced by permission of BMJ Publishing Group from Hakama *et al.*, 1999, J Med Screen 6:209-216)

22

The material first analysed to evaluate the effect consisted of the women born in 1927–1939 and residing in the municipalities screened by the cancer organisations. The invitations and screening mammograms in 1987–1989 were recorded, and each invitee was subsequently classified as a screenee or a non-responder.

The expected mortality was estimated, first, on basis of the preceding national mortality from breast cancer as in the study on cervical cancer. There were no indications of effect on the age-specific breast cancer mortality in the age group 50 to 59 years (Figure 2) or in the cohort-specific breast cancer mortality in cohorts born in 1928, 1932 or 1936 (Figure 3). As demonstrated later, this approach was too crude to disclose any effect.
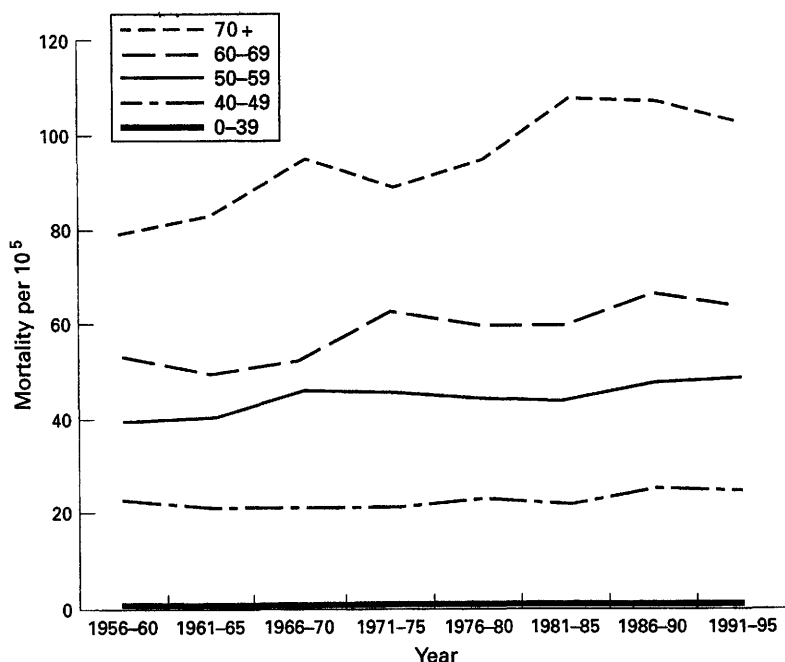
Figure 2. Age specific mortality rates from breast cancer in Finland in 1956–1995 (reproduced by permission of BMJ Publishing Group from Hakama et al., 1999, J Med Screen 6:209-216)

In the second approach, individual controls were selected. These controls were women in the same municipalities as those screened and they were born in 1927, 1929, 1931, 1933, 1935, 1937 or 1939 and individually identified at the time of identification of the invitees. Because the women born in 1931 and 1937 were recommended to be screened for the first time in 1989 they potentially provided

few person years only and short follow-up, and they were excluded from the analyses. Women born in 1936 were also excluded in order to balance the ages of cases and controls.

Altogether 89 893 women were invited to participate the programme in 1987–1989 and 76 389 accepted the invitation (Table 6). The number of controls was 68 862. Altogether 907 breast cancers were diagnosed among the invitees and 677 among the controls by the end of 1992.
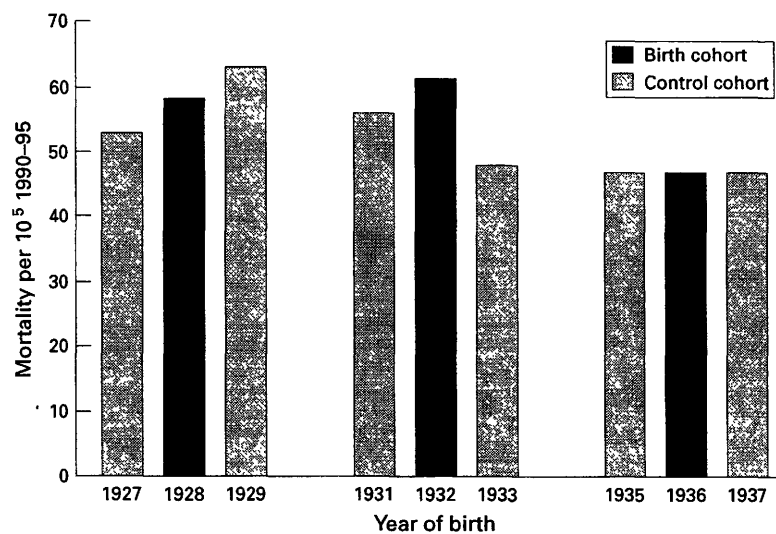

Figure 3. Birth cohort specific mortality rates from breast cancer in Finland among women born in 1927–1937 (reproduced by permission of BMJ Publishing Group from Hakama *et al.*, 1999, J Med Screen 6:209-216)

The total breast cancer mortality was first evaluated by comparing the numbers of breast cancer deaths in 1987–1992 among the invitees (210) to that expected for the whole of Finland. Among those screened, standardised mortality ratio for breast cancer was only 0.63 (Table 7). The bias due to self selection of women into those participating and into non-responders was eliminated by comparing the mortality by the intention to screen. The mortality among non-responders was very high (SMR = 3.67). Mortality from breast cancer among the invitees – those screened and those who did not attend combined – was equal to that expected on the basis of the Finnish rates (SMR = 1.01). This would point to ineffectiveness of the screening programme.

24

Table 6. Nation-wide screening programme for breast cancer organised by the cancer organisations in Finland in 1987–1989. Cumulative number of cases of breast cancer in 1987–1992 (reproduced by permission of the BMJ Publishing Group from Hakama *et al.*, 1997, Br Med J 314:864-867)

	Invitees			Controls	Total
	Screenees	Non-responders	Total		
Women	76 389	13 504	89 893	68 862	158 755
Breast cancers	774	133	907	677	1 564

Table 7. Total number of deaths from breast cancer with standardised mortality ratios (SMR): Nation-wide screening programme for breast cancer organised by the Cancer Society of Finland in 1987–1989 with follow-up for mortality to 31.12.1992 (reproduced by permission of the BMJ Publishing Group from Hakama *et al.*, 1997, Br Med J 314:864-867)

Breast cancer, total	Invitees			Controls
	Screenees	Non-responders	Total	
Numbers	114	96	210	· 175
SMR	0.63	3.67	1.01	1.12

However, the risk of death among the controls was higher than expected (SMR = 1.12), indicating that the municipalities with high risk of breast cancer were more likely to have been included in the programme. The problems inherent in selecting an appropriate referee population to provide "expected" rates, against which to quantify the effect of screening, is well demonstrated. The use of "national rates" was inappropriate (too low). In addition, the effect of screening was underestimated by the inclusion of deaths which could not possibly have been prevented (occurring in women diagnosed with breast cancer before screening began). Overall, there was a small difference in the total breast cancer mortality,

25

which was favourable to those invited for screening compared to those not invited (RR = 1.01/1.12 = 0.90). Finally, the effect was estimated on the basis of "refined mortality". There were 64 deaths among the invitees and 63 deaths among the controls from breast cancer diagnosed after the start of screening, with follow-up to the end of 1992 (Table 8). The refined breast cancer mortality was lower among the invitees than among the controls, RR = 0.76 (SMRs 0.31/0.41), indicating a 24 per cent protective effect due to screening. The public health programme was effective. The effect was small and was not evident in the national mortality rates (not even if classified by year of birth).

Table 8. Refined numbers of deaths from breast cancer with standardised mortality ratios (SMR) from breast cancer: Nation-wide screening programme for breast cancer organised by the Cancer Society of Finland in 1987–1989 (reproduced by permission of the BMJ Publishing Group from Hakama *et al.*, 1997, Br Med J 314:864-867)

Breast cancer, refined	Invitees			Controls
	Screenees	Non-responders	Total	
Number	49	15	64	63
SMR	0.27	3.58	0.31	0.41

The role of the cancer registry

It appears from the above that design is more important than analysis in demonstrating effectiveness of screening. Design issues are epidemiological (whereas many of the quality issues are not). Epidemiological expertise should be found at the cancer registries, which should be involved in the planning of a screening programme from the very beginning. In Finland the screening for breast cancer was initiated by the cancer registry which permitted many optimal details in the design.

The effectiveness of screening cannot be evaluated solely on the basis of the cancers found at screening. The cancer registry is essential in identifying the cancers diagnosed among the screenees outside the programme (interval cancers),

among the non-responders, and among the controls or – in absence of a control group – in providing the expected incidence and mortality estimates assuming no screening.

References

Day NE & Brown CC (1980). Multistage models and primary prevention of cancer. *J Natl Cancer Inst* **64**:977–989

Day N, Walter S, Tabár L, Fagerberg C & Collette H (1988). The sensitivity and lead time of breast cancer screening: A comparison of the results of different studies. In: Day NE, Miller AB (eds) *Screening for breast cancer*. Toronto, Hans Huber, pp. 105–109

Engeland A, Haldorsen T, Tretli S, Hakulinen T, Hörte LG, Luostarinen T, Magnus K, Schou G, Sigvaldason H, Storm HH, Tulinius H & Vaittinen P (1993). Prediction of cancer incidence in the Nordic countries up to the years 2000 and 2010. *APMIS* Suppl., **38**:1-124

Farewell VT (1977). The combined effect of breast cancer risk factors. *Cancer* **40**:931–936

Hakama M (1984). Selective screening by risk groups. In: Prorok PC and Miller AB (eds) *Screening for cancer*. I – General principles on evaluation of screening for cancer and screening for lung, bladder and oral cancer. Geneva, UICC Technical Report Series **78**:71–80

Hakama M (1982). Trends in the incidence of cervical cancer in the Nordic countries. In: Magnus K (ed) *Trends in Cancer Incidence: Causes and Practical Implications*. Washington, Hemisphere, pp. 279–292

Hakama M & Räsänen-Virtanen U. (1976) Effect of a mass screening program on the risk of cervical cancer. *Am J Epidemiol* **103**:512–517

Hakama M, Pukkala E & Saastamoinen P (1979). Selective screening: Theory and practice based on high-risk groups of cervical cancer. *J Epidemiol Community Health* **33**:257–261

Hakama M, Pukkala E, Heikkilä M & Kallio M (1997). Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *Br Med J* **314**:864–867

Hakama M, Pukkala E, Söderman B & Day N (1999). Implementation of screening as a public health policy: issues in design and evaluation. *J Med Screen* **6**:209-216

Hakulinen T, Andersen A, Malker B, Pukkala E, Schou G & Tulinius H (1986). Trends in cancer incidence in the Nordic countries. *Acta Pathol Microbiol Immunol Scand* **94A** Suppl. 288:1–151

Hristova L & Hakama M (1997). Effect of screening for cancer in the Nordic countries on deaths, costs and quality of life up to the year 2017. *Acta Oncol* **36** Suppl. 9:1-60

IARC Working Group on Evaluation of Cervical Cancer Screening Programmes (1986). Screening for squamous cervical cancer: the duration of low risk after negative result of cervical cytology and its implication for screening policies. *Br Med J* **293**:659–664

Läärä E, Day N & Hakama M (1987). Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. *Lancet* **1**:1247–1249

Lynge E, Madsen M & Engholm G (1989). Effect of organised screening on incidence and mortality of cervical cancer in Denmark. *Cancer Research* **49**:2157–2160

Peto R (1977). Epidemiology, multi-stage models and short term mutagenicity test. In: Hiat HH, Watson JD, Winsten JA (eds) *Origins of Human Cancer.* New York, Cold Spring Harbor Laboratory, pp. 1403–1428

Soini I &, Hakama M (1978). Failure of selective screening for breast cancer by combining risk factors. *Int J Cancer* **22**:275–281

28

Chapter 3. Evaluation and monitoring of cancer screening: theoretical issues

S.M. Moss

Methods of evaluation

Randomised controlled trials

Ideally, screening for a disease should be evaluated by means of a randomised controlled trial to establish both effectiveness and cost-effectiveness before a population-based screening programme is introduced. Randomised controlled trials (RCTs) of cancer screening compare mortality in populations offered and not offered screening, although sometimes the randomisation is of volunteers rather than the general population (Mandel et al., 1993), and sometimes cluster rather than individual randomisation is used (Roberts et al., 1990). It is important to note that it is disease-specific mortality which is compared, and that screening for any individual cancer site should not be expected to have a demonstrable effect on all-cause mortality (Hardcastle, 1997). Comparisons of all-cause mortality are however useful to provide evidence on lack of bias in randomisation. Subjects diagnosed with cancer before entry to the trial are generally excluded, either prior to randomisation or at the point of analysis, since they could not have benefited from screening.

This evaluation process has taken place for breast cancer screening, where a number of randomised controlled trials were carried out in the 1960-80's and have shown reductions in breast cancer mortality in the population offered screening of the order of 25% (Wald et al., 1994) for women aged 50 and over. Population screening programmes for breast cancer are now established or being developed in a number of countries. The effectiveness of screening for colorectal cancer by faecal occult blood (FOB) testing in reducing mortality from the disease has also been demonstrated by randomised controlled trials (Mandel et al., 1993; Hardcastle et al., 1996; Kronborg et al., 1996). Trials are in progress for screening for other cancer sites such as prostate and ovarian cancer.

Randomised controlled trials usually involve the collection of detailed data in order to establish a population register, and to record information on screening intervention, which is generally beyond the scope of cancer registries, although registries will have a role, for example, in providing follow-up data on cancer incidence.

Alternative methods of evaluation

However, other methods of evaluation are sometimes necessary. For example, in the case of screening for cervical cancer, no randomised controlled trial has ever been performed, yet population screening is widely recommended and carried out. Evidence for the effectiveness of cervical cancer screening has come mainly from comparisons of trends, both in the incidence of invasive disease and in mortality from cervical cancer, as well as from case-control studies (see Chapters 6-8 in this volume). Comparisons of trends are made either between time periods before and after the introduction of screening, or between similar but geographically separate regions with and without organised screening (Louhivouri, 1991), or between areas with different levels of intensity of screening (Miller *et al.*, 1976). The problems associated with such comparisons of trends are discussed below under evaluation of population screening.

Survival analysis

An improvement in disease-specific survival is a necessary, but not sufficient, indicator of a beneficial effect of screening. Cancer registries are generally able to produce survival analyses for series of cases registered in different time periods. If data on screening history/method of diagnosis are held on the cancer registry database, then comparisons of survival by method of diagnosis as well as between time periods will be possible. More commonly, some form of record linkage between registry and screening databases will be required. However, there are a number of biases, which will affect such comparisons.

<u>Lead-time</u> is the length of time by which the diagnosis of a case is advanced by screening. The survival time of such a case will be increased by the lead-time L (Figure 1), even if the subject in fact dies at exactly the same point in time at which they would have died if diagnosed clinically.
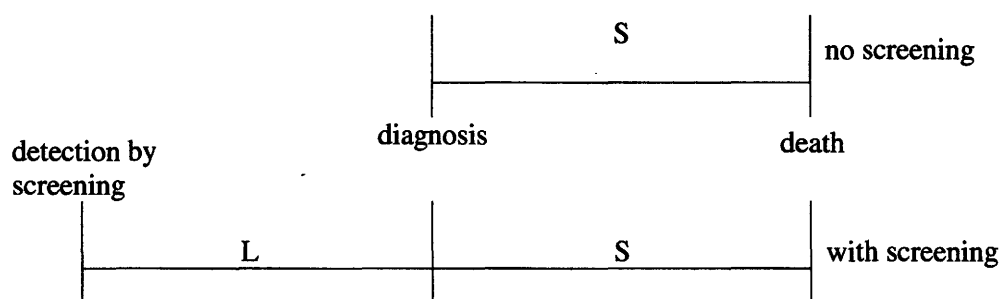

Figure 1. The effect of lead-time (L) on the total survival time of screen-detected patients (L+S)

Such a case wills not benefit from screening in terms of life-years gained, but screening is likely to have a detrimental effect in terms of quality of life, since the subject lives longer with the knowledge that they have the disease, and with the possible adverse effects of any treatment. However, it is also possible that quality of life may be improved. For cases where death from the cancer is delayed or prevented, survival will still be increased by length L over and above the gain due to screening.

Length bias can also affect comparisons of survival. The aim of screening is to detect cancers in the 'pre-clinical detectable phase' (PCDP) before they become symptomatic and are diagnosed clinically. For any given cancer site, there will be a wide variation in the growth rate of tumours, and hence in the length of the PCDP. This in turn is likely to be associated with the prognosis of the case, with the slower-growing, less aggressive, tumours having relatively better prognosis. Screening at routine intervals is more likely to detect slow-growing tumours, which spend longer in the PCDP, and hence screen-detected cases may be associated with a better prognosis (Figure 2). This is likely to be particularly true for first or prevalent screens, where there is a large prevalent pool of undiagnosed slow-growing tumours. Conversely, interval cancers occurring between routine screens may be more aggressive and have poorer prognosis.

It is worth noting that while these biases will influence comparisons between screen-detected and other cases, they will also affect comparisons of survival

Length bias

Slow growing -
good prognosis

Fast growing -
poor prognosis
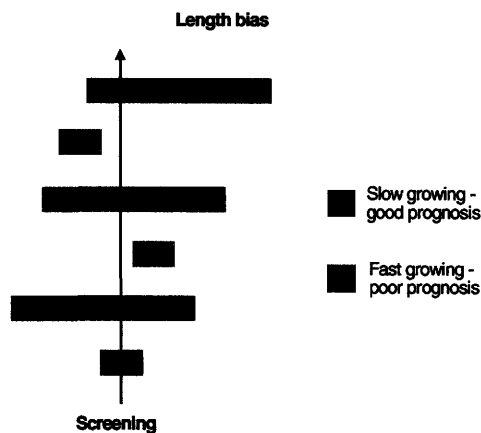
Screening

Figure 2. The effect of length bias on the proportions of slow and fast growing tumours detected by screening

figures for all cases in time periods before and after the introduction of screening. In such comparisons the extent of the bias will depend on the coverage of screening and the proportion of screen-detected disease.

Selection bias affects primarily comparisons of cancers detected in acceptors and non-acceptors of screening. This can affect comparisons both of incidence and of survival or mortality. The effect of selection bias may be in either direction, i.e. those accepting screening may be at either increased or decreased risk of developing and/or dying from the disease, compared with the general population. For disease-incidence, selection bias may occur due to an association of a risk factor such as social class with both incidence rate and with the probability of attending for or accepting screening. For example, in the HIP randomised trial of screening for breast cancer carried out in New York in the 1960's (Shapiro *et al.*, 1982), the non-attenders for screening had a 20% lower incidence of breast cancer than the women in the control group, implying a 11% greater risk in the 65% of women accepting screening. In this case, the selection bias meant that the actual benefit of screening in those accepting was greater than would otherwise be estimated. In the UK Trial of Early Detection of Breast Cancer (1988), there was little difference in the observed incidence in the non-attenders for screening and the comparison centres. However, the non-attenders had a significantly higher mortality from breast cancer. This effect of selection bias on mortality can be due to those not attending for screening being those more likely in any event to present at a late stage of disease and hence with poorer prognosis. This bias is also sometimes referred to as the 'healthy screenee effect' or, where screening is based on volunteer participation, as the 'healthy volunteer effect'.

As discussed below, socio-economic status can also potentially cause bias when cluster randomisation is used, if different arms of the trial have different underlying risk. This was observed in the Edinburgh RCT of breast cancer screening, where the control arm had a higher all-cause mortality, and lower risk of breast cancer (Alexander *et al.*, 1989).

Selection bias is a particular cause for concern when a case-control study approach is used to estimate the effectiveness of screening. In such studies, screening history in cases and controls are compared. Although a number of such studies have attempted to adjust for potential confounding, nevertheless many appear to overestimate the benefit of screening (Moss, 1991).

Selection bias will also affect cohort studies in which mortality (or incidence of invasive disease) in subjects with different screening histories is compared (Lynge and Poll, 1986).

32

Case-control studies

Case-control studies are being increasingly used to estimate the effectiveness of screening, either in the absence of data from randomised controlled trials, or to provide additional information. They are thus a potential application of cancer registry data, since they will rely on complete ascertainment of cases, however defined. However, both the design and interpretation of such studies require some care, and there is now a considerable literature on the methodology of such studies (Sasco et al., 1986; Moss, 1991; Cronin et al., 1998; Weiss, 1983). Problem areas include establishing eligibility criteria to ensure that both cases and controls had the opportunity to be screened, and were at risk for the study endpoint.

Sasco et al. (1986) identified two different settings which have implications for the selection of cases and controls:
Type A refers to the situation where the aim of screening is to detect cancer at an earlier stage than that at which it would present clinically, and thus to reduce the risk of dying from the disease. An example of this is screening for breast cancer.
Type B refers to the situation where the aim of screening is to detect disease in a precancerous state, and to prevent the incidence of invasive disease. An example of this is screening for cervical cancer, where the detection and treatment of dysplasia will prevent the onset of invasive cervical cancer.

However, it should be noted that some screening (e.g. for colon and oral cancer) is aimed at detecting both preinvasive and early invasive disease, and that all screening aims ultimately at reducing disease-specific mortality.

Selection of cases and controls

The definition of case subjects should be based on the event which screening is trying to prevent. In Type A screening, cases should be selected as deaths from the disease in question (although some studies have used patients with late-stage cases of disease as a surrogate). For case subjects to have been able to benefit from screening, they must have been diagnosed with the disease after they had the opportunity to be screened. As in randomised controlled trials, no benefit would be expected to be observed until several years of follow-up after the introduction of screening. In this situation, the screen at which screen-detected cancers are diagnosed should be included in the screening history. However, it needs to be clear that any such test is a screening test rather than a diagnostic procedure. This may not always be clear (e.g. in determining the reason for a PSA test for prostate cancer). However, it has been pointed out (Cronin et al., 1998) that exclusion of all symptomatic tests may also cause bias if such tests are likely to be carried out for diagnosis of an associated disease. For example, exclusion of PSA tests for

benign prostate hyperplasia might overestimate the efficacy of the screening test for prostate cancer.

Controls should ideally be drawn from the general population, since other sources (e.g. hospital controls) are likely to have different patterns of screening, and not be representative of the source population (Weiss, 1983). Cancer registries may therefore be a source of cases, but they will not be appropriate for selection of controls. Controls are generally matched for age, but further matching may result in 'over-matching'. Controls should be free of invasive disease at the time of diagnosis of the case in order to ensure that the case and control have the same opportunity of 'exposure' to screening (Sasco *et al.*, 1986). The screening history of both cases and controls should be measured over the same time period.

In Type B screening, cases may be selected as patients with invasive disease. The screen at which cancer is detected should not be included in the screening history. A potential difficulty is in differentiating between symptomatic and 'screening' diagnostic tests, if the former are included in the screening history, a bias will result against an effect of screening, since cases will appear to have a greater number of screening tests. One solution is to exclude all tests within a given period (e.g. 6 months) prior to the diagnosis of the case.

Controls should again be drawn from the general population. For cases of invasive disease detected by screening, controls should be selected from subjects screened at the same time as the case, since they are likely to have a different screening history (e.g. in a 5-yearly cervical screening programme, their time since previous screen is likely to be approximately 5 years (IARC Working Group on the Evaluation of Cervical Cancer Screening Programmes, 1986).

Evaluation of population screening programmes

Once a decision to introduce population screening has been made, monitoring trends in both incidence, and mortality from the disease in question becomes an important means of evaluating the effect of screening, since usually no uninvited control group is available.

A number of theoretical issues in the evaluation of cancer screening require different interpretation according to whether screening is aimed at detecting invasive disease at an earlier stage, or a precancerous state (e.g. cervical smears) as discussed above. It is clear that the impact of these two types of screening on trends in cancer incidence will be different. The former will cause an initial increase in disease incidence. In the absence of overdiagnosis, rates should fall to prescreening levels, except in the youngest age-group being invited for the first time. Mortality rates should eventually decrease. The latter will theoretically reduce both incidence

34

and mortality rates, although screening is likely also to detect some early invasive as well as in situ disease.

Screening for other sites may work in both ways, for example, screening for colorectal cancer may produce an initial effect by the detection of invasive disease at an earlier stage, and a longer-term effect by the detection and removal of adenomas, a proportion of which would subsequently have developed into cancer.

An initial increase in the age-group invited for screening was observed in England and Wales where the incidence of invasive breast cancer increased by 40% between 1979 and 1992, with the steepest increases in the screened age-group (Quinn and Allen, 1995). In the United States increases in breast cancer incidence were found in line with increased use of mammography (Wun et al., 1995). The size and timing of the observed increase will depend on the speed with which the population is covered by screening, the prevalence of the disease (reflecting the underlying incidence and the duration of the preclinical phase of the disease), and the sensitivity of screening.

Disease-specific mortality should eventually fall, and again the monitoring of mortality trends is essential to the evaluation of population screening. However, as observed in a number of randomised trials, it will be several years after the start of screening before an effect on mortality appears. For example, in the Swedish Two County study, a difference in breast cancer mortality between the study and control groups emerged 4-5 years after the start of the trial. With population screening, this is likely to be compounded by a staggered introduction of screening, so that the full impact will not be expected until several years after full coverage has been achieved. In addition, mortality rates in the general population will for some time include deaths in cases diagnosed before the introduction of screening, which have been excluded in the estimate of benefit from randomised controlled trials. This will tend to dilute the benefit of screening for a number of years.

As for disease incidence, a further problem is the existence of underlying trends in mortality unrelated to screening. Such trends may be related to changes in risk factors. For example, Hermon and Beral (1996) have observed a levelling-off or decline in breast cancer mortality rates in many western countries, which appears in part to be due to both cohort effect, for example a reduction in childlessness, and a reduction in age at first birth among women born after about 1920. There may also be period effects due, for example, to improvements in the treatment. The observed fall in breast cancer mortality in the UK in the period 1987-1994 is believed to be due to the increased use of tamoxifen (Quinn and Allen 1995).

All the above factors make the estimation of the effect of population screening on mortality difficult. Record linkage with screening history is an option, but

comparison of deaths in cases with different screening histories will be subject to a number of the biases discussed above.

Another use of cancer registry data is to study trends in the incidence of late stage disease. For example, the introduction of screening for breast cancer would be expected to produce a fall in the rate of late stage disease preceding any fall in mortality. A reduction in cumulative advanced stage disease was shown, for example, in the Swedish Two County study (Tabár *et al.*, 1989). However, such analyses require complete and accurate data on stage, including data for the period before the introduction of screening.

Interval cancers and sensitivity

Interval cancers are generally defined as those cancers occurring following a negative screen, in the interval before the next routine screen is due. They are potentially a useful means of evaluating the performance of a screening programme and the appropriateness of the screening interval being used.

Cancer registries are of great importance in the monitoring of interval cancers, since accurate estimates of sensitivity require complete ascertainment of interval cancers. Reliance on cases which become known to the screening programme from other means is likely to be incomplete since, for example, they may not include cases diagnosed at a different hospital to that where a screening programme is based. There may also be a lack of interest for screening programmes to search out interval cases. However, cancer registries alone cannot use interval cancer data to produce estimates of sensitivity, since not only is good record linkage with screening history required to identify interval cancers, but data are needed to provide the denominator.

Estimating the sensitivity of screening

The sensitivity of screening is the ability of a screening test to detect true cases of the disease. A number of investigators now differentiate between test sensitivity and programme sensitivity. The former is the probability that a tumour in the preclinical detectable phase will be diagnosed after a positive screening test, whilst the latter has been defined as the probability that a case in the PCDP at any time during an ongoing screening programme (and ending at the last screen) will be diagnosed following a positive test (Church *et al.*, 1997).

There are a number of ways in which the sensitivity can be estimated (Day, 1985). All require knowledge of the rate of interval cancers following negative screens, since these will include those cases missed by screening. However, interval cancers

36

will also include some cancers newly arising since the last negative screen, and the proportion of the latter will increase with increasing time since a negative test.

The 'traditional' method, in which sensitivity is estimated as $\dfrac{s}{(s + i)}$ where s is the number of cancers detected by screening and i the number of cancers appearing in a given interval after the screen, is likely to overestimate sensitivity since a number of the cancers detected by screening may not otherwise have appeared in this interval (Day, 1985).

The 'proportional incidence method' for estimating sensitivity calculates the rate of interval cancers as a proportion (p%) of the expected cancer incidence in the absence of screening. The sensitivity is then estimated as (100 − p%). In randomised controlled trials of screening, the incidence in the control group can be used to estimate the expected incidence (taking account of selection bias as necessary). For population screening programmes, it is necessary to estimate the expected incidence from the historical incidence rates in the period prior to the introduction of screening (or for similar geographical areas without a screening programme).

Mathematical models are now often used to make joint estimates of sensitivity and the length of the PCDP.

The potential for overdiagnosis

There are three possible sources of overdiagnosis − taken here to mean the diagnosis (and registration) of a cancer which would not have emerged clinically in the absence of screening. One is the diagnosis of slow-growing disease which would not have progressed to clinical significance in the person's lifetime. The extent of such overdiagnosis will depend on the natural history of the cancer. For example, it is apparent that screen detected prostate cancers include some proportion of slow-growing tumours, although it is not clear whether there is in fact a latent form of the disease or a wide range of disease growth. Evidence for overdiagnosis comes from autopsy studies, which have shown for example that a percentage of men at autopsy have undiagnosed prostate cancer. In addition, high prevalence rates at first screen indicate a long mean sojourn time.

A second source is the inclusion of non-progressive disease (clinically benign) amongst the cancer cases. It is apparent from observer variability studies among pathologists that agreement is poorest in borderline categories. Diagnosis of a proportion of benign cases as malignant will affect incidence rates (but not mortality).

The diagnosis of in situ disease is another area of potential overdiagnosis. In situ cancer is mainly diagnosed by screening. For example, in the UK NHS Breast Screening Programme 17% of tumours diagnosed by screening are in situ (Moss *et al.*, 1995). Again the natural history DCIS is uncertain, one study has estimated that 50% may progress. Thus screening will increase the incidence of in situ disease, and potentially decrease the incidence of invasive breast cancer.

The effect of overdiagnosis will show both in incidence rate and in survival analysis, resulting in artificially high survival rates, particularly for early stage disease.

Surrogate outcome measures

In recent years there has been considerable interest in the use of surrogate outcome measures, principally in the evaluation of randomised controlled trials. Prognostic variables such as tumour size, nodal status and grade in cancers in the two arms of the trial are used to predict the outcome in terms of mortality. The advantage of using surrogate outcome measures is that the results of the trial become available much more quickly, than if it is necessary to wait for mortality results. In addition, the sample size required is considerably lower, although this assumes that there is no variation in the mortality prediction. A number of potential problems have been identified with surrogate outcome measures. Firstly, depending on the number of categories used for the prognostic indicators, there will be variation in prognosis within each category (e.g., within 'node positive' tumours the number of positive nodes will vary). There is the possibility of measurement error, together with the possibility that data may be less complete for non-screen-detected cases. Also, the possibility of overdiagnosis, discussed above, is a potential problem if non-progressive cases detected by screening are not correctly identified. Lastly, treatment differences, either changes over time or a correlation with mode of detection, may affect the results.

In order for surrogate outcome measures to be useful in monitoring population screening, accurate information on the prognostic variables for all cancers would be required. Observer variation between pathologists is such that, in most trials so far, all pathology is reviewed in order to produce standardised results. This will clearly not be feasible on a population basis.

The UK trial of one year vs. three year screening interval for breast cancer is using surrogate outcome measures, based on the Nottingham Prognostic Index (Haybittle *et al.*, 1982), although mortality follow-up is also included. Cancer diagnosed at the prevalent screen in each arm of this trial are excluded in order to remove 'length bias'.

Analyses of survival by stage and mode of detection in a number of screening trials suggest that the pathological information collected has not been sufficient to allow accurate prediction of mortality (UK Trial of Early Detection of Breast Cancer Group, 1993).

References

Alexander FE, Roberts MM, Lutz W & Hepburn W (1989). Randomisation by cluster and the problem of social bias. *J Epidemiol Community Health* **43**:29-36

Church TR, Ederer F & Mandel JS (1997). Fecal occult blood screening in the Minnesota study: sensitivity of the screening test. *J Natl Cancer Inst* **89**:1440-1448

Cronin KA, Weed DL, Connor RJ & Prorok PC (1998). Case-control studies of cancer screening: theory and practice. *J Natl Cancer Inst* **90**:498-504

Day NE (1985). Estimating the sensitivity of a screening test. *J Epidemiol Community Health* **39**:364-366

Hardcastle JD (1997). Screening for colorectal cancer (letter). *Lancet* **349**:358

Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, James PD & Mangham CM (1996). Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* **348**:1472-1477

Haybittle JL, Blamey RW, Elston CW, Johnson J. Doyle PJ, Campbell FC, Nicholson RI & Griffiths K (1982). A prognostic index in primary breast cancer. *Br J Cancer* **45**:361-366

Hermon C & Beral V (1996). Breast cancer mortality rates are levelling off or beginning to decline in many western countries: analysis of time trends, age-cohort and age-period models of breast cancer mortality in 20 countries. *Br J Cancer* **73**:955-960

IARC Working Group on Cervical Cancer Screening Programmes (1986). Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policies. *Br Med J* **293**:659-664

Kronborg O, Fenger C, Olsen J, Jorgensen OD & Sondergaard O (1996). Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet* **348**:1467-1471

Louhivouri K (1991). Effect of a mass screening programme on the risk of cervical cancer. *Cancer Det Prev* **15**:471-474

Lynge E, Poll P (1986). Risk of cervical cancer following negative smears in Maribo County, Denmark, 1966-1982. In: Hakama M, Miller AB, Day NE (eds) *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No.76), Lyon, International Agency for Research on Cancer, pp. 69-86

Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM & Ederer F (1993). For the Minnesota Colon Cancer Control Study. Reducing mortality from colorectal cancer by screening for fecal occult blood. *N Engl J Med* **328**:1365-1371

Miller AB, Lindsay J & Hill GB (1976). Mortality from cancer of the uterus in Canada and its relationship to screening for cancer of the cervix. *Int J Cancer* **17**:602-612

Moss SM (1991). Case-control studies of screening. *Int J Epidemiol* **20**:1-6

Moss SM, Michel M, Patnick J, Johns L, Blanks R & Chamberlain J (1995). Results from the NHS breast screening programme 1990-1993. *J Med Screening* **2**:186-190

Quinn M & Allen E, on behalf of the United Kingdom Association of Cancer Registries (1995). Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. *Br Med J* **311**:1391-1395

Roberts MM, Alexander FE, Anderson TJ, Chetty U, Donnan PT, Forrest P, Hepburn W, Huggins A, Kirkpatrick AE, Lamb J, et al (1990). Edinburgh trial of screening for breast cancer: mortality at seven years. *Lancet* **335**:241-246

Sasco AJ, Day NE & Walter SD (1986). Case-control studies for the evaluation of screening. *J Chronic Dis* **39**:399-405

Shapiro S, Venet W, Strax P, Venet L & Roeser R (1982). Ten-to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst* **69**:349-355

Tabár L, Fagerberg G, Duffy SW & Day NE (1989). The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health* **43**:107-114

UK Trial of Early Detection of Breast Cancer Group (1988). First results on mortality reduction in the UK Trial of Early Detection of Breast Cancer. *Lancet* II:411-416

UK Trial of Early Detection of Breast Cancer Group (1993). Breast cancer mortality after 10 years in the UK trial of early detection of breast cancer. *The Breast* 2:13-20

Wald NJ, Chamberlain J & Hackshaw A (1994). European Society of Mastology Consensus Conference on Breast Cancer Screening, Paris, 4-5 February 1993: Report of the Evaluation Committee. *Clin Oncol* 4:261-268

Weiss NS (1983). Control definition in case-control studies of the efficacy of screening and diagnostic testing. *Am J Epidemiol* 118:457-460

Wun LM, Feuer EJ & Miller BA (1995). Are increases in mammographic screening still a valid explanation for trends in breast cancer incidence in the United States? *Cancer Causes Control* 6:135-144

Chapter 4. Use of cancer registry data: prerequisites, limitations and solutions

J.A.A.M. van Dijck and L.J. Schouten

Introduction

Periodic screening of asymptomatic persons to detect cancer at an early or even pre-malignant stage has become an important tool in cancer control. For several malignant diseases, screening programmes have been introduced. Cancer registries play a substantial role in the evaluation of screening programmes.

First of all, cancer registries, especially the population-based ones, provide background information on the site-, sex- and age-specific incidence rates of cancer in a defined geographical area. These regional incidence data may form the basis for setting priorities for cancer control activities. Further, the existence of a cancer registry in a geographical area where a screening programme is ongoing makes it relatively easy to monitor incidence rates of that specific cancer, before and during the screening programme. However, monitoring the incidence rates will give no insight into the effectiveness of the screening. For that purpose, it will be necessary to identify all cases of cancer in the target population and separate them into screen-detected and interval cancers and cancers among the non-responders (Schouten et al., 1998). Screen-detected cancers, diagnosed as a result of a positive screening test, may be relatively easy to discover even in the absence of a cancer registry. However, interval cancers, which are diagnosed clinically after a negative screening test and before the next scheduled screening examination, may be very difficult to identify without the existence of a cancer registry. Only population-based cancer registries have complete coverage of the screened population.

The aim of most cancer screening programmes is to detect the disease as early as possible. Up to now, the cervix uteri is the only site for which prevention of invasive cancer has been demonstrated to be possible. For other sites, the aim is to prevent deaths from cancer. Therefore, monitoring of the incidence according to disease stage, as an early indicator of screening effectiveness, may be very useful. As an example, Figure 1 shows the incidence of breast cancer according to tumour size in the Netherlands, in the years 1989-1995. The introduction of the screening programme started in 1989 and was completed in 1998. The incidence of small tumours rose between 1989 and 1993, and seems to be stable thereafter. However, the incidence of larger tumours started decreasing only after 1994.

43

The detection of cancer at an early stage may influence the choice of treatment. One of the secondary benefits of breast cancer screening is that a greater proportion of the patients can undergo breast-conserving therapy. For those registries that do collect data on therapy, monitoring trends in treatment may be useful.
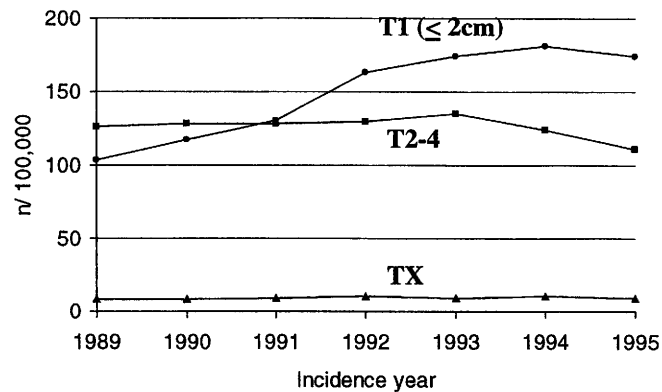

Figure 1. The incidence of breast cancer according to tumour size in women aged 50-69 years (Source: Netherlands Cancer Registry, 1998)

Cancer registries must fulfil several prerequisites in order to be valuable in the evaluation of screening programmes. In this chapter, the following issues will be discussed: quality of cancer registry data (completeness and validity of the data), necessary items, record linkage and timeliness. Some limitations of the cancer registry data and possible solutions will also be discussed.

Quality of cancer registry data

One of the aims of population-based cancer registries is estimating site-specific incidence rates by sex, age, and stage, etc. For this purpose, accurate enumeration of all incident cases of cancer in the target population is a basic prerequisite. Further requisites are the correct and reproducible classification and coding of the cancer cases. In order to be able to calculate (relative) survival, the complete follow-up until emigration out of the target population, or death is necessary (Parkin *et al.,* 1994). In addition, accurate and frequent estimates of the population at risk must be available, broken down by as many variables of interest to the registry as possible (the absolute minimum is by sex and age group). The general aspects of quality control in cancer registration have been published before and will not be repeated here (Parkin *et al.,* 1994). However, some aspects of completeness and validity deserve particular

44

attention when cancer registry data are used for the evaluation of a screening programme.

Completeness

Completeness of case ascertainment in relation to screening status may be selective. If so, the degree of completeness is usually higher for screen-detected cases than for interval cancer cases, which will lead to an over-estimation of the sensitivity of the screening. This may occur when all screened persons with a positive test result or the diagnosis of cancer are notified to the cancer registry by the screening organisation. In this situation, it is very unlikely that any case of screen-detected cancer will be missed. Interval cancer cases, for whom no such additional source of notification exists, may have a lesser degree of completeness. Further, when screened-detected cancers are more likely than interval cancers to be present in the data source which is most important for the cancer registry, the completeness may be better, e.g., when screen-detected cases are more likely to have histological confirmation of the cancer, and pathology reports are an important source. This could occur, e.g., in screening for lung cancer. Finally, one should remember that because of the screening some cancer cases may be detected that would never have become clinically manifest (see also section on validity and comparability below).

Completeness of detail is another concern. For screen-detected patients the information recorded in the medical records may be better, or more extensive than for clinically detected cancer patients. This form of selective completeness was present in the Maastricht Cancer Registry data of the years 1991-1996 (Schouten *et al.*, 1998). For breast cancer in women aged 50-69 years, disease stage was unknown in 2.0 % of the screen-detected patients. Disease stage appeared to be unknown in 4.4 % of the interval cancer patients and in 4.6 % of the unscreened patients. So, although a difference in completeness was present, it was small. However, in other situations large differences in the proportions of missing information may be present between screen-detected and other cancer patients.

Validity and comparability

Validity refers to the extent to which the information recorded on the different variables is true, or accurate. A screening programme may have a great impact on the information to be recorded. Many borderline lesions will be detected, such as ductal and lobular carcinoma in situ of the breast or micro-invasive prostate cancer. The clinical significance of these lesions is often not clear. It is very well possible that many of these cancers would never have been detected in the absence of screening. It can be discussed if these lesions should be called cancer. Furthermore, there may be diagnostic problems. The pathologic classification of cancer and its pre-invasive

analogues may change because of the screening programme. The boundary between borderline benign and borderline malignant may change. This will hamper the evaluation of time trends. Coding these borderline lesions in an identifiable way by using specific codes for stage, morphology or histologic grade makes it possible to regard them separately when analysing the data.

Another problem with the validity of the data is that outcome variables may change because of screening. One of the most important prognostic factors for breast cancer is whether the patient has axillary lymph node metastases or not. For the evaluation of breast cancer screening, the incidence of cancers with positive regional lymph nodes is considered one of the (early) outcomes. It is used as an indicator for a future mortality reduction. At the start of the national breast cancer screening programme in The Netherlands, the pathologists have been instructed to take larger samples of lymph nodes and to look at the resected lymph nodes more intensively. Therefore, lymph node metastases can be found in a larger proportion of the cases. In the Maastricht Cancer Registry in 1993-1996, the incidence rate of node positive breast cancer was roughly 10% higher when lymph nodes with micro-invasion (N1a) were included. The clinical implications of N1a are not clear, however. Patients with micro-invasion in the axillary lymph nodes may have a prognosis similar to that of patients with negative lymph nodes. Already 15 years ago attention was paid to this phenomenon called stage migration (Feinstein *et al.*, 1985). With the introduction of new diagnostic tests, patients with any stage may have a better prognosis than patients with the same disease stage who had been staged with old methods. Had patients been staged with the new methods, many of them would have been classified into a worse disease stage. This problem could also arise because the introduction of a screening programme has changed the use and interpretation of old diagnostic tests. If screening programmes are evaluated by analysing trends in disease stage, one should find out whether new diagnostic tests have been introduced since the start of the screening.

Necessary items

A discussion of basic and optional items of patient information to be collected by cancer registries can be found in Chapter 6 of the Monograph "Cancer Registration: Principles and methods" (Jensen *et al.*, 1991).

For purposes of evaluating screening programmes, the items recorded by the cancer registry and the screening programme should be coded in a similar way. Identification items should be recorded in such a way that record linkage is possible. Date of birth, sex, name (may be encrypted) and address belong to the basic information to be recorded. Names can easily contain errors, because names, which are pronounced in the same way, may be spelled differently. Further, people may change their name, e.g.

46

because of marriage or divorce. This will depend on cultural and legal background. If a unique personal number is available that is generally used, it should be recorded. Especially when this number is permanent and does not change when people get married or divorced it is very useful. Tumour information, such as topography, morphology, behaviour and grade are preferably coded according to the ICD-O classification. However, the ICD-O coding of topography may not be sufficient for screening of skin melanoma, because it is not very detailed. For the definition of date of incidence and the most valid basis of diagnosis, the IACR guidelines are recommended.

Optional items

Besides the necessary items mentioned in the previous section, recording of the items differentiation grade, stage, method of detection, and treatment are highly recommended, especially if the design of the screening evaluation includes them. If any extra items are collected, they should be available for all patients.

Differentiation grade could be useful as a prognostic factor. For example, the optimal management of ductal carcinoma in situ of the breast has not yet been established. Many clinical trials are underway. Differentiation grade is used as a marker to identify subgroups requiring different treatment (Holland *et al.*, 1994).

Stage

Stage is regarded as one of the early indicators of the effectiveness of cancer screening for most sites. Therefore, data on stage should be collected by all cancer registries that are used for the evaluation of screening programmes. In Table 1, several of the available classifications are listed. A classification that is used broadly and that is available for all sites is the extent of disease (EoD). For cancer registries, it is relatively easy to collect. A major drawback of the EoD is, however, that it contains only one axis. If metastases are present, no information is available with respect to the regional lymph nodes and tumour size. If positive lymph nodes have been diagnosed, the EoD contains no information on the local tumour extension. Therefore, the information coded in this classification is limited. In the evaluation of breast cancer screening, information on tumour size is important. Further, the EoD is not frequently used by clinicians.

The TNM classification is available for many sites. It has been widely accepted by clinicians. A major advantage of the TNM classification is that local extension, regional lymph node involvement and distant metastases are coded separately. A disadvantage is that the accuracy and validity are relatively low. For cancer registries that may not be able to obtain all the information necessary for the TNM

classification, the European Network of Cancer Registries has provided guidelines on a condensed TNM classification.

Table 1. Classification for the coding of disease stage

Site	Available classifications		
Breast	EoD	TNM	Tumour size
Cervix	EoD	TNM	FIGO
Melanoma	EoD	TNM	Clark /Breslow
Prostate	EoD	TNM	
Colon	EoD	TNM	Dukes

As an example, Figure 2 gives the incidence of breast cancer by TNM stage for women aged 50-69 years in the Netherlands, were the screening programme started in 1989. In the years 1990 to 1995, 11%, 25%, 48%, 69%, 77% and 88% of the target population had been invited. The attendance was roughly 75%. In the period from 1989-1995, the incidence of breast cancer in situ increased threefold, that of stage I doubled, that of stage II increased by 16% up to 1993 and decreased almost to the level of 1989. Stage III or higher decreased by 22% in the observation period.
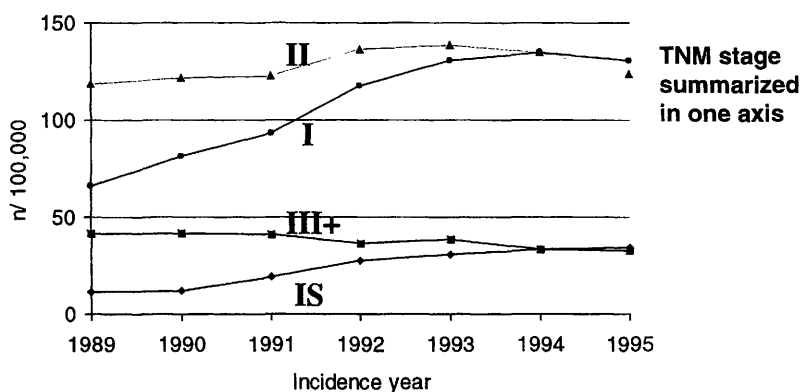

Figure 2. **Trends in stage distribution for breast cancer, ages 50-69 years** (Source: Netherlands Cancer Registry, 1998)

48

Figure 3 shows the same data summarised in a different way. The incidence of distant metastases (M1) is shown separately, and if M0, the incidence of node negative (N0) and node positive (N+) tumours is shown. The incidence of M1 tumours decreased by 24% (from 16.9 to 12.9 per 100 000). The incidence of node positive tumours was stable, whereas the incidence of node negative tumours increased by almost 60% up to 1993, and decreased thereafter. Figure 3 gives a somewhat less favourable impression of the effects of screening than Figure 2. One of the reasons for this may be the increased frequency of positive axillary lymph nodes with micro-invasion, N1a, as mentioned before.
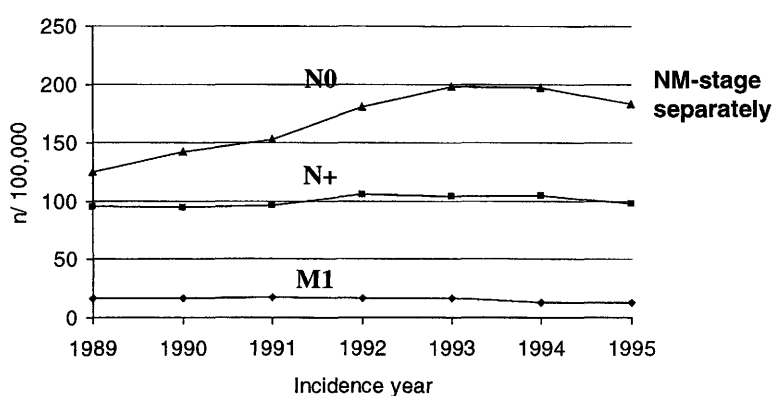
Figure 3. Trends in stage distribution for breast cancer, ages 50-69 years (Source: Netherlands Cancer Registry, 1998)

The TNM classification provides a categorisation of tumour size in rather broad classes, when the registration of the exact tumour size may be more useful in the evaluation of cancer screening. This item would be rather easy to collect for most sites, at least the pathological size. For melanoma screening, the thickness of the tumour should be recorded (Breslow) in millimeters. For some sites it may not be useful and/or possible to record the tumour size (e.g. prostate cancer), and the reproducibility and precision of the recorded tumour size can often be questioned. The clinical tumour size depends on the diagnostic tests used to estimate the size, and on the reproducibility of these tests, and even the pathological tumour size cannot be measured very precisely in many circumstances. Figure 4 illustrates this for the pathologically measured tumour size of invasive breast cancers (Peer et al., 1996).

The tumour size often seems to be rounded to the nearest 5 mm, and for larger tumours even to the nearest 10 mm. An explanation for this is that with the varying shapes of the tumours, it may be difficult to decide which dimension to measure.

Recording the exact tumour size will give an illusion of precision. For tumours larger than 10 mm rounding to the nearest 5 mm may be precise enough.

Number

120 mm: 2*
150 mm: 2*
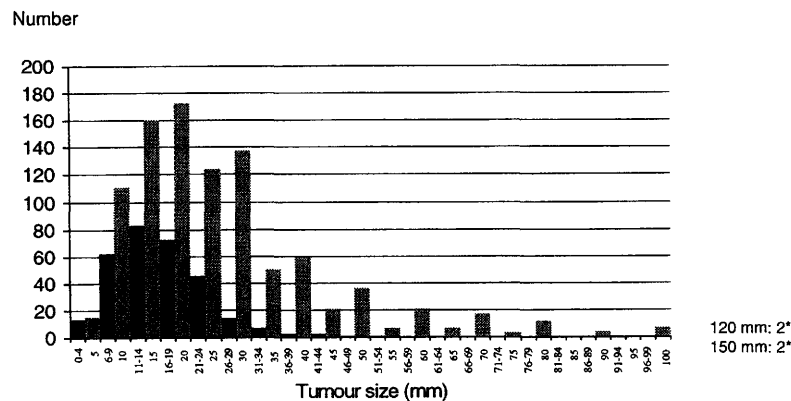
Tumour size (mm)

Figure 4. Precision of the pathological tumour size for breast cancer (Nijmegen Breast Cancer Screening Study)

Detection

The availability of screening tests will decrease the proportion of patients with symptoms, or it may increase the incidence of a cancer without decreasing the number of patients with symptoms. Registering the 'method of first detection can be very helpful in the evaluation of time series, especially when the diagnostic possibilities for a specific tumour or primary site have changed over time. In many countries, the availability of PSA and mammography have given rise to opportunistic screening. To evaluate the consequences of this kind of screening, the method of first detection would be essential. Since the wide acceptance as PSA as a marker for the presence of prostate cancer, its use has increased significantly. Even for the evaluation of an organised screening programme, the item will be helpful in separating screen-detected and interval cancers and cancers among non-responders. This would be very important when no linkage between the screening records and the cancer registry records can be performed. Also, patients can be identified who have been diagnosed outside an official screening programme, for example because they have a high cancer risk based on a positive family history. In some situations the coding of this item may be difficult due to the lack of detailed history or, e.g., when a patient came to the screening examination while having symptoms.

The experience from a field trial performed at the cancer registries in Maastricht and Nijmegen on breast, cervix and prostate cancer was that the method of detection could

be collected very easily from the medical files. However, a hierarchical order of the codes is necessary:

1. Incidental finding at autopsy
2. Screening examination
3. Incidental finding (on examination, at surgery)
4. Clinical presentation (with symptoms)
8. Other
9. Unknown

In this hierarchical order the higher code is preferred over the lower code, unless there is proof for the lower code. So, a patient with symptoms who came to the screening would be coded as 4: clinical presentation.

Although the information could be collected very easily in the Netherlands, it may be difficult in other countries. The information necessary to code the item may not be in the medical files. Further, many cancer registries do not collect their data from the medical files, but obtain their data electronically or from forms filled in by physicians in the hospitals. For these cancer registries it may be difficult to generally register this item. However, for special purposes, i.e., for a limited time to evaluate a screening programme, additional data can be registered from the clinicians.

Treatment

The existence of a screening programme may have a great impact on treatment. Because of the screening, tumours should be detected at a smaller size. For example, a larger proportion of the breast cancer patients will undergo breast-conserving therapy. This can also be seen as a favourable outcome of screening. However, breast cancer therapy is only useful if it is coded with enough detail. So it would be necessary not only to code surgery, but also the type of surgery. Also, the frequency of adjuvant therapies such as radiotherapy, chemotherapy and hormonal therapy may change. Data on treatment may be valuable when expenses of the screening programme are to be evaluated.

Record linkage

Record linkage is necessary to match the records of screening organisations with those of the cancer registry, to evaluate the effects of screening on an individual basis. In some countries, e.g. the Nordic countries, a unique personal identifying number is used to link personal data between different registers.

In most countries, there are no universal personal identifying numbers, so that their usefulness is limited. Then, several other identifiers, such as date of birth, sex and family name, have to be used in the record linkage. However, these identifiers are not unique and they may contain errors. A name may be misspelled, a date of birth may be written incorrectly, even sex may be coded wrong. Sometimes the patients may refuse to have their name registered. In some countries the name and address have to be encrypted, further hampering the linkage of records. Small differences in the spelling of names may remain unnoticed because of the encryption. Further, screening records and cancer registry records will be encrypted differently. Because of these problems, record linkage will almost never be perfect.

A probabilistic method was developed based on the calculation of the odds in favour of a correct match associated with a specific combination of identifiers (Newcombe, 1988). The odds of a very common name will be much lower than the odds of a rare name. The calculation of the odds can be refined to accommodate weights associated with identifier values and coding errors. A drawback of this method is that it requires detailed prior knowledge about the frequency of the values of identifiers that will be used.

In the Netherlands, a protocol was developed for use in the Netherlands Cancer Registry (Van den Brandt *et al.*, 1990). The procedure is a two-stage process. First, a computerised linkage is performed with the following identifiers: date of birth, sex, family name (first 4 letters) and a part of the postal code. Next, all possible matches are inspected visually with additional information. The particular case for which the protocol was developed was linkage between the Netherlands Cohort Study on Diet and Cancer and the Netherlands Cancer Registry (Van den Brandt *et al.*, 1990). The linkage was performed with the use of several identifiers. When only family name was used, the number of possible matches was almost 50 000. The number of missed matches was two, which gave a sensitivity of 98.9%. However, only 0.4 % of all matches was true positive (predictive value of positive match [PV+] = 0.4%). Date of birth also gave a high number of matches, with a sensitivity of 98.9% and PV+ = 3.5%. When date of birth, family name and sex were combined, sensitivity was slightly lower, 97.9%, and PV+ rose to 97.9% (Table 2). Including the first initial decreased sensitivity to 91.5% (16 matches missed).

Is it possible to increase the sensitivity to 100%? For linkage of family name, Soundex could be used. Soundex was developed to index the United States' censuses. It codes surnames of the same and similar sounds but of variant spellings. This may avoid the problem that differently written names do not match. However, this software is developed for the English language, and it is unknown how useful it may be in other languages. Moreover, it cannot be used with encrypted data.

52

Table 2. Record linkage between the Netherlands Cohort Study (n=8081) and the Maastricht Cancer Registry (n=8917) (reproduced by permission of the author and Oxford University Press from Van den Brandt et al., Int J Epidemiol 19:553-558, 1990)

Identifiers	Matches	True positives	False negatives	Sensitivity (%)	PV+ (%)
Date of birth (DOB)	5276	186	2	98.9	3.5
F4[1]	102 070	186	2	98.9	0.2
Family name (F)	49 808	183	5	97.3	0.4
DOB+F+Sex (S)	188	184	4	97.9	97.9
DOB+F+S+Initial	172	172	16	91.5	100.0
DOB+F4+S+Postal code (P4)	183	183	5	97.3	100.0

[1] F4= first four characters of F

Another solution to increase the sensitivity of linkage is to take into account known patterns of errors that may occur. For example, in the date of birth, month and day may be transposed, such as 01-11 (1 November) versus 11-01 (11 January). One has to keep in mind that all arrangements to increase sensitivity will decrease specificity.

The example in Table 3 shows results from a computerised linkage of the Maastricht Cancer Registry and files with the regional breast cancer screening programme (Schouten et al., 1998). The linkage was restricted to women. The identifiers - date of birth, the first 4 letters of the family name and the 4 numbers of the postal code - were evaluated two by two, to identify all possible links. 3031 possible matches were verified manually, with the use of additional items such as name of the husband and initials. The result was 360 true positive links. The sensitivity of the combination of date of birth, 4 letters of the family name and postal code was 83%, PV+ was 100%. The combination of DOB and F4 gave another 61 matches, of which 51 were correct. DOB and P4 gave 39 matches, 8 were correct. The combination of F4 and P4 gave only two additional matches at the cost of 2632 records to be verified.

Table 3. Record linkage between the Maastricht Cancer Registry and the Breast Cancer Screening Programme (Source: Maastricht Cancer Registry)

Identifiers	Matches	True positives	False negatives	Sensitivity (%)	PV+ (%)
DOB+F4+P4	299	299	61	83 .1	100.0
DOB+F4	360	350	10	97.2	97.2
DOB+P4	338	307	53	85.3	90.8
F4+P4	2931	301	59	83.6	10.3

Abbreviations: see Table 2

Timeliness

Depending on the method of data collection, it may take several years before all records of one incidence year have been completed. Thus, for the evaluation of the screening programme routine cancer registry data will be available with a considerable delay.

Figure 5 shows for the Maastricht Cancer Registry the time required to complete the records of the incident cases in 1996. Notifications were received up to February 1997 from the national pathology registry, the most important source. The small increase in the number of notified records after that date was due to notifications from the national hospital discharge registry. Final data from this source is received after a fair delay. All Dutch cancer registries collect their data from the medical records of the hospitals and outpatient clinics. The delay is considerable; in 1996 in Maastricht, it was at least 6 months. At the end of 1998, the data of 1996 had been completed. In December 1997, when the data set for the evaluation of the breast cancer screening programme was produced, 85 % of the records had been completed. So it may take up to three years before all data of a specific year are completely registered. Then it will take some time before results of the record linkage are reported and the indicators of the effectiveness of the screening programmes are available. Only after this delay is it possible to assess whether changes in the screening process are necessary.

Number of
all cancers
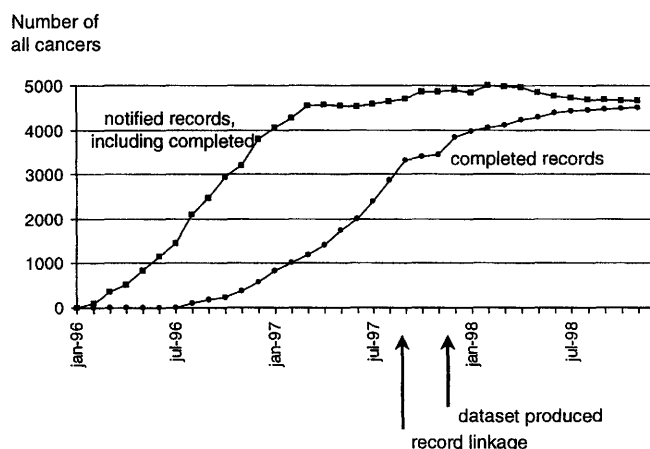
Figure 5. Maastricht Cancer Registry for the incidence year 1996

How to overcome this long delay? The first option is to use the notified but not yet completed records for the evaluation (Figure 5). In the Netherlands, it will depend on the regional situation and on tumour site whether data from the pathology departments are reliable enough for this purpose. For breast cancer screening, where the pathology report gives reliable localisation and morphology, this may be a good option to speed up the evaluation. For other sites, such as the prostate, the clinical examinations may be necessary to make a reliable diagnosis. In this situation, fast-tracking registration may be a solution. This means that priority is given to the registration of the tumour site for which the evaluation will be performed. For some sites, such as lung cancer, the number of missed cancers may be relatively large when only histologically and cytologically notified cancers are used. When a cancer registry obtains notifications electronically, the delay may be much less than for registries that collect data from the hospital records. When cancer screening programmes are evaluated with rapid reporting systems, it should be kept in mind that the indicators for screening effectiveness may be selectively biased. As discussed before, the completeness may be selective, and this problem will be much larger when rapid reporting systems are used.

Mortality data

The first outcome of screening programmes should be evaluated by using cause specific mortality data. In cancer registries which use death certificates as one source for identifying cases, the causes of death of deceased patients are known. This information can be linked with the screening files. Many cancer registries, however, cannot use this source of notification, because in their country, the death registration is anonymised, and it may thus be very difficult to complete the follow up of registered patients. Then, mortality data should be analysed on a population level.

Conclusion

Cancer registries play an important role in the evaluation of screening programmes. In many countries, the existence of screening programmes is used as one of the justifications for the existence of cancer registries. Cancer registries should be prepared to adapt procedures to facilitate these evaluations.

References

Feinstein AR, Sosin DM & Wells CK (1985). The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med* **312**:1604-1608

Holland R, Peterse JL, Millis RR, Eusebi V, Faverly D, Van de Vijver MJ & Zafrani B (1994). Ductal carcinoma in situ: a proposal for a new classification. *Sem Diagnostic Pathol* **11**:167-180

Jensen OM, Parkin DM, MacLennan R, Muir CS & Skeet RG (eds) (1991). *Cancer registration: principles and methods* (IARC Scientific Publications No. 95), Lyon, International Agency for Research on Cancer

Krag D, Weaver D, Ashikaga T, Moffat F, Klimberg VS, Shriver C, Feldman S, Kusminsky R, Gadd M, Kuhn J, Harlow S & Beitsch P (1998). The sentinel node in breast cancer - a multicenter validation study. *N Engl J Med* **339**:941-946

Newcombe HB (1988). *Handbook of record linkage: methods for health and statistical studies, administration and business*. Oxford, Oxford University Press

Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH, Whelan SL (1994). *Comparability and Quality Control in Cancer Registration* (IARC Technical Report No. 19), Lyon, International Agency for Research on Cancer

Peer PGM, Verbeek ALM, Mravunac M, Hendriks J, Holland R (1996). Prognosis of younger and older patients with early breast cancer. *Br J Cancer* **73**:382-385

Schouten LJ, De Rijke JM, Schlangen JT, Verbeek ALM (1998). Evaluation of the effect of breast cancer screening by record linkage with the cancer registry, the Netherlands. *J Med Screening* **5**:37-41

Van den Brandt PA, Schouten LJ, Goldbohm A, Dorant E, Hunen PMH (1990). Development of a record linkage protocol for use in the Dutch Cancer Registry for epidemiological research. *Int J Epidemiol* **19**:553-558

Chapter 5. Evaluation and monitoring of screening for cervix cancer: time trends

J. Smith and D.M. Parkin

Introduction

Cancer of the cervix is an important cause of mortality and morbidity worldwide (371 000 new cases in 1990 and 190 000 deaths (Parkin et al., 1999)). Although often perceived as a relatively rare disease in Europe, it certainly was not in the era before screening – age standardised (world) incidence was 28.3 in Denmark in 1953-57 and 36.5 in Hamburg in 1960-62 (Doll et al., 1966); these rates are not dissimilar from those in high-risk countries today. Cervix cancer was the first malignancy for which an effective method of screening was introduced, and although never subjected to evaluation via a randomised controlled trial, the benefits in terms of reduction in invasive cancer (and subsequent mortality) have been quite clear. The data provided by cancer registries were an important component of this evaluation, and they remain so in the monitoring of existing screening programmes, for several reasons. The very success of screening has rendered cervix cancer a relatively rare disease, so that some commentaries often query the benefits in relation to the service infrastructure, and costs, involved in maintaining a screening programme. The optimal deployment of screening resources need to be reviewed, to determine whether different population groups (defined geographically or otherwise) are obtaining equivalent benefit (principles of equity).

Much has been made of the importance of organised screening (programmes in which tests are delivered to invited women, in predefined age groups, at predefined intervals) in contrast to 'opportunistic' screening (which takes place at the initiative of the subjects themselves, or of individual health care providers). In fact, the effectiveness of the screening test itself in detecting preclinical lesions is no different in the two settings (Gustafsson et al., 1995). However, it can be easily demonstrated that optimisation of a programme (in terms of cost effectiveness) requires a high level of coverage, with regular tests at defined intervals in the age groups at greatest risk (Parkin et al.,1985; Van Oortmarssen et al., 1982; Gustafsson and Adami, 1992), and that these are features of organised screening rather than of opportunistic testing (Bos et al., 1998).

Screening programmes may be monitored in terms of their level of activity (process measures), or in terms of the outcome sought. Monitoring the activity of

cervix cancer screening programmes is not the role of the cancer registry. At its simplest level, the level of activity is expressed in terms of the number of examinations performed (in relation to the size of the target population), and the percentage of which prove to be positive. This type of statistic is readily available from the laboratories carrying out the cytological examinations. But it is obvious (and easy to demonstrate) that effectiveness of a programme depends also on the distribution of tests – ensuring a high coverage, regular testing, follow-up of abnormal smears. Information on these aspects requires a more sophisticated information system, with linkage between a register of women eligible for screening, and the results of testing.

Cancer registries may have information on one aspect of the screening process, however - that is, on the numbers of certain pre-invasive lesions which are detected by screening programmes. Many cancer registries have recorded *in situ* carcinoma of cervix, along with invasive disease. This used to provide a useful insight into the intensity of screening, or even of its effectiveness, given that a proportion of *in situ* cancers would have become invasive if they had remained undetected. However, the changing terminology of pre-invasive disease has meant that the utility of the information for monitoring screening has been severely compromised. In the UK, there was a change in definition in 1984, such that the *in situ* cancers included all the CIN3 cases and in England and Wales the incidence rate increased by 50% compared to 1983 (OPCS, 1993).

Studies of age-specific incidence

The underlying principle is to describe trends in the incidence and mortality from cervical cancer and relate these to the uptake, or, if this is not known, the availability of screening. Since the objective of a cervical screening programme is to prevent the development of invasive cervical cancer, changes in the incidence of the disease should both be seen prior to and be more directly related to screening than any effect on mortality.

The problem in interpreting trends in incidence (or mortality) to screening is the uncertainty of what would have happened if there had been no screening programme, against which to compare the observations. It will generally be over simplistic to assume that the underlying risk of disease has remained unchanged over time. It is certainly most useful to examine age-specific trends. This is partly because screening effects should be seen in specific age groups (the targets of the programme), and affect all of these from the same period of time (when screening was introduced). Conversely, changes in disease incidence due to differences in the underlying risk (brought about by different exposures to aetiological agents) will normally affect successive generations of women, and appear as a "cohort effect".

60

In order to evaluate trends in age-specific incidence of disease, it is essential to appreciate the pattern of age-specific incidence in the absence of screening, in a single generation (birth cohort) of women. The general pattern of the incidence curve is a rapid rise to a peak at a comparatively young age, followed by a plateau and a variable decline thereafter. This is a somewhat unusual curve for an epithelial cancer, and we are not aware of any obvious explanation for this pattern of risk with age. This profile is readily distorted by screening, and, if cross-sectional data (from a single time period) are examined, by birth cohort specific changes in risk (Ashley,1966; Hakama and Penttinen,1981).

Since the basis of screening for cervical cancer was recognised in the 1940's, and implemented in the 1950s (USA) and 1960s (in Europe), few cancer registries are able to analyse trends in incidence and mortality before, during and after the introduction of screening. In an attempt to define the age-specific incidence patterns of cervical cancer without any influence from screening activity, Gustafsson et al. (1997a) compiled incidence data for 28 different populations, for long periods of time between 1935 and 1989, and analysed age-specific incidence rates from populations unaffected by screening. After scaling the rates (to permit direct comparison between countries with incidence rates of differing orders of magnitude), most populations fitted to 1 of 2 reference curves used for descriptive purposes (Figure 1). The first group, comprising Denmark, the former GDR, the former FRG, the Netherlands, Norway, Slovenia and Sweden, had an onset at about age 25, a rapid rise between 30 and 40 and a peak at ages 44 to 49 years. After the peak, the decline was fairly rapid and the 'half peak value' was reached at 70-75 years. The second group, comprising most American, Asian, and African registries, plus Finland and Poland, had onset at approximately the same age, but a slower rise to a peak at ages 50-65, followed by a decline similar to that in the first group. Data from the UK and China did not fit these curves; for the UK this is almost certainly the result of long term variation in risk by birth cohort (Hill and Adelstein, 1967, Osmond et al., 1982), and in China the consequence of a low level, and late onset, of exposure to aetiological factors, especially HPV. When temporal changes in the curves were analysed, in the Nordic countries, the peak incidence shifted with time towards earlier ages. This also probably represents the effect of increasing risk with succesive birth cohorts, since cross-sectional analysis of age-specific incidence showed that a 3% annual increase in successive birth cohorts would move the shape of the curves seen in the second group of countries above closer towards the shape of that seen in the first group. This adds further weight to the other evidence that there are strong cohort effects which need to be taken account of in any analysis of incidence with respect to time.

Modelling time trends

Modelling techniques have been widely used as a means of determining the nature of any temporal changes. Assuming that the effect of age on risk remains constant over time, the simplest procedure is to determine whether an age-period model, or an age-cohort model, provides the best description of the observed data. Although the interpretation of effects observed with using models is difficult and fraught
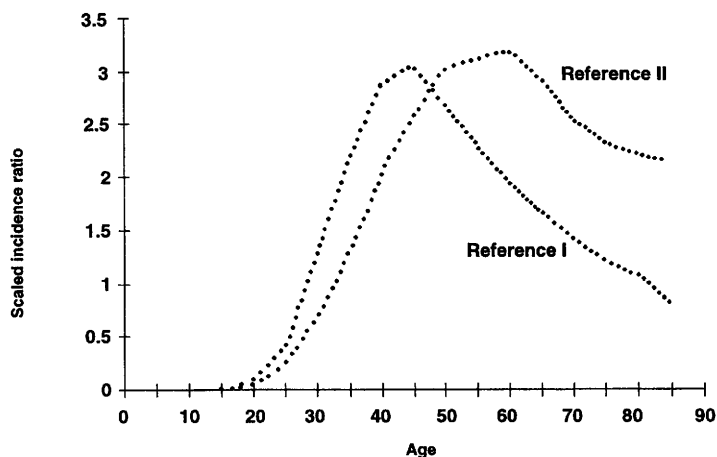
Figure 1. Scaled age-specific incidence ratios for cervical cancer for time periods prior to screening. Reference I: weighted average from Denmark, Germany, Netherlands, Norway, Slovenia and Sweden. Reference II: weighted average from Finland, Estonia, Latvia, Lithuania, Poland, Connecticut, Brazil, Colombia, Jamaica, Puerto Rico, Hong Kong, India, Israel, Japan, New Zealand, Singapore, Thailand and Africa. Scaling is by dividing each value by the world-standardized rate for the same population (reproduced by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc., from Gustafsson *et al.*, 1997a)

with error (Clayton and Schifflers, 1987) they nevertheless allow the investigation of time trends when data may not be present for all desirable points in the study. Age period models use log-linear models to describe variations in rates by age according to calendar period of observation whereas age cohort models use log-linear models to describe variation within a cohort or generation to which the rates apply. The methodology has been extended to develop age-period-cohort models. These are widely used in comparing data across countries, as described by Coleman *et al.*, (1993). Even at a national or regional level, the cohort effects are so strong in some countries, notably the UK, that the beneficial effects of a

62

screening programme can be masked. (See Chapter 10 in this volume for further discussion on the use of models).

Studies of trends in incidence

There have been many analyses of cancer registry data in relation to the possible influence of screening programmes on the incidence of cancer, both in developed countries (e.g., Anderson et al.,1988) and developing countries (e.g., Aristizabal et al., 1984).

Pompe Kirn et al. (1992) examined trends for invasive and in situ cancers in Slovenia from 1950 to 1986. Increasing registrations of in situ cases was paralleled by a decline in invasive disease from to early 1960's until about 1981. The age-specific trends suggest, however, that interpretation may not be entirely straightforward, since there appears to have been an increasing risk of disease in recent birth cohorts.

Marked changes in birth-cohort specific risk of disease have been a particular problem in interpreting time trends in the UK. Quinn et al. (1999) recently completed a comprehensive analysis of incidence in England and Wales, in which age-specific trends were compared with the coverage of the screening programme (percent of women aged 25-64 screened in the preceding five years) and registrations of in situ cancer. The introduction of the national call and recall system in 1988 had had a clear effect in increasing coverage and was associated with a decline in incidence in all age groups under 74. However, quantifying the fraction of disease being prevented remains contentious.

Perhaps the best known studies of time trends in incidence are those undertaken for the Nordic countries, where it was possible to compare the trends in incidence (and mortality) across countries with their different policies in relation to screening (Hakama, 1982; Hakulinen et al., 1986). The extent of the decline in incidence and mortality is related to the coverage and extent of the organised programmes in the respective countries (Figure 2). It was not possible to take account of the opportunistic screening occurring in all countries except Iceland. Nevertheless, the decline in incidence appeared to be clearly proportional to the intensity of organised screening (Sigurdsson, 1999). In Finland, Iceland and Sweden organised mass screenings have been conducted countrywide since the mid-1960's. In Finland, women between the ages of 30 and 55 years have been invited, women aged 30-49 in Sweden, while in Iceland the age range has been wider. In Denmark, about 40% of the population have been subject to organised mass screening, in Norway, fewer than 5%. The declines in incidence were most marked in the age groups targetted by the organised programmes (Table 1).
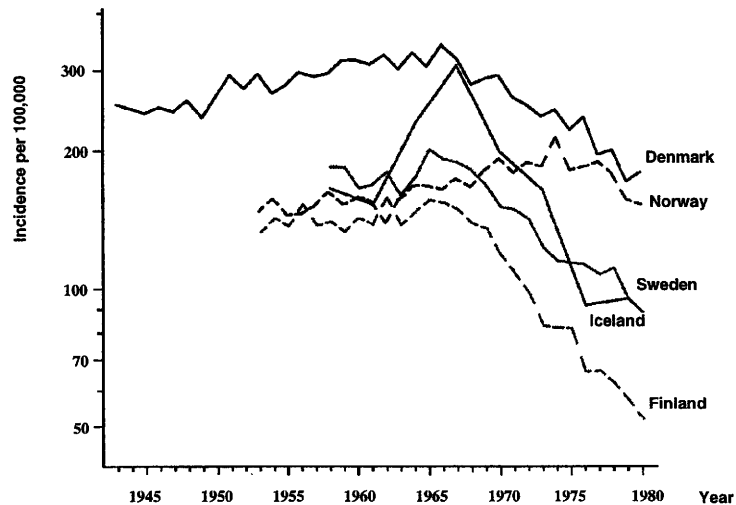
Figure 2. Trends in the age-adjusted incidence rates: cancer of the cervix uteri (reproduced by permission of APMIS from Hakulinen *et al.*, 1986)

Trends by histological subtype

The importance of studying time trends by histological subtype of cancer has been stressed by some workers (e.g. Stockton *et al.*, 1997). Several studies have shown rising incidence rates of adenocarcinoma, in populations where - presumably as a result of screening - incidence rates from squamous cell carcinomas are declining (Kjaer and Brinton, 1993). The increasing risk of adenocarcinoma appears to affect relatively recent generations of women from many countries (Vizcaino *et al.*, 1998), and although there has been no increase in Finland, even in that country the relatively constant incidence of adenocarcinoma contrasts with declines in squamous cell tumours (Nieminen *et al.*, 1995). The cytological detection of carcinoma or precursor lesions is undoubtedly less efficient than for squamous cell tumours (Fu *et al.*, 1987; Sigurdsson, 1995) and a case-control study (Mitchell *et al.*, 1995) has shown that the risk of adenocarcinoma is not reduced by screening.

64

Table 1. **Screening for cervix cancer in the Nordic countries (to 1995)** (reproduced by permission of ©2000 Munksgaard International Publishers Ltd. Copenhagen, Denmark, from Sigurdsson, 1999)

Cervical cancer screening	Iceland	Finland	Sweden	Demark	Norway
Start of organized screening (fully developed in year...)	1964 (1969)	1963 (1970)	1964 (1973)	1962 (1975)	1994
Age range of targeted population up to 1985	25-69	30-55	30-49	30-50	
Screening interval in years up to 1985	2-3	5	4	4	
Change in screening policy (year) since 1985	1988	1987	1985	1986	
Target age group	20-69	30-60	20-59	23-75	25-70
Screening interval	2-3	5	3	3-5	3
Targeted coverage of national population in 1991	100%	100%	100%	45%	
% of smears taken outside organized screening after 1980	16-24	66-80	75-80	>80	100
Five-year period with highest incidence rate	1966-70	1961-65	1961-65	1961-65	1971-75
Reduction of overall world-adjusted incidence rate through 1986-1995	67%	75%	55%	54%	34%

Trends by stage of disease

Although the primary objective of cervical screening is to prevent the development of invasive cervical cancer, the screening test also detects asymptomatic invasive disease, and hence the introduction of cervical screening will also affect the stage distribution of invasive disease. However analysis of temporal trends in stage are always beset with the difficulties of changes in the precision with which stage is determined over time, and the high proportion of unknown stage in earlier years.

Johannesson *et al.* (1978) reported on the incidence of cervical cancer in Iceland by stage of disease at diagnosis between 1965 and 1974 and Sigurdsson (1993) has updated the results to 1991. The screening programme led to a considerable increase in early, micro-invasive (Stage 1A) cancers, with a fall in incidence of more advanced stage disease (stages II-IV). The authors note that micro-invasive disease, like carcinoma *in situ*, has an excellent prognosis (no cases had died since 1964) so that detection of these cases should be counted as a beneficial effect of the programme (rather than counted along with the invasive cancers, which the programme aims to reduce).

In British Colombia, screening was introduced early (1949-50) and careful records were maintained on screening histories of individual women. The progressive fall in the incidence of invasive cancer from the mid-1950's to 1977 occurred despite an increase in the incidence of micro-invasive and occult invasive carcinomas (Boyes *et al.*, 1981). Comparing cancers registered in 1969-91 with those from 1955-59, there was a higher proportion of later stage disease among the invasive cancers (despite a fall in absolute numbers). This was because the invasive cancers were occurring in older women who had, for various reasons, escaped the screening programme and also to the possibly more rapid growth on non-screen detected cases. Christopherson *et al.* (1976) had noted the same phenomenon in Louisville, Kentucky, where, after 15 years of screening, the increase in the proportion of incident cancers, which were in older women (less intensively screened), resulted in a relative increase in later stage disease.

Herbert *et al.* (1998) studying the effects of screening in the south of England, demonstrate the importance of following time trends by stage of disease, in their evaluation of the effects of improvements to the screening programme in Southampton and SW Hampshire (UK). Increased registration of screen-detected cancers either microinvasive (FIGO Stage Ia1) or with minimal invasion (FIGO Stage Ib) resulted in very little overall change in incidence in the first six years, despite a marked decline in incidence of symptomatic invasive disease (Figure 3).
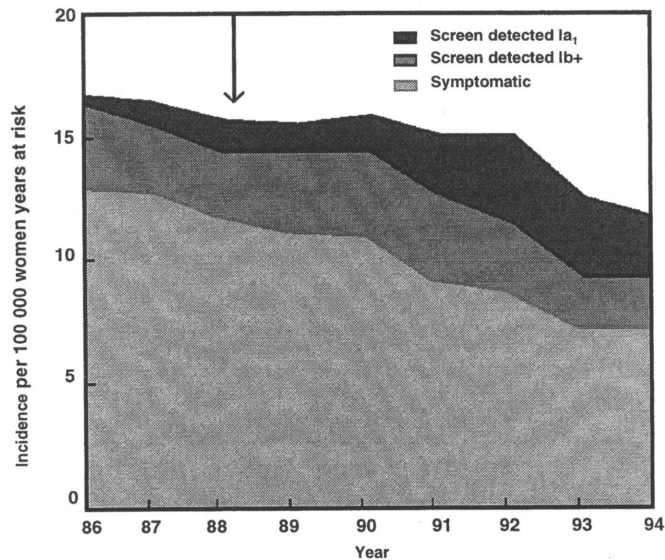
Figure 3. **Trends in cervical cancer incidence, by stage, Southampton and SW Hampshire (England). Introduction of national screening programme in 1988 (arrow)** (reproduced by permission of BMJ Publishing Group from Herbert *et al.*, J Med Screen 5:92-98, 1998)

Studies of trends in mortality

Although a reduction in disease-specific mortality is the ultimate goal of cancer screening, in the case of cervix cancer, this is achieved through a reduction in the incidence of the disease. A study of the latter is thus a more direct indicator of the outcome of screening. In general, although mortality data are often available over a longer historical time period than incidence data, and for national populations, they are less satisfactory for the evaluation of cervical screening programmes for the following reasons:

(a) Mortality can be affected by improvements in treatment (reflected by better survival) as well as changes in incidence. It could be argued that improved survival should follow the detection of invasive cancer at earlier stages (which, as noted above, is a consequence of screening programmes), so that improvement in mortality may be greater that the reduction in incidence. However, survival <u>within stage</u> has also improved over time (Sparén *et al.*, 1995) implying better results from treatment of invasive disease.

67

(b) Death certification is less precise in terms of cause of death than incidence data recorded by cancer registries. As well as erroneous cause of death statements, a varying proportion of cervix cancer deaths are coded to "uterus not otherwise specified". A change in the proportions of "Uterus NOS" deaths over time can lead to spurious time trends in mortality from cervix cancer (Figure 4). Some form of "reallocation" of these deaths to more specific categories is required, a process which is always to some extent arbitrary.

(c) There is a time delay of approximately 10 years before changes in mortality are apparent.

There have been many reviews of trends in mortality rates from cervix cancer, including international comparative studies (La Vecchia *et al.* 1992; Beral *et al.*, 1994; Coleman *et al.*, 1993; Robles *et al.*, 1996), as well as reviews confined to single countries. The objective of international comparative studies is generally simply descriptive − to draw attention to inconsistencies in the pattern between countries, as an indicator of where more focussed research would be of interest, rather than to interpret the trends in terms of the efficiency (or otherwise) of specific screening programmes. An exception to this are the studies of mortality in the Nordic countries which, like the studies of incidence trends, have been interpreted in terms of the extent of organised screening (Läärä *et al.*, 1987).
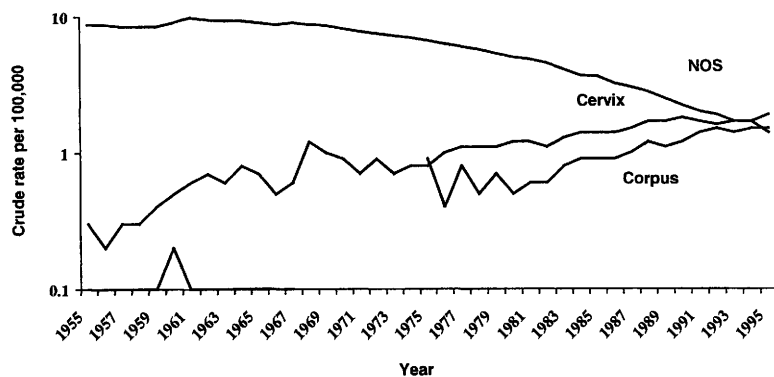
Figure 4. Trends in recorded mortality of cancer of the cervix uteri, cancer of the corpus uteri and uterus cancer, NOS, in Spain (1955-1995)

Cook and Draper (1984) examined mortality trends for England and Wales for the period 1950-1982. They observed the strong birth-cohort effects, described earlier by Hill and Adelstein (1967), and noted that, because of these, it was difficult to quantify the effect of screening on mortality. However, by reviewing also the trends in incidence (of invasive cancer and carcinoma *in situ*), they could conclude

68

that screening was probably responsible for part of the mortality decline at ages 35-54. MacGregor and Teper (1978), also drew attention to the rise in mortality in young women (25-34) in England and Wales between 1968 and 1976; however, in Scotland, there was a decline in this age group in the two regions with well organised screening (but a rise elsewhere). No quantitative data were provided on screening intensity, however.

Devesa *et al.* (1989) reviewed trends in mortality (1950-1985) and incidence (1969-1985) in the United States. Both had showed marked declines, with a moderation in the fall in younger women in the more recent years. They noted that, considering that the prevalence of most other risk factors (related to sexual lifestyles, contraception, smoking) might have been expected to increase risk of the disease, the observed declines might reasonably be attributed to screening.

De Schryver (1989) noted that the falls in mortality in Belgium had been relatively modest, confined to the middle age groups (35-54), and had begun before screening was introduced. He drew attention to the similarity to the trends in Norway, where screening was also opportunistic, and not very intense.

Vlasak *et al.* (1991) observed increasing mortality and incidence (especially in younger women) in Slovakia (1968-1987); the few cases of carcinoma *in situ* registered suggested that screening was failing to prevent much disease.

Quantifying the impact of screening:

The problem of lack of "expected" rates of disease in the absence of screening, when interpreting observations of trends in incidence (or mortality), has been alluded to earlier. This aspect, often overlooked, has been approached in several ways.

In investigating the effect of cervical screening programmes globally, Gustafsson and colleagues (1997a) determined age-specific incidence patterns in several populations, as described above. They then (Gustafsson *et al.*, 1997b) compared observed incidence in 17 countries (with 15 or more years of follow up) during a period after screening was introduced with the rates observed pre-screening. To estimate changes from the baseline period before screening, incidence rates directly age standardised to the world standard population were calculated for all ages and to the world population truncated to 35-64 years. Where age standardised rates decreased at least 25% from the baseline rates, age-specific rate ratios were calculated. Age-standardised incidence rates decreased by at least 25% in 11 of the 17 studied populations, with an age-specific pattern demonstrating some common features. There was no change in incidence rates below 30 years of age, a strong

69

reduction around the peak incidence leading to a flattening of the incidence curve with age, followed by a less pronounced reduction with increasing age. The U-shaped curves of age-specific rate ratios diverged from 1 (no effect) at ages 25-35, reached a minimum (maximum reduction) at 45 to 55, and approached 1 at higher ages (Figure 5). In 12 of these populations, examination of data from periods before screening had demonstrated the presence of pre-existing trends. But since they were mainly positive, the authors noted that this could only have resulted in an underestimate of the effects of screening. The authors note, however, that it is difficult to determine the precise effect of screening without knowledge of the intensity of screening.
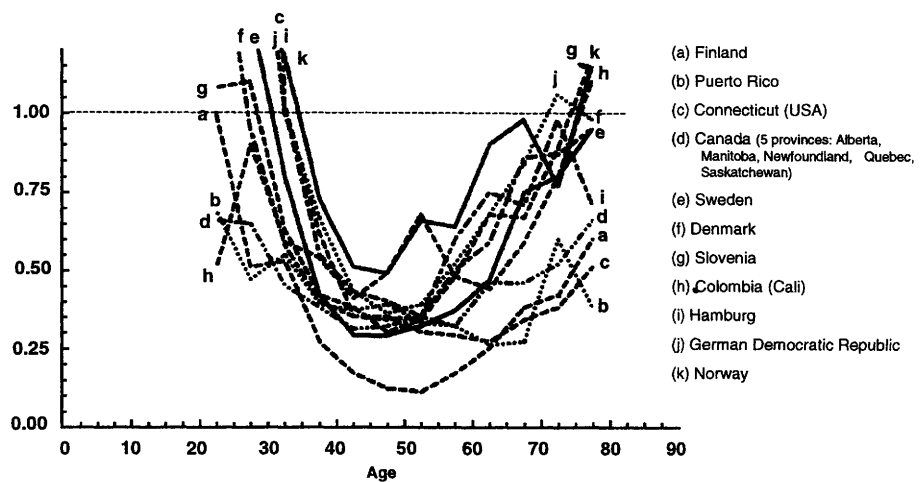
Figure 5. Rate ratios between age-specific incidence rates of cervix cancer for the last follow-up and baseline periods for successfully screened populations: (a) Finland, (b) Puerto Rico, (c) Connecticut (USA), (d) Canada (5 provinces: Alberta, Manitoba, Newfoundland, Quebec, Saskatchewan), (e) Sweden, (f) Denmark, (g) Slovenia, (h) Colombia (Cali), (i) Hamburg, (j) German Democratic Republic, (k) and Norway. For each age, the age-specific incidence rate in the last follow-up period is divided by the corresponding age-specific incidence rate in the baseline period (reproduced by permission of Kluwer Academic Publishers from Gustafsson et al., 1997b)

The approach taken by the East Anglia Cancer Registry in England (Gibson et al., 1997) takes into account short term pre-existing time trends within age groups (and allows for regional variation in such trends) in preparing "expected" values against which to compare observed incidence. The "pre-screening" period (1981-1990) had a rather lower percentage coverage of screening than the observation period (1991-

70

93), and the authors considered that the magnitude of the difference in Observed-Expected cases of cancer, and the absence of effect in women aged over 70, strongly implicated a beneficial effect of screening.

The problem of estimating expected incidence (or mortality) has been particularly acute in UK, because of the marked variation in risk of disease in different birth cohorts. This effectiveness of screening has therefore had to be judged against a risk of disease increasing in successive generations of women born since about 1935 (Osmond *et al.*, 1982; Sasieni and Adams, 1999). In fact, evaluating underlying risk of disease from observed trends in mortality will result in an underestimate, since mortality will itself have been reduced by screening. The approach taken by Parkin *et al.* (1985), was to "correct" the observed incidence rates, making use of the information on the number of cases of *in situ* cancer detected as a result of screening activity, and the proportion of these which would
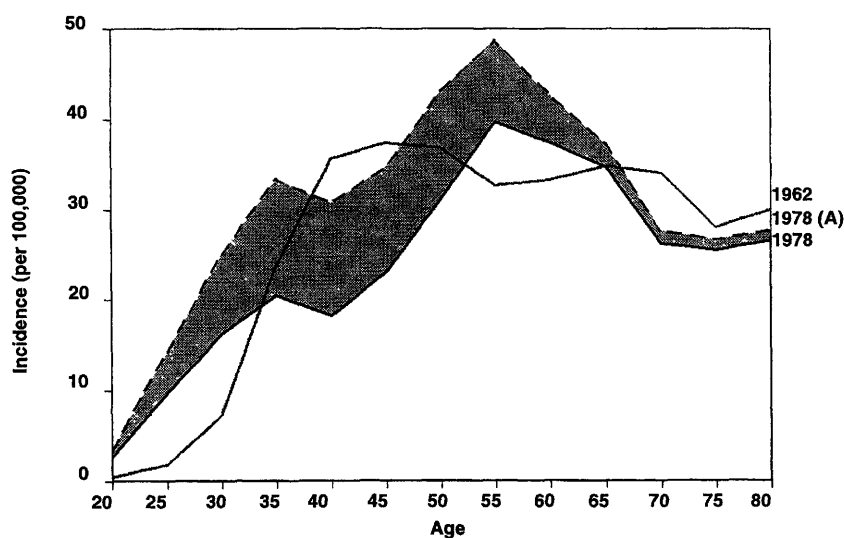
Figure 6. Annual age-specific incidence rates (per 100 000 uteri) of carcinoma of the cervix uteri in England and Wales in 1962 and 1978. For 1978, the estimated incidence in the absence of screening (A) is also shown. The shaded area represents the reduction in incidence due to screening (reproduced by permission of Blackwell Science from Parkin, 1997, in Franco & Monsonego (eds) New Developments in Cervical Cancer Screening and Prevention; originally from Parkin *et al.*, 1985, Br J Obstet Gyn 92:150-157)

have been expected to progress to invasive cancer. The rates were also corrected for the women who would cease to be at risk of the disease by virtue of them

71

having a hysterectomy. In this way the expected incidence of invasive cancer was estimated and compared to the observed rates. This demonstrated that the reported increase in younger women (aged under 35) between 1963 and 1978 was the result of cohort specific increases in risk for generations born since 1931; screening had reduced the potential increase in incidence by 50% and prevented a significant increase in older age groups (Figure 6). This at least partly explains the observations of the UK data in Gustafsson's paper. However, using this model to quantify the effect of well established programmes of long standing would require some estimates of both the incidence and the rate of progression of CIN2 lesions which is not easy to acquire on a national basis.

Conclusions

Cancer registries can make an important contribution to the evaluation and monitoring of cervical cancer screening programmes by the analysis of trends over time in the incidence of, and to a lesser extent mortality from, cervical cancer within defined populations. Such analysis has identified strong cohort effects which, if not taken into account in the evaluation of a screening intervention, can lead to incorrect conclusions about the effectiveness of a screening programme. Ideally trends in cancer incidence should be analysed over the period before, during and after the introduction of a screening programme and are of maximum value when screening is undertaken as part of an organised, structured programme. The usefulness of cancer registry data can be enhanced by statistical modelling techniques, both if cancer registry data is not available for all of the period required, and by facilitating comparison with the observed cancer incidence and that expected in the absence of screening. Cancer registries should maximise the opportunity for using registration data in the evaluation and monitoring of cervical screening programmes by developing links with the screening programme itself, modifying the cancer registration dataset to include detailed staging information, and analysing time trends in cancer incidence alongside screening activity data.

References

Anderson GH, Boyes DA, Benedet JL, Le Riche JC, Matisic JP, Suen KH, Worth AJ, Millner A & Bennett OM (1988). Organisation and results of the cervical cytology screening programme in British Columbia, 1955-85. *Br Med J* **296**:975-978

Aristizabal N, Cuello C, Correa P, Collazos T & Haenszel W (1984). The impact of vaginal cytology on cervical cancer risks in Cali, Colombia. *Int J Cancer* **34**:5-9

Ashley DJ (1966). Evidence for the existence of two forms of cervical carcinoma. *J Obstet Gynaecol Br Commonw* **73**:382-389

Beral V, Hermon C, Munoz N & Devesa SS (1994). Cervical Cancer. *Cancer Surveys* **19-20**:265-285

Bos AB, van Ballegooijen M, van Gessel-Kabekaussen AA & Habbema JD (1998). Organised cervical cancer screening still leads to higher coverage than spontaneous screening in the Netherlands. *Eur J Cancer* **34**:1598-1601

Boyes DA, Worth AJ & Anderson GH (1981). Experience with cervical screening in British Columbia. *Gynaecol Oncol* **12**:S143-S155

Christopherson WM, Lundin FE, Mendez W & Parker JE (1976). Cervical cancer control. A study of mortality and morbidity trends over a 21-year period. *Cancer* **38**:1357-1366

Clayton D & Schifflers E (1987). Models for temporal variation in cancer rates. I. Age-period and age-cohort models. *Stat Med* **6**:449-467

Coleman MP, Estève J, Damiecki P, Arslan A & Renard H (eds) (1993). *Trends in Cancer Incidence and Mortality* (IARC Scientific Publications No.121), Lyon, International Agency for Research on Cancer

Cook GA & Draper GJ (1984). Trends in cervical cancer and carcinoma in situ in Great Britain. *Br J Cancer* **50**:367-375

de Schryver A (1989). Does screening for cervical cancer affect incidence and mortality trends? The Belgian experience. *Eur J Cancer Clin Oncol* **25**:395-399

Devesa SS, Young JL Jr, Brinton LA & Fraumeni JF Jr (1989). Recent trends in cervix uteri cancer. *Cancer* **64**:2184-2190

Doll R, Payne P & Waterhouse J (eds) (1966). *Cancer Incidence in Five Continents*, Vol. 1, Union Internationale Contre le Cancer, Geneva, Springer Verlag

Fu YS, Berek JS & Hilborne LH (1987). Diagnostic problems of in situ and invasive adenocarcinomas of the uterine cervix. *Appl Pathol* **5**:47-56

Gibson L, Spiegelhalter DJ, Camilleri-Ferranti C & Day NE (1997). Trends in invasive cervical cancer incidence in East Anglia from 1971-1993. *J Med Screen* **4**:44-48

Gustafsson L & Adami HO (1992). Optimization of cervical screening. *Cancer Causes Control* 3:125-136

Gustafsson L, Sparén P, Gustafsson M, Wilander E, Bergström R & Adami HO (1995). Efficiency of organised and opportunistic cytologic screening for cancer in situ of the cervix. *Br J Cancer* 72:498-505

Gustafsson L, Pontén J, Bergström R & Adami HO (1997a). International incidence rates of invasive cervical cancer before cytological screening. *Int J Cancer* 71:159-165

Gustafsson L, Pontén J, Zack M & Adami HO (1997b). International incidence rates of invasive cervical cancer after introduction of cytological screening. *Cancer Causes Control* 8:755-763

Hakama M (1982). Trends in the incidence of cervical cancer in the Nordic countries. In: Magnus K (ed) *Trends in Cancer Incidence: Causes and Practical Implications.* Washington, Hemisphere, pp. 279-292

Hakama M & Penttinen J (1981). Epidemiological evidence for two components of cervical cancer. *Br J Obstet Gynaecol* 88:209-214

Hakulinen T, Andersen A, Malker B, Pukkala E, Schan G & Tulinius H (1986). Trends in cancer incidence in the Nordic countries. A collaborative study of the five Nordic cancer registries. *APMIS* Suppl. 288:1-151

Herbert A, Bryant TN, Campbell MJ & Smith J (1998). Investigation of the effect of occult invasive cancer on progress towards successful cervical screening. *J Med Screen* 5:92-98

Hill GB & Adelstein AM (1967). Cohort mortality from carcinoma of the cervix. *Lancet* 2:605-606

Johannesson G, Geirsson G & Day N (1978). The effect of mass screening in Iceland 1965-1974, on the incidence and mortality of cervical cancer. *Int J Cancer* 21:418-425

Kjaer SK & Brinton LA (1993). Adenocarcinoma of the uterine cervix: the epidemiology of an increasing problem. *Epidemiol Rev* 15:486-495

Läärä E, Day NE, Hakama M (1987). Trends in mortality from cervical cancer in the Nordic countries: association with screening programmes. *Lancet* 1:1247-1249

La Vecchia C, Lucchini F, Negri E, Boyle P, Maisonneuve P & Levi F (1992). Trends of cancer mortality in Europe, 1955-1989: III, Breast and genital sites. *Eur J Cancer* **28A**:927-998

MacGregor J & Teper S (1978). Mortality from carcinoma of cervix uteri in Britain. *Lancet* **2**:774-776

Mitchell H, Dedley G, Gorden I & Giles G (1995). Cervical cytology reported as negative and risk of adenocarcinoma of the cervix: no strong evidence of benefit. *Br J Cancer* **71**:894-897

Nieminen P, Kallio M & Hakama M (1995). The effect of mass screening on incidence and mortality of squamous and adenocarcinoma of cervix uteri. *Obstet Gynaecol* **85**:1017-1020

Office of Population Censuses and Surveys (1993) Series MBI No 20. *Cancer Statistics registration: Registration of cancer diagnosed in 1987, England and Wales*. London, Her Majesty's Stationary Office

Osmond C, Gardner MJ & Acheson ED (1982). Analysis of trends in cancer mortality in England and Wales during 1951-80 separating changes associated with period of birth and period of death. *Br Med J* **284**:1005-1008

Parkin DM (1997). The Epidemiological Basis for Evaluating Screening Policies. In: Franco E and Monsonego J (eds). *New Developments in Cervical Cancer Screening and Prevention*. Blackwell Science Ltd, Oxford, pp. 51-69

Parkin DM, Nguyen-Dinh X & Day NE (1985). The impact of screening on the incidence of cervical cancer in England and Wales. *Br J Obstet Gynaecol* **92**:150-157

Parkin DM, Pisani P & Ferlay J (1999). Global Cancer Statistics. *CA Cancer J Clin* **49**:33-64

Pompe Kirn V, Kovacic J & Primic Zakelj M (1992). Epidemiological evaluation of cervical cancer screening in Slovenia up to 1986. *Eur J Gynaecol Oncol* **13**:75-82

Quinn M, Babb P, Jones J & Allen E (1999). Effect of screening on incidence of and mortality from cancer of cervix in England: evaluation based on routinely collected statistics. *Br Med J* **318**:904-908

Robles SC, White F & Peruga A (1996). Trends in cervical cancer mortality in the Americas. *Bull Pan Am Health Organ* **30**:290-301

Sasieni P & Adams J (1999). Effect of screening on cervical cancer mortality in England and Wales: analysis of trends with an age period cohort model. *Br Med J* **318**:1244-1245

Sigurdsson K (1993). Effect of organised screening on the risk of cervical cancer. Evaluation of screening in Iceland 1964-1991. *Int J Cancer* **54**:563-570

Sigurdsson K (1995). Quality assurance in cervical cancer screening: the Icelandic experience 1964-1993. *Eur J Cancer* **31A**:728-734

Sigurdsson K (1999). The Icelandic and Nordic cervical screening programs: trends in incidence and mortality rates through 1995. *Acta Obstet Gynecol Scand* **78**:478-485

Sparén P, Gustafsson L, Friberg LG, Pontén J. Bergström R & Adami HO (1995). Improved control of invasive cervical cancer in Sweden over six decades by earlier clinical detection and better treatment. *J Clin Oncol* **13**:715-725

Stockton D, Cooper P & Lonsdale RN (1997). Changing incidence of invasive adenocarcinoma of the uterine cervix in East Anglia. *J Med Screening* **4**:40-43

Van Oortmarssen GJ, Habbema JDF, Lubbe JThN, De Jong GA & Van der Maas PJ (1982). Predicting the effects of mass screening for disease - a simulation approach. *Eur J Oper Res* **6**:399-409

Vizcaino AP, Morenzo V, Bosch FX, Muñoz N, Barros-Dios XM & Parkin DM (1998). International trends in the incidence of cervical cancer: 1. Adenocarcinoma and adenosquamous cell carcinomas. *Int J Cancer* **75**:536-545

Vlasak V, Plesko I, Dimitrova E & Hudakova G (1991). Recent trends in uterine cervix cancer in Slovakia, 1968-1987. *Neoplasma* **38**:533-540

Chapter 6. Cervix cancer: geographical correlations

A. Anttila and E. Läärä

Background

There is nowadays no big divergence from the opinion that adequately performed Pap smear screening can markedly reduce cervical cancer incidence and mortality. In spite of widespread application, there is still disagreement on the size of the overall effects of Pap-screening programmes. Factors affecting the overall effectiveness include a high and evenly-spread coverage, regular compliance of the target population, including high-risk populations at the defined intervals, adequate sample taking and handling practices, combined with well-controlled diagnostic work in the cytological analysis of the smears; and the adequacy and completeness of the subsequent follow-up and treatment of the precancerous lesions detected (Coleman et al., 1993). Lack of one or more of these characteristics might explain the relative failure of some screening programmes.

The evaluation of a screening programme should not only be organised in terms of the screening activities and outcomes, but also include follow-up of incidence and mortality data among attenders and non-attenders as well as at the level of the overall target population (Hakama et al., 1985; IARC, 1986; Coleman et al., 1993). Variation in opinion has resulted in differences in screening policies adopted in different countries or centres − between different options for organised screening programmes, and between organised and spontaneous screening modalities. The scope of this review is to summarise the results of the studies on the overall impact of Pap-screening activities in decreasing cervical cancer incidence and mortality. Estimation is based on the associations between screening indices and invasive cervical cancer endpoints in different geographical areas. Variation in screening policies has made it possible to compare their overall effectiveness in different settings. We also describe briefly the methods and designs used in these studies and discuss their limitations.

Methodological considerations

Studies of geographical correlations as an epidemiological approach can be subsumed under the so-called ecological studies (Morgenstern, 1998). In these studies, the units of study are groups, aggregates or populations of individuals, often citizens living within the same administrative region. This is in contrast with

the common cohort and case-control designs, in which the study units are individuals themselves. In individual-based studies, the measurements of both the outcome and the exposure variables can be linked at individual level within the entire study population.

In ecological studies, the measurements of exposure and outcome variables refer to a whole group or population of individuals. The outcome is represented by the incidence or mortality rate aggregated over the individual occurrences of the disease in the group during the same follow-up period. The ecological exposure variables can be of three types: (a) aggregate or summary measures (averages or proportions) of individual values, (b) environmental characteristics shared by the members of the group, or (c) global measures that are attributes of the group, organisation, or place for which there is no distinct analogue at the individual level, e.g., population density, type of health care system. As a consequence, in an ecological study, information exists only on the marginal distributions of the exposures and the outcome within a group. Another important statistical feature present in these studies concerns the nature of random variation. In addition to the unsystematic variation across individuals in a given population group affecting the statistical precision of the incidence rate in that group, there is an important variance component in an ecological study due to the random variation between the groups. If one takes into account only the individual variation within a group, the resulting standard errors and confidence intervals of the ecological estimates are likely to be too narrow.

Ecological studies can be divided into the following designs: (I) multiple-group design, in which the disease occurrence is studied in different populations during the same limited period in time, (II) time-trend design, in which the disease occurrence is investigated in one selected population over various periods in time, and (III) mixed design, in which several populations are followed-up over the same periods of time.

Geographical correlation studies are generally considered within design category (I), and they are sometimes described as "descriptive", i.e., quantifying the statistical associations between aggregate exposure variables and disease incidence, in order to give clues to aetiological relationships to be further investigated in "analytical" case-control and cohort studies at the individual level. The above view on geographic or ecological studies is too restricted. Many interesting aetiological questions can readily be studied on group level. In fact, there are important causal hypotheses which can only be tested at this level. The effectiveness of a given health care policy, like a screening programme targeted to a demographically defined population, is such a causal question, the answer of which demands an ecological design, because a population-based intervention should also be analysed in terms of population-based outcome.

78

Cause and effect with respect to population interventions (like screening policies) is generally studied by investigating the outcome incidence in a target population under the implemented screening policy (like a defined organised screening programme) as compared with the incidence one would have observed, if an alternative policy (like no organised screening programme at all) were adopted. At best an 'analytical' ecological study is a randomised experiment, in which the different aggregate units forming the meta-population of the study are randomly divided into two sets: in the intervention set of units, the screening programme is implemented to cover the whole target population, and in the control set of units, there is no such organised intervention.

The label 'geographical correlation' should not be interpreted too narrowly to mean that only correlation coefficients are used to describe the statistical association between ecological exposure variables and the outcome. A correlation coefficient (whether Pearson or Spearman) may not be informative at all about the causal association of interest. It merely tells whether an exposure variable is linearly related to the outcome. Certain other statistical measures and methods, like appropriate contrasts (absolute or relative) of the incidence rates between screened and non-screened units, may better describe the effect of an intervention.

In this paper, we focus on non-experimental ecological studies, in which the intervention programme is not randomised across the groups forming the meta-population; i.e., the characteristics of the screening programme are varying between the groups. As in individual-based studies, biases due to misclassification and confounding are also present in geographical studies. Therefore, the interpretation of the results of these studies is not straightforward at all. A special form of bias prevalent in ecological studies is the ecological bias: the association at the group level is qualitatively different from that at the individual level, which is a consequence of heterogeneity of exposure level and covariate levels within groups. This heterogeneity cannot be captured with ecological data because of missing information on the joint distributions of the exposures and outcome at individual level. An example of an ecological bias is the association between the income level and incidence of cancer of the cervix (Hakama, 1982). At the level of municipalities one can observe a positive association: the higher the average income in a municipality the higher is the incidence. On the other hand, at individual level, the association is negative: women with higher than average income have a lower incidence than women do at lower income levels. The results imply that a woman with a low standard of living in a well-to-do environment has a high risk of cervical cancer.

In a non-experimental ecological evaluation of a screening programme one would preferably use the mixed design (III). For several populations defined by geographical region there should be outcome data from a period before there was screening programme, and from a period after an appropriately long time had

passed since the initiation of a screening programme in some of the regions. This design allows a before-after analysis (Morrison, 1998) relating the change in incidence or mortality to the existence and properties of a screening policy. Each region, thus, serves as its own control. However, one cannot rely on before-after analysis of incidence trend in one population only, so there should also be control regions or populations with different screening policies at the outset.

Screening policy is not an all-or-none variable. When evaluating the effectiveness of an organised screening programme, one should have data at least on the following important characteristics: (i) target age range, (ii) screening interval, (iii) certain organisational aspects (information, invitation, attendance, and referral procedures), (iv) diagnostic and clinical aspects and (v) quality control procedures in the whole screening process. Changes in these characteristics over time should also be taken into account.

In the context of evaluating a cervical (or any) screening programme, coverage and compliance are of utmost importance: what proportion of the target population are actually invited or offered screening, and what proportion of those invited have smears taken according to the scheme. These are of course important determinants of the ultimate outcome, i.e., the incidence of (or mortality from) the invasive disease. However, coverage and compliance are not intrinsic fixed properties of a programme, but they are dependent on the characteristics of the programme. Thus, they can be viewed as an intermediate outcome. Compliance is also, to a large extent, dependent on certain socio-economic features of the target population. It is also related to the availability of spontaneous smear taking in the region. These aspects may vary over time. The actual compliance and coverage must of course be taken into account in the interpretation of ecological analyses of screening programmes. A straightforward 'blind' adjustment for them might sometimes give a misleading estimate of the actual effectiveness of the programme, however, if e.g. the individual screening histories in the population were not considered adequately.

Materials and methods

We found 27 references reporting data on geographical correlation or ecological association between Pap-screening activities and invasive cervical cancer incidence or mortality rates (summarised in Table 1). In a single report the multiple-group design (I) was adopted such that the incidence and mortality figures were correlated with the intensity of screening in different districts during a single period only (Fouquet and Gage 1996). Moreover, in this study the outcome was obtained from a simultaneous or even a partly earlier time period than the screening data. The time-trend design (II) was applied in 13 studies in which the incidence or mortality rates before and after the initiation of screening activity were compared

in one population only (country or region in a country). Another 13 studies used the mixed design (III) allowing evaluation of the temporal changes of incidence or mortality related to screening in many populations. Yet, most of these were restricted to comparing different regions in one country, but in three studies several countries were included (Hakama, 1982; Läärä et al., 1987; Gustafsson et al., 1997; additional follow-up material available in Engeland et al., 1993; Hristova and Hakama, 1997; Parkin et al., 1997).

In most of the studies, the association between screening and incidence or mortality was analysed in a rather informal way. In comparisons between the Nordic countries (Hakama, 1982; Läärä et al., 1987) summary statistics on certain key characteristics (target age range, screening interval, coverage, attendance) of the screening programmes in the five countries were described to aid the interpretation of changes in the outcome. More common were studies in which important details on the organisational aspects, coverage, attendance and other properties of screening in the study populations were inadequately described. For example, in the article by Gustafsson et al. (1997) the incidence of cervical cancer after "the introduction of a screening programme" was compared to that before the era of screening in several populations without any information on what these different "screening programmes" actually contained. Interpretations of the observed changes of incidence are therefore difficult to make from this study. In contrast to that, Gibson et al. (1997) performed a more focused before-after analysis in which the trends of incidences in the districts of East Anglia were related to the timing of the well-documented changes in the organisation and management of the national screening programme in England occurring in 1988-1989. A special methodological feature in this study was the appropriate statistical treatment of the variability between districts.

A few studies performed a formal analysis in terms of correlation coefficients between measures describing the screening activity and percent change in the incidence or mortality rate since the onset of the screening activity (Cramer, 1974; Miller et al., 1976; Murphy et al., 1988). In general, though, besides the information on the extent or intensity of screening, not much historical details on the organisation and clinical aspects of the different screening centres were given. Usually the proportion (rate) of screenings done per female population during a one- to five-year period were described, the length of the period depending on the targeted screening interval within the region. These do not necessarily describe well the screening histories of individual women. In some study areas there might have been screening activities at least in a limited extent also during the time considered 'unscreened' (noted by Miller et al., 1976). So far only one study has reported systematically the details on the quality assurance in cervical cancer screening (Sigurdsson, 1995).

A couple of the correlation studies (e.g. Miller *et al.*, 1976) attempted to estimate the impact of screening by simultaneously adjusting for other ecological-level variables (for example, on the social and demographic characteristics, population flows, prevalence of hysterectomies, proxies on sexual life, smoking habits, or some other behavioural risk factors) using multiple regression methods. A log-linear Poisson model was applied by Lynge *et al.* (1989) in estimating the impact of the initiation of an organised screening programme in Denmark where the various administrative regions adopted different policies concerning organised screening.

Results

After the introduction of the cervical cytological smear in the early 1940s, it took several years before the test became widely available (Ayre, 1964; Cramer, 1974; Gardner and Lyon, 1977). There were some early screening programmes launched in limited areas in the United States and Canada already in late 1940s, but it was not until the late 1950s or in 1960s that cytological screening became widespread in these countries. The cervical cancer mortality rates had started to decrease to a minor degree already before the introduction of the screening programmes, with little if any evidence of a further contribution of screening (Miller *et al.*, 1976; Gardner and Lyon, 1977). In limited areas in the United States, decreases between 40% and 66% in the incidence of invasive cervical cancer and in mortality were reported, following an expansion of the programmes to one-year coverage of 25% to 45% in the late 1960s (Dickinson *et al.*, 1972; Cramer, 1974; Christopherson *et al.*, 1976, Table 1). In Canada, the intensity of screening was related to the fall in death rate between 1960-1962 and 1970-1972 at the provincial, county, and census division levels (Miller et al. 1976). The first large-scale centrally organised screening programme had been launched in the province of British Columbia, Canada in 1949. The lifetime coverage of smears increased up to 85% from 1970s onwards (Boyes *et al.,* 1981; Anderson *et al.*, 1988). The age-specific incidence rate of squamous-cell carcinoma of the cervix uteri in the British Columbia decreased from 28.4 cases per 10^5 woman-years in 1955 to 6.4 cases per 10^5 woman-years in 1985, i.e., 78%, among women over age 20 years. The corresponding mortality rate decreased from 11.4 deaths per 10^5 woman-years in 1958 to 3.1 deaths per 10^5 woman-years (decrease 72%).

In the United Kingdom, the screening activities started to expand in the early 1960s (Parkin *et al.*, 1985; Fouquet and Gage, 1996; Gibson *et al.*, 1997; Table 1). There was some decrease in the registered incidence and mortality rates related to screening intensity also in the UK, at least among the 35-54 years old target population (Parkin *et al.*, 1985, Gustafsson *et al.*, 1997). The rates in younger age groups had increased in a birth cohort-wise manner. There appeared to be no clear

82

effect on the incidence at all ages between 1960 and 1986 associated with the screening activities (Fouquet and Gage, 1996; Gustafsson *et al.*, 1997). The screening programme has been re-modelled since 1988 by introducing a computerised call and recall system for invitations, new quality standards, and new guidelines for the follow-up of women with abnormal smears. According to a report, using data from East Anglia, the coverage of smears has increased after these changes rapidly to 80% of the target population and the incidence of invasive cervical cancer has started to decrease (Gibson *et al.*, 1997). The decrease in incidence was 34% in the age group of 20 to 69 years in 1991-93, compared to the expected incidence based on the age- and district-specific trends in the period 1971-90. In whole England, the age-standardised mortality from cervical cancer has decreased from 6.1 deaths per 10^5 woman-years in 1987 to 3.7 deaths per 10^5 woman-years in 1997 (Quinn *et al.*, 1999).

In the Nordic countries, organised screening programmes based on personal invitations were started in the early 1960s. Spontaneous smears were not in widespread use before the introduction of the organised Pap-screening programmes, but since then they became increasingly common. In Finland, Sweden, and Iceland the programmes became nation-wide in the early 1970s. The coverage of invitations was close to 100% at the target age groups and the participation rate was 70% to 80% in each invitational round in these countries. In Norway, the organised programme was introduced in one county only. In Denmark, the organised programme was introduced in some counties, leading to a population coverage of about 40% at the national level. In Denmark in 1986, however, only 25% of women in the target age range had actually been covered by an organised screening programme (Lynge, 1998). The coverage of the organised screening programmes has increased in Denmark, as well as in Norway, during the 1990s. In Finland and Iceland, the nation-wide programmes were administered centrally, whereas in Sweden and Denmark the organisation has been de-centralised within autonomous counties. Finland has been the only country over the decades where the opportunistic and programme smears have not been 'coordinated', i.e., where all the women at target age groups have been invited irrespective of any recent spontaneous smears. There have also been differences in the screening interval and target age range between these countries.

Before the introduction of the programmes there had been slight increases in the incidence of cervical cancer in each of the Nordic countries, reflecting more intensive detection or registration, or an increased background risk (Hakama, 1982). The programmes were started in the period 1963 to 1967 in most of these countries. Up to the mid-1970s, soon after the introduction of the programmes, there were large decreases in cervical cancer incidence and mortality rates in each of the countries with a large-scale screening programme (Denmark, Finland, Iceland, and Sweden; Hakama, 1982; Läärä *et al.*, 1987; Figure 1). The decrease was confined largely to the age groups 30 to 59 years, i.e., to those groups targeted

83

in the programmes (Figure 2). In Norway – almost completely without organised activities – incidence rates continued to increase, particularly among those below 60 years of age (i.e., among the potential target population). The slope of the decreasing trend in incidence, as well as in mortality, was steepest in those countries with the highest coverage of the organised programme (Finland, Iceland, and Sweden). During that time in Denmark, comparison of the county-based incidence and mortality rates suggested an effect of about 33% associated with the presence of organised screening (Lynge, 1983, Lynge et al., 1989).

The five-year period-specific age-adjusted rates of cervical cancer published from all the Nordic countries can be extended up to the year 1992 (Engeland et al., 1993; Hristova and Hakama, 1997; Parkin et al., 1997). Large differences in the trends have emerged (Figure 1). The decrease in age-adjusted cervical cancer incidence has been largest in Finland, 77% overall, whereas it was 68% in Iceland and 56% in Sweden (Table 1). In Denmark, the overall decrease has been 52%, and in Norway, during the latter half of the follow-up period between 20% and 30%. Interestingly, the slope in the incidence trend in Denmark and Sweden has been almost identical. The relative reductions in the incidence rates have extended in the Finnish patterns also to older women with the aging of the screened population. The incidence rates have increased in Finland during the 1990s in the youngest target ages (Anttila et al., 1999). A high incidence appears to correlate with a low attendance rate in the screening programme within the municipality level. The percent decreases in the age-adjusted mortality in the Nordic countries have been almost of similar magnitudes to those seen in the incidence (Hristova and Hakama, 1997; Figure 3).

There is a unique report available on the impact of the screening programme in East Berlin, former German Democratic Republic, during 1973 to 1982 (Ebeling and Nischan, 1986). The coverage of the programme appeared high, up to 90%. The cervical cancer rates had been very high at the start of the programme. In the ten-year period up to 1982, the incidence had decreased by about 33%. In another nation-wide programme in Cuba, no clear decrease in cervical cancer incidence has been observed (Fernandez et al., 1996). The coverage of the well-controlled screening activities appeared to be limited in the latter programme.
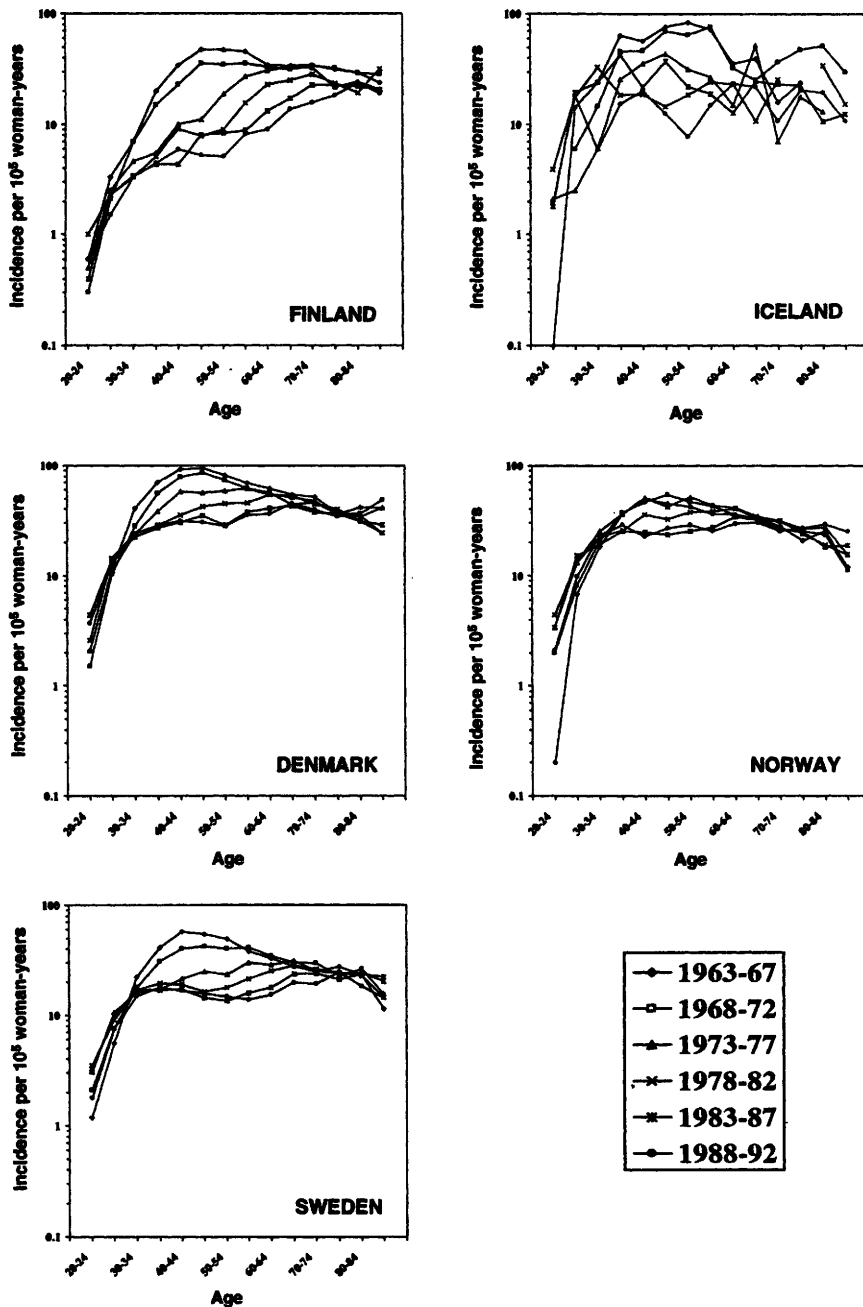
Figure 1. Cervical cancer incidence rates in the Nordic countries, 1958-1992. Whole female population, five-year period specific rates adjusted for age to the world standard population (Hakama, 1982; Engeland *et al.*, 1993; Parkin *et al.*, 1997)
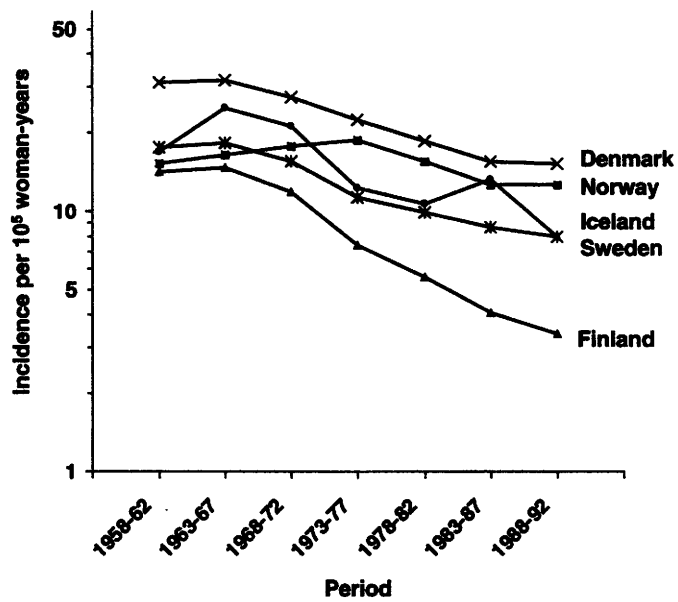
Figure 2. Age- and period-specific cervical cancer incidence rates in the Nordic countries, 1963-1992 (Engeland *et al.*, 1993; Parkin *et al.*, 1997)
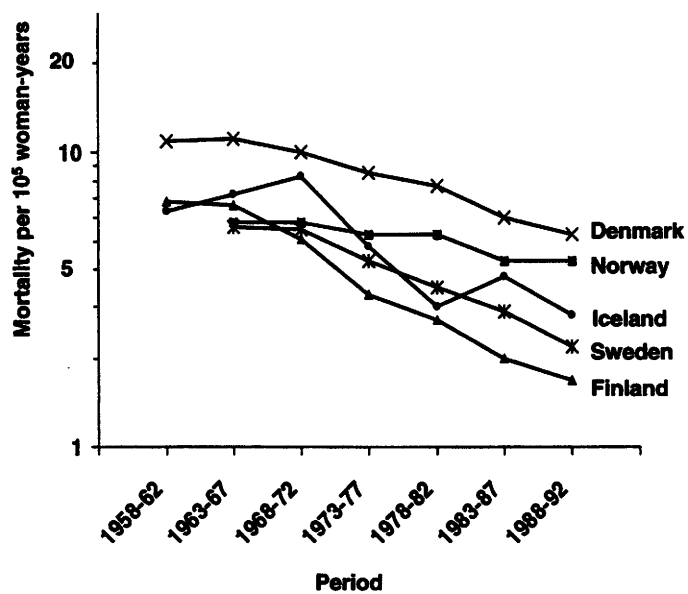
**Figure 3. Mortality rates of cervical cancer in the Nordic countries, 1958-1992. Whole female population, adjusted for age to the world standard population** (Läärä *et al.*, 1987; Hristova and Hakama 1997)

86

Discussion

We reviewed 27 studies on the impact of organised activities for cervical cancer screening. All but one of these studies showed clear decreases in cervical cancer incidence or mortality after introduction of Pap-screening activities. Evidence of the largest effects in terms of percent decrease in the age-adjusted cervical cancer incidence rates (68% to 78%) come all from areas with high coverage of smears taken in a centrally organised screening programme, i.e., Finland and Iceland, and British Columbia, Canada (in the latter the squamous cell carcinoma rates were reported). As these relatively large effects emerged rapidly, within few years, among women in the target ages of the organised activity, it is unlikely that any other factor would explain the findings. Some organisational differences between these most effective programmes, e.g., in the target ages and screening interval, apparently affect the overall effectiveness in a minor degree. Those differences may affect the efficiency in a greater extent. The impact of less centrally organised activities vary roughly from 10% to 60% depending on the period and length of follow-up, and age groups reported. All in all, the present-day evidence based on geographical correlations of the effect of large-scale organised Pap-screening programmes can be considered reassuring.

In addition to the organisational design in screening (considered the major determinant for the rapid reductions in incidence and mortality patterns) there are fluctuations which might also have affected the trends in a minor degree. Any increased intensity of a spontaneous activity, possibly simultaneously with the organised programmes, might contribute or modify part of the overall impact. Changes in sexual life, e.g., in age at first intercourse, in the number of sexual partners, or in the proportion of unprotected intercourses may relate to the prevalence of some potentially oncogenic infections (such as oncogenic HPV types). Changes in smoking habits among women might conceivably also affect the trends. If these risk factors have become more prevalent during the last few decades, the changes would favour an increase rather than decrease in the background risk. This could lead to underestimation of the overall screening effect in some areas (think, for example, highly urbanised areas in comparison with more remote areas).

Inadequacies in the coding and coverage of the registration of invasive cervical cancer incidence or mortality might have affected country-specific data, particularly concerning the effects of the early programmes. For example, a proportion of cervical cancer cases might have been coded to 'uterus, unspecified or NOS' in earlier years. The invasive tumours might not have been separated from the *in situ* tumours. There were efforts to correct for these kinds of errors, e.g., by re-analysing the diagnoses from hospital records in limited areas (Dickinson *et al.*, 1972; Cramer, 1974; Christopherson *et al.*, 1976). Concerning the present-day diagnostic accuracy and registration practice, one would expect only minor

87

differences between the developed countries. There are not many data available on direct comparisons of histological samples, and therefore potential differences in the diagnostic criteria may complicate international comparisons. For example, the age-adjusted incidence to mortality ratios appear to be somewhat higher in Denmark than in Finland, suggesting that the diagnostic criteria in incidence may differ between these countries. Nevertheless, the country-specific percent decreases have been the same whether estimated from incidence or mortality figures in these two countries.

There are differences in the prevalence of hysterectomised women between countries. In the US, where hysterectomy rates are higher than in most countries in Europe, cervical cancer incidence rates, if left uncorrected, have lead at least 10% errors in the trends since 1960s (Lyon and Gardner, 1977; Marrett, 1980). These errors vary between age groups, being most prominent in the postmenopausal age groups.

One of the main limitations in the evidence on how to attain the maximal effectiveness in Pap-screening programmes is that not much detail is available on the organisational, diagnostic, and clinical aspects in screening centres in relation to the trends. Diagnostic criteria in the cytological Pap-screening test or in the histological diagnosis of the precancerous lesions may differ greatly between different centres or countries. In those countries where there exist well-established screening and cancer registries and archives it should be a rather simple procedure to document the quality assurance, including a systematic audit of the screening histories, and to compare directly the diagnostic or clinical aspects between different areas or countries. These could add much new evidence on the overall effectiveness of the Pap-screening activities.

References

Anderson GH, Boyes DA, Benedet JL, LeRiche JC, Matisic JP, Suen KC, Worth AJ, Millner A & Bennett OM (1988). Organisation and results of the cervical cytology screening programme in British Columbia, 1955-85. *Br Med J* **296**:975-978

Anttila A, Pukkala E, Söderman B, Kallio M, Nieminen P & Hakama M (1999). Effect of organised screening on cervical cancer incidence and mortality in Finland, 1963-1995: A recent increase in cervical cancer incidence. *Int J Cancer* **83**:59-65

Ayre JE. The impact of cytology and cytogenetics upon gynecology and obstetrics (1964). *Obstet Gynecol Surv* **19**:799-837

Boyes DA, Worth AJ & Anderson GH (1981). Experience with cervical screening in British Columbia. *Gynecologic Oncology* **12**:S143-S155

Christopherson WM, Lundin FE, Mendez WM & Parker JE (1976). Cervical cancer control. A study of morbidity and mortality trends over a twenty-one-year period. *Cancer* **38**:1357-1366

Coleman D, Day N, Douglas G, Farmery E, Lynge E, Philip J & Segnan N (1993). European guidelines for quality assurance in cervical cancer screening. *Eur J Cancer* **29A** Suppl 4:S1-S38

Cramer DW (1974). The role of cervical cytology in the declining morbidity and mortality of cervical cancer. *Cancer* **34**:2018-2027

Dickinson L, Mussey E, Soule EH & Kurland LT (1972). Evaluation of the effectiveness of cytologic screening for cervical cancer. *Mayo Clin Proc* **47**:534-544

Ebeling K & Nischan P (1986). Organization and results of cervical cancer screening in the German Democratic Republic. In: Hakama M, Miller AB, Day NE (eds). *Screening for cancer of the cervix uteri* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, pp. 251-266

Engeland A, Haldorsen T, Tretli S, Hakulinen T, Hörte LG, Luostarinen T, Magnus K, Schou G, Sigvaldason H, Storm HH, et al. (1993). Prediction of cancer incidence in the Nordic countries up to the years 2000 and 2010. APMIS Suppl., **38**:1-124

Fernandez Garrote L, Lence Anta JJ, Cagezas Cruz E, Romero T & Camacho R (1996). Evaluation of the cervical cancer control program in Cuba. *Bull Pan Am Health Organ* **30**:387-391

Fouquet R & Gage H (1996). Role of screening in reducing invasive cervical cancer registrations in England. *J Med Screen* **3**:90-96

Gardner JW & Lyon JL (1977). Efficacy of cervical cytologic screening in the control of cervical cancer. *Prev Med* **6**:487-499

Gibson L, Spiegelhalter DJ, Camilleri-Ferrante C & Day NE (1997). Trends in invasive cervical cancer incidence in East Anglia from 1971 to 1993. *J Med Screen* **4**:44-48

Gustafsson L, Pontén J, Zack M & Adami HO (1997). International incidence rates of invasive cervical cancer after introduction of cytological screening. *Cancer Causes Control* **8**:755-763

Hakama M (1982). Trends in the incidence of cervical cancer in the Nordic countries. In: Magnus K (ed) *Trends in Cancer Incidenc: Causes and Practical Implications.* Washington, Hemisphere, pp. 279-292

Hakama M, Miller AB & Day NE (eds) (1986). *Screening for cancer of the uterine cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer

Hakama M, Chamberlain J, Day NE, Miller AB & Prorok PC (1985). Evaluation of screening programmes for gynaecological cancer. *Br J Cancer* **52**:669-673

Hakama M, Hakulinen T, Pukkala E, Saxén E & Teppo L (1982). Risk indicators of breast and cervical cancer on ecologic and individual levels. *Am J Epidemiol* **116**:990-1000

Hristova L & Hakama M (1997). Effect of screening for cancer in the Nordic countries on deaths, costs and quality of life up to the year 2017. *Acta Oncol* **36** Suppl 9:1-160

Lynge E (1983). Regional trends in incidence of cervical cancer in Denmark in relation to local smear-taking activity. *Int J Epidemiol* **12**:405-413

Lynge E (1998). Cervical cancer screening in Denmark: can cervical cancer screening be cost-effective? In: Ottesen B & Tabor A (eds). *New Insights in Gynecology and Obstetrics.* Lanos, Parthenon Publishing, pp. 251-256

Lynge E, Madsen M & Engholm G (1989). Effect of organised screening on incidence and mortality of cervical cancer in Denmark. *Cancer Research* **49**:2157-2160

Lyon JL & Gardner JL (1977). The rising frequency of hysterectomy: its effect on uterine cancer rates. *Am J Epidemiol* **105**:439-443

Läärä E, Day NE & Hakama M (1987). Trends in mortality from cervical cancer in the Nordic countries: Association with organised screening programmes. *Lancet* 1:1247-1249

Mählck CG, Jonsson H & Lenner P (1994). Pap smear screening and changes in cervical cancer mortality in Sweden. *Int J Gynecol Obstet* **44**:267-272

Marrett LD (1980). Estimates of the true population at risk of uterine disease and an application to incidence data for cancer of the uterine corpus in Connecticut. *Am J Epidemiol* **111**:373-378

Miller AB, Lindsay J & Hill GB (1976). Mortality from cancer of the uterus in Canada and its relationship to screening for cancer of the cervix. *Int J Cancer* **17**:602-612

Morgenstern H (1998). Ecologic studies. In: Rothman KJ, Greenland S (eds). *Modern epidemiology.* Philadelphia, Lippincott – Raven Publishers, pp. 459-480

Morrison AS (1998): Screening. In: Rothman KJ & Greenland S (eds). *Modern Epidemiology,* Philadelphia, Lippincott – Raven Publishers, pp. 499-518

Murphy MF, Campbell MJ & Goldblatt PO (1988). Twenty years' screening for cancer of the uterine cervix in Great Britain, 1964-84: further evidence for its ineffectiveness. *J Epidemiol Community Health* **42**:49-53

Nieminen P, Kallio M & Hakama M (1995). The effect of mass screening on incidence and mortality of squamous and adenocarcinoma of cervix uteri. *Obstet Gynecol* **85**:1017-1020

Parkin DM, Nguyen-Dinh X & Day NE (1985). The impact of screening on the incidence of cervical cancer in England and Wales. *Br J Obstet Gynaecol* **92**:150-157

Parkin DM, Whelan SL, Ferlay J, Raymo/nd L & Young J (eds) (1997). *Cancer Incidence in Five Continents,* Vol. VII. (IARC Scientific Publications No. 143), Lyon, International Agency for Research on Cancer

Quinn M, Babb P, Jones J & Allen E (1999). Effect of screening on incidence of and mortality from cancer of cervix in England: evaluation based on routinely collected statistics. *Br Med J* **318**:904-908

Sigurdsson K (1993). Effect of organised screening on the risk of cervical cancer. Evaluation of screening activity in Iceland, 1964-1991. *Int J Cancer* **54**:563-570

Sigurdsson K (1995). Quality assurance in cervical cancer screening: The Icelandic experience 1964-1993. *Eur J Cancer* **31A**:728-734

van der Graaf Y, Zielhuis GA & Vooijs GP (1988). Cervical cancer mortality in the Netherlands. *Int J Epidemiol* **17**:270-276

Table 1. Studies evaluated and their designs, methods, and main findings (Mählck *et al.*, 1994; Nieminen *et al.*, 1995; van der Graaf *et al.*, 1988)

Country/area, and the screening period studied	Screening coverage or extent [desired screening interval]	Screening options contrasted	Design for the correlation analysis	Follow-up period and source data for incidence (and for mortality data)	Cervical cancer incidence, cases per 100,000 woman-years; percent difference; ages included in the follow-up (information on mortality if available)	Reference
USA						
Olmsted County, Minnesota, 1947-1967	1-y coverage 0-7% in 1947-1956; 15-43% in 1957-1967; 7-y cov. <63% in 1960-1967 [1-y interval]	Expanding programme vs. no or low coverage screening	Trend, crude	1935-1967, incidence from medical records; (1935-1967, mortality records)	17.4 in 1947-1956; 9.4 in 1957-1967; -40% to -46%; all ages (similar effect in mortality)	Dickinson et al. 1972
US, seven cities, 1961-1968	1-y coverage 4.6-24.5% [1-y interval]	Expanding programme vs. no screening	Correlation coefficient & age-adjusted trend	Surveys in 1947 and 1969-1970, incidence (1950-1969, mortality records)	32-67 in 1947, 14-29 in 1969-1970; -54% to -66%; all ages[1]	Cramer 1974
Louisville, Jefferson County, Kentucky; 1956-1973	3-y coverage <58% in 1971-1973; lifetime coverage 90% [1-y interval]	Expanding programme vs. no screening	Age-adjusted trend	1953-1973, incidence from medical records (1953-1973, mortality records)	29.7 in 1953-1955, 12.7 in 1971-1973; -57%; ages 20+ (corresponding decrease in mortality)	Christopherson et al. 1976
Connecticut	No data given	(Any vs. no screening?)	Age-adjusted trend	1983-1987, incidence, compared to 1935-1949	17.7 in 1935-1949, 7.2 in 1983-1987; -59%; all ages	Gustafsson et al. 1997

Country/area, and the screening period studied	Screening coverage or extent [desired screening interval]	Screening options contrasted	Design for the correlation analysis	Follow-up period and source data for incidence (and for mortality data)	Cervical cancer incidence, cases per 100,000 woman-years; percent difference; ages included in the follow-up (information on mortality if available)	Reference
Canada						
10 provinces, and 147 census divisions or counties; 1962-1967	1-y coverage 0.6-23% to 13-41% [1-y interval?]	Expanding programme	Correlation coefficient and age-standardised trend	(1950-1952; 1960-1962; 1970-1972, mortality records)	(R=0.72 for percent decrease in mortality between 1960-1962 and 1970-1972, and screening rate in 1966; ages 30-64)	Miller et al. 1976
British Columbia, from 1949 onwards	1-y coverage up to 44%, lifetime up to 85% from 1970s and onwards [1-y interval]	Programme vs. no screening	Age-specific trend	1955-1985, sq. ca. incidence (1958-1985, sq. mortality)	28.4 in 1955, 6.4 in 1985; -78%; ages 20+ (-72% in mortality rate)	Boyes et al. 1981, Anderson et al. 1988
5 provinces: Alberta, Manitoba, Newfoundland, Quebec, Saskatchewan	No data given (1-y coverage 5-19% in 1962-1967, from Miller et al. 1976) [1-y interval?]	Programme vs. no/any screening?	Age-adjusted trend	1963-1966, 1983-1987, incidence	27.1 in 1963-1966, 11.1 in 1983-1987; -59%; all ages	Gustafsson et al. 1997
UK						
England and Wales, from early 1960s to 1980	700,000 smears in 1965; 2,900,000 in 1980 [3-5.5 -y interval?]	Expanding early programme	Age-adjusted trend	1963-1978, incidence	-26% to -58%; ages 35-54; corrected for CIN3 incidence; increase in registered incidence in age below 35	Parkin et al. 1985

93

Country/area, and the screening period studied	Screening coverage or extent [desired screening interval]	Screening options contrasted	Design for the correlation analysis	Follow-up period and source data for incidence (and for mortality data)	Cervical cancer incidence, cases per 100,000 woman-years; percent difference; ages included in the follow-up (Information on mortality if available)	Reference
Great Britain, 16 regions, from 1964 to 1984	Mean (annual?) smear rates 10%-25% [3-5.5 –y interval?]	Expanding early programme	Spearman's correlation coefficient	1963-1984, mortality records	Inverse correlations with screening effort and age-adjusted mortality; ages 15-64	Murphy et al. 1987
England, within 145 districts	5.5-y coverage 77%-80% in 1985-1991 [3-5.5 –y interval?]	Programme	Correlation coefficient, and adjusted RR[2]	1985-1987, incidence	No effect; –0.07 in ages 20-34; +0.07 in ages 35-64, per smear rate; positive correlation	Fouquet & Gage 1996
England, Birmingham	No data	Early programme vs. no/any screening?	Age-adjusted trend	1983-1986, incidence, compared to 1960-1966	+ 6%; all ages –14%; ages 35-64	Gustafsson et al. 1997
England, Scotland	No data	Early programme vs. no/any screening?	Age-adjusted trend	1983-1987, incidence, compared to 1963-1966	+ 7%; all ages –16%; ages 35-64	Gustafsson et al. 1997
East Anglia, 1971 onwards	5-y coverage 36-49% in 1988-1989, increased to 83-89% in 1992-1993 [3-5.5 y interval]	Activated[3] vs. early programme	Adjusted trend	1971-1993, incidence	–34% since 1990; ages 20-69	Gibson et al. 1997

Country/area, and the screening period studied	Screening coverage or extent [desired screening interval]	Screening options contrasted	Design for the correlation analysis	Follow-up period and source data for incidence (and for mortality data)	Cervical cancer incidence, cases per 100,000 woman-years; percent difference; ages included in the follow-up (information on mortality if available)	Reference
England, 1987 onwards	5-y coverage increased to 85% between 1988 and 1994 [3-5.5 y interval]	Activated[3] vs. early programme	Age-adjusted trend	1987 to 1995, incidence 1987 to 1997, mortality	15.6 in 1987; 10.3 in 1995; -34%, ages 20+ (mortality 6.1 in 1987; 3.7 in 1997; -39%; ages 20+)	Quinn et al. 1999
Nordic countries						
Finland, Iceland, Sweden, Denmark, Norway; from early 1960s onwards	Invitational coverage 100% at target ages (Fin, Ice, Swe), 40% (Den) and 5% (Nor) in early 1970s [Fin 5-y; others 2-4 y interval]	Programmes vs. no/ spontaneous screening	Age-adjusted trend	1943 (1958) - 1992, incidence from cancer registries (1953-1992 mortality records)	5-y average incidences 14.1 to 31.1 in 1963-1967; 3.4 to 15.2 in 1988-1992; -77% Fin, -68% Ice, -56% Swe, -52% Den, -22% Nor; all ages (similar decreases in mortality)	Hakama 1982, Läärä et al. 1987; further follow-up in: Engeland et al. 1993, Nieminen et al. 1995, Hristova & Hakama 1997, WHO 1997, Anttila et al. 1999

Country/area, and the screening period studied	Screening coverage or extent [desired screening interval]	Screening options contrasted	Design for the correlation analysis	Follow-up period and source data for incidence (and for mortality data)	Cervical cancer incidence, cases per 100,000 woman-years; percent difference; ages included in the follow-up (information on mortality if available)	Reference
Denmark, 22 counties; from 1960s onwards	Average annual number of smears 0.20 to 0.31 in 1974/75 [3-4 y interval]	Programme vs. no/ spontaneous screening	Age-adjusted trend	1943-1982, incidence and mortality records	-33% for organised screening activity, 5-y lapse; ages 30-59 (similar effects in mortality)	Lynge 1983, Lynge et al. 1989
Sweden, 24 counties; early 1960s onwards	Screening programme in operation in all counties by 1973 [3-4 y interval]	Programme vs. no/ spontaneous screening (score based on age, period, and county)	Age-adjusted trend	1962-1986, mortality records	Effect of the screening score variable = -53%; ages 30-59	Mählck et al. 1994
Iceland, screening programme started in 1964	3-y coverage of smears 83% in 1990-1992; [2-3 -y interval]	Programme vs. no/ spontaneous screening	Age-adjusted trend	1955-1993, incidence and mortality records	>25.0 in 1965-1969, <10.0 in 1990-1993; -62%; all ages	Sigurdsson 1993 & 1995
Finland, 450 municipalities, 1990-1995	5-y coverage of smears 90% in 1995; attendance rate 71% [5-y interval]	Attendance rate in municipality ≤66%, 67-79%, or ≥80%	Age-adjusted rate, 25-54 years old women	1990-1995, incidence records	RR=1.5 (95% CI 1.1-2.3), 1.3 (0.9-1.8) and 1.00 over the three categories in attendance rate	Anttila et al. 1999

96

Country/area, and the screening period studied	Screening coverage or extent [desired screening interval]	Screening options contrasted	Design for the correlation analysis	Follow-up period and source data for incidence (and for mortality data)	Cervical cancer incidence, cases per 100,000 woman-years; percent difference; ages included in the follow-up (information on mortality if available)	Reference
Other countries						
East Berlin, GDR; from 1973 to 1982	2-y coverage 74% in 1977-1978; 10-y cov. 90% in 1973-1982 [2-y interval?]	Programme vs. no/ expanding screening	Age-adjusted trend	1970-1972, 1980-1982, incidence and mortality records	49.7 in 1970-1972, 33.4 in 1980-1982; - 33% in ages 20+; - 58% in ages 40-49; (-37% in mortality rate)	Ebeling and Nischan 1986
The Netherlands, from 1970s onwards	Organised programme since 1976 [5-y interval]	Programme vs. no/early activities	Age-adjusted trend	1936-1983, mortality	-10% between 1960 to 1970, up to 50% decrease thereafter	Van der Graaf et al. 1988
Cuba, a nation-wide programme since 1968	1-y coverage 27% in 1993-1994 (2-y cov. <54%?); high amounts of unsatisfactory smears [2-y interval?]	Expanding programme	Trend, crude and age-adjusted	1980-1990, incidence and mortality records	Crude incidence 13.0 in 1980, 19.2 in 1990; +30%; ages 20+	PAHO 1996

[1] A weak positive correlation between one-year coverage and percent decrease (correlation coefficient 0.27 with mortality decrease in 50 states) were also reported.

[2] Mortality data taken partly from an earlier period than the screening data.

[3] A computerised call and recall system for invitations, new quality standards, and guidelines for the follow-up of women with abnormal smears were adopted.

97

Chapter 7. Cervix cancer: case-control studies on screening

M. Zappa and S. Ciatto

Introduction

Case-control studies (as well as other observational studies) had a crucial role in assessing the efficacy of Pap smear as no randomised controlled trial has ever been conducted. The case-control approach allows screening efficacy to be measured and the natural history of cervical cancer to be investigated. In the present paper we will describe the main characteristics and results of case-control studies published since 1986 (The IARC monograph on the evaluation of cytological screening was published in 1986 and we assume that after that publication screening efficacy was widely accepted) (Hakama et al., 1986). These studies were aimed at evaluating the effectiveness of specific screening programmes, essentially for monitoring purposes. The importance of these studies will be discussed, as well as their limits. We will also describe 'audit' studies of cervical cancer, an approach analysing the natural history of cases of invasive cancer, particularly useful in identifying the weak phases of a screening programmes: 'audit' studies do not use the case-control approach, but use a similar retrospective reconstruction of screening history. Finally, we describe a recent attempt to monitor the screening process by a case-control study which considers the different phases of the process as determinants of outcome: although such a methodology has its limits, the approach is quite interesting as it allows the identification of areas of practice susceptible to improvement.

Case-control studies: some methodological aspects

Case-control studies have been used in the past to evaluate screening efficacy, particularly for screening programmes for which randomised field trials had not been performed either for ethical or logistic reasons. The rationale of these studies is the comparison of the prior use of a screening test in two groups of subjects, namely (1) those who have suffered the adverse outcome that the screening is expected to prevent (cases), and (2) subjects sampled from the source population from which cases were drawn, whose screening history reflects that of the population (controls). Such a study design estimates the occurrence of an adverse outcome in subjects who had a screening test as compared to those who had not. The case-control approach has a number of advantages over other observational studies. First, screening history is not required for the whole population, but only

99

for cases (developing invasive cancer or dying of it) and controls. Furthermore, limiting the study to a relatively small number of subjects, collection of information on relevant confounding variables is also feasible.

Two study variants have been described, according to whether the screening test is aimed at detecting early stage cancers or precancerous lesions (Sasco *et al.*, 1986). In the case of cytological screening for cervix cancer (as well as screening for colorectal cancer), both outcomes may be considered: in fact screening is aimed at preventing the occurrence of invasive cervical carcinoma (CC) and thus at reducing mortality from CC, but it is also possible that death from CC may be prevented by detecting early and curable invasive CC. It is also possible that mortality reduction is determined by a better quality of medical care of screen-detected cases with respect to not screen-detected cases: it could be argued that this fact is also a favourable effect of screening, but the assumption that the benefit is due to screening could be biased. The validity of this type of study in the evaluation of screening has been the object of a wide debate (Morrison, 1992; Parkin and Day, 1985; Sasco *et al.*, 1986; Moss, 1991; Weiss, 1994; Gill and Horwitz, 1995; Cronin *et al.*, 1998).

Definition of cases and controls

Cases are subjects who experienced the adverse outcome. If the source population is population-based, cases are identified by means of a cancer registry or of a pathological archive collecting all cases occurring within a geographical boundary. If cases are identified from hospitals or clinics (hospital-based) the corresponding source population is defined as all the subjects who would attend that clinic if they had the disease of interest (hospital catchment population). The choice of hospital series may be convenient for practical reasons but it is a source of possible bias in the selection of controls.

If cases (as well as controls) are drawn only from a screening programme archive (nested case-control study within a cohort) the effect of screening could be underestimated, since being screened at a given date is generally correlated with previous screening attendance. This approach may be useful for evaluating the relative benefits of a previous smear performed at different times in the past.

If the adverse outcome of interest is death, some additional problems may occur. One is related to the quality of death certification. For example, in Florence District more than 30% of deaths from CC were coded as ICD-9 179 (= cancer of uterus not otherwise specified) (Barchielli *et al.*, 1990). Moreover, it is difficult to judge only from the death certificate if the woman died from or with CC. To avoid the latter bias it has been suggested that medical records should be reviewed in addition to death certificates and a panel of experts should be set up for blinded case review. (Weiss, 1994; Gill and Horwitz, 1995). The inclusion of screen-

100

detected cases tends to underestimate the benefit of screening since the effect of lead-time bias may lead to a surplus of such cases (Moss, 1991).

Controls must be an unbiased sample of the source population. In the selection of controls an equal chance of 'exposure' to screening must be ensured for cases and matched controls. If migration to and from the study area is relevant, it must be verified that cases and controls have been resident for the same period in that area (and/or received the same number of invitations to screening). In fact, a bias may arise if migrating people are at different baseline risk.

Definition of screening history

The definition of screening history presents several problems. All screening tests (negative or positive) must be included in the analysis of screening efficacy. Considering only negative tests makes sense only if the objective is to evaluate the negative predictive value of a screening test with respect to the adverse outcome (Weiss, 1994).

A method should be established to distinguish true 'screening' tests from tests performed for diagnostic purpose in symptomatic subjects, although it may not be easy to decide this retrospectively. If smears taken because of symptoms are included, this will lead to an underestimate of the screening benefit, as cases are more likely than controls to have had such tests. Within an organised programme, with active invitation, this is not a major problem as only smears performed as a consequence of an invitation may be considered.

When this information is not available (opportunistic screening and organised screening without invitation at a fixed date) some studies have tried to identify and exclude smears taken because of symptoms. Gill and Horwitz (1995) suggest that this uncertainty is addressed by using 'clinical' confidence intervals (a sort of sensitivity analysis), by varying the criteria for excluding possible diagnostic or follow up tests. Many case-control studies have excluded smears taken within 6 or 12 months before the date of diagnosis of matched case in an attempt to exclude diagnostic tests. This choice causes a bias of its own, as cases cannot have a true screening smear for the whole length of the excluded period, whereas controls can. If the preclinical invasive detectable phase is longer than the excluded period, then the effect of screening will be overestimated (Weiss, 1998). To minimise such a bias, some studies included only advanced carcinomas as cases (van der Graaf *et al.*, 1988) or estimated the length of the preclinical invasive phase and excluded tests performed within that interval (Berrino *et al.*, 1986). Unfortunately it is difficult to obtain a good estimate of that phase on either an individual or group basis. Matching screen detected cases with controls screened at the same date has been suggested (Moss, 1991). This is a good choice if we admit that the interval between tests is longer than the preclinical detectable phase.

101

Another problem arises when timing cancer diagnosis in relation to the availability of the screening test. Cases occurring immediately after the start of the programme may be useless to evaluate screening effectiveness, as the benefits of screening are expected only after several years. This is particularly true if the outcome is death.

Another problem is represented by the so-called 'healthy screenee bias' i.e. the potential bias in using the frequency of screening tests as measure of 'exposure' to screening. Indeed, a subject receiving multiple screening tests is necessarily disease free at several points in time, whereas cases had less opportunity to perform screening tests. To avoid such bias it has been proposed to use as 'exposure' the ever/never screened condition during a given time period (i.e. the period in which a smear can detect a precancerous lesion). If we are interested in the study of the predictive value of a negative test (i.e. how long the subject is protected after a negative test), we must take into account the relatively low sensitivity of cytology. After two negative tests the probability of being a false negative is quite low. In fact, we observed a considerably longer protection for subjects with two or more negative smears with respect of those with one smear only (IARC, 1986).

Comparability of information and comparability of populations

The comparability of information is always a relevant issue in retrospective studies. In particular a serious bias (recall bias) can arise if the recall of the screening history is correlated with the current status of the subject (case or control) - differential misclassification. As far as the screening history is concerned, this problem can be avoided using an independent archive (or administrative data) of screening tests. In this instance, however, the problem remains for the collection of confounding variables. If a complete screening archive is not available, studies on the validity of the collected information (by interview, questionnaire, etc.) must be undertaken.

Problems in 'comparability of population' (selection bias) may arise if non-attenders are *per se* at higher or lower risk of developing (and or dying of) CC as compared to attenders.

All the above mentioned methodological problems can be minimised if
(a) the screening programme under evaluation is centralised and an active invitation at a fixed date is used;
(b) a cancer registry (or pathological archive) as well as a complete archive of all invitations (for matching cases and controls to date of invitations) and all preventive smears (i.e. the tests performed as a consequence of a screening invitation) is available;
(c) the screening programme has been going on for some time.

102

Otherwise, the case-control approach must be handled with caution; a careful scrutiny of all design aspects as well as of result interpretation is needed.

The role of case-control studies in assessing the efficacy of cytological screening in reducing CC incidence

Case-control studies (as well as other observational studies) had a crucial role in assessing the efficacy of screening as no randomised controlled trial has ever been carried out. In 1986 the conclusions of the IARC Working Group on Cervical Cancer Screening were drawn from the overview of several studies performed in a number of countries with widely different approaches. Among them, the results of five case-control studies were analysed; two (Macgregor et al., 1985; Geirsson et al., 1986) carried out within organised programmes in Aberdeen and Iceland and three (Clarke and Anderson, 1979; Berrino et al., 1986; Raymond et al., 1984) from areas where screening was not centrally organised (in Toronto, Milan and Geneva respectively).

The studies in Iceland and Aberdeen were case-control studies nested within a cohort and were designed to investigate the reduction in risk of invasive CC among women with a previous negative smear, in terms of time elapsed since the smear was taken (Macgregor et al., 1985; Geirsson et al., 1986). In fact in such studies cases and controls were both derived from the screening archive and controls must have had their first negative smear prior to the date of diagnosis of the case. The combined analysis of two studies showed a relative protection (RP) of more than ten-fold (RP=11.1, 95% confidence intervals (95% CI) 2.4-52.2) for women with their last negative smear performed 0-11 months before the diagnosis of the case as compared to women who had had their last negative smear ten years or more before. The RP remained significant if a smear was performed within 71 months (RP=2.0 95% CI 1.1-3.7) The number of previous negative smears increased the RP by approximately 1.5-fold. No clear effect of age was observed.

The other three case-control studies were addressed to evaluate the effectiveness of cytological screening for invasive CC. Differences existed among these studies in the criteria for case and control selection and in the definition of screening history. Nevertheless, such differences did not affect the results. The odds ratios (OR) for ever vs. never screened tended to be similar in these studies, ranging from 0.26 to 0.37. Furthermore, the effect was similar to that observed in the other studies when time elapsed since the last screening and number of previous tests were analysed.

Case-control studies published after the IARC monograph

Since the publication of the IARC monograph (Hakama *et al.*, 1986) on Pap smears efficacy in 1986, several case-control studies have been carried out within a process of evaluating current screening programmes. In the present review only studies published after 1986 and aimed at evaluating screening effectiveness are considered. Three were carried out in North America: Manitoba, Canada (Cohen, 1993), Maryland, USA (Celentano *et al.*, 1988), and Washington, USA, (Shy *et al.*, 1989); two in Central America: Panama and Costa Rica (Herrero *et al.*, 1992) and Mexico city, Mexico (Hernandez-Avila *et al.*, 1998); four in Asia: Thailand (Wangsuphachart *et al.*, 1987), Jingan, China (Zhang *et al.*, 1989); Osaka, Japan (Sobue *et al.*, 1990), Miyagi, Japan (Sato *et al.*, 1997); four in Europe: Nijmegen, The Netherlands (van der Graaf *et al.*,1988), Denmark (Olesen, 1988), Florence, Italy (Palli *et al.*, 1990), Aberdeen, Scotland (Macgregor *et al.*, 1994), United Kingdom (Sasieni *et al.*, 1996). Table 1 summarises the main characteristics of these studies.

About half of them were carried out within organised programmes (i.e. programmes with an active invitation of women). In all studies CC incidence was chosen as the outcome. Two studies also considered in separate analysis CC mortality. Most studies limited the selection of CC to the squamous subtype, whereas two studies made a separate analysis for squamous- or adenocarcinomas (Herrero *et al.*, 1992; Sato *et al.*, 1997). In some studies the histological type of CC was not specified. In some studies attention was paid to the stage of CC and/or only advanced stages were considered (van der Graaf *et al.*, 1988; Zhang *et al.*, 1989) or separate analysis was carried out for different grades of invasion (Herrero *et al.*, 1992). In fact, we would expect the strongest protective effect for advanced carcinomas. About half of the studies were designed on a population basis, whereas the others were hospital-based.

In one study (Herrero *et al.*, 1992) controls were selected partly from population census lists and partly from hospital admission lists; in other studies (Zhang *et al.*, 1989; Sato *et al.*, 1997) cases and controls were selected from screening archives.

In population-based studies, the source of CC cases was the local cancer registry or archives, which were supposed to collect all the cases occurring in a given area. In hospital-based studies the source of information on cases were the hospital admission records.

Table 2 summarises the main results of these studies. The proportion of controls 'ever screened' give us an overview of the coverage of Pap smear practice in the general population (although it must be considered that in most studies controls were matched to cases for several covariates so that the actual coverage cannot be directly estimated from such figures). The comparison of the proportion of ever-

104

screened controls with the proportion of recently screened controls (in general the last three years, if not otherwise specified) gives us an idea of how regularly Pap smears are performed. In fact, in a situation where all subjects have a Pap smears every three years (or more frequently) these values would tend to be similar. It is worth noting that these data differ substantially from one study to another. In fact, the proportion of ever-screened controls ranges from 20 % in Osaka, and 37% in Bangkok to 88% in Miyagi and 93% in Maryland. Such a finding is important when evaluating the result of a case-control study; in fact, if the proportion of never screened is small (or large) the problem of comparability of populations becomes less trivial. The observed differences can be partly explained by different attitudes in different periods and geographical areas, as well as by different criteria in selection of controls or in collecting screening histories. Most studies tried to identify and exclude smears taken because of symptoms by means of exclusion of smears taken within 6 or 12 months before the index date. In some studies, controls were selected from subjects with a negative smear at the date of diagnosis of matched case (Macgregor et al., 1994); in another study (Sobue et al., 1988) controls for screen-detected cases were selected among subjects with a negative smear taken the same year of the index date; otherwise controls were subjects with no smear during the same year.

In spite of the above listed differences (eligibility criteria for cases and controls, modality of screening history collection, adjustments for confounding variables), the results are quite similar (with respect to reduction in risk of invasive CC), ORs ranging from 0.27 in the Danish study to 0.43 in the Canadian study. Some studies are out of this range, with lower risks in the studies in Miyagi, Japan and Jingan, China (OR=0.16) and the higher risk observed in Mexico City (OR=0.76). In the latter study the OR becomes 0.38 (95% CI= 0.28-0.52) when only smears performed without gynaecological symptoms are considered.

When death from CC is assumed as the end point, the protective effect of screening tends to be slightly higher. In fact, in the Scottish experience the OR for mortality was 0.25 (95% CI 0.11-0.48) and for incidence 0.35 (95% CI 0.25-0.50), and in the Osaka study 0.22 (95% CI 0.03-1.95) and 0.41 (95% CI 0.13-1.29) respectively. Of course, a direct comparison with incidence is not possible as these figures refer to different groups of women, but nevertheless such results are to be expected if we assume a beneficial effect of early detection of invasive CC.

The combined evidence from these studies confirms the IARC working group statement on screening effectiveness in reducing CC incidence and mortality and emphasises the strength of the association between exposure to Pap smears and (reduced) risk of developing CC.

Do case-control studies have a role in monitoring screening programmes?

The above mentioned case-control studies are focused on the epidemiological relationship between screening history and invasive CC incidence and/or mortality. However, it is worth noting that these studies answered also some other interesting questions. The first concerns the magnitude of the protective effect of screening from cancer other than squamous cervical carcinoma, i.e. adenocarcinoma or mixed cancer. The first results (Herrero et al., 1992, Sato et al., 1997) seem to indicate a lower impact (but anyway an impact; OR= 0.5; 95% CI 0.3-0.8 and OR=0.45; 95% CI 0.1-3.7 respectively) for adenocarcinoma as compared to squamous cell carcinoma. Also in the Danish study (Olesen, 1988) a reduction of OR was observed when only squamous cell carcinoma patients and their controls were analysed. When evaluating these results one must keep in mind the low statistical power of such analysis, as adenocarcinomas are relatively rare. A recent large case-control study carried out in Victoria, Australia (Mitchell et al., 1995) on the relationship between negative cervical cytology and risk of adenocarcinoma (113 cases and 452 controls) showed a very modest benefit from screening. Again, when evaluating this result, one must consider that endocervical brushes were introduced into routine use in that area only after 1989, and that most cases in the study occurred before that date. Further studies are needed, but anyway the finding of a lower protection of cervical cytology for adenocarcinomas may be explained by problems in cell sampling and by difficulties in cytological assessment. The evaluation of the screening history of such carcinomas may justify efforts to improve specific screening (sampling) techniques.

The second issue concerns the role of risk factors (in particular human papilloma viruses (HPV) type 16-18) for CC in screening policies. The presence of HPV does not modify the protective effect of screening because virtually all cancers are HPV positive (Herrero et al., 1999). Only one of the above mentioned studies (Herrero et al., 1992) has considered the presence of HPV. In that study a smear was obtained from all control women. Although details are not given in the paper, no effect modification was observed according to the presence of HPV infection. However, a different screening policy (with shorter screening intervals for women with HPV and longer intervals for HPV negative women) might be proposed in order to increase the efficacy of the programme.

To date the protective effect of a well conducted screening programme is totally accepted and further confirmations are of limited interest for the evaluation of specific screening programmes. On the other hand, serious methodological problems exist with the case-control approach and even if recommended methods were followed, several difficulties arise in practice. Thus results obtained in different situations, with different methodological assumptions, may be not easily

comparable. Probably in large areas where screening programmes are not centralised and most smear taking is opportunistic, this approach may still play an important role. It could provide information on the coverage, stratified by different covariates. Furthermore, if an unexpectedly high OR were observed, it could indicate (as in the study carried out in Mexico City) problems in quality control of smear taking, smear reading or in other phases of the screening process. In this perspective, it is important to obtain a population-based recording of CC cases and a cancer registry or a pathological archive must be available. For monitoring and evaluating a screening programme the analysis of trends in incidence and mortality are essential. But, particularly in small areas, these trends may be difficult to interpret due to statistical fluctuations of figures. Furthermore, such analysis could become quite complicated if we suspect a spontaneous, not screen-related change in baseline risk of CC, as suggested by some recent studies (Macgregor *et al.*, 1994, Cecchini *et al.*, 1995).

For monitoring a screening process several issues must be considered. The occurrence of CC (any incident CC should be regarded as a potential failure of the programme) could be the consequence of several factors, namely:

a) Inadequate coverage of the population
   - non-attendance
   - too long interval since last smear (more than five years might be the cut-off?)
b) False negative cytology
   - inadequate smear taking
   - inadequate smear processing (fixation and staining)
   - errors in smear interpretation
c) Inadequate follow-up of abnormal smears
   - requiring repeat smear
   - requiring colposcopy
d) Inadequate colposcopy and/or biopsy
e) Inadequate treatment

To monitor all these aspects, a list of process indicators (with the relative reference standards) has been proposed (Coleman *et al.*, 1993). The most important indicators to control for are the coverage of the population and the screening history of CC cases. With respect to the latter, there have been several investigations. In the Florence (Italy) experience (Ciatto *et al.*, 1993) 42 CC cases were considered. No or irregular attendance (negative smear-to-diagnosis interval >5 years) was recorded in 25 women (59%); in seven (17%) cases assessment or treatment inadequacy was found; false negative smears by the local screening programme were recorded in four cases (10%). In the Swedish experience (Stenkvist and Söderström., 1996), 38 cases reported as CC by the central Cancer Registry were examined. Out of them, 11 cases had not been classified correctly as

invasive squamous carcinomas due to data transfer mistakes. Moreover, five cases were revised as carcinoma *in situ* after double blind examination of the specimens. After linkage with the computerised archive 55% of patients with invasive squamous carcinomas had no previous smear, 23% had had only one smear more than five years before the diagnosis, and 36% had had no diagnostic work-up after an abnormal smear. In Scotland (Macgregor *et al.*, 1994), 83% (234/282) of invasive cancers occurred in unscreened or inadequately screened women; an inadequate smear interpretation was found in 32 (11%) slides, and recurrence after treatment at colposcopy occurred in eight women (3%); for nine women (3%) information was not available. In the Icelandic study (Sigurdsson, 1995) among 120 women aged 25-64 registered at the Cancer Registry with invasive cervical cancer during the period 1980-1989, 49% were screen detected, 21% were interval cancers (i.e., cancers occurred within three years of a negative tests) and 30% had never been screened. For squamous and adeno/adenosquamous carcinomas, these percentages were 54%, 15%, 31% and 29%, 46%, 25% respectively. In the experience of Connecticut (Janerich *et al.*, 1995), the screening history of 481 women with CC was ascertained by means of direct interview to the patient or physician. 28.5% of CC occurred among women who had never been screened, and another 23.5% occurred among women whose last Pap test was taken more than five years before diagnosis, 7% had a misread negative smear and 9% were not followed properly. These results give us relevant information about the weak points of the programmes but do not tell us the weight of each cause of failure in terms of preventable fraction.

An interesting monitoring modality has been established recently in England (Sasieni *et al.*, 1996). The screening histories of all 348 women diagnosed with invasive CC (included 89 microinvasive cases - FIGO stage1A) in 24 Districts of England, Wales and Scotland in 1992 were examined and compared with those of 677 age- and residence-matched controls. Cases were obtained from local pathology laboratories. Controls were randomly chosen from the computerised registry held by the local family health services authority. Women with previous hysterectomy were excluded. The screening histories were determined, for cases and controls, from the same computerised archive. The analysis excluded (in both cases and controls) any smear performed 6 months before the case's diagnosis. A conditional logistic regression was made and the population attributable risk was calculated using the techniques appropriate to matched case-control data (Kuritz and Landis, 1988). Forty one percent of invasive CC cases vs. 30% of microinvasive and 24% of controls had never had a Pap smear or had no test within five years. As far as the adequacy of follow up is concerned, 13 % of all CC cases (17% among microinvasive and 11% among invasive) had screening histories indicative of inadequate follow up of smears requiring colposcopy. The same percentage among controls was 1%. Analysis using this method of the adequacy of treatment of precancerous lesions is in progress. The interesting aspect of this study is that it is oriented towards monitoring (rather than evaluation) and several

determinants of cancer are taken into consideration (not only screening yes or no) such as the quality of cytology, the adequacy of follow-up and treatment. More generally, this kind of approach allows identification of areas of current practice for improvement.

Some problems emerged from this study: the authors do not discuss selection bias and, as mentioned above, the choice of excluding smears performed within six months of diagnosis is questionable. Furthermore, the statistical power of the study should have been taken into account for the precision of the estimate, e.g., we would have expected quite a low percentage inadequately treated or followed among controls.

The role of cancer registry in monitoring cervical screening programme

In order to monitor a screening programme, it is essential to have:

- a complete registration of all CC cases which occurred in the target population;
- a complete archive of all performed preventive smears and all further assessment performed after a positive test in the target population;
- a valid key for linking the two archives.

Cancer registries provide a universal registration of all cases occurring in a defined area and almost all cases occurring in the target population (if strong migratory effect is not present) will be available. The importance of the universality of the registration must be stressed when monitoring screening programmes. The correlation between type of hospital where women are admitted for diagnosis and socio-economic status is well known; and the latter could be correlated with screening history, so that a bias could occur if only a selected proportion of incident cases is considered. Furthermore, interval cancers (i.e. cancers occurring some years after a negative test) or cancers occurring after inadequate colposcopy or treatment, may be referred to gynaecological practices different from those for cases managed by the screening authority. Such types of cancers are of key importance in evaluating screening activity, and any lack of information of the screening history of such women deprives us of a lot of information on the weak points of the programme. A cancer registry collects all cases occurring among the residents of a given area. From a public health point of view, a screening programme should be addressed to all subjects living in that area. As a matter of fact, especially in the great metropolitan areas, an increasing proportion of the population has no legal residence (students, temporary workers, immigrated people and so on). It is suggested that storage (in a separate file from that used for

calculating incidence rates) of the information about all cancers diagnosed in the area (resident or not) would provide better monitoring of all screened people.

One problem is common when a cancer registry is used as a database for monitoring screening: cancer registry data should be available within a relatively short time after incidence date. Otherwise if, for example, data on interval cancers were available only several years after their occurrence, retrieving negative slides for slide review might be rather difficult and the consequent corrective action could be ineffective.

Delay in reporting official figures is a problem in some cancer registries. In such a situation, the development of processes whereby the pathology archives can be rapidly utilised, to provide a raw list of all incident CC cancers is strongly recommended. The histological subtyping of CC is important, since the protective effect of screening differs for squamous or adenocarcinomas, and this information should be available from cancer registries. Nevertheless, misclassification and misinterpretation have been reported (Sigurdson *et al.*, 1995).

Finally, problems may occur in recording of stage. In fact, microinvasive CC (FIGO stage 1A) cannot be considered as a failure of the programme since its detection is often a consequence of a preventive smear and may be regarded as early detection, and it is associated with extremely high cure rates. Information on histologically confirmed cervical intraepithelial neoplasia (CIN) is not directly important in evaluating the effectiveness of a screening programme. The occurrence of CIN is highly dependent on features of the screening programme, such as referral rates for colposcopy, and the colposcopy directed biopsy rate (CIN are asymptomatic lesions), as well as on screening frequency. Nevertheless, their collection by cancer registries could be useful if an archive of screening smears is also available. In fact, the detection rates of CIN2 or more severe lesions in women at their first smear give us an estimate of the baseline risk of the disease and allow a better analysis of trends in CC incidence and mortality.

References

Barchielli A, Buiatti E & Carli S (1990). Mortalitá per collo dell'utero nella provincia di Firenze negli anni 1985-1987. *Epidemiol Prev* 12:33-38

Berrino F, Gatta G, d'Alto M, Crosignani P & Riboli E (1986). Efficacy of screening in preventing invasive cervical cancer: a case-control study in Milan. IARC Working Group on Evaluation of Cervical Cancer Screening. In: Miller AB, Hakama M & Day NE (eds) *Screening for Cancer of the Uterine Cervix* (IARC

Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, pp. 111-123

Cecchini S, Ciatto S, Zappa M & Biggeri A (1995). Trends in the prevalence of cervical intraepitelial neoplasia grade 3 in the district of Florence, Italy. *Tumori* **81**:330-333

Celentano DD, Klassen AC, Weisman CS & Rosenshein NB (1988). Cervical cancer screening practices among older woman: result from the Maryland Cervical Cancer Case-Control Study. *J Clin Epidemiol* **41**:531-541

Ciatto S, Grazzini G, Cecchini S & Iossa A (1993). Screening history of incident cases of invasive carcinoma of the cervix. Florence district 1988-1989. *Tumori* **79**:311-313

Clarke EA & Anderson TW (1979). Does screening by 'Pap' smears help prevent cervical cancer? A case-control study. *Lancet* **2**:1-4

Cohen MM (1993). Using administrative data for case-control studies: the case of the Papanicolau smear. *Ann Epidemiol* **3**:93-98.

Coleman D, Day N, Douglas G, Farmery E, Lynge, E, Philip J & Segnan N (1993). European guidelines for quality in cervical cancer screening. Europe Against Cancer Programme. *Eur J Cancer* **29A**:S1-S38

Cronin KA, Weed DL, Connor RJ & Prorok PC (1998). Case-Control Studies of Cancer Screening: Theory and Practice. *J Natl Cancer Inst* **90**:498-504

Geirsson G, Kristiansdottir R, Sigurdsson K, Moss S & Tulinius H (1986). Cervical cancer screening in Iceland: a case-control study. IARC Working Group on Evaluation of Cervical Cancer Screening. In: Hakama M, Miller AB & Day NE (eds) *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, pp. 37-41

Gill TM & Horwitz RI (1995). Evaluating the efficacy of cancer screening: clinical distinctions and case-control studies. *J Clin Epidemiol* **48**:281-292

Hakama M, Miller AB & Day NE (eds) (1986). *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer

Hernandez-Avila M, Lazcano-Ponce EC, Alonzo de Ruiz P & Romieu I (1998). Evaluation of the cervical cancer screening programme in Mexico: a population-based case-control study. *Int J Epidemiol* **27**:370-376

111

Herrero R, Brinton LA, Reeves WC, Brenes MM, de Britton RC, Gaitan E & Tenorio F (1992). Screening for cervical cancer in Latin America: a case-control study. *Int J Epidemiol* **21**:1050-1056

Herrero R, Hildesheim A, Bratti C, Sherman ME, Hutchinson M, Morales J, Balmaceda I, Greenberg MD, Alfaro M, Burk RD, Wacholder S, Plummer M & Schiffman M (2000). A population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. *J Natl Cancer Inst* **92**:464-474

IARC Working Group on Evaluation of Cervical Cancer Screening Programmes (1986). Screening for squamous cervical cancer: the duration of low risk after negative result of cervical cytology and its implication for screening policies. *Br Med J* **293**:659–664

Janerich DT, Hadjimichael O, Schartz PE, Lowell DM, Wister Meigs J, Merino MJ, Flannery JT & Polednak AP (1995). The screening histories of women with invasive cervical cancer, Connecticut . *Am J Public Health* **85**:791-794

Kuritz SJ & Landis JR (1988). Attributable risk estimation from matched case-control data. *Biometrics* **44**:355-367

Macgregor JE, Moss SM, Parkin DM & Day NE (1985). A case-control study of cervical cancer screening in north east Scotland. *Br Med J* **290**:1543-1546

Macgregor JE, Campbell MK, Mann EMF & Swanson KY (1994). Screening for cervical cancer in north east Scotland shows fall in incidence and mortality from invasive cancer with concomitant rise in pre-invasive disease. *Br Med J* **308**:1407-1411

Mitchell H, Medley G, Gordon I & Giles G (1995). Cervical cytology reported as negative and risk of adenocarcinoma of the cervix: no strong evidence of benefit. *Br J Cancer* **71**:894-897

Morrison AS (1992). *Screening in Chronic Disease.* 2nd edition. New York, Oxford University Press

Moss SM (1991). Case-Control Studies of Screening. *Int J Epidemiol* **20**:1-6

Olesen F (1988). A case-control study of cervical cytology before diagnosis of cervical cancer in Denmark. *Int J Epidemiol* **17**:501-508

Palli D, Carli S, Cecchini S, Venturini A, Piazzesi G & Buiatti E (1990). A centralised cytology screening programme for cervical cancer in Florence. *J Epidemiol Community Health* **44**:47-51

Raymond L, Obradovic M & Riotton G (1984). A case-control study to estimate the detection of cancer of the cervix uteri by cytology. *Rev Epidemiol Sante Publique* **32**:10-15

Report by the Evaluation Committee (1989). Population screening for cervical cancer in The Netherlands. *Int J Epidemiol* **18**:775-781

Sasco AJ, Day NE & Walter SD (1986). Case-control studies for the evaluation of screening. *J Chron Dis* **39**:399-405

Sasieni PD, Cuzik J & Lynch-Farmery E, and the National Co-ordinating Network for Cervical Screening Group (1996). Estimating the efficacy of screening by auditing smears histories of women with and without cervical cancer. *Br J Cancer* **73**:1001-1005

Sato S, Makino H, Yajima A & Fukao A (1997). Cervical cancer screening in Japan. A case-control study. *Acta Cytol* **41**:1103-1106.

Shy K, Chu J, Mandelson M, Greer B & Figge D (1989). Papanicolau smear screening interval and risk of cervical cancer. *Obstet Gynecol* **74**:838-843

Sigurdsson K (1995). Quality assurance in cervical cancer screening: The Iceland experience 1964-1993. *Eur J Cancer* **31A**:728-734

Sobue T, Suzuki T, Fujimoto I, Yokoi N & Naruke T (1990). Population-based case-control study on cancer screening. *Environ Health Perspect* **87**:57-62

Stenkvist B & Söderström J (1996). Reasons for cervical cancer despite extensive screening. *J Med Screen* **3**:204-207

van der Graaf Y, Zielhuis GA, Peer GM & Vooijs PG (1988). The effectiveness of cervical screening: a population-based case-control study. *J Clin Epidemiol* **41**:21-26

Wangsuphachart V, Thomas DB, Koetsawang A & Riotton G (1987). Risk factors for invasive cervical cancer and reduction of risk by Pap smears in Thai women. *Int J Epidemiol* **16**:362-366

Weiss NS (1994). Application of the case-control method in the evaluation of screening. *Epidemiol Rev* **16**:102-108

Weiss NS (1998). Analysis of case-control studies of the efficacy of screening for cancer: how should we deal with tests done in persons with symptoms? *Am J Epidemiol* **147**:1099-1102

Zhang ZF, Parkin DM, Yu SZ, Estève J, Yang XZ & Day NE (1989). Cervical screening attendance and its effectiveness in a rural population in China. *Cancer Detect Prev* **13**:337-342

Table 1. Main characteristics of case-control studies on cervical cancer screening (Studies published after 1986)

Area/author/date of publication	Outcome	Type of screening programme	Cases diagnosed in	Number and source of cases	Number and selection of controls
Bangkok, Thailand *Wangsuphachart et al. 1987*	Incidence	Not organised	1979-83	189, aged 15-54 All histological types Hospital records	1023, subjects admitted in the same hospital.
Denmark *Olesen, 1988*	Incidence	Organised	1983	428 All histological types Average age 52.6 years Cancer registry	428, matched for age and GP's practice.
Nijmegen, The Netherlands *Van der Graaf et al. 1988*	Incidence	Organised	1979-85	36, aged <70 Advanced CC (FIGO >1A) Cancer registry	120, married or ex married, matched for age and districts. From registrar's office of the community.
Maryland, USA *Celentano et al. 1988*	Incidence	Not organised	1982-84	153, aged 22-84 Hospital admission records	153, matched for age, race, and residence. Identified through case nomination or neighbourhood canvassing.
Osaka, Japan *Sobue et al. 1990*	Incidence	Organised	1965-87	28, aged 30-79 Cancer registry	272, alive, matched for age and dwelling history. If case screen-detected or with symptoms, controls with negative screen the same year, otherwise control without screening the same year.
Washington, USA, *Shy et al. 1989*	Incidence	Not organised	1979-1983	92, aged 31-75 Symptomatic at diagnosis All histological types Advanced CC (FIGO >1B-occult). Cancer registry	178, random-digit dialing matched for age. Women who had not had hysterectomy.
Jingan, China *Zhang et al. 1989*	Incidence	Organised	1965-1974	109, aged 35-85 Advanced squamous cell CC (FIGO >1A) Screening archive	545, matched for age, residence and attendance at the same screening round at which the case was diagnosed. Screening archive.

115

Area/author/date of publication	Outcome	Type of screening programme	Cases diagnosed in	Number and source of cases	Number and selection of controls
Osaka, Japan *Sobue et al. 1990*	Mortality	Organised	1965-87	15, aged <80 dead from CC Cancer Registry	150, alive matched for age and dwelling history.
Florence, Italy *Palli et al. 1990*	Incidence	Organised	1982-1985	191, aged < 75 Cancer registry	540, matched for age and residence. Residents list.
Bogota, Mexico City, Panama, Costa Rica *Herrero et al. 1992*	Incidence	Not organised	1986-87	759, aged <70 Cancer treatment centres	1433, most from hospital admission list, partly from census list.
Manitoba, Canada *Cohen 1993*	Incidence	Not organised	1981-84	415, aged 25-64 Cancer registry	29269, random sample from residents list.
South east Scotland, UK *Macgregor et al. 1994*	Incidence	Organised	1982-91	282, squamous CC	564, matched for age, with a negative smear at the date of diagnosis of case.
South east Scotland, UK *Macgregor et al. 1994*	Mortality	Organised	1982-91	108, dead from CC	216, matched for age, with a negative smear at the date of diagnosis of case.
United Kingdom *Sasieni et al.1996*	Incidence	Organised	1992	348, aged >20 Pathology laboratories	677, matched for age. For each case one control was registered with the same GP. Excluded women who had had hysterectomy. Registry of local health authority.
Miyagi, Japan *Sato et al. 1997*	Incidence	Organised	1984-89	119, aged 35-79 Screening Archive	218, matched for age, time and place of screening. From screening archives.
Mexico City, Mexico *Hernandez-Avila et al. 1998*	Incidence	Not organised	1990-92	397 Hospital admission records	1005, sample of residents, age stratified, aged 25-80, residents for >2 years

Table 2. Main results of case-control studies on cervical cancer screening (Studies published after 1986)

Area/author/date of publication	Proportion of cases ever screened	Proportion of controls ever screened	Proportion of cases screened within 3 years	Proportion of controls screened within 3 years	OR ever vs. never (95% CI)	Definition and source of information on screening history
Bangkok, Thailand *Wangsuphachart et al. 1987*	30%	37%	NA	NA	0.39[1] (0.21 - 0.74)	Smears taken >6 months before diagnosis for case or hospitalisation for control. Questionnaire.
Denmark *Olesen 1988*	45%	67%	28%	42%	0.27 (0.18 - 0.42)	Smears taken >6 months before suspicion of CC. Questionnaire to GPs.
Nijmegen, The Netherlands *van der Graaf et al. 1988*	47%	68%	NA	NA	0.22 (0.1 - 0.81)	Preventive' smears >12 months before diagnosis. Questionnaire.
Osaka, Japan *Sobue et al. 1988*	25%	39%	NA	NA	0.41 (0.13 - 1.29) (incidence)	Smears taken within 10 years of diagnosis excluding the year of diagnosis. Only negative smears. Screening archive.
Osaka, Japan *Sobue et al. 1988*	7%	20%	NA	NA	22 (0.03 - 1.95) (mortality)	Smears taken within 10 years of diagnosis, including diagnostic test. Screening archive.
Maryland, USA *Celentano et al. 1988*	72%	93%	37%	72%	0.29[2] (0.15 - 0.58)	Smears taken >12 months before diagnosis. Personal interview for cases, phone interview for controls.
Washington, USA *Shy et al. 1989*	85%	93%	61%	79%	0.21[3] (0.09 - 0.50)	Smears collected in the follow-up of an abnormal test or at the diagnosis of CC were excluded.Telephone interview.
Jingan, China *Zhang et al. 1989*	NA	NA	NA	NA	0.16[4] (0.05 - 0.58)	Smears taken at the time of diagnosis were excluded.
Florence, Italy *Palli et al. 1990*	19%	48%	7%	21%	0.29 (0.15 - 0.55)	Smears taken >6 months before diagnosis. Screening archive.
Bogota, Mexico City, Panama, Costa Rica *Herrero et al. 1992*	50%	72%	32% (within 4 years)	54% (within 4 years)	0.40 (0.31 - 0.48)	Smears taken >12 months before diagnosis. Personal interview.

117

Area/author/date of publication	Proportion of cases ever screened	Proportion of controls ever screened	Proportion of cases screened within 3 years	Proportion of controls screened within 3 years	OR ever vs. never (95% CI)	Definition and source of information on screening history
Manitoba, Canada *Cohen 1993*	76 % (within 10 years)	87% (within 10 years)	67% (within 5 years)	74% (within 5 years)	0.43[5] (0.32 - 0.57)	Smears taken >12 months before diagnosis. Computerised health care files.
South east Scotland, UK *Macgregor et al. 1994*	45%	73%	14%	23%	0.35 (0.25 - 0.50) (incidence)	Smears taken before the presentation of the case. From cytopathology database.
South east Scotland, UK *Macgregor et al. 1994*	35%	73%	15% (within 5 years)	42% (within 5 years)	0.25 (0.11 - 0.48) (mortality)	Smears taken before the presentation of the case. From cytopathology database.
United Kingdom *Sasieni et al. 1996*	73%	85%	19%	48%	0.26[6] (0.14 - 0.47)	Smears taken >6 months before diagnosis. Screening archive.
Miyagi, Japan *Sato et al. 1997*	55%	88%	40%	80 %	0.16[7] (0.09 - 0.28)	Smears taken before the presentation of the case. Personal interview and screening archive.
Mexico City, Mexico *Hernandez-Avila et al. 1998*	42%	51%	21% (within 4 years)	27% (within 4 years)	0.76[8] (0.59 - 0.98)	Smears taken >12 months before diagnosis for cases and before interview for controls. Personal interview with a questionnaire.

NA= not available. RR = relative risk, OR = odds ratio

[1] Screened every 2 to 5 years vs never.
[2] Screened within 3 years vs never.
[3] Calculated from data provided in paper.
[4] Smears performed within last 2 years vs smears performed 6 or more years previously.
[5] Screened within 10 years vs not screened within 10 years.
[6] Smears performed 24-35 months before vs no screened or screened more than 66 months before; microinvasive cancers were excluded.
[7] Screened within 5 years vs not screened within 5 years.
[8] OR = 0.38 (95% CI 0.28-0.52) when tests due to gynaecological symptoms were excluded.

118

Chapter 8. Cohort studies in evaluation of cervix cancer screening

E. Lynge

Introduction

In 1941, Papanicolaou and Traut (Papanicolaou and Traut, 1941) presented a "simple, inexpensive method of diagnosis ... which could be applied to a large number of women in the cancer-bearing period of life". The authors clearly stated that they were "not yet in a position to offer a statistical proof of the reliability of this method of diagnosis, but we can say that in our experience it yields a high percentage of correct diagnoses when checked by tissue biopsies. There is evidence that a positive diagnosis may also be obtained in some cases of early disease".

If the Pap smear worked for early detection of invasive disease, a decrease in cervical cancer mortality was to be expected in screened populations. If the Pap smear worked also for detection of preinvasive lesions, a decrease was to be expected in cervical cancer incidence (Lynge, in press). However, the understanding of the need for randomised controlled trials of medical procedures was new (Chalmers, 1998), and no one waited for the statistical proof, as the Pap smear was adopted widely throughout the world in the post war decades.

Only observational data have therefore been used for evaluating the outcome of the widespread cervical cancer screening. Evaluations have been based on both time trends, geographical comparisons, cohort studies, and case-control studies. The purpose of the present chapter is to give an overview of the cohort studies.

Cohort study data requirements

In most historical cohort studies, the observed number of deaths or incident cancer cases in the exposed cohort has been compared with the number expected, given that the exposed cohort had experienced the same mortality or incidence as the national population, i.e., standardised mortality or incidence ratios (SMR or SIR, respectively) have been calculated. This comparison is of course a valid estimation of the risk associated with the exposure only if the cohort members are similar to the national population, apart from the exposure considered.

119

In cohort studies for evaluation of cervical cancer screening, the exposure has been defined as either invitation to screening or as participation in screening. Both approaches have their advantages and disadvantages. Invited women will normally be a representative sample of the national population, the observed cervical cancer incidence or mortality among the invited women will, however, reflect both their background risk, the effect of the screening, and their participation rate. Participating women will on the other hand be a biased sample, but the observed cervical cancer incidence or mortality among these women will reflect only their background risk and the effect of the screening. The appropriate study design will depend on the questions addressed in the cohort study.

Both types of cohort studies of cervical cancer screening require that the individual screening histories are known and can be linked to the individual follow-up for incident cases of invasive cervical cancer and/or deaths from cervical cancer. Cohort studies can thus be undertaken only when:

- all Pap smears and biopsies are registered by date and personal identification for all women for a well defined cohort in a given time period;

- all women in the cohort can be followed up for deaths, emigrations, cause of death and/or incident cases of invasive cervical cancer;

- the two sets of data can be linked at the individual level.

In most circumstances, the use of Pap smears started spontaneously within the clinical setting and not as organised screening programmes. Pap smears from a given area were often read in many different private laboratories with no requirement for systematic and long-term record keeping. The follow-up for occurrence of invasive cervical cancer requires the presence of comprehensive cancer registration or other case ascertainment systems. The linkage of screening history data with follow-up data requires use of personal identification numbers or other unique personal identification. With this background, it is not surprising that the effect of cervical cancer screening has been evaluated in relatively few cohort studies.

Included here are only the long-term cohort studies originally set up for evaluating the effect of screening on the cervical cancer incidence or mortality. Not included are short-term follow up studies aimed at evaluating, e.g. progression of cervical dysplasia (Richart and Barron, 1969) or importance of endocervical cells in a smear (Mitchell and Medley, 1991).

120

Cohort studies

British Columbia, Canada

The British Columbia cervical cancer screening project started in 1949. In 1959, about 8% of women above the age of 20 were screened annually, and this proportion increased to 44% in 1971. Two cohort studies have been undertaken here. First, based on the population data and the individual screening records from 1958 to 1966, women in British Columbia were divided into those previously screened or not. In 1965, 13 clinical invasive carcinomas were detected among the screened women where 81.2 cases were expected based on the incidence rates from 1955-1957 (SIR 0.16). For the unscreened women the numbers were 67 and 62.1 respectively (SIR 1.08). The effectiveness of the programme could thus be estimated for the total population: 80 observed vs. 143.4 expected (SIR 0.56). It should be noted, however, that preclinical invasive cases were not included among the observed cases (Fidler *et al.*, 1968).

Second, two cohorts of women born 1914-1918 and 1929-1933 were formed. The size of the cohorts over time was estimated from the population data. Data on Pap smears used by the cohort members during the period 1949-1969 were extracted from the British Columbia Central Cytology Laboratory. The original publication from this study reported the results for detected cases of "carcinoma in situ and worse", complicating comparison with other data sources (Boyes *et al.*, 1982).

Ostfold, Norway

Organised screening was offered in Ostfold County, Norway, from the first round in 1959-1965 to the fifth and last round in 1974-1977. A cohort study has been undertaken of all 45 960 women invited to the first screening in the age-group 25-59 years and not previously diagnosed with cervical cancer. The cohort was followed up to the end of 1982, and the observed incidence and mortality were compared with that of women in five neighbouring counties not offered organised screening. During the period 1959-1982, 267 new cases of invasive cervical cancer were observed in the cohort, where 341.5 cases were expected, SIR 0.78, and 103 deaths from cervical cancer were observed where 124 were expected, SMR 0.83 (Magnus *et al.*, 1987).

Finland

In Finland, an organised screening programme started in 1963 and it gradually developed to become nation-wide. Pap smears were offered every fifth year to women aged 30 to 55. In 1971, 414 164 women had been offered at least two rounds of the programme, and the incidence of cervical cancer was followed in this cohort to the end of 1972. In the age group 30-59, the probability of contracting microinvasive carcinoma after the first Pap smear was 0.002 and the probability for frankly invasive carcinoma was 0.002, compared with a probability of 0.010 for invasive cervical cancer in Finnish women before the screening, and with 0.016 among women who did not attend the screening. The participation rate in Finland was 85%, and the effectiveness of the programme was thus SIR 0.42, if only frankly invasive carcinomas are considered, and SIR 0.58 if both microinvasive and frankly invasive carcinomas are considered (Hakama and Räsänen-Virtanen, 1976).

Manitoba, Canada

A province-wide cervical cytology screening programme with a registry was initiated in Manitoba in 1963, and the registry was used to estimate the percentage of women in the province who had been screened during the years 1963-1972. Cases of cervical cancer were known from the Manitoba Cancer Registry (Choi and Nelson, 1986). The data were included in the study performed by the International Agency for Research on Cancer (IARC study), see below.

Sweden

From 1964, several counties in Sweden gradually introduced organised screening for women aged 30-49 every fourth year. By 1973, the programme covered all of Sweden, except the municipality of Gothenburg. All smears taken inside the organised programme were reported to the National Board of Health and Welfare, where 930 127 women were registered with at least one smear taken during the period 1967-1975. This cohort was followed up for incidence of invasive cervical cancer to the end of 1980 (Pettersson et al., 1986). The data have been included in the IARC study, see below.

In a later series of studies incorporated into the doctoral thesis of Sparén (Sparén, 1996), an open cohort of 118 890 women in the county of Uppsala was followed up from 1969 to 1988 for in situ cancer of the cervix. The authors reported that the difference in efficiency between organised and opportunistic screening in the detection of cancer in situ was slight, if any, based on the comparison of detection rates. Their conclusion was that the dogma that organised screening is significantly

122

more efficient than the opportunistic type needs reconsideration (Gustafsson *et al.*, 1995a). It should be noted that 75% to 80% of all Pap smears taken during the study period were opportunistic, i.e. taken outside the organised programme. Their results have been challenged by other researchers (e.g. Nieminen *et al.*, 1999)

They further reported a low efficiency of cytological screening for cancer *in situ* of the cervix in older women based on decreasing detection ratio by age in the same cohort (Gustafsson *et al.*, 1995b). They also combined the women resident in the Uppsala and Gävleborg counties into a cohort of 386 990 women. This cohort was followed up for invasive squamous cell cancer of the cervix. The women's screening history was ascertained in computerised registers of Pap smears taken in the area and record-linkages enabled complete follow-up with regard to cancer incidence, out-migration and survival from 1968 to 1992. They compared the age-specific incidence of squamous cell cancer in ever vs. never-screened women and found a relative risk (RR) of 0.55 with 95% confidence interval of 0.51-0.61. The age-specific relative risks followed a U-shaped curve with lowest risks among the screened women in the age group of 40 to 59 years (RRs from 0.27 to 0.38) (Sparén, 1996).

Iceland

Cervical cancer screening started in Iceland in 1964 and was extended to the entire country in 1969. Women aged 25-59 were invited every 2-3 years, later extended to women aged 25-70. In 1974, close to 90% of women aged 30-49 had had at least one Pap smear. In all Icelandic women aged 25-59, the mortality from cervical cancer changed from 20 per 100 000 in 1955-1959, to 21 in 1960-1964, 32 in 1965-1969, and 15 in 1970-1974. The rates in never screened women were 30 in 1965-1969 and 23 in 1970-1974 (Johannesson *et al.*, 1978). In a later study, the mortality was reported to have fallen by 60% between 1959-1970 and 1975-1978, with a corresponding fall in the incidence of advanced tumours. The mortality rates among the unscreened population were more than ten-fold greater than among the screened. The greater part of the fall in mortality was attributed to the mass screening programme (Johannesson *et al.*, 1982).

Maribo, Denmark

An organised screening programme started for women aged 30-49 in Maribo County, Denmark, in 1967. The 16 187 women invited to the first round in 1967-1970 have been followed up for incident cases of cervical cancer to the end of 1984, and the observed numbers have been compared with the expected number based on rates for all Danish women. In the 87% of the invited women who participated, 115 cervical cancer cases were observed compared with 217 expected (SMR 0.53), whereas the numbers were 63 and 35.96 (SMR 1.75) among the 13%

of women who did not participate (Berget, 1979; Mellemgaard et al., 1990; Mellemgaard et al., unpublished data, 1990). The effectiveness of the programme estimated by the SIR for the total population was 0.70 (178/253). It should be remembered that the comparison group was the total population of Denmark, where there was extensive spontaneous screening and some other organised programmes as well. The Maribo cohort was later extended to include all women screened in the area 1967-1982, see IARC study below.

Use of cohort data

Although several of the cohort studies were originally established to evaluate the effect of the Pap smear screening on the cervical cancer incidence and mortality, the best known evidence for this effect in the end came not from the cohort studies, but from the analyses of time trends in cervical cancer incidence and mortality (Christopherson et al., 1970; Miller et al., 1976; Hakama, 1982; Läärä et al., 1987). The reasons for the limited impact of the cohort studies in assessing the effect of screening were probably the often small numbers of cases in the cohort studies, the selection bias implied in the comparison of rates for participants with rates in the total population not offered screening, and the complicated ways in which the cohort data were often presented.

The importance of the cohort studies therefore comes mainly from the secondary use of the data in the study of the natural history of cervical cancer. It is desirable to know the regression, persistence and progression of preinvasive lesions in order to optimise the screening for these lesions. As observation without treatment of preinvasive lesions has been considered unethical for many years, such data on the natural history of cervical cancer are available only for small groups of women (Östör, 1993). Modelling studies of the cohort data have been used therefore to fill this information gap. These studies were the International Agency for Research on Cancer study on risk of cervical cancer following a negative Pap smear, and the 'MISCAN' modelling studies on transition probabilities.

International Agency for Research on Cancer, IARC, study

In the 1980's, the data on time trends indicated a beneficial effect of screening for cervical cancer. Few data were available, however, on the effect of different screening schedules, the key question being whether screening was needed annually or whether a similar protection could be obtained with less frequent screens (Editorial, 1981). The IARC study was established to evaluate the incidence of invasive cervical cancer by number of previous negative smears and

124

time elapsed since last negative smear, Figure 1 (Day, 1986; IARC Working Group on Cervical Cancer Screening, 1986a).
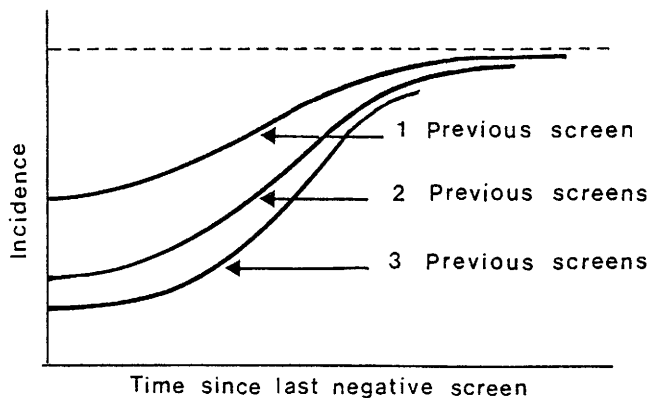
Figure 1. Anticipated incidence of invasive cervical cancer following one or more negative smears, by time since last negative smear, assuming sensitivity <100% (from Walter and Day, 1983)

Data were collected from ten areas with well established screening programmes (IARC Working Group on Evaluation of Cervical Cancer Screening Programmes, 1986b). The data collection was organised either as case-control or cohort studies. The cohort studies being the above listed from British Columbia (van Oortmarssen and Habbema, 1986), Manitoba (Choi and Nelson, 1986), Sweden (Petterson et al., 1986), Ostfold, Norway (Magnus and Langmark, 1986) and Maribo, Denmark (Lynge and Poll, 1986a; Lynge and Poll, 1986b). The cohort studies estimated the incidence of cervical cancer in women with negative smears by time since the last negative smear and by number of negative smears. Comparison was made with the incidence expected given that no screening had taken place. While the observed incidence could be calculated accurately from the screening and the population data, the expected incidence given no screening was estimated based on the incidence in the area prior to implementation of the screening programme.

The IARC study showed that the cumulative rate of invasive cervical cancer in women aged 35-64 could be reduced by 93.5% with screening every year, by 92.5% with screening every second year, and by 90.8% with screening every third year, see Table 1.

125

Table 1. Per cent reduction in cumulative rate of invasive cervical cancer in women aged 35-64 with different frequencies of screening (IARC Working Group on Evaluation of Cervical Cancer Screening Programmes, 1986b)

Interval between screening (years)	Per cent reduction in cumulative incidence	Number of tests
1	93.5%	30
2	92.5%	15
3	90.8%	10
5	83.6%	6
10	64.1%	3

Assuming that a woman is screened at age 35 and that she had at least one negative screen previously.

These results could not directly be applied to any public health policy as they were based on the conditions, that

1) the background incidence used in the analysis was a valid reflection of what the incidence would have been given no screening;
2) a 100% participation rate in the screening programme;
3) all detected non-negative smears were appropriately followed up;
4) the women entering the programme were essentially disease free, having had one or more previous negative tests.

However, the data provided a useful guide, as they indicate that almost the same protective effect could be achieved with three year screening intervals as with annual screening. The IARC study has formed the basis in many countries for the recommendation of three or five year screening intervals.

Transition probability from preinvasive to invasive disease

The 'MISCAN' simulation model of cancer screening has used the cohort data from British Columbia (van Oortmarssen and Habbema, 1991; van Oortmarssen et al., 1992; van Oortmarssen and Habbema, 1995) and from Maribo, Denmark (Bos et al., 1997) as the basis for estimation of the proportion of non-progressing cervical intraepithelial neoplasias and estimation of the duration of the preclinical phase in progressive lesions. Data on these parameters are needed in order to avoid overtreatment of cervical intraepithelial neoplasias detected in screening programmes.

Of the women screened in Maribo County, 1.5% had a positive smear, and one third of these women had a negative diagnosis at follow-up, whereas two thirds had a histologically confirmed cervical intraepithelial neoplasia. The analysis of the Maribo data showed that at least half of the confirmed cases of cervical intraepithelial neoplasia would not have progressed into clinical cancer in the women's lifetime.

The modelling of the cohort data has indicated that screening with a short time interval and with a high proportion of smears classified as non-negative will imply treatment of many preclinical lesions which, if undetected and untreated, would not have progressed to invasive disease.

Conclusion

Collection of cohort data for evaluation of cervical cancer screening requires that high quality register data are available on the screening history, the incident cancers and the deaths. Relatively few cohort studies have therefore been undertaken. The cohort studies were not decisive in establishing the effect of Pap smear screening on cervical cancer incidence and mortality, as the time trend data from national populations were less biased and easier to understand. Data from the cohort studies have, however, later formed an important basis for evaluating the optimal interval in cervical cancer screening and for studying the natural history of cervical cancer.

References

Berget A (1979). Influence of population screening on morbidity and mortality of cancer of the uterine cervix in Maribo Amt. *Dan Med Bull* **26**:91-100

Bos AB, van Ballegoijen M, van Oortmarssen GJ, van Marle ME, Habbema JDF & Lynge E (1997). Non-progression of cervical intraepithelial neoplasia estimated from population-screening data. *Br J Cancer* **75**:124-130

Boyes DA, Morrison B, Knox EG, Draper GJ & Miller AB (1982). A cohort study of cervical cancer screening in British Columbia. *Clin Invest Med* **5**:1-29

Chalmers I (1998). Unbiased, relevant and reliable assessment in health care. *Br Med J* **317**:1167-1168

Choi NW & Nelson NA (1986). Results from a cervical cancer screening programme in Manitoba, Canada. In: Hakama M, Miller AB & Day NE (eds) *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, pp. 61-67

Christopherson WM, Parker JE, Mendez WM & Lundin FE Jr (1970). Cervix cancer death rates and cytologic screening. *Cancer* **26**:808-811

Day NE (1986). The epidemiological basis for evaluating different screening policies. In: Hakama M, Miller AB & Day NE (eds) *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, pp. 199-209

Editorial (1981). Intervals between Pap tests: a compromise. *World Health Forum* **2**:533-540

Fidler HK, Boyes DA & Worth AJ (1968). Cervical cancer detection in British Columbia. *J Obstet Gynecol Br Commonw* **75**:392-404

Gustafsson L, Sparén P, Gustafsson M, Wilander E, Bergström R & Adami HO (1995a). Efficiency of organised and opportunistic cytological screening for cancer in situ of the cervix. *Br J Cancer* **72**:498-505

Gustafsson L, Sparén P, Gustafsson M, Pettersson B, Wilander E, Bergström R & Adami HO (1995b). Low efficiency of cytological screening for cancer in situ of the cervix in older women. *Int J Cancer* **63**:804-809

Hakama M (1982). Trends in the incidence of cervical cancer in the Nordic countries. In: Magnus K (ed) *Trends in Cancer Incidence*. Washington, Hemisphere, pp. 279-292

Hakama M & Räsänen-Virtanen U (1976). Effect of mass screening program on the risk of cervical cancer. *Am J Epidemiol* **103**:512-517

IARC Working Group on Cervical Cancer Screening (1986a). Screening for squamous cervical cancer - the duration of low risk following negative results in cervical cytology tests: introduction. In: Hakama M, Miller AB & Day NE (eds) *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, pp. 15-24

IARC Working Group on Evaluation of Cervical Cancer Screening Programmes (1986b). Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policies. *Br Med J* **293**:659-664

Johannesson G, Geirsson G & Day N (1978). The effect of mass screening in Iceland, 1965-74, on the incidence and mortality of cervical carcinoma. *Int J Cancer* **21**:418-425

Johannesson G, Geirsson G, Day N & Tulinius H (1982). Screening for cancer of the uterine cervix in Iceland 1965--1978. *Acta Obstet Gynecol Scand* **61**:199-203

Läärä E, Day N & Hakama M (1987). Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. *Lancet* **1**:1247-1249

Lynge E (in press). Cancer screening. In: Souhami RL, Tannock I, Hochenberger P & Horiot JH (eds) *Oxford Textbook on Oncology*, Second Edition. Oxford, Oxford University Press

Lynge E & Poll P (1986a). Incidence of cervical cancer following negative smear. A cohort study from Maribo county, Denmark. *Am J Epidemiol* **124**:345-352

Lynge E & Poll P (1986b). Risk of cervical cancer following negative smears in Maribo County, Denmark, 1966-1982. In: Hakama M, Miller AB & Day NE (eds) *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, pp. 69-86

Magnus K & Langmark F (1986). Cytological mass screening in Østfold County, Norway. In: Hakama M, Miller AB & Day NE (eds) *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, pp. 87-90

Magnus K, Langmark F & Andersen A (1987). Mass screening for cervical cancer in Østfold County of Norway 1959-77. *Int J Cancer* **39**:311-316

Mellemgaard A, Ewertz M, & Lynge E (1990). The association between risk of breast cancer and age at first pregnancy and parity in Maribo County, Denmark. *Acta Oncol* **29**:705-708

Miller AB, Lindsay J & Hill GB (1976). Mortality from cancer of the uterus in Canada and its relationship to screening for cancer of the cervix. *Int J Cancer* **17**:602-612

Mitchell H & Medley G (1991). Longitudinal study of women with negative cervical smears according to endocervical status. *Lancet* **337**:265-267

Nieminen P, Kallio M, Anttila A & Hakama M (1999). Organised vs. spontaneous Pap smear screening for cervical cancer: A case-control study. *Int J Cancer* **83**:55-58

Östör AG (1993). Natural history of cervical intraepithelial neoplasia: a critical review. *Int J Gynecol Path* **12**:186-192

Papanicolaou GN & Traut HF (1941). The diagnostic value of vaginal smears in carcinoma of the uterus. *Am J Obstet Gynecol* **42**:193-206

Pettersson F, Näslund I & Malker B (1986). Evaluation of the effect of Papanicolaou screening in Sweden: Record linkage between a Central Screening Registry and the National Cancer Registry. In: Hakama M, Miller AB & Day NE (eds) *Screening for Cancer of the Uterine Cervix*. Lyon, International Agency for Research on Cancer, pp. 91-105

Richart RM & Barron BA (1969). A follow up study of patients with cervical dysplasia. *Am J Obstet Gynecol* **105**:386-393

Sparén P (1996). *Early Detection and Screening for cancer of the Cervix in Sweden during the 20th Century*. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 599. Uppsala, Acta Universitatis Upsaliensis

Van Oortmarssen GJ & Habbema JDF (1986). Cervical cancer screening data from two cohorts in British Columbia. In: Hakama M, Miller AB, Day NE (eds) *Screening for Cancer of the Uterine Cervix* (IARC Scientific Publications No. 76), Lyon, International Agency for Research on Cancer, pp. 47-60

Van Oortmarssen GJ & Habbema JDF (1991). Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. *Br J Cancer* **64**:559-565

Van Oortmarssen GJ & Habbema JDF (1995). Duration of preclinical cervical cancer and reduction in incidence of invasive cancer following negative Pap smears. *Int J Epidemiol* **24**:300-307

Van Oortmarssen GJ, Habbema JDF & van Ballegooijen M (1992). Predicting mortality from cervical cancer after negative smear test results. *Br Med J* **305**:449-451

Chapter 9. Cancer registries in evaluation of breast cancer screening programmes

S.H. Wait

Introduction

Over the past 30 years, the effectiveness of mammography as a screening tool for breast cancer has been established through over 11 randomised clinical trials and case-control studies (Wald *et al.*, 1994). In these studies, screening mammography allowed for a reduction of up to 30% in breast cancer mortality in women aged 50-65 or 50-69 years. Screening programmes have been set up in Europe, North America and parts of Asia with the objective of reproducing this breast cancer mortality reduction in their target populations.

The ultimate measure of the effectiveness of screening programmes is the reduction in breast cancer mortality; however this benefit may only become apparent some 7-10 years after the introduction of screening (Day *et al.*, 1989). This time period is incompatible with the needs of public policy makers and screening units alike, who require early, reliable measures for the purposes of programme evaluation. Thus intermediary indicators have been developed to help assess the quality of mammography as a screening test and the efficacy and potential impact of screening programmes on breast cancer mortality (Day *et al.*, 1989). Many of these measures rely on the existence of a cancer registry which systematically records all cancer cases in the population targeted by screening.

The purpose of this chapter is to describe the role of cancer registries in the evaluation of existing breast cancer screening programmes. This paper first highlights some general issues related to data linkage between cancer registries, screening databases and other potential sources of data for the purposes of programme evaluation. Secondly, measures that require a cancer registry to be assessed are described in terms of how they are derived and what they may potentially contribute to the assessment of screening.

It should be noted that this chapter is meant as a general overview of potential measures for the evaluation of breast screening programmes within the specific context of cancer registries. Thus it offers only limited critical comment on the comparative value of each measure in evaluating screening programmes. Moreover, target rates for these measures, and the comparative success of different screening programmes in achieving them, are not discussed. For further guidance

on this topic, readers are recommended to consult the Europe Against Cancer epidemiological guidelines (Broeders et al., 1996).

Data linkage in the evaluation of breast cancer screening programmes

The evaluation of breast cancer screening programmes requires the combination of data from various sources and thus raises important issues of data linkage. Aspects of screening which need to be evaluated on a routine basis include: adherence to screening by the target population (compliance rate); the quality of screening as a preventive measure; the efficacy of screening and its impact on the epidemiology of breast cancer. Measures that may be used to evaluate each of these facets of screening programmes are listed in Table 1. Quality of screening should include quality of diagnostic procedures used in the assessment of screen-positive cases, and is measured in terms of false positive rates, sensitivity and specificity. Efficacy parameters allow determination of how effective screening is at finding cancers and detecting them early. Measures of the impact of screening allow assessment of the potential effect of screening on reducing breast cancer mortality, and require the positioning of screening results within the overall epidemiological picture for breast cancer in the target population.

As is described in Table 1, cancer registry data are needed in order to evaluate the impact of screening on the entire target population. Screening programmes must have a comprehensive dataset of all persons targeted by the screening programme, as well as comprehensive follow-up information on the outcome of the persons screened, and analysis of these data may allow determination of the quality of screening itself. However, the limitation of this perspective is that it provides no indication of the overall impact of screening on the epidemiology of breast cancer in the target population and restricting the evaluation of screening to screen-detected cancers may overestimate the actual effectiveness of screening.

Cancer registries and screening programme databases are usually run completely independently, so that linkage may only be possible if a unique identifier is used. Moreover, data from the screening units may or may not be contained in a single database. The easiest and most reliable identifier for record linkage remains a unique national identifier (e.g., social security numbers). However, some countries prohibit its usage in cancer registries or screening databases due to issues of confidentiality and data anonymity. If a unique identifier cannot be used, linkage requires computer matching of files from the screening database by last name, first name, and date of birth with cancer cases recorded in the cancer registry. In some studies, postal code is used as an additional identifier (van Dijck et al., this volume). Manual checking of positive matches between the two data sources is necessary, to rule out any double occurrences, check for multiple cancer diagnoses per individual, and ascertain correct diagnosis dates. Some authors have advocated

134

Table 1. Early indicators that may be used in the evaluation of breast cancer screening programmes (Adapted from Schaffer et al., 1996, Rev Epidém et Santé Publ, ©Masson Editeur)

Measure	Screening database[1]	Cancer registry
Impact of screening		
attendance rate[2]	✓	
Quality of screening process		
recall rate	✓	
false positive rate at screening	✓	
screening sensitivity[3]	✓	✓
screening specificity	✓	✓
biopsy rate	✓	
positive predictive value of biopsy	✓	
cancer detection rate	✓	
Screening efficacy		
screening prevalence/expected incidence ratio	✓	✓
proportion of in situ cancers[4]	✓	
proportion of invasive cancers <10 mm[4]	✓	
proportion of invasive cancers >20 mm[4]	✓	
proportion of cancers with positive nodal status (N+)[4]	✓	
proportion of advanced cancers (Stage II or more)[4]	✓	
Impact of screening on cancer epidemiology		
interval cancer rate	✓	✓
absolute rate of advanced cancer detection	✓	✓
absolute rate of small cancer detection	✓	✓
evolution of prognostic factors in target population over time		✓

[1] For the measurement of certain indicators, additional data linkage with population registers, census population lists or other data sources may be required.

[2] In programmes where women may also have access to screening outside of the organised programme, rates of individual or spontaneous screening should also be assessed.

[3] This indicator may also be used as a measure of the impact of screening on cancer epidemiology, namely in terms of the rates of interval cancers.

[4] Although the optimal description of these counts is as cases per 1000 women screened, it is customary to also present these data as a percentage of the total number of cancers detected by screening.

135

doing this by manually checking the validity of women's address details in the source files used (McCann *et al.*, 1998, Woodman *et al.*, 1995).

Screening programmes may also require access to other sources of data such as target population lists derived from local health authorities, mayor's offices or other population census bureaux. In the UK national programme for example, the names of addresses of women within target age groups for screening are obtained from local family health service authority listings. In France, local sickness fund lists are used (Wait and Allemand, 1996). Problems reported with such data sources are frequent and include the finding of "ghosts" (persons who no longer live at the indicated address) (Chamberlain *et al.*, 1993) or of missing women from invitation lists. All efforts should be made to correct for inaccuracies in the invitation lists, as these translate into under- or over-estimation of the target population for screening. Although these calculations do not directly rely on cancer registry data, the rates obtained (screening uptake rate, cancer detection rate,) are important measures of the impact of the screening programme and understanding of all other screening parameters hinges on their accuracy.

Counting cancer cases

Cancer registries typically measure rates in terms of the number of cancers, as opposed to the number of individuals affected. If one is linking cancer registry data to screening attendee data, one may argue that the denominator should reflect the number of women with breast cancer, as opposed to the number of cancers detected. Thus women who present with more than one cancer should only be counted once. Conventionally, one would then record data reflecting the cancer with the worst diagnosis or stage (Schouten *et al.*, 1998) or detected using the most invasive diagnostic technique. This convention is especially relevant when measuring the impact of screening on stage or size distribution.

Measures dependent on cancer registry data

As was presented in Table 1, a subset of evaluation measures requires a cancer registry for calculation. Table 2 briefly describes the significance of these measures to the evaluation of a screening programme's overall quality and effectiveness. The remaining part of this paper will describe each of these indicators in more detail.

Prevalence/incidence ratio of screen-detected cancers

One of the most important contributions of cancer registries to the evaluation of screening programmes is to show the evolution of breast cancer incidence following the introduction of screening. The underlying premise of screening is

136

Table 2. Description and significance of selected effectiveness measures requiring cancer registries in breast cancer screening programmes

Measure	Definition	Significance
Screening prevalence/expected incidence ratio	Ratio of number of screen-detected cancers at 1st round to the expected incidence in the absence of screening	Measures the number of cancers detected by screening which (in the absence of overdiagnosis) should have been detected clinically at some later point in time.
Interval cancer rate	Number of cancers appearing in between two screening rounds in women who screened negative at the previous screening round (as % of expected incidence)	Measures ability of screening to advance the diagnosis of cancer as well as screening sensitivity. Interval cancer rates are inversely proportional to the expected reduction in breast cancer mortality.
Absolute rate of small cancer detection	Percent change in proportion of small cancers (≤10 mm) detected as compared to before the onset of screening	Evidence of ability of screening to advance the diagnosis of cancer to an earlier stage. May indicate overdiagnosis due to screening.
Absolute rate of advanced cancer detection	Percent change in proportion of advanced cancers (Stages II-IV) as compared to before the onset of screening	Early predictor of the impact of screening on breast cancer mortality. The assumption is that a decrease in the rate of advanced cancers will allow for a reduction in breast cancer morbidity and mortality.
Evolution of prognostic factors in target population over time	Evolution of tumour size, nodal involvement, stage, and presence of metastases in the screened and unscreened population over time.	The more favourable the distribution of prognostic factors in screen-detected cancers with respect to clinically detected cancers, the larger the potential impact of the programme.

that it allows for the detection of more cancers at an earlier stage, thus causing an apparent increase in cancer incidence at the "prevalent" or initial rounds of screening. Day (1989) devised a helpful formula to demonstrate the link between prevalence at 1st round of screening, expected incidence in the absence of screening, screening sensitivity and sojourn time of cancers, that is the preclinical time period during which a cancer is detectable by screening The formula is as follows:

$$\text{prevalence at 1st screen/expected incidence rate} =$$
$$\text{screening sensitivity} \times \text{sojourn time}$$

The above formula highlights the relationship between the quality of screening, its effectiveness and the natural history of breast cancer. This relationship is further illustrated in the following diagram (Figure 1). The efficacy of mammography hinges on its ability to detect cancers during their preclinical phase, thus before they would have been detected clinically. Hence one uses the rate of small cancers (≤ 10 mm) as an important efficacy measure for breast cancer screening, although one must be conscious of the fact that high rates may reflect overdiagnosis through screening and overestimate the actual efficacy of the screening process.
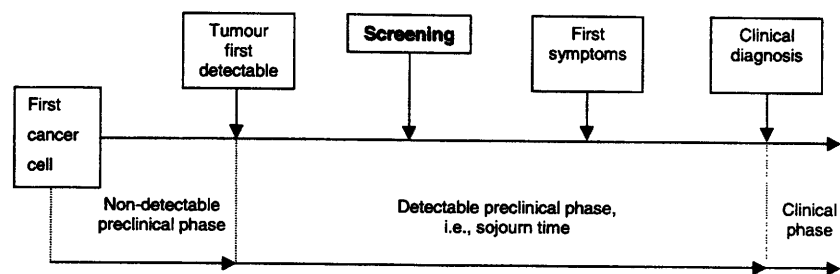
Figure 1. Early detection of breast cancer through screening (adapted from de Koning 1993)

Interval cancers

Significance of interval cancers

Interval cancers may be defined as cancers that present symptomatically in between two screening rounds in women who were screened negative at the previous screening mammogram. Interval cancer rates are an important indicator of the quality of mammography, as well as of the potential impact of screening programmes on breast cancer mortality. Based on the promising results of the Swedish Two-County trial (Tabar *et al.*, 1987), Day and colleagues (1995) established a formula for estimating the expected reduction in breast cancer mortality from interval cancer rates. When they applied this formula to their data from East Anglia, they found paradoxical results, namely that their programme met targets for cancer detection, yet reported unexpectedly high rates of interval cancers. These results were also reported in the analysis of interval cancer rates in the Northwest region screening programme (Woodman *et al.*, 1995). The authors attributed the high interval cancer rates, especially those in the 24-36 month period after screening, to the long interval between screens and to the single reading of screening mammograms. These results have sparked the debate over the need to shorten the screening interval from 36 months to 24 months in the UK (Field *et al.*, 1995) and in France (S. Wait, personal communication). They suggested that the much lower interval cancer rates observed in the Swedish trial was partly explained by the fact that interval cancers were calculated based on an interval period of 33 months on average, as opposed to 36 months in the UK. However, this explanation does not shed any light on differences observed in the periods 0-11 or 12-23 months after screening. Moreover, analysis from the Limburg and Nijmegen programmes in the Netherlands, both based on a 2-year screening interval and with systematic double reading of all mammograms, demonstrated similarly high rates of interval cancers to the UK programme 12-24 months post-screening (Schouten *et al.*, 1998). A special working group on interval cancers has been set up in the UK to explore means of reducing interval cancer rates and to reassess targets set for existing screening programmes (Moss and Blanks, 1998). Further research needs to be conducted to determine the optimal balance between interval cancers, screening interval and resource levels (e.g., double reading) required to achieve screening targets.

Data linkage with interval cancers

The ascertainment of interval cancer cases may be done using various sources, namely population registers, cancer registries, pathological registries, clinical records, and death certificates. The most recommended source remains the cancer registry, as this avoids both the reporting bias problematic in clinical records, and

allows for more completion than death certificates, which vary significantly from one country to another. In districts in which no cancer registry exists, a pathology registry may provide an acceptable alternative, although its completeness will depend on the compliance of pathology laboratories, the quality and uniformity of their data recording and diagnostic patterns in the region.

There is an inevitable lag in the recording of cancer cases in registries, which poses some challenges to assessing specific interval cancer rates and may lead to underestimating of the most recent year. Several programmes have tried to overcome this lag by instituting a system of "fast tracking" breast cancers in their districts (McCann et al., 1998, Woodman et al., 1995). Nonetheless, some interval cancer cases may be absent from the registry either because they are too recent to be registered or they occur in women who may have moved away from the area or are treated outside of the district. Record linkage between screening unit databases and registries may be complemented by specific searches at the level of screening units for interval cancers missed by the cancer registry. McCann et al. (1998) as well as Woodman et al. (1995), who retrieved some data directly from screening units before the cases had been recorded in the registry, took this approach. In screening programmes which have both a cancer registry and a pathology register, cross-verification of these two sources of data may also allow for increased completeness of recording of interval cancers. Schouten et al. (1998) chose to limit their dataset to those cancers recorded in the registry, running the risk of underestimating interval cancer rates in the latter year of their analysis. However, they argue that using data from one source for evaluation provides better comparability of data, namely in terms of the way staging or other prognostic information is recorded.

Presentation of interval cancer data

Interval cancers may be presented using a range of indicators (Table 3). When looking at interval cancers per 1000 women screened, it is conventional to present data by screening round and to separate the initial (prevalent) round from subsequent (incident) rounds of screening. Interval cancer rates should also distinguish between *in situ* cancers and invasive cancers. In the UK, it is specifically recommended to limit the analysis of cancer detection rates and interval cancer rates to invasive cases only, as only these would have been considered in the estimate of underlying incidence (Chamberlain et al., 1993). Moreover, the age used for describing interval cancers should always be the age at the time of screening.

140

Table 3. Presentation of interval cancers

Measure	Significance
Interval cancers per 1000 women screened.	May be compared to screening detection rate to measure share of interval with respect to screen-detected cancers.
Interval cancers per 1000 women screened at 0-11, 12-23 and 24-36 months post-screening.	Allows for comparison of interval cancer occurrence within 1, 2 or 3 years after screening. Provide an indication of sojourn time of cancers and the potential need to shorten or lengthen screening interval.
Proportionate incidence of interval cancers, i.e. total number of interval cancers divided by the total expected incidence during that period.	Allows for comparison of interval cancer rates across different programmes.

Schouten *et al.* (1998) recommend the age standardisation of the target population used as the denominator in the detection rates. No age standardisation was performed by any of the UK programme authors in the presentation of their interval cancer data. Adjustments for migration or death to the total number of women screened as a denominator were also not done by McCann *et al.* (1998) or by other authors.

Calculation of the underlying incidence of breast cancer

The proportionate incidence rates for interval cancer allows comparison of interval cancer data across different programmes, regardless of detection rates or the size of the target population. To obtain this rate, one needs to ascertain the underlying incidence of breast cancer, that is the incidence of invasive breast cancer that would have been observed in the absence of screening. Calculation of this rate is simple in randomised clinical trials, as it can be based on the incidence rates observed in age-matched controls. For example, in the Swedish two-county trial, underlying incidence was taken from the control population and data from the literature for women aged 50-59 and 60-69 in the trial control arm were used to estimate the expected incidence rate in the absence of screening for women aged 50-69 years (Tabar *et al.*, 1992). In non-experimental situations, however, the expected incidence may only be estimated based on projections from historical data, with potential errors. There exists no clear consensus on what the best method

141

is to estimate the underlying incidence, however, several methods have been described in the literature. In their analysis of the Nijmegen programme interval cancer rates, Verbeek *et al.* assumed the underlying incidence in Nijmegen to be that of the adjacent population in Arnhem (Verbeek *et al.,* 1984). In the study in the Northwest Region (England), Woodman *et al.* (1998) estimated the mean annual incidence rate of invasive cancer in the three years preceding the onset of screening and used this average rate as the underlying incidence rate. This method was also used by Schouten *et al.* (1998). Day *et al.* (1995) used linear extrapolation of the invasive cancer rates observed before the onset of screening (1976-1988), by age group, to estimate the expected incidence rates during the screening period (1990-1993). The expected incidence in women 50-64 was taken as the summed average incidence rate over the period 1990-1993 of three 5-year age bands (50-54, 55-59, 60-64 years).

The adjustment for breast cancer incidence increases over time used by Day *et al.* (1995) has been advocated as the preferred method by several authors (Prior *et al.,* 1996). However, it is questionable that adjustment will account for large differences in observed proportionate incidence rates. Schouten *et al.* (1998) only observed a difference of 1-2% in estimated proportionate incidence rates when they applied this adjustment (0.8% annual increase in incidence for women aged 50-69) to their data. The authors of the Northwest Region study made similar observations.

In a publication by a UK special working group on interval cancers, Moss and Blanks (1998) assumed a log-linear relationship between age and incidence to calculate underlying incidence rates for the year before the onset of screening. They then extrapolated these data based on trends from 1987-1995 and used the midpoint of the two estimates (from 1987 and 1995) to determine age-specific underlying incidence during the years following screening.

Classification of interval cancers

Interval cancer rates may certainly be decreased with improved sensitivity of the screening process (Day *et al.,* 1995), however, there will inevitably remain a certain number of true interval cancers that cannot be detected at the time of screening. A number of programmes have run blinded retrospective chart review exercises in an effort to determine the characteristics of false negative cancer cases missed by radiologists. However, the following classification scheme has been devised to distinguish cancers based on whether they could or could not have been detected at the time of screening:

Table 4. Classification of interval cancers (Adapted from Broeders et al., 1996)

Classification	Definition
True interval	Cancer that appeared negative on the screening mammogram, yet was radiologically apparent on the presenting mammogram at diagnosis.
Radiologically occult	Cancer that appeared negative on the screening mammogram and negative on the presenting mammogram at diagnosis.
Minimal signs	Cancer that upon rereading shows minimal signs, but was classified as negative on the screening mammogram; however, it appeared positive on the presenting mammogram at diagnosis.
False negative	Cancer that is apparent upon rereading the screening mammogram and is visible on the presenting mammogram at diagnosis.
Unclassifiable	Cancer for which no screening or presenting mammogram is available.

The impact of screening on stage distribution

The rationale underlying breast cancer screening is that it will allow for the detection of cancers at a more precocious stage. This should translate into (i) favourable distribution of prognostic factors in screen-detected cancers as compared to cancers detected outside of the screening programme, and (ii) an overall increase in the total share of cancers of favourable prognosis and a decrease in advanced cancers as compared to these proportions before the introduction of screening.

Obtaining staging information

Cancer registries do not always record staging information of cancers; moreover the quality of existing data has been found to vary significantly from one registry to another. In their comparison of cancer survival rates across European cancer registries, the EUROCARE research group originally intended to provide stage-specific survival data, however, they retracted when faced by the formidable discrepancies in quality and completeness of staging data (Berrino et al., 1995). As a result, a second study effort was created (EUROCARE-2) in an attempt to set out

143

a methodology for comparing staging data across cancer registries. Even within a single registry, the comparison of staging information from one institution to another may be biased, should one institution have more sophisticated diagnostic techniques. These differences will have an important bearing on stage-specific survival analyses. Moreover, they may bias the comparison of recent staging data with that from historical controls, as cancers which may have been described as localised with less sophisticated diagnostic techniques would now appear as metastatic. This phenomenon is known as "stage migration". A solution chosen by several authors is to limit analysis to histological size as a proxy for stage.

Stage distribution of screen-detected versus other breast cancers

In order to verify that screening detects cancers at an earlier stage, one may compare the staging properties of screen-detected cancers, cancers detected in non-attenders, interval cancers, and cancers detected in lapsed attenders (women whose last mammogram was negative and preceded the last screening round). This grouping of cancers by detection mode requires the matching of files from the cancer registry and screening databases, as dates of screening invitation (if applicable) and diagnosis are required in addition to detection mode and prognostic information. Computer matching follows the same principles outlined above for the analysis of interval cancers.

Of particular interest is the comparison of prognostic characteristics of interval cancers to those of screen-detected cancers. A predominance of poor prognosis interval cancers indicates the failure of the screening programme to detect cancers at an earlier stage. It is also assumed that interval cancers are of similar or worse prognosis compared with cancers detected in non-attenders. Interval cancers were of worse prognosis than non-attender cases in the Malmö (Anderssen *et al.*, 1988) and in the Bas-Rhin screening programmes (S. Wait, data on file). In East Anglia, the prognostic factors of interval cancers were similar to those of the unscreened population (McCann *et al.*, 1998).

One important point to consider is that the comparison of prognostic factors should ideally be presented as rates per population screened, as opposed to percentages amongst screen-detected or non-screen detected cancers. Rates allow one to account for the actual share of cancers detected through screening. For example, even if there is a higher proportion of small cancers amongst screen-detected cancers than amongst cancers detected in non-attenders, the overall impact on staging distribution and breast cancer mortality will be small, if screen-detected cancers only make up 30% of all cancers detected. The importance of cancer registry data in the evaluation of screening programmes thus becomes paramount, as this linkage allows for a comprehensive perspective on the evolution of breast cancer prognosis in the target population for screening.

144

The reduction of advanced cancer incidence

If screening programmes are successful at advancing the diagnosis of cancer, one should observe an overall reduction in advanced cancer rates in the target population, which in turn translates into a reduction of mortality from advanced disease. Day described a sequential impact of screening on stage distribution in the following schema (Figure 2, Day *et al.*, 1989). He suggested that differences in stage distribution by mode of detection will appear immediately, however one only begins to see an impact on advanced cancer rates approximately four years after screening initiation, followed by an impact on mortality some two years later, namely six or seven years following the onset of screening.

Day *et al.* (1989) suggested that screening should cause a decrease of at least 30% in the rate of advanced (Stages II-IV) tumours after four years. In the Limburg programme, a decrease of 10% in advanced cancers and of 15% in node positive tumours was shown one to four years after the onset of screening (Schouten *et al.*, 1998). A study in East Anglia found that the increase in small cancer incidence in the early years of screening was much larger than the subsequent decrease in advanced cancer incidence, thus suggesting that the reduction in mortality may also be somewhat less than targeted (McCann *et al.*, 1998). An important factor is that the incidence of interval cancers is increasing regularly in many screening programmes, and that prognostic characteristics of interval cancers are similar to those of non-attenders. Moreover, the actual proportion of cancers detected through screening does not exceed one-third in many programmes, even with high compliance rates (McCann *et al.*, 1998). These factors combined suggest that the potential of screening programmes to significantly reduce the rates of advanced cancer detection may be less than expected, with consequent impact on reductions in breast cancer mortality.

Estimating the expected rate of advanced cancers

The calculation of the reduction in advanced cancer incidence due to screening requires a reliable estimation of advanced cancer incidence before the introduction of screening. McCann *et al.* (1998) explore three different methods for arriving at this estimate. First, they project from the advanced cancer incidence observed in 1976-1986 to estimate this rate in 1995. Secondly, they take the average rate of advanced cancers observed in 1987-1988, immediately before the onset of screening, and compare this to the 1995 rate. Finally, they take the ratio of advanced to early cancers in 1989-1994 in women who had not yet received an invitation to screening to generate an expected incidence rate of advanced cancers, and multiply this by actual number of cases in order to ascertain actual numbers of cases expected. Schouten *et al.* (1998) used rates of advanced and node positive tumours in the three years preceding the screening programme as their reference

Years since start of programme

0 1 2 3 4 5 6 7 8 9 10 ...

First Screening Round

Second Screening Round

Third Screening Round

Attendance rate at 1st round
Prevalence/incidence ratio
Stage distribution at 1st round

Attendance rate at 2nd round
Prevalence/incidence ratio
Stage distribution at 2nd round

Rate of interval cancers after 1st round

Rate of interval cancers after 2nd round

Evaluation of impact on advanced cancer rates

Evaluation of impact on mortality

Figure 2. Sequential measures of effectiveness of a breast cancer screening programme
(by permission of BMJ Publishing Group, adapted from Day *et al.*, Br J Cancer 59:954-958, 1989)

rates and calculated a rate ratio of the incidence of advanced and node positive tumours detected in each year since the onset of screening as compared to the pre-screening incidence.

Discussion

Despite years of experience with breast cancer screening, the true impact of existing screening programmes on breast cancer mortality remains to be elucidated and has been the topic of some controversy in the recent literature. Existing screening programmes have succeeded in reaching most targets for quality and efficacy. However, targets for the impact on mortality (e.g. interval cancer rates) based on clinical trial results may need to be revised to account for evolving screening practices and the realities of exercising screening in uncontrolled settings (Moss et al., 1998).

Randomised clinical trials for screening are designed to detect a statistically significant difference in breast cancer mortality between two age-matched populations (screening attenders and non-attenders). The controlled setting implies that treatments will be identical in the trial arms and, thus, that any difference in the trial endpoint (ultimately mortality) may definitely be attributed to screening. In operational screening programmes, it is elusive to try to isolate the effect of screening on breast cancer mortality, as one cannot control for the differences in the patient risk profiles and in the treatment modalities received. Moreover, increases in incidence rates will also impact on mortality trends, thus further complicating the interpretation of mortality trends in the target population. Any comparison between screen-attenders and non-attenders is unlikely to have sufficient power to detect meaningful differences in mortality between groups while controlling for all these other factors. While this limitation may also apply to the comparison of earlier measures of the impact on cancer epidemiology, the measure of these outcomes is not subject to the long lag necessary to measure mortality data. All efforts should be made to allow for an unbiased comparison of these measures based on comprehensive data analysis for screening attenders and non-attenders of ongoing programmes.

In summary, cancer registries play an essential role in the evaluation of screening effectiveness, both through early indicators such as prevalence, incidence, staging distribution and interval cancer rates, as well as through later indicators such as breast cancer survival and mortality. Most importantly, registry data allow for the calculation of the expected incidence in the absence of screening programme and the comparison of actual to expected trends. The development of a cancer registry specifically for the purposes of evaluating a screening programme is not recommended due to its incurring high cost. However, programmes located in

areas with a reliable population-based cancer registry may benefit greatly from the data provided for the purposes of screening evaluation.

Although the first screening trials began over 30 years ago, there remain many challenges in the evaluation of the impact of screening on breast cancer epidemiology. Further studies based on existing screening programmes are needed to fully understand the value of early indicators of screening efficacy and to elucidate the true impact of screening on breast cancer incidence and mortality. These studies will rely on the existence of high quality, longitudinal data from cancer registries.

References

Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, Ljungberg O, Ranstam J & Sigfusson B (1988). Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *Br Med J* **297**:943-948

Berrino F, Sant M, Verdecchia A, Capocaccia R, Hakulinen T & Estève J (eds) (1995). *Survival of Cancer Patients in Europe: The EUROCARE Study* (IARC Scientific Publications No.132), Lyon, International Agency for Research on Cancer

Broeders MJM, Codd MB, Ascunce N, Linos A & Verbeek ALM (1996). *Quality assurance in the epidemiology of breast cancer screening.* 2nd revision, Europe Against Cancer Programme, European Commission

Chamberlain J, Moss S, Kirkpatrick AE, Michell M & Johns L (1993). National Health Service breast screening programme results for 1991-2. *Br Med J* **307**:353-356

Day N, McCann J, Camilleri-Ferrante C, Britton P, Hurst G, Cush S &, Duffy S (1995). Monitoring interval cancers in breast cancer screening programmes: the East Anglian experience. Quality Assurance Management Group of the East Anglian Breast Screening Programme. *J Med Screen* **2**:180-185

Day NE, Williams DRR & Khaw KT (1989). Breast cancer screening programmes: the development of a monitoring and evaluation system. *Br J Cancer* **59**:954-958

De Koning HJ (1993). *The effects and costs of breast cancer screening.* (Doctoral thesis), Rotterdam, Erasmus University

Field S, Mitchell MJ, Wallis MG & Wilson AR (1995). What should be done about interval breast cancers? [editorial]. *Br Med J* **310**:203-204

McCann J, Stockton D & Day N (1998). Breast cancer in East Anglia: the impact of the breast cancer screening programme on stage at diagnosis. *J Med Screen* **5**:42-48

Moss S & Blanks R (1998). Calculating appropriate target cancer detection rates and expected interval cancer rates for the UK NHS Breast Screening Programme. *J Epidemiol Community Health* **52**:111-115

Prior P, Woodman CB, Wilson S & Threlfall AG (1996). Reliability of underlying incidence rates for estimating the effect and efficiency of screening for breast cancer. *J Med Screen* **3**:119-122

Schaffer P, Renaud R, Gairard B & Guldenfels C (1996). Le rôle des registres dans l'évaluation des campagnes de dépistage du cancer du sein. (Cancer registries and evaluation of breast cancer screening). *Rev Epidém et Santé Publ* **44**:S15-S21

Schouten LJ, de Rijke JM, Schlangen JT & Verbeek AL (1998). Evaluation of the effect of breast cancer screening by record linkage with the cancer registry, the Netherlands. *J Med Screen* **5**:37-41

Tabár L, Fagerberg G, Duffy S, Day NE, Gad A & Gröntoft O (1992). Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am* **30**:187-210

Tabár L, Faberberg G, Day NE & Holmberg L (1987). What is the optimum interval between mammographic screening examinations? - An analysis based on the latest results of the Swedish two-county breast cancer screening trial. *Br J Cancer* **55**:547-551

Verbeek AL, Hendriks JH, Holland R, Mravunac M, Sturmans F & Day NE (1984). Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegen project, 1975-1981. *Lancet* **1**: (8388):1222-1224

Wait S, Allemand H (1996). The French breast cancer screening programme: epidemiological and economic results of the first round of screening. *Eur J Public Health* **6**:43-48

Wald NJ, Chamberlain J & Hackshaw A (1994). European Society for Mastology. Consensus Conference on Breast Cancer Screening, Paris, 4-5 February 1993: report of the Evaluation Committee. *Clin Oncol (R Coll Radiol)* **6**:261-268

Woodman CBJ, Threlfall AG, Boggis CRM & Prior P (1995). Is the three year breast screening interval too long? Occurrence of interval cancers in NHS breast screening programme's north western region. *Br Med J* **310**:224-226

Chapter 10. Breast cancer: modelling of mortality trends

T. Hakulinen

Introduction

The main and ultimate indicator of a success of a cancer screening programme is a decrease in the mortality of the cancer in question. It is therefore natural that routine statistics on cancer mortality could be used for the evaluation purpose. Changes in cancer incidence may also be utilised if the cancer had a sufficiently long pre-clinical phase and if the treatment of a pre-invasive lesion prevented its development into a truly invasive cancer. For example, cervical cancer has such a pre-invasive stage, whereas *in situ* lesions of the breast are less well defined. Thus, in breast cancer screening, mostly invasive but preclinical lesions are searched for in order to decrease breast cancer mortality.

Unfortunately, it is not always possible to evaluate population-based mass-screening activities within controlled trials. Even if a proper trial has been introduced it is possible that public health service screening will differ from screening in a trial in terms of quality and hence in its effectiveness (Hakama, 1982). Compromises may have to be made in service screening, and the enthusiasm and experience of the field staff may be decreased compared to a trial. Thus, even when a specific trial has not been designed, it is important for quality assurance of any ongoing screening programme that an evaluation can be made of the activity.

Bases for comparison

Ideally, the screening policy should be specifically designed to facilitate evaluation. The Finnish breast cancer screening (Figure 1) provides a good practical example. For practical reasons and due to shortage of funding it was not possible to provide the screening service for the entire target population at the beginning. This lead to a national recommendation to start screening the cohorts born in even years earlier than those born in odd years (Hakama *et al.*, 1999). A non-significant 24% reduction in breast cancer mortality, based on deaths among the incident breast cancer cases in the period 1987-1992, was observed for the birth cohorts screened early, compared to

those screened later. The reduction was 44% and significant among women under 56 years of age at the beginning of the study (Hakama *et al.*, 1999).

Year of screening	Year of birth											
	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939
1987	■	□	□	□	■	□	□	□	■	□	□	□
1988	□	□	■	□	□	□	■	□	□	□	■	□
1989	■	□	□	■	■	□	□	□	■	■	□	□
1990	□	□	■	□	□	■	■	■	□	□	■	■
1991	■	■	□	■	□	□	□	□	■	□	□	□
1992	□	□	■	□	□	■	■	■	□	□	■	■
1993	□	■	□	■	□	□	□	□	■	□	□	□
1994	□	□	■	□	□	■	■	■	□	□	■	■
1995	□	□	□	■	□	■	■	■	□	□	■	■
1996	□	□	□	□	□	■	■	□	□	■	■	□

Figure 1. Finnish National Board of Health's recommendations of screening rounds in organised screening programme for breast cancer, by birth cohort and calendar year (reproduced by permission of BMJ Publishing Group from Hakama *et al.*, J Med Screen 6:209-216, 1999)

The routine mortality statistics were not sufficient for such an evaluation. The nationwide Finnish Cancer Registry provided the dates of diagnosis of the incident cancer cases so that an evaluation could be based on mortality related to incident breast cancers diagnosed during the period of study.

If the screening policy has not been designed to facilitate the evaluation, a prerequisite for an empirical evaluation is some difference in the screening policies, either over time (Gibson *et al.*, 1997) or between geographical regions (Törnberg *et al.*, 1994). Differences in the temporal and geographical patterns of disease incidence or mortality should subsequently reflect the differences in screening patterns. Thus, consistent and comparable information on disease occurrence and deaths among the patients should also be available. A population-based cancer registry may also be able to provide – in addition to reliable incidence statistics – improved data on cancer mortality compared

with official statistics on causes of death (Saxén, 1982). The cancer registry receives usually a number of notifications on a cancer case prior to death and may be in a better position to know the appropriate cause of death than the person signing the death certificate.

Confounders may seriously complicate the evaluation of a screening policy, unless there are documented differences in the screening policy that could be regarded as a non-experimental design for a study. For example, treatment is a potential confounder which should be taken into account when using routine mortality data to evaluate mammographic screening (Figure 2). Treatment with tamoxifen was introduced in England and Wales approximately at the time when mammographic screening was initiated (Quinn and Allen, 1995). Thus, it is very difficult to estimate the independent effect of screening.
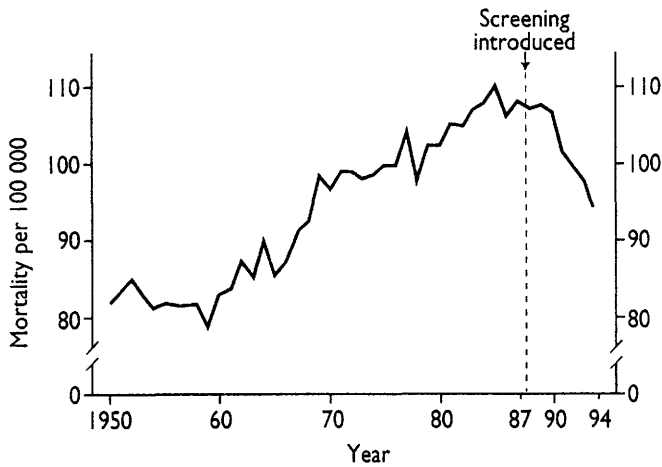

Figure 2. **Age-adjusted breast cancer mortality rates in women 55-69 years of age in England and Wales in 1950-1994** (reproduced by permission of BMJ Publishing Group from Quinn *et al.*, BMJ 311:1391-1395, 1995)

Historical developments

The incidence and mortality rates may have existing trends and differentials before the screening activity started. The areas to be compared may also be internally heterogeneous with respect to disease trends and differentials. These factors should be taken appropriately into account using a statistical model for the incidence or mortality. The part of the incidence or mortality data that is not related to a period when screening was practised may, using an appropriate model, be extrapolated into the period of screening in order to give a hypothetical expected number of cases or deaths in the absence of screening (Prior et al., 1996; Gibson et al., 1997). If the observed number of cases or deaths is very much below the expected one, screening may be considered successful. Information on the screening activity may be incorporated in the model in order to estimate as detailed numerical effect of the screening as possible (Törnberg et al., 1994).

Trends in cancer incidence and mortality may also be caused by technical issues (Hakulinen, 1996). Definitions, and diagnostic criteria and facilities change over time and may create or mask trends. It may be difficult to predict for how long these changes may continue.

Statistical modelling

Although breast cancer mortality would be the most useful target to model, also incidence rate models may give useful information (Prior et al., 1996). Breast cancer screening leads to an increased incidence of the disease, at least temporarily, due to the fact that diagnoses are made earlier than in the absence of screening. It may also be possible that in some cases small tumours are detected that otherwise would never have surfaced as clinical cancers. Incidence trend modelling may help in estimating how much of an increase in incidence may be related to the screening activity.

A screening is usually targeted to certain age groups only. If the observed incidence is compared with a predicted incidence in the absence of screening, it is expected that the incidence is increased in these age groups. The age-specific predicted and observed incidence rates of breast cancer in Finland in 1989-1993 agreed fairly well, given that the prediction base was 1954-1983, when there was no mammographic screening in Finland. There was one major exception: Mammographic screening took place in ages 50-64 years in the years following the prediction base, and the observed incidence of

154

breast cancer exceeded that predicted in those ages (Figure 3). Of course, the increased incidence is not as such beneficial. In the years after the start of the screening, the advancement of breast cancer diagnoses should lead to a decrease in breast cancer mortality attributable to a smaller lethality of these early detected tumours (Hakama et al., 1997).

Statistical model building should lead to a tailor-made product for the particular problem considered. The models in the comparisons are usually based on Poisson regression (Breslow and Day, 1987). They should take into account the uncertainty in the model parameters due to the randomness in the historical data and the random variation of the observations that are made in the period of the evaluation (Hakulinen and Dyba, 1994; Gibson et al., 1997).
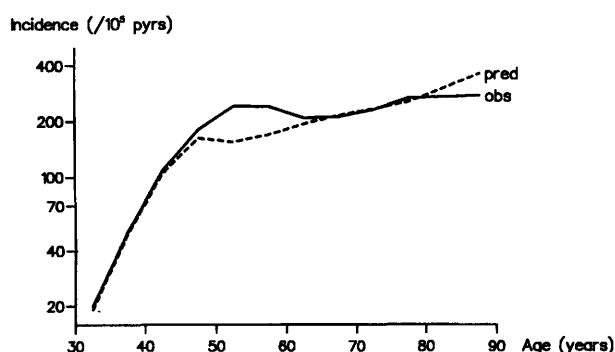
Figure 3. The predicted age-specific breast cancer incidence in Finnish females (pred) and that observed (obs) in 1989-1993, by age (reproduced by permission of Scandinavian University Press from Hakulinen, Acta Oncol 35:665-710, 1996)

Under the assumption that the model chosen is correct, the range of likely outcomes is described by a confidence interval. The confidence intervals, the middle intervals of the predictions in Figure 4, have been actually built in such a way that only the uncertainty in the historical data has been taken into account. Prediction intervals, the total (outer) intervals of the predictions in Figure 4, also account for the randomness in

the numbers of cases themselves. Consequently, the prediction intervals are somewhat wider than the corresponding confidence intervals.

Theoretically, if screening for melanoma and cancers of the colon, stomach or lung could be based on early detection of preclinical lesions and if this screening had been started after 1980-1985, one simple way to evaluate its effect would be to check whether the observed number of clinical cancers in 2000-2004 would fall below the
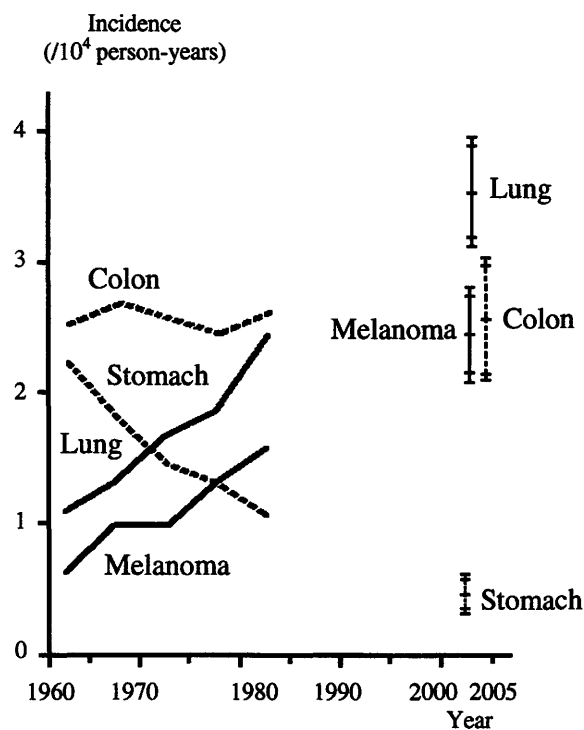
Figure 4. **Age-adjusted incidence rates of cancers of the lung, stomach and colon and melanoma of the skin in females in the Stockholm-Gotland Oncological Region in Sweden in 1960-1984, by five-year periods, with predictions for 2000-2004 (the total interval: 95% prediction interval for the future observation; the middle interval: 95% confidence interval for the expected value of the future observation)** (Hakulinen T & Dyba T, 1994. Precision of incidence predictions based on Poisson distributed observations. Stat Med 13:1513-1523. Copyright John Wiley & Sons Ltd. Reproduced with permission)

156

prediction intervals (Figure 5). Special attention should, nevertheless, be given to possible confounding and technical factors that may also be involved.

It is possible to use models that preserve the age-incidence or age-mortality pattern of the disease in the period of prediction when calculating the theoretical expected or predicted incidence or mortality in the absence of screening (Dyba et al., 1997). To guarantee this, Prior et al. (1996) used a model of the form

$$\ln (EM_{it}) = \alpha_i + \beta t$$
(1)

where M_{it} is the incidence (or mortality) of disease in age group i and period t, EM_{it} is its expected value, and α_i and β are unknown parameters. The problem with this model in longer-term predictions is that it specifies an implausible exponential incidence growth with time. A model without the log transformation of the EM_{it} precludes this property but with cancers, the incidence (or mortality) rates cannot have the same parameter of absolute change in all ages, as the rates vary strongly by age. Thus, the growth parameter has to become age-dependent:

$$EM_{it} = \alpha_i + \beta_i t.$$
(2)

This model, however, as the first model tried by Prior et al. (1996), ln $EM_{it} = \alpha_i + \beta_i t$, does not guarantee a plausible age-incidence pattern in the future.

A model that both precludes the exponential future growth and guarantees a plausible future age-incidence pattern was recently proposed by Dyba et al. (1997):

$$EM_{it} = \alpha_i(1 + \beta t).$$
(3)

This model is non-linear in its parameters but linear with respect to time. Figure 5 shows a comparison of models (2) and (3) for the skin melanoma data of Figure 4. Clearly, the future age-incidence pattern is more plausible and the age-specific confidence intervals are shorter for model (3). Model (3) also fits the historical data, a feature that is very important to check.

Whatever model is used, it would be important to make it simple for reasonable prediction intervals for future observations. These intervals are important in showing how much confidence a particular prediction has provided that the model being used can be relied upon. Of course, it is quite possible that the model used for predictions is

wrong even if it fits the historical data, and the confidence interval in such cases has a limited value only.

Incidence
(/10⁴ person-years)

Incidence
(/10⁴ person-years)

Age group

Age group
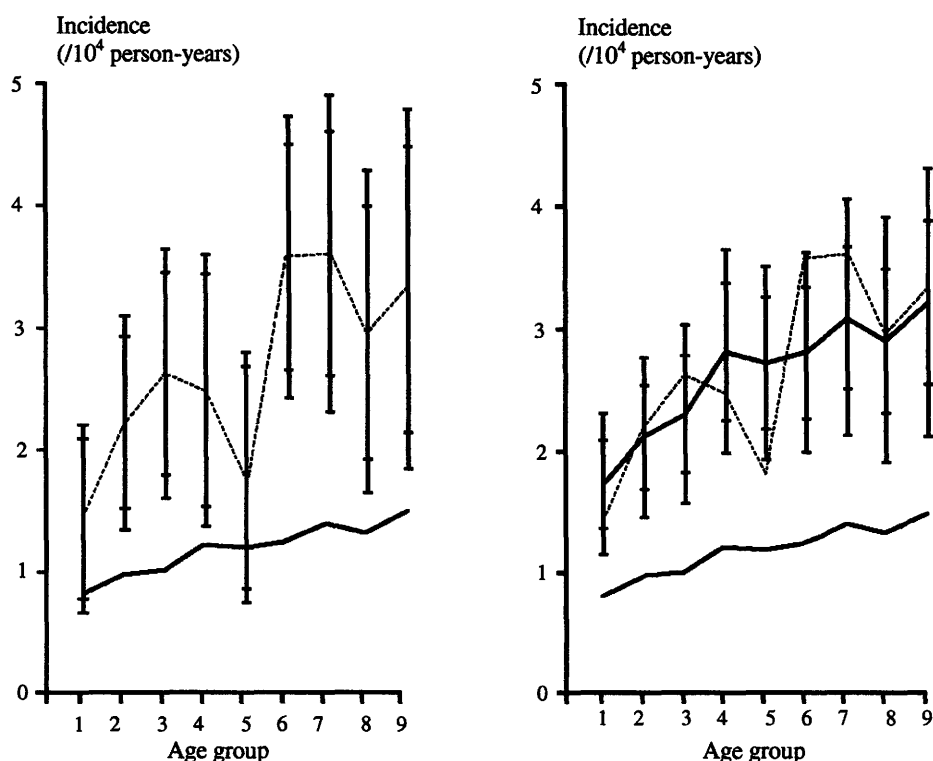
Figure 5. Age-specific incidence rates of skin melanoma in females in the Stockholm-Gotland Oncological Region in 1960-1984 (lower solid line) and incidence rates predicted for 2000-2004 as estimated by model (2) (dotted line) and model (3) (upper solid line) with 95% prediction intervals for the future observations (total intervals) and confidence intervals for the expected values of the observations (middle intervals) (Dyba *et al.*, 1997. A simple non-linear model in incidence prediction. Stat Med 16:2297-2309. Copyright John Wiley & Sons Ltd. Reproduced with permission)

Incorporation of exposure variables

The screening variable can be also explicity included in the model as an indicator (Luostarinen *et al.*, 1995) or as a more elaborated score variable (Törnberg *et al.*,

158

1994). This facilitates and increases the efficiency of the formal testing of the effect of the screening programme and helps in estimating its effect.

For the quality assurance of the Swedish mammographic screening (Törnberg et al., 1994) it was of interest to check, whether the effect of the large Swedish trials could be demonstrated using routine breast cancer mortality data and details about the screening. The evaluation of the service screening could be considered to be based on routine mortality statistics. Ideally the screening would, as a routine activity, be as successful as the Swedish mammographic screening trials.

A successful screening should reduce the breast cancer mortality from the predicted level, based on trend extrapolation. However, a large proportion of breast cancer mortality during the observation period is attributable to cases diagnosed earlier, i.e., before the screening was introduced. Consequently, the statistical model applied in the Swedish study contained an assumption that the full effect of screening can be seen starting from 10 years after the period the screening was initiated. From 5 to 10 years after the start, the effect was assumed to be only half of the full effect. No effect was assumed during the first five years after the initiation of the screening programme. When the whole age group was not subject to screening, the effect of the programme was assumed to concern only the proportion screened.

All the 26 Swedish counties, with the exception of Gävleborg county with a long history of service screening, were used as geographical units in the evaluation. The age groups considered were those of the screening, five-year categories between 50 and 75 years in five-year time periods between 1971 and 1990. Deviations from mortality trends were predicted to occur in the last two periods only (Table 1). The expected reduction was largest, 60% of the theoretical maximum effect, in Kopparberg in 1986-1990. These proportions were called mammography scores.

A satisfactory fit to the breast cancer mortality data was given by a Poisson regression model including the categorical variables age, period and county, and the numerical mammography score but no interaction terms. The p value for removing the mammography score from the model was 0.08. The estimate for the mammography score indicated a 19% protective effect, with 95% confidence interval from –3% to 37%. There was no evidence that the effect of mammography score would have been different for different counties.

Table 1. Fractions of theoretical maximum effects (mammography scores) for the Swedish counties with mammographic screening trials in 1981-1990 (reproduced by permission of BMJ Publishing Group from Törnberg *et al.*, J Med Screen 1:184-187, 1994)

County	Period	
	1981-1985	1986-1990
Stockholm	0	0.08
Östergötland	0.15	0.40
Malmö	0.25	0.50
Gothenburg	0	0.15
Kopparberg	0.27	0.60

The 19% protective effect, even if non-significant, is compatible with the 24% effect estimated in an overview of the Swedish trials (Nyström *et al.*, 1993). The modelling based on routine mortality data is of course a cruder alternative than a proper evaluation of the trials. In any case, an effect could be disclosed also using routine mortality data when modelling had been done. Nevertheless, it can be expected that the effects may end up being somewhat underestimated depending on the appropriateness of the model and on confounding effect of deaths of patients with breast cancer diagnosed before the start of the trial.

On the other hand, it would have been possible to sharpen the model employed by Törnberg *et al.* (1994). For example, separate mammography scores could have been used for different age groups in a county and the differences in the trial design could have been taken explicitly into account.

Trend extrapolations were based on a simultaneous modelling of the rates in all counties, not just making a model for each county separately. When the number of counties increases, multi-level modelling should be considered (Gibson *et al.*, 1997).

In some countries, trends in breast cancer mortality in the age groups subject to mammographic screening were already decreasing before any screening programmes were introduced (Coleman *et al.*, 1993). When evaluating any beneficial effects of mammographic screening on mortality trends, the existing trends and particular screening policies should be properly taken into account.

In any case, this kind of evaluation is very crude. Techniques exist (Chen *et al.*, 1998) to base the evaluation on a number of intermediate endpoints describing the severity of the disease of the individual patients. Thus, there is no need to wait until the breast cancer patients have died and contributed to breast cancer mortality. However, an information system, a mass-screening register (Hakama *et al.*, 1997), is needed for a successful accomplishment of such a task. It is advantageous to have a population-based cancer registry to guarantee the completeness and to study the representativeness of the breast cancer patient series needed in the evaluation and monitoring of mammographic screening activities in a population.

References

Breslow NE & Day NE (eds) (1987). *Statistical Methods in Cancer Research. Volume II. The Design and Analysis of Cohort Studie* (IARC Scientific Publications No. 82), Lyon, International Agency for Research on Cancer

Chen HH, Thurfjell A, Duffy SW & Tabár L (1998). Evaluation by Markov chain models of a non-randomised breast cancer screening programme in women aged under 50 years in Sweden. *J Epidemiol Community Health* **52**:329-335

Coleman MP, Estève J, Damiecki P, Arslan A & Renard H (eds) (1993). *Trends in Cancer Incidence and Mortality* (IARC Scientific Publications No. 12), Lyon, International Agency for Research on Cancer

Dyba T, Hakulinen T & Päivärinta L (1997). A simple non-linear model in incidence prediction. *Stat Med* **16**:2297-2309

Gibson L, Spiegelhalter DJ, Camilleri-Ferrante C & Day NE (1997). Trends in invasive cervical cancer incidence in East Anglia from 1971 to 1993. *J Med Screen* **4**:44-48

Hakama M (1982). Trends in the incidence of cervical cancer in the Nordic countries. In: Magnus K (ed) *Trends in Cancer Incidence:Causes and Practical Implications*, Washington, Hemisphere, pp. 279-292

Hakama M, Pukkala E, Söderman B & Day N (1999). Implementation of screening as a public health policy: issues in design and evaluation. *J Med Screen* **6**:209-216

Hakulinen T (1996). The future cancer burden as a study subject. *Acta Oncol* **35**:665-670

Hakulinen T & Dyba T (1994). Precision of incidence predictions based on Poisson distributed observations. *Stat Med* **13**:1513-1523

Luostarinen T, Hakulinen T & Pukkala E (1995). Cancer risk following a community-based programme to prevent cardiovascular diseases. *Int J Epidemiol* **24**:1094-1099

Nyström L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Rydén S, Andersson I, Bjurstam N, Fagerberg J, Frisell J, Tabár L & Larsson LG (1993). Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* **341**:973-978

Prior P, Woodman CBJ, Wilson S & Threlfall AG (1996). Reliability of underlying incidence rates for estimating the effect and efficacy of screening for breast cancer. *J Med Screen* **3**:119-122

Quinn M & Allen E (1995). Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. *Br Med J* **311**:1391-1395

Saxén E (1982). Trends: Facts or fallacy. In: Magnus K (ed) *Trends in Cancer Incidence: Causes and Practical Implications*, Washington, Hemisphere, pp. 5-16

Törnberg S, Carstensen J, Hakulinen T, Lenner P, Hatschek T & Lundgren B (1994). Evaluation of the effect on breast cancer mortality of population based mammography screening programmes. *J Med Screen* **1**:184-187

Chapter 11. Cohort studies on breast cancer screening

R. Sankila and E. Lynge

Introduction

The first randomised controlled trial to evaluate the efficacy of mammographic screening for breast cancer was launched in 1963 in New York State (Shapiro *et al*, 1966). Subsequently, major randomised screening trials have been performed in, e.g., Canada, Scotland and Sweden (Miller *et al.*, 1992; Roberts *et al.*, 1984; Nyström *et al.*, 1993). The results have been fairly consistent, suggesting that a reduction of up to 30% in breast cancer mortality can be achieved among the screened women.

Based on the results of the trials, organised screening programmes have been set up, first as pilot programmes, and later expanded into regional or nationwide public health interventions. For example, in Sweden, the first pilot study was performed in 1974, followed by screening trials (Nyström *et al.*, 1993). In 1986, the National Board of Health published national guidelines on mammographic screening and, eventually, in 1997, all counties in Sweden were providing mammography service screening (Olsson *et al.*, 2000).

Why does any doubt still exist?

There are several reasons why the full benefits of screening, as demonstrated in the randomised trials, may not be accomplished in real-life settings. Quantifying the benefits in a service setting is only possible through population-based studies, using data before and after the launching of an organised screening programme. Due to the prospective nature of such studies, the relative rarity of breast cancer in the screening age groups, and the high survival rates among breast cancer patients, the study populations need to be very large for statistically significant effects on the mortality rates to be seen. Further, collection of data may need to continue for 10 years at least, as most of the randomised trials began to show a decrease in mortality only after six or seven years of follow-up.

Cohort studies

Although the notion of a cohort study is fairly clear to most epidemiologists in the context of investigating possible aetiological factors, in reports of studies evaluating screening, different types of studies have been referred to as cohort studies - or at least the terminology 'study cohorts' has been used. For example, the two groups of women in screening trials originally randomised to be screened or not may be referred to as 'cohorts' (Alexander et al., 1999). These cohorts may continue to be monitored even after the end of the trial itself (UK Trial of Early Detection of Breast Cancer group, 1999). Possibly more conventional is the nested case-control study, in which exposure (in cases and controls) is defined as having actually been screened, within the populations of a screening trial, in an organised screening programme or simply in opportunistic screening activities in the general population.

In this latter instance (and in this chapter), the study cohort is the entire female population which is targeted for screening. The population should be geographically defined, but it can be stratified according to the research needs. In the analysis phase, it is possible, e.g., to exclude non-attenders from the entire cohort that was intended to be screened. There should be a clear and explicit reason for any exclusions and stratifications, however.

The principle constraint upon the classical approach of prospectively observing women who have and have not been screened (possibly separating screenees by the setting of the test (organised vs. opportunistic, and the number of tests and intervals between them) is the difficulty in determining outcomes. The screening registry must be organised to this effect, or, more plausibly, linked to a population-based registry, which can determine outcomes in the population comprising the cohort. Record linkage (between screening records and cancer registers) faces challenges from misconceived regulations on confidentiality. Furthermore, if unique personal identification numbers do not exist, the linkages can be technically very challenging.

These complexities explain why few cohort studies have been performed and why the existing data fall short of producing a comprehensive picture of the entire screening programme.

Although a decrease in breast cancer mortality due to the screening programme should be the natural outcome of a population-based cohort study, only one such study has been published so far (Hakama et al., 1997). Other such studies have produced results on a variety of quality indicators of the screening process itself or on intermediary outcomes in the screened populations.

164

Examples of cohort studies involving population-based cancer registries

An example of a published study evaluating the effectiveness of screening for breast cancer as a public health policy is the one performed in Finland (Hakama *et al.*, 1997). Hakama has described the study in detail elsewhere in this Monograph (see Chapter 2).

Another example of a population-based cohort study evaluating mammography screening for breast cancer in Copenhagen will be described in more detail here (Mammography Screening Evaluation Group, 1998).

Mammography screening for breast cancer in Copenhagen April 1991 - March 1997

Based on the results of randomised controlled trials, women aged 50-69 have since 1991 been offered biennial mammography screening in the municipality of Copenhagen, Denmark. The target population is approximately 40 000 women. As relatively short time has passed since start of the programme, only the short-term indicators of its effectiveness have been evaluated and compared with the outcomes in other service screening programmes.

Organisation of mammography screening in Copenhagen

All mammography screening in Copenhagen takes place at a special clinic. Two-view mammography is used in the first screening round. The radiographer checks the image quality before the woman leaves the clinic. Women with dense breast tissue will continue to have two-view mammography, whereas other women will have single-view mammography in the subsequent rounds.

Invitations to the screening are issued by Københavns Kommunes IT-Service, based on the updated population register for the municipality. All women aged 50-69 at the beginning of the invitation round and living in the municipality of Copenhagen are invited. The invitation register is updated daily with movements in and out of the municipality and with deaths (using the Central Population Register). The personal invitation gives an appointment at the screening clinic. This appointment can be changed by telephone. It is also possible to inform the clinic by telephone if a woman does not want to participate in this round and/or in any future rounds. A reminder with a new appointment is sent to women who have not contacted the clinic.

Data sources

Mammography screening data in København Kommunes IT-Service are stored in three files; two with information on the target population from the current and previous invitation rounds, and one with all the screening data. Women developing breast cancer in this population were identified through the Danish Cancer Registry (which has information on all incident invasive cancer cases in Denmark 1943-1994) and from two registers of the Danish Breast Cancer Cooperative Group, 'DBCG-invasive' with information on most incident invasive breast cancer cases diagnosed in Denmark 1978-1996, and 'DBCG-*in situ*' with information on most cases of carcinoma *in situ* (CIS) in the breast in Denmark 1978-1996. It proved rather difficult to determine the final diagnoses of women who screened positive and the number of interval cancers. All the available data were reviewed manually on a case by case basis, and any inconsistencies were resolved by retrieving further information from the pathology register and the clinical records.

First invitation round

The target population consisted of all women born between 31 March 1921 and 1 April 1941 (N = 43 092). The first date of screening mammography was 4 April 1991 and the last date was 28 March 1995. The final diagnosis of the last screen-detected case was made in November 1995. A total of 29 966 women tested negative in the first invitation round, with 28 303 women with a negative mammography, 1432 women with a positive mammography but with a negative assessment, and 231 women with a positive assessment but with a negative surgery. Among 363 women with malignant diagnosis at surgery, 316 had invasive breast cancer, 44 had carcinoma *in situ* and three had other cancer diagnoses. A flowchart of the results of the first screening round is provided in Figure 1.

Analysis of interval cancers

The population at risk for calculation of interval cancer rates comprised women who tested negative in a given screening round. The negative 'test' could occur at mammography, at assessment or at surgery. Theoretically, the risk period for each woman started from the date of the negative 'test' until the date of death, emigration, next screening date or two years since the previous screening date, whichever came first. However, as the dates for assessment and surgery were not systematically recorded, the start of the risk period was defined from the date of mammography.

166
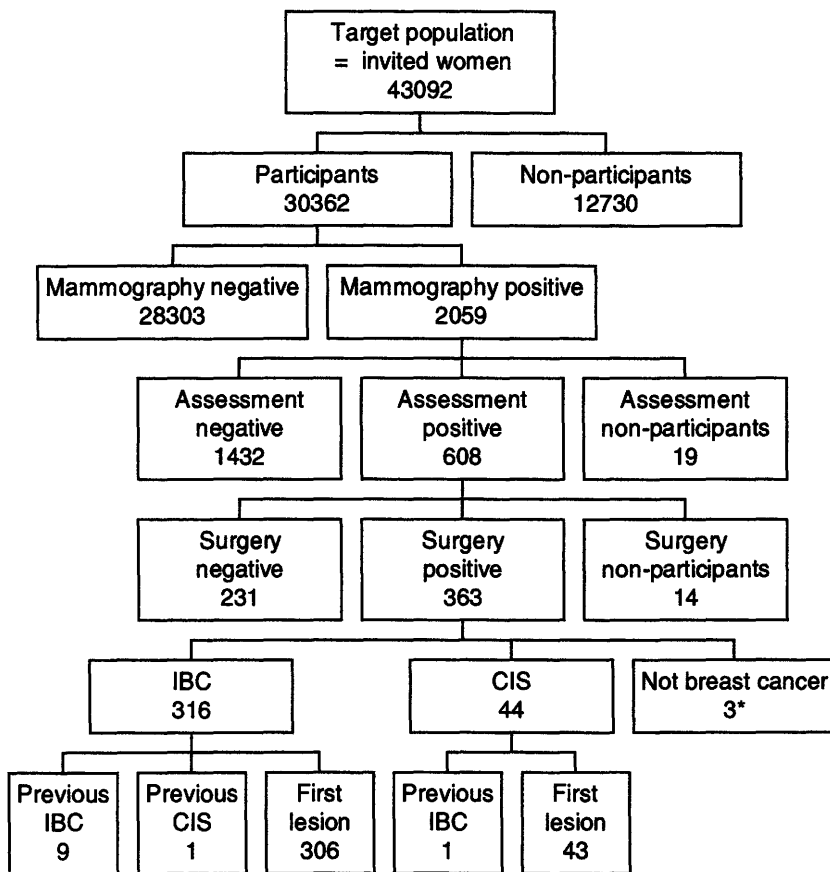
Figure 1. Mammography screening in Copenhagen first invitation round 4 April 1991 -
23 April 1993, women born 31 March 1921 - 1 April 1941 (modified from Lynge, APMIS
Suppl. 83, 106, 1998)

The expected number of breast cancer cases in women who tested negative was
calculated using Danish Cancer Registry incidence rates for invasive breast cancer
among women living in the municipality of Copenhagen in 1986-90, i.e., in the period
immediately before the start of the screening programme. The results are presented as
proportionate interval cancer rates where the observed number of interval cancers is
divided by the expected number.

A total of 52 invasive breast cancers were diagnosed after the first screening round. Of these, 16 occurred within the first year after the screening date, and 36 occurred within the second year after the screening date. The expected number of invasive breast cancer cases was 152.25. The proportional interval cancer rate was thus 52/152.25 = 0.34. The rate was 16/77.42 = 0.21 within the first year, and 36/74.84 = 0.48 within the second year (Table 1).

Table 1. Mammography screening in Copenhagen. Interval cancers after first invitation round (modified from Lynge, APMIS Suppl. 83, 106, 1998)

	Negative at mammography	Negative at assessment	Negative at surgery	Total
No. of women	28 303	1432	231	29 966
After 0-11 months				
Obs	13	3	0	16
Exp	73.10	3.72	0.60	77.42
O/E	0.18	0.81	-	0.21
95% CI	0.09-0.30	0.17-2.36	-	0.1-0.34
After 12-23 months				
Obs				
Exp	34	2	0	36
O/E	70.65	3.61	0.58	74.84
95% CI	0.48	0.55	-	0.48
	0.34-0.67	0.07-2.00	-	0.35-0.67
After 0-23 months				
Obs	47	5	0	52
Exp	143.75	7.33	1.18	152.25
O/E	0.33	0.68	-	0.34
95% C	0.25-0.40	0.22-1.60	-	0.26-0.45

Comparison with results from randomised controlled trials and other service programmes

Key indicators for the outcome of the first invitation round are presented in Table 2 together with data from other service screening programmes (Thurfjell and Lindgren, 1994; Hakama *et al.*, 1995; Lenner and Jonsson, 1997; NHS Breast screening programme, 1997; Sundhedsstyrelsen, 1997; Fracheboud *et al.*, 1998). The participation rate was low in the Copenhagen programme compared with that of other service programmes. The Copenhagen municipality is a completely urbanised area. A regional gradient in participation rates is observed in other countries as well. In the Netherlands, the participation rate was 69% in strongly urbanised areas compared with 82% in not urbanised areas (Fracheboud *et al.*, 1998). In England, a marked geographical gradient from a participation rate of 60% in North West Thames Region to 83% in Northern Region during the prevalence screening round almost disappeared during the incident screening rounds (NHS Breast screening programme, 1997). The gradient in the participation rate in Denmark from 71 % in Copenhagen to 88% in Fyn county during the prevalence screening round is, thus, in accordance with the experience in other countries.

There is a marked difference between the programmes in the percent of participants recalled for assessment, with England and Copenhagen at the top of the list. There is clearly a wider difference between the recall rates than between the incidence of breast cancer in the various regions. The differences in recall rates will therefore result in differences in the proportion of women with false positives results, and potentially also in different interval cancer rates.

Copenhagen has the highest detection rate of 11.8 per 1000 in the first invitation round. However, the detection rate in Fyn, at 9.8 per 1000, is close to that in Copenhagen, and the difference between these two Danish areas is compatible with the overall 15% regional difference in breast cancer incidence (Andreassen *et al.*, 1994). Twelve per cent of the cases detected in the Copenhagen programme were CIS. This proportion is relatively low compared with other screening programmes, and it reflects a deliberately conservative attitude towards supposedly benign microcalcifications.

Outcome measures

While data on breast cancer mortality continue to be unavailable, the potential outcome of a screening programme must be assessed from short-term surrogate measures (Day *et al.*, 1989). The detection rate in the screening programme compared with the background breast cancer incidence is a first indicator of the success of a programme.

169

Table 2. Results from the first invitation round in selected populations based on service screening programmes with mammography (modified from Lynge, APMIS Suppl. 83, 106, 1998)

	Copen-hagen	England	Holland	Uppsala	Finland	Nordbotten Väster - Norrland	Fyns Amt
Screening age group (years)	50-69	50-64	50-69	40-69	50-59	50-69	50-69
First invitation round	1991-93	1994-95	1990-95	1988-89	1987-88	1991-93	1993-95
Reference		1	2	3	4	5	6
Participants in percent of invited	71%	75%	79%	87%	88%	89%	88%
Percent of participants recalled for assessment	6.8%	7.2%	1.3%	4.6%	4.5%	2.1%	2.7%
Percent of participants with surgery	2.0%	0.7%	1.0% [1]	0.9% [2]	0.9%	NA	1.3%
Detected IBC + CIS per 1000 participants	11.8	5.9	6.8 [1]	4.8	4.7 [3]	NA	9.8
CIS in percent of IBC+ CIS	12%	19%	14% [1]	11%	-	NA	16%
Percent of recalled with IBC+ CIS	17%	8%	48% [1]	10%	8%	NA	36%
Percent of women with surgery with IBC+ CIS	60%	59%	66% [1]	53%	42%	NA	74%
Percent of false positive	5.6%	6.6%	0.7% [1]	4.1%	-	NA	1.8%

[1] These results refer to first screened women coming from different invitation rounds
[2] Aspiration cytology only
[3] Defined in original paper only as 'cancers'.

References :   1 NHS Breast Cancer Screening Programme, 1997
              2 Fracheboud *et al.*, 1998
              3 Thurfjell and Lindgren, 1994
              4 Hakama *et al.*, 1995
              5 Lenner and Jonsson, 1997
              6 Sundhedsstyrelsen, 1997

In the WE-trial these ratios were 3.09 for women aged 50-59 years and 4.59 for women aged 60-69 years (Day *et al.*, 1989). Table 3 shows the detection rates for invasive breast cancers and carcinoma *in situ* cases in Copenhagen compared with the incidence of invasive breast cancer in the period prior to the screening programme. The ratios between the two sets of rates varied from about 3 to about 6 for all age groups of women screened for the first time.

Table 3. Detection rates in the Copenhagen mammography screening programme compared with the breast cancer incidence in Copenhagen prior to screening (modified from Lynge, APMIS Suppl. 83, 106, 1998)

Age group (years)	50-54		55-59		60-64		65-69	
	Rate (per 10^4)	O/E	Rate (per 10^4)	O/E	Rate (per 10^4)	O/E	Rate (per 10^4)	O/E
Incidence of IBC Copenhagen 1986-90	21	(1)	24	(1)	29	(1)	27	(1)
Detection IBC +CIS First screen - first IR 1991-93 - second IR 1993-95 - third IR 1995-97	72 41 52	3.4 2.0 2.5	100 112 92^1	4.2 4.7 3.8^1	127 172 131^1	4.4 5.9 4.5^1	154 114^1 203	5.7 4.2^1 7.5^1
Detection IBC+ CIS Second screen - second IR 1993-95 - third IR 1995-97	47 25	2.2 1.2	38 79	1.6 3.3	70 56^1	2.4 1.9^1	77 214	2.9 7.9
Detection IBC+ CIS Third screen -third IR 1995-97	67^1	3.2^1	53	2.2	78	2.7	51	1.9

IBC: Invasive breast cancer
CIS: Carcinoma *in situ*
IR: Invitation round
[1]Based on less than 5 observed cases

The participation rate is a second key indicator for the potential success of a programme. It is common practice to calculate the participation rate as the number of screened women as a percentage of those invited. However, in Copenhagen it is

possible for women to ask not to be invited. The participation rate of concern for the potential reduction in breast cancer mortality in Copenhagen is therefore rather the number of screened women as a percent of the target population. Table 4 shows that the participation rate calculated, using the invited women as the denominator, has remained stable at about 70% throughout the three invitation rounds.

Table 4. Participation rates in the Copenhagen mammography screening programme (modified from Lynge, APMIS Suppl. 83, 106, 1998)

Invitation round	Participants as percent of target population	Participants as percent of invited women	Participants as Percent of invited, 'regularly screened' [1]
1	71 %	71 %	71 %
2	65 %	69%	83 %
3	63%	70%	90%

[1] 'regularly screened': those who participated in all previous rounds

Table 5 shows the proportionate interval cancer rates between the first and second invitation rounds from selected programmes. The data from the WE study in Sweden (Tabár *et al.*, 1992) are from the randomised trial, while the data from the two areas in England (Day *et al.*, 1995; Woodman *et al.*, 1995; McCann *et al.*, 1997) and from the Netherlands (Schouten *et al.*, 1998) are, like the Copenhagen data, from service screening programmes. The WE trial had a proportionate interval cancer rate of 0.24 within the first two years after the screening, at a 'cost' of 4.4% women with false positive results. The Copenhagen service programme did worse on both indicators, as the proportionate interval cancer rate was 0.34, at a 'cost' of 5.6 % false positives. However, compared with the English service programmes, Copenhagen did relatively well.

Population flow

In the evaluation of the prospects for a later reduction in breast cancer mortality in Copenhagen, the continuous flow in and out of the target population should be taken into account. The major part of this is of course due to the inclusion and exclusion of different birth cohorts in successive invitation rounds. In the long term, this would,

172

Table 5. Proportionate interval cancer rate and false positive percent in first invitation round of selected mammography screening programmes (modified from Lynge, APMIS Suppl. 83, 106, 1998)

Type of programme	Location	Ref.	Age years	Proportionate interval cancer rate			False positive %
				0-11 months	12-23 months	0-23 months average	
Service	Copenhagen		50-69	0.21	0.48	0.34	5.6
Trial	Sweden WE	1	50-69	0.17	0.30	0.24	4.4
Service	England, East Anglia	2	50-64	0.24	0.59	0.42	5.0
Service	England, N. Western	3	50-64	0.31	0.52	0.42	-
Service	Limburg, The Netherlands	4	50-64	0.31	0.60	-	-

References: 1 Tabár *et al.*, 1992
2 Day *et al.*, 1995, McCann *et al.*, 1997
3 Woodman *et al.*, 1993
4 Schouten *et al.*, 1998

however, not affect the programme as all women would be offered 11 screens between their 50th and 70th birthdays. There is, however, an additional population flow due to movements in and out of Copenhagen and due to deaths. To illustrate the size of this, we have looked at women born between 1 April 1925 and 1 April 1941 and, thus, potentially included in the target population for all three invitation rounds.

In total 34 405 women were born between 1 April 1925 and 1 April 1941 and were included in the target population for either the first, the second or the third invitation rounds. Among these women, 27 894, equal to 81%, were included in all the three

target populations. The remaining 6511 women were present in the municipality of Copenhagen for only part of the six years from April 1991 to March 1997. Of the 27 894 women, 15 898, equal to 57%, were screened three times. At the third invitation round, the 15 898 women who had been screened three times constituted 53% of the target population of women born between 1 April 1925 and 1 April 1941.

The relatively low participation rate, combined with the population flow explain why many of the breast cancer cases diagnosed in women aged 50-69 in Copenhagen after April 1991 do not come from the screening programme. Table 6 shows a cross tabulation of the incident invasive breast cancer cases detected in the screening programme and the incident invasive breast cancer cases diagnosed in Copenhagen in the age groups and periods which were the target of the screening programme. We have included here only the two first invitation rounds, because we wanted to use the data from the Danish Cancer Registry in order to ensure that all incident cases were included in the comparison. By the time of the study, the Danish Cancer Registry was only fully updated for 1994.

Table 6. Incident invasive breast cancer cases in women aged 50-69 in Copenhagen April 1991 to May 1995 by status in the screening programme (modified from Lynge, APMIS Suppl. 83, 106, 1998)

Status in the Danish Cancer Register	IBC in first invitation round	IBC in second invitation round	Screen detected CIS	Not screen detected	Total
Before April 1991	7	6	..	..	..
April 1991-April 19913	290	3	6	188	487
May 1993-May 1995	16	122	5	116	259
After May 1995	0	6	..	..	..
Missing	3	9	..	..	..
Total screen detected IBC	316	146	..	..	..

IBC: Invasive breast cancer
CIS: Carcinoma in situ
.. : Not relevant

Cancer register data

Table 6 shows that the majority of the screen detected cancers were found among the cases recorded in the same time period in the cancer register. A small number of women with screen detected cancers were registered in the cancer register with a date of diagnosis before screening. These are women with two primary breast cancers, and their second cancer detected at screening. As expected, a small number of women with screen detected cancers were found in the cancer register with dates of diagnosis after the end of their invitation round, due to delay in assessment and surgery. Finally, some screen-detected cancers were missing in the Danish Cancer Registry especially from the second invitation round which ended in May 1995. In total, 61 % of the 487 incident invasive breast cancers diagnosed in women aged 50-69 in Copenhagen at the time of the first invitation round were screen detected, and 55% of the 259 cases diagnosed at the time of the second invitation round.

Data from the Danish Cancer Registry was used to identify second primaries among the screen detected cancers, to calculate detection rates for each screening round, to calculate proportionate interval cancer rates and to measure the screen detected cancers as proportion of all breast cancers diagnosed in the area. Thus, this cohort study would not have been possible without access to long term data from a population-based cancer registry. However, the relatively slow process of producing complete cancer registry data for a calendar year was a problem and independent clinical databases with more up to date data were used in addition.

A cohort study of breast self-examination

A cohort of women who took part in the 'Mama' breast self-examination (BSE) screening program in Finland from 1973 through 1975 (with BSE used for screening, and mammography for diagnosis) was studied (Gastrin *et al.*, 1994). In total, 28785 women who returned calendars recording their practice of BSE over a 2-year period were followed by linkage with the records of the Finnish Cancer Registry through 1986. The incidence of and mortality from breast cancer was compared with that expected in the Finnish population, based on a model incorporating Finnish national data for breast cancer incidence and case fatality.

In the study cohort, breast cancer incidence was higher than expected (a rate ratio of 1.19 over all ages). The stage distribution of cases was not different from that expected from Finnish Cancer Registry data for 1980, but breast cancer mortality was lower than expected (a rate ratio of 0.75).

The authors state that the reduction in mortality from breast cancer in the study cohort was consistent with an effect of BSE, though selection bias, inherent in any observational study of screening, provides a plausible alternative explanation for the findings. The participants (as in other studies of breast cancer screening compliers) came from higher social classes and had a higher educational status than the non-participants or the general population. All cause mortality among the participants was lower than that among comparable general population (rate ratio 0.70). Thus, the lower breast cancer mortality among the participants (rate ratio 0.75) is difficult to ascribe to the screening. The conclusions have been challenged by another Finnish study using the survival analysis approach (Auvinen *et al.*, 1996). Another large study reported negative results in relation to BSE practice, as reported by 548 000 US women in 1959 (Holmberg *et al.*, 1997)

Conclusion on cohort studies assessing breast cancer screening

Breast cancer screening programmes are being implemented and planned in several countries. Population-based cancer registries will provide essential background data and later they will be involved in the monitoring of the programmes. The launching of a new programme provides an opportunity to plan large-scale population-based cohort studies. Thus, cancer registries should be active in the early design and planning phases of screening programmes, and seriously examine the possibilities of prospective (cohort) studies of outcome at the population level. Such studies can provide important information on the performance of the programme in a service setting, thereby suggesting the need for modification or improvement of different aspects, and justifying (or otherwise) the expenditure of funds on the programme. However, in addition to complete registration (so that all incident breast cancer cases are detected), and effective linkage to the screening register (to distinguish screen detected from interval cancers), cancer registries will generally have to improve upon the routine mechanisms for recording size and stage of registered cases. They will, in addition, have to devise methods for producing timely results with enhanced case-finding mechanisms for the cancer(s) of interest, and preparation of interim analyses before the usual "Annual Report", often delayed until reporting of all cancers is virtually complete.

References

Alexander FE, Anderson TJ, Brown HK, Forrest APM, Hepburn W, Kirkpatrick AE, Muir BB, Prescott RJ & Smith A (1999). 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* **353**: 1903-1908

Andreassen A, Andersen KW, Madsen M, Mouridsen H, Olesen KP & Lynge E (1994). Regional trends in breast cancer incidence and mortality in Denmark prior to mammography screening. *Br J Cancer* **70**: 133-137

Auvinen A, Elovainio L & Hakama M (1996). Breast self-examination and survival from breast cancer: a prospective follow-up study. *Breast Cancer Res Treat* **38**:161-168

Day NE, Williams DR & Khaw KT (1989). Breast cancer screening programmes: the development of a monitoring and evaluation system. *Br J Cancer* **59**: 954-958

Day N, McCann J, Camilleri-Ferrante C, Britton P, Hurst G, Cush S & Duffy S (1995). Monitoring interval cancers in breast screening programmes: the East Anglian experience. *J Med Screen* **2**:180-185

Ellman R, Moss SM, Coleman D & Chamberlain J (1993). Breast self-examination programmes in the trial of early detection of breast cancer: ten year findings. *Br J Cancer* **68**:208-212

Fracheboud J, De Koning HJ, Beemsterboer PMM, Boer R, Hendriks JHCL, Verbeek ALM, van Ineveld BM, de Bruyn AE & van der Maas PJ (1998). National Evaluation Team for Breast Cancer Screening (NETB). Nationwide breast cancer screening in the Netherlands: results of initial and subsequent screening 1990-95. *Int J Cancer* **60**:694-698

Gastrin G, Miller AB, To T, Aronson KJ, Wall C, Hakama M, Louhivuori K & Pukkala E (1994). Incidence and mortality from breast cancer in the Mama Program for Breast Screening in Finland, 1973-1986. *Cancer* **73**:2168-2174

Hakama M, Holli K, Isola J, Kallioniemi OP, Kärkkäinen A, Visakorpi T, Pukkala E, Saarenmaa I, Geiger U, Ikkala J, Nieminen T, Godenhjelm K & Koivula T (1995). Aggressiveness of screen-detected breast cancers. *Lancet* **345**:221-224

177

Hakama M, Pukkala E, Heikkila M & Kallio M (1997). Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *Br Med J* **314**:864-867

Holmberg L, Ekbom A, Calle E, Mokdad A & Byers T (1997). Breast cancer mortality in relation to self-reported use of breast self-examination. A cohort study of 450,000 women. *Breast Cancer Res Treat* **43**:137-140

Lenner P & Jonsson H (1997). Excess mortality from breast cancer in relation to mammography screening in northern Sweden. *J Med Screen* **4**:6-9

Mammography Screening Evaluation Group (1998). Mammography screening for breast cancer in Copenhagen April 1991-March 1997. *APMIS* **106** (Suppl. 83):1-44

McCann J, Wait S, Seradour B & Day N (1997). A comparison of the performance and impact of breast cancer screening programmes in East Anglia, UK, and Bouches du Rhône, France. *Eur J Cancer* **33**:429-435

Miller AB, Baines CJ, To T & Wall C (1992). Canadian National Breast Screening Study: 2. Breast cancer detection and death rates women aged 50 to 59 years. *Can Med Assoc J* **147**:1477-1488

NHS Breast screening programme (1997). *NHS Breast Cancer Screening Programme Review 1996.* Sheffield: NHS Breast screening programme

Nyström L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Rydén S, Andersson I, Bjurstam N, Fagerberg G, Frisell J, Tabár L & Larsson LG (1993). Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* **341**:973-978

Olsson S, Andersson I, Karlberg I, Bjurstam N, Frodis E & Hakansson S (2000). Implementation of service screening with mammography in Sweden: from pilot study to nationwide programme. *J Med Screen* **7**:14-18

Roberts MM, Alexander FE, Anderson TJ, Forrest AP, Hepburn W, Huggins A, Kirkpatrick AE, Lamb J, Lutz W & Muir BB (1984). The Edinburgh randomised trial of screening for breast cancer: description of method. *Br J Cancer* **50**:1-6

178

Schouten LJ, de Rijke JM, Schlangen JT & Verbeek AL (1998). Evaluation of the effect of breast cancer screening by record linkage with the cancer registry, The Netherlands. *J Med Screen* **5**:37-41

Shapiro S, Strax P & Venet L (1966). Evaluation of periodic breast cancer screening with mammography. Methodology and early observations. *JAMA* **195**:731-738

Sundhedsstyrelsen (1997). *Tidlig opsporing af brystkraaft. [Early detection of breast cancer]*. Kobenhavn, Sundhedsstyrelsen (in Danish)

Tabár L, Fagerberg G, Duffy SW, Day NE, Gad A &, Gröntoft O (1992). Update of the Swedish two-country program of mammographic screening for breast cancer. *Radiol Clin North Am* **30**:187-210

Thurfjell EL & Lindgren JAA (1994). Population-based mammography screening in Swedish clinical practice: Prevalence and incidence screening in Uppsala county. *Radiology* **193**:351-357

UK Trial of Early Detection of Breast Cancer group (1999). 16-year mortality from breast cancer in the UK Trial of Early Detection of Breast Cancer. *Lancet* **353**:1909-1914

Woodman CB, Threlfall AG, Boggis CR & Prior P (1995). Is the three year breast screening interval too long? Occurrence of interval cancers in NHS breast screening programme's north western region. *Br Med J* **310**:224-226

Chapter 12. Cancer registries in monitoring, evaluating and planning breast cancer screening programmes

D. Stockton and J. McCann

The aim of breast screening programmes is to reduce breast cancer mortality through early detection of breast cancers, which has been shown to improve the chance of successful treatment. Cancer registries can help ensure that a programme has every opportunity to achieve the mortality reduction by being involved in identifying problems in the programme at an early stage so that remedial action can be taken.

When a screening programme is introduced the anticipated mortality reduction will not be seen for a number of years but will depend largely on three aspects of the programme (Day et al., 1989):

a)   the compliance rate
b)   the rate of interval cancers
c)   the distribution of prognostic indicators among screen-detected and interval cancers

The routine monitoring of screening performance cannot easily be undertaken by cancer registries because of the time lag between diagnosis and registration. However, cancer registries can play a large role in evaluating the programme both looking at interim indicators and end point indicators of success (Figure 1).

The importance of the cancer registry is that it provides information on the whole population of women invited for screening - not just those who accept the invitation. Thus it provides information on the population impact of screening.

The following chapter looks at tools used to monitor and evaluate breast cancer screening programmes, with particular attention to areas where cancer registries can make a contribution.
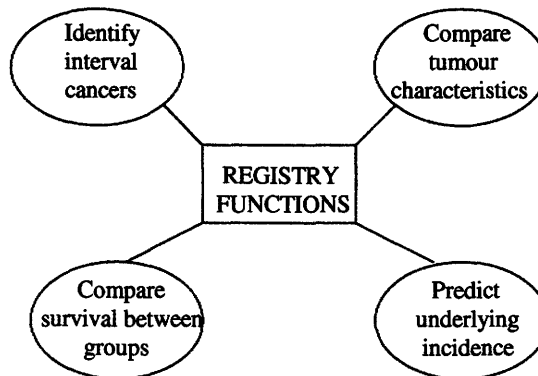
Figure 1. The key cancer registry functions in evaluating screening programmes

Monitoring compliance

Even though it is usually not the role of the cancer registry to monitor compliance of women attending screening, it is important when evaluating a programme to be aware of compliance rates as programme success relies on consistently high compliance over progressive screening rounds.

The non-compliers (or non-attenders) will include women who did not receive an invitation to come for screening due to administrative errors (wrong address details etc.) and those who chose not to attend. In the latter group there are those who have never attended screening and those who have attended a screening test but subsequently stopped attending.

Administrative errors
These can be minimised, particularly by checking population lists for accuracy.

Never attender
It has been widely seen that non-attenders have a worse prognosis than unscreened control populations (Duffy *et al.*, 1991). This may be due, in part, to women who already have symptoms not attending screening due to denial or fear. Other non-attenders may not appreciate the importance of regular mammography and they should be encouraged by providing more information on the benefits of screening, sending out appointment reminder letters, and by having a more flexible appointment system (Lidbrink *et al.*, 1995).

It has been shown that compliance can be improved by re-inviting, either with an open or fixed appointment, women who did not attend their initial invitation in a particular screening round, particularly if the letter is sent via or signed by their personal physician (Dinnes et al., 1997). Specific groups of women may need to be targeted in different ways, for example, one study found that variations in compliance by general practice were closely related to social deprivation and there was some evidence that contact with a female GP was beneficial (Gatrell et al., 1998).

Monitoring interval cancers

Interval cancers are related both to the sensitivity of a screening programme (i.e. how good the programme is at picking up cancers) and the length of the inter-screening interval (frequency with which screening is performed). The monitoring of interval cancers requires the results of, at least, the second round of screening in order to define the results over an entire screening round, i.e. interval cancers from just after the prevalence screen to just before the second screening, plus the cancers detected at the second screening. This group of cancers is termed the 'unbiased set' (Tabár et al., 1992). The initial prevalence screen must be excluded due to length bias, i.e., the detection, at screening, of a large number of good prognosis cancers which might never present symptomatically in a woman's lifetime (Morrison, 1992).

The cancer registry is essential for identifying interval cancers. Sensitive record linkage between the cancer registry and screening programme is necessary to ensure that cases are not missed. It should be possible to categorise all cancers registered at the cancer registry, in women in the screening age range, into one of the following categories of interest (McCann et al., 1998):

a) Screen-detected cancer
b) Interval cancer
c) Cancer in non-attender
d) Cancer in lapsed attender
e) Cancer in woman not yet invited

The sensitivity and specificity of the record linkage should be continually assessed, as it is important to accurately categorise the cancers, particularly the screen-detected and interval cancers, otherwise monitoring and evaluation of the programme will be compromised. Problems will arise if the registry does not have high case ascertainment or does not cover the whole screened population. For example, if the registry case ascertainment is in general quite low but the

ascertainment of screen-detected cases is high (due to good information exchange between the registry and screening programme), bias will be introduced making the rate of screen-detected cancers look artificially high in comparison to the rate of cancer in the other categories. The screening programme may also link to death records, hospital admission and discharge records or pathology laboratories to ensure that no cases are missed.

Once all the identified cancers have been assigned to one of the above-mentioned categories, the rate of interval cancers can be calculated using woman-years at risk as the denominator. Three different methods for calculating the woman-years at risk have been compared by the Scottish breast screening programme (Everington *et al.*, 1999):

Method 1: Number of women screened negative.
Method 2: Adjusting the number of women screened negative at three time periods (1, 2 and 3 years after the last screen) by removing women who had been re-screened or diagnosed with cancer before that time period. Thus, all women contribute one, two or three years to the woman-years at risk.
Method 3: Person-time elapsed between the date of the last negative screen and the date of re-screening or diagnosis of cancer, truncated at 3 years if this occurs later than 3 years.

They estimated the underlying incidence rate by (a) extrapolating 1978-87 trends within each age band up to 1991 and then using the 1991 estimate for all subsequent years, and (b) using an age-period Poisson regression model. The number of cancers expected in the absence of screening was then calculated using the three estimates of the person-years and the two estimates of the underlying incidence rates.

In their analysis, they found that the expected number of cancers estimated by the computationally intense methods (methods 2 and 3 with age-period Poisson regression models) were not substantially different from those obtained by method 1 with trend extrapolation (Table 1).

The definition of an interval cancer can have an important effect on the calculated interval cancer rates. If a programme is falling behind in the invitation schedule and re-inviting women beyond the agreed screening interval, should interval cancers that occur during this "slippage time" (i.e. beyond three years) be included? In evaluating the overall success of a screening programme, it would seem appropriate to include interval cancers arising in slippage time. This is because slippage may be seen as a failure of the programme.

184

Table 1. Comparison of the Scottish results for women screened in 1991-1992 using the different methods of calculating the underlying incidence (method a or b) and person-years at risk (methods 1, 2 or 3)

Methods (as described in text)	0 - <12 months since last negative screen		12 - <24 months since last negative screen		24 - <36 months since last negative screen	
	Rate	*Exp.*	*Rate*	*Exp.*	*Rate*	*Exp.*
1 and a	**4.6**	*182*	**11.5**	*182*	**12.3**	*182*
2 and a	**4.6**	*181*	**11.8**	*178*	**12.7**	*176*
2 and b	**4.6**	*183*	**11.8**	*183*	**12.7**	*186*
3 and a	**4.6**	*179*	**11.8**	*176*	**13.5**	*165*
3 and b	**4.6**	*179*	**11.8**	*181*	**13.5**	*175*

The interval cancer rates per 10 000 women years at risk are shown in bold and the expected number of cancers are shown in *italics*

However, when evaluating the effectiveness of a screening test in detecting cancers, it may be important not to include interval cancers arising outside the screening interval. Round slippage will vary between screening programmes (Faux *et al.*, 1998) and so it is important to study the interval cancers occurring both in the agreed interscreening interval and in slippage time when evaluating the appropriateness of the screening interval.

If interval cancer rates are higher than expected, based on set standards[1], special studies should be performed to evaluate the extent to which this is due to low sensitivity of the programme (e.g., the quality of mammography, the experience of the radiologist, or the problems when women are recalled for assessment) or the screening interval being too long. Alternatively, it may be necessary to revise the standards (Moss and Blanks, 1998)

---

[1] Current standards are derived from screening trials rather than population-based screening programmes and may be artificially high. Now that population-based screening programmes have been running a number of years these standards should be refined.

Monitoring tumour characteristics

If cancer registries are to help monitor a screening programme, it is very important that they collect information on tumour pathology including histological type, grade, size and nodal status. It is advantageous if this has also been collected for cases diagnosed before the screening programme began.

Tumour characteristics that determine breast cancer survival have been identified. Comparison of the distribution of these characteristics between the cancers diagnosed over a screening cycle and those diagnosed in an unscreened population gives a direct estimate of the effect of screening in improving prognosis. If there is a shift towards a more favourable distribution of these tumour characteristics, then it is likely that screening will have the desired effect of reducing mortality.

Duffy et al. (1991) investigated which prognostic factors recorded in the Swedish two-county study had the largest effect on survival, using multivariate analysis. They showed that the favourable prognosis of screen-detected (not including prevalent screen) cancers could be largely accounted for by three tumour characteristics: size, nodal status and grade.

High interval cancer rates indicate that a screening programme is not performing as well as it could be. However, comparing the tumour characteristics of interval cancers with those in the other categories, especially those among women in the 'not yet invited' category, should indicate whether a screening programme with high interval cancer rates might still be successful in reducing mortality. Burrell et al. (1996) compared the size, grade and lymph node status of interval cancers in their programme with cancers in an unscreened control group and with screen-detected cancers over the same time period. They found that the interval cancers had a significantly worse distribution of tumour characteristics than the screen-detected cancers but a similar distribution to the control group cancers, indicating that the interval cancers are no worse than those that would arise in the absence of screening. Their programme should, therefore, have some impact on mortality, but not as great as would be the case with a lower interval cancer rate. Frisell et al. (1992) reported that interval cancers in their programme also had a similar distribution of prognostic indicators to those in an unscreened control group, but the overall survival by stage was consistently higher among the patients diagnosed with interval cancers compared with those in the control group. In the Swedish two-county study, the patients with interval cancers had slightly better survival rates than the controls (Duffy et al., 1991).

Collins et al. (1998) investigated whether the survival rate for women with interval cancers in their programme was different from the rates for women diagnosed with

186

cancer in an unscreened population. Interval cancer rates had been higher than expected and it was of interest to determine what effect this would have on the reduction in mortality in the population invited for screening. When choosing an unscreened population to use as controls, they first ruled out two groups (1) non-attenders, because they had been shown to have a worse outcome than controls in the Swedish two-county study, and (2) historical cases, because of recent advances in managing breast cancer. Due to the phased introduction of screening in their programme, they were able to use women not yet invited to screening as their control group. They found no significant difference in survival between the patients with interval cancers and the controls, again indicating that, in their programme, the interval cancers were not a subset of more aggressive tumours.

As well as monitoring interval cancers, it is important to study the distribution of tumour characteristics in other categories. However, care must be taken when reporting analyses by prognostic categories. For example, in the first round of screening a wide variety of tumour types and sizes are picked up and it is expected that many small and fairly benign tumours will be detected that would either not have presented clinically until a long time into the future or might never have presented at all (so called 'length bias'). For this screen, studying the size distribution of tumours would give an artificial appearance of a big reduction in large tumours. Until the screening programme is stable, i.e. the overall incidence rate has come back down to a level approaching the expected underlying incidence rate (although it might be a bit higher due to a "detection age shift" - older women with higher incidence being detected at an earlier age), tumour characteristics should not be monitored using proportions.

Furthermore, when investigating the distribution of prognostic indicators, one needs an overall picture of how the programme is performing. Therefore, in evaluating a specific screening round, sufficient time should be allowed, i.e. up to the next screening round, so that all interval cancers can be included in the analysis. There may be a very favourable distribution of prognostic indicators in the screen-detected cancers, but this will not give a good indication of how the programme is performing, if interval cancer rates are high or these cancers have a particularly poor prognosis. As screening will detect slow-growing cancers with good prognosis, whereas fast-growing cancers will become apparent between screens, we would expect the distribution of prognostic indicators to be very different between the screen-detected and interval cancers.

In Table 2, the incidence rates by TNM stage are shown, both prior to and during screening, for the East Anglian programme, UK (McCann et al., 1998). The programme was phased in from 1989 and most women had been invited for screening for the first time by the end of 1993. By 1995, virtually the entire population of women aged 50-69 would have received at least one invitation to

screening. As the table shows, during the first screening round the rates of early and advanced stage cancers are difficult to interpret, particularly with the effect of the staggered introduction; thus, comparing the proportions would be misleading. In 1995 most women were undergoing their second screening but the total incidence for the 50-69 year olds was still quite a bit higher than the expected underlying incidence (see next section) so the data should still be interpreted with caution.

The rate of advanced cancer obtained with the screening programme in place can be compared with the expected underlying incidence rate of advanced cancer (see next section) to evaluate how the programme is performing. If the programme is to be successful in reducing mortality, an early indicator would be a lower rate of advanced cancer compared with the estimated underlying incidence rate of advanced cancer.

Table 2. Incidence rates of invasive breast cancer by TNM stage in the 50-69 year age group. Rate per 100 000 women in the East Anglian region

Year of diagnosis	1981-1986	1987-1988	1989	1990	1991	1992	1993	1994	1995
Early stage (stage 1)	55.5	76.1	89.1	141.0	159.8	141.9	123.6	141.3	118.8
Advanced (stages 2,3,4)	131.4	162.1	158.9	140.5	129.2	109.9	133.7	135.4	131.8
Total (incl. unknown stage)	196.4	245.0	256.2	288.8	297.3	257.4	262.8	281.7	253.3
Proportion of advanced	70%	68%	64%	50%	45%	44%	48%	49%	53%

Schouten *et al.* (1998) found that, immediately after the introduction of screening, the annual number of breast cancer diagnoses increased by almost 50% and then

188

decreased to previous levels after completion of the first screening round. Evaluating cases diagnosed in the second screening round, they compared incidence by prognostic factor with that seen in a period directly before screening began (1987-1990). They found that the incidence of node positive tumours was 1% lower in 1994 and 15% lower in 1995, indicating that the screening programme was having the desired effect of improving prognosis, which should lead to a mortality reduction.

Hakama et al. (1995) compared size, nodal status and histological type of breast cancers in a population of women invited to screening with those diagnosed among unscreened controls. They designed the study to avoid length bias by excluding cancers detected in the first screening round. They found that the cancers detected in the population invited to screening (screen-detected, interval and non-attender cancers combined: the unbiased set) had a better prognosis than the controls, indicating that their programme should also bring about a reduction in breast cancer mortality.

Monitoring and interpreting time trends

An estimate of what the incidence of breast cancer would have been in the absence of screening (the underlying incidence rate) is essential in order to evaluate the performance of a breast screening programme. For example, in order to estimate the magnitude of a reduction in late stage cancers, an estimate of what the rate of late stage cancers would have been in the absence of screening is needed. In order to calculate the proportionate incidence of screen-detected and interval cancers, an estimate of what the overall rate of breast cancer would have been in the absence of screening is required. Cancer registries are very important for evaluating screening programmes that do not have a contemporaneous unscreened control group since they can estimate the "underlying" incidence rates. The three most widely used methods for estimating the underlying incidence are:

a) Extrapolating pre-screening trends in incidence using statistical modelling procedures.
b) Using, as fixed, the rate of breast cancer seen directly before screening began, i.e. assuming that incidence in the absence of screening would have been static.
c) For programmes with a phased introduction of screening, using the rate of cancer seen in women of screening age who have not yet been invited to screening.

Of these, the most commonly used method is (1) as it can be adopted by any registry that has collected numerous years of accurate data prior to the start of

189

screening. If trends by tumour characteristics are to be extrapolated, then complete information about these must also have been collected historically by the registry. If the incidence rates do not increase equally in all age groups, the trends must be analysed separately for each age group or a more complicated model should be adopted.

Before the introduction of screening, an upward trend in the incidence of breast cancer was observed in many countries. Some studies suggested that cohort, rather than calendar period, effects are the cause of this upward trend (Quinn et al. 1995). If it is known that strong cohort effects were in force, trend extrapolation will not be valid. However, many years of data are needed to demonstrate cohort effects and, as this quantity of data is not always available, most studies still use trend extrapolation to estimate the underlying incidence rate.

Prior et al. (1996) used an age-period model to predict the underlying incidence rate. The age-specific rates were fitted using Poisson regression models with the population as a weighting factor, and a year by age interaction term was added. Prediction intervals were calculated to reflect the statistical uncertainties that arise when extrapolating beyond the range of available data. From their analysis, they concluded that the rate of breast cancer was increasing prior to the introduction of screening and that the use of one incidence rate for all age groups was not appropriate since the rate of increase in incidence varied between age groups. The observed and predicted underlying incidence of breast cancer from their analysis is shown in Figure 2.
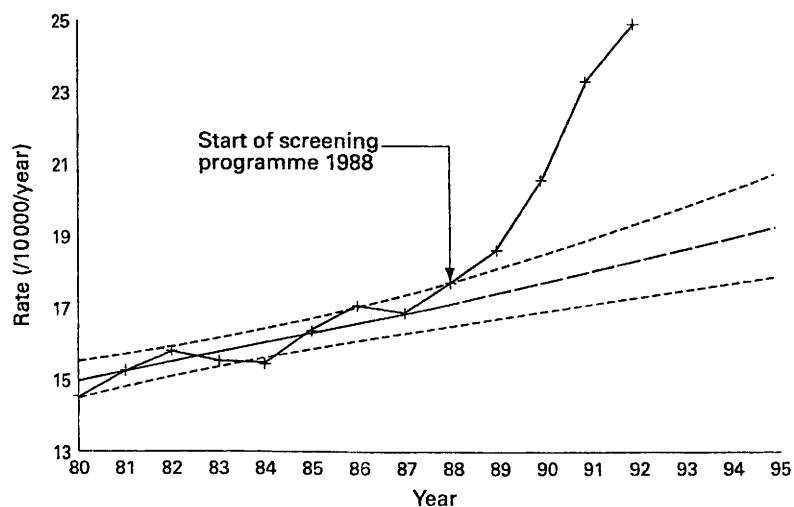

Figure 2. Oberved and predicted underlying incidence of breast cancer in the UK (Prior et al., 1996, J Med Screen 3:119-122; by permission of BMJ Publishing Group)

190

The importance of "exploring" the data before performing trend analyses cannot be stressed enough. The genuineness of increasing trends in incidence must be established since observed increases could actually be due to improved reporting to the registry or a change in the diagnostic procedures. If there is any doubt about the validity of the data, trend extrapolation must not be used.

McCann et al. (1998) found a marked and unexplained increase in the incidence of advanced and total breast cancer in the two years prior to the introduction of screening in East Anglia, UK, compared with only a slight upward trend seen over the preceding 10 years. This complicated their trend extrapolation. The difference between the trends when including and excluding the data for these two years is shown in Figure 3.
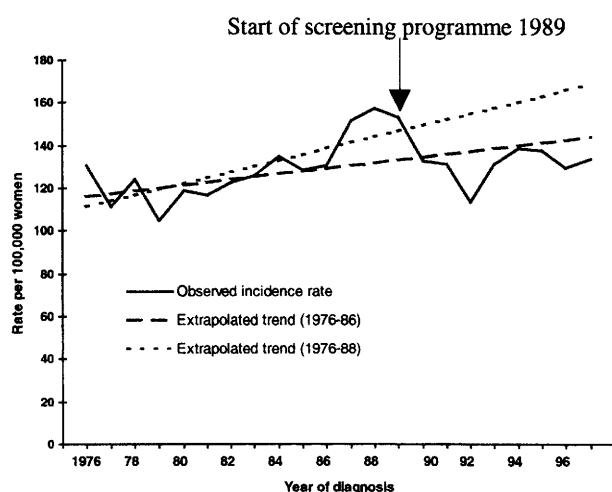

Figure 3. Extrapolated trends (1976-86 and 1976-88) showing the possible range of the underlying incidence rate of advanced breast cancer in women aged 50-69 in East Anglia

In this study, the analysis of the data using methods (2) or (3) described earlier was also complicated by the increase in incidence in 1987 and 1988 since it was unclear whether the increase was real (and so would have continued) or was an artefact. Indeed, looking at the predicted rate of advanced cancers using the three

methods above, the predicted rates using data for 1987 and 1988 are very similar, whilst the predicted rate excluding 1987 and 1988 shows a different picture (Table 3). Which is correct? Interpretation of results is not always simple and analyses should be embarked upon with caution!

Table 3. Predicted rate of advanced cancer in 1997 in East Anglia using the three methods to calculate the estimates

Method	Description of method	Observed rate	Expected rate	Percent reduction (%)
1	Extrapolating advanced cancer incidence trends from 1976-86	129.3	136.5	5.3
1	Extrapolating advanced cancer incidence trends from 1976-88	129.3	162.8	20.6
2	The proportion of advanced cancer in 1988 applied to the 1997 cases	129.3	162.5	20.4
3	The proportion of advanced cancer in women not yet invited to screening applied to the 1997 cases	129.3	163.5	20.9

In conclusion, routine monitoring of breast screening programme performance is largely undertaken by the programme itself due to the inevitable time lag between cancer diagnosis and registration, and the restricted number of data items collected by cancer registries. However, cancer registries can and do play a very important role in evaluation, particularly of the impact of screening in the whole of the target population.

References

Burrell HC, Sibbering DM, Wilson AR, Pinder SE, Evans AJ, Yeoman LJ, Elston CW, Ellis IO, Blamey RW & Robertson JF (1995). Screening interval breast cancers: mammographic features and prognosis factors. *Radiology* 3:811-817

Collins S, Woodman CB, Threlfall A & Prior P (1998). Survival rates from interval cancer in NHS breast screening programme. *Br Med J* 316:832-833

Day NE, Williams DRR & Khaw KT (1989). Breast cancer screening programmes: the development of a monitoring and evaluating system. *Br J Cancer* 59:954-958

Dinnes J, Smith C, MacIntyre L & Cooper G (1997). Breast Screening: Factors which influence attendance and strategies to increase uptake. Scottish Health Purchasing Information Centre Report

Duffy SW, Tabár L, Fagerberg G, Gad A, Gröntoft O, South MC & Day NE (1991). Breast screening prognosis factors and survival - results from the Swedish two-county study. *Br J Cancer* 64:1133-1138

Everington D, Gilbert FJ, Tyack C & Warner J (1999). The Scottish breast screening programme's experience of monitoring interval cancers. *J Med Screen* 6:21-27

Faux AM, Lawrence GM, Wheaton ME, Wallis MG, Jeffery CL & Griffiths RK (1998). Slippage in the NHS breast screening programme: an assessment of whether a three year screening round is being achieved. *J Med Screen* 5:88-91

Frisell J, von Rosen A, Wiege M, Nilsson B & Goldman S (1992). Interval cancer and survival in a randomised breast cancer screening trial in Stockholm. *Breast Cancer Res Treat* 24:11-16

Gatrell A, Garnett S, Rigby J, Maddocks A & Kirwan M (1998). Uptake of screening for breast cancer in south Lancashire. *Public Health* 112:297-301

Hakama M, Holli K, Isola J, Kallioniemi OP, Karkkainen A, Visakorpi T, Pukkala E, Saarenmaa I, Geiger U & Ikkala J *et al.* (1995). Aggressiveness of screen detected breast cancers. *Lancet* 345:221-224

Lidbrink E, Frisell J, Brandberg Y, Rosendahl I & Rutqvist LE (1995). Non-attendance in the Stockholm mammography screening trial: relative mortality and reasons for non-attendance. *Breast Cancer Res Treat* **35**:267-275

McCann J, Stockton D & Day NE (1998). Breast cancer in East Anglia: the impact of the breast screening programme on stage at diagnosis. *J Med Screen* **5**:42-48.

Morrison AS (1992). *Screening in Chronic Disease.* 2nd edition. New York, Oxford University Press

Moss S & Blanks R (1998). Calculating appropriate target cancer detection rates and expected interval cancer rates for the UK NHS Breast Screening Programme. Interval Cancer Working Group. *J Epidemiol Community Health* **52**:111-115

Prior P, Woodman CB, Wilson S & Threlfall AG (1996). Reliability of underlying incidence rates for estimating the effect and efficiency of screening for breast cancer. *J Med Screen* **3**:119-122

Quinn M & Allen E (1995). Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. *Br Med J* **311**:1391-1395

Schouten LJ, de Rijke JM, Schlangen JT & Verbeek AL (1998). Evaluation of the effect of breast cancer screening by record linkage with the cancer registry, The Netherlands. *J Med Screen* **5**:37-41

Tabár L, Faberberg G, Duffy SW, Day NE, Gad A & Gröntoft O (1992). Update of the Swedish two-county programme of mammographic screening for breast cancer. *Radiol Clin North Am* **30**:187-210

Chapter 13. Population-based trends of prostate cancer in the United States before and after widespread use of PSA

R.A. Stephenson

Introduction

In the late 1980s, physicians in the United States and elsewhere began to recognise the potential usefulness of prostate specific antigen (PSA) in the detection of prostate cancer (Catalona *et al.*, 1991; Crawford and DeAntoni., 1993; Catalona *et al.*, 1994). Based on the promise of enhanced prostate cancer detection, use of PSA testing rapidly gained acceptance and widespread use in the United States. A substantial literature now clearly demonstrates improved prostate cancer detection with the addition of PSA testing to previously available detection methods (Catalona *et al.*, 1991; Crawford and DeAntoni, 1993; Catalona *et al.*, 1994). PSA based detection is now widely prevalent in the United States (Potosky *et al.*, 1995). In some regions of the United States by the mid 1990s, nearly 50% of all men over the age of 40, and 70% of men over age 70 had undergone PSA testing (Mansfield *et al.*, 1997; Close *et al.*, 1998).

Subsequent to widespread use of PSA based detection unprecedented changes in prostate cancer incidence rates were observed. These changes in incidence had substantial effects on the characteristics of diagnosed prostate cancer cases through lead time and perhaps length time effects. The magnitude of these changes in incidence are unique in the history of modern oncology. While PSA based community screening was a part of early prostate cancer detection efforts, most PSA based detection is now part of the routine pattern of care for men in the US Clearly, despite these trends in the United States, no convincing evidence is available to support or refute PSA based detection strategies for prostate cancer. Population-based mortality effects, and survival analyses from randomised screening trials are not currently available. Unfortunately, we are left with a situation in which the appropriate role of PSA testing remains to be defined. In this setting of persisting uncertainty, different philosophies for detection and management of prostate cancer have emerged from different regions of the world. The United States is notable for proactive PSA based detection and aggressive treatment, while Scandinavian countries are notable for minimal use of PSA detection and largely conservative treatment approaches.

In this paper, population-based temporal trends in prostate cancer incidence, diagnosis, grade, stage, age, and mortality data obtained from the Surveillance,

195

Epidemiology and End Results (SEER) Program are being used to describe the effect of widespread use of PSA testing in the United States (Stephenson, 1998; Stanford *et al.*, 1999).

Incidence

As seen in Figure 1, the incidence of prostate cancer rose gradually in the United States from the inception of the SEER Program in 1973 through the late 1980s (Stephenson, 1998; Stanford *et al.*, 1999). This gradual rise in incidence can be attributed either to changing environmental risk factors within the US population, or to gradual improvement or increased use of diagnostic methods. Rates of TURP (transurethral resection of the prostate) diagnosed prostate cancer rose gradually through 1987 in the United States (Merrill *et al.*, 1999). The rise in TURP diagnosed prostate cancer was due to increasing use of TURP for treatment of benign prostatic hypertrophy (BPH). Historically, 10 to 20% of BPH cases treated with TURP will have incidental prostate cancers identified in the removed pathological material (Newman *et al.*, 1982). Merrill and colleagues indicate that TURP diagnosed cancers account for nearly all of the increasing incidence rates of
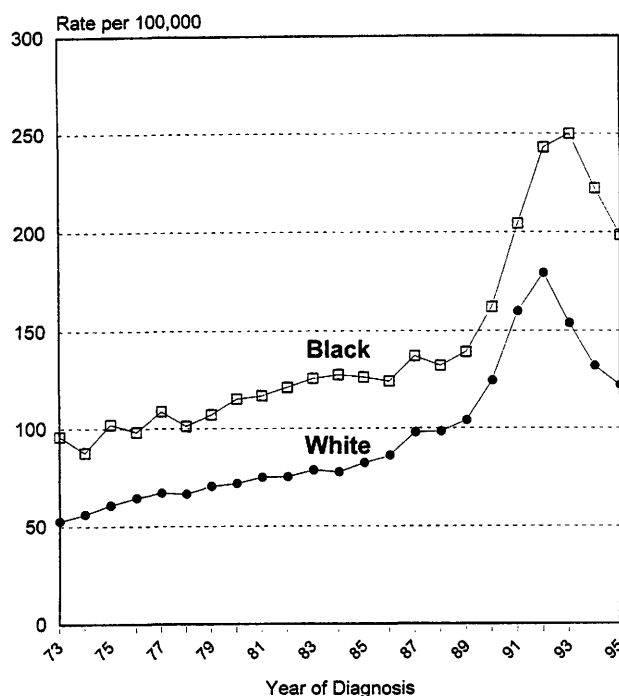

Figure 1. Prostate cancer. Incidence rates. SEER Program 1973-1995

prostate cancer from 1973 through 1987 (Merrill *et al.*, 1999). After 1987 the role of TURP in the diagnosis of prostate cancer has steadily declined. The rise and decline in TURP diagnosed prostate cancer rates are probably related to the rise and decline of TURP utilisation for treatment of BPH. While the rise in TURP utilisation through 1987 likely relates to increasing urological manpower in the United States, the decline in TURP utilisation probably relates to partial replacement of TURP by effective pharmacological treatments for BPH or less invasive surgical procedures that do not generate pathological material for examination.

Beginning in 1988 or 1989 prostate cancer incidence rates rose dramatically in the United States (Figure 1). These years are coincident with the rapid acceptance and use of PSA in screening and routine clinical detection. Incidence rates peaked in 1992 and declined rapidly to rates approaching those prior to the widespread use of PSA. A change in incidence of this magnitude has never been previously observed for any cancer. While the increased prostate cancer incidence rates of 1988 to 1992 are probably related to increased use of PSA, the reasons for the decline in incidence after 1992 are less clear (Figure 1). Factors related to declining PSA rates appear to include: (1) a cull phenomenon where fewer PSA detectable cancers are identified in repeatedly screened individuals and populations; (2) enthusiasm for continuing screening among practitioners may have declined as they recognised declining prostate cancer case yields among populations which were substantially contaminated by previously screened, low prostate cancer yield individuals; (3) screening activity in the United States may also have been inhibited by publications in the lay and professional media which questioned the wisdom of PSA based prostate cancer detection; and (4) little may have been done to reach individuals who were not inclined to seek PSA testing or other forms of health care, resulting in a large fraction of the population in whom PSA testing was never performed. The relative contributions of these four factors and others are yet to be adequately measured and described.

Grade

Figure 2 shows temporal trends in prostate cancer grade (Stephenson, 1998, Stanford *et al.*, 1999). In the SEER Program Gleason Scores are assigned in the following fashion: Gleason Scores 2-4 - good differentiation, Gleason Scores 5-7 - moderate differentiation and Gleason Scores 8-10 - poor differentiation. A remarkable trend toward moderate differentiation is seen as the PSA era moves forward. Approximately 75% of the increased incidence of the PSA era are accounted for by moderately differentiated tumours, while only 8% of the increase were due to well-differentiated tumours. Even after 1992 when incidence rates
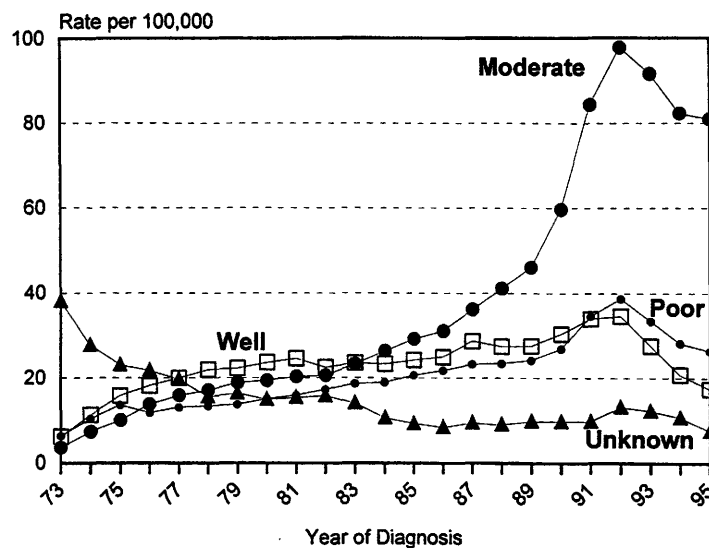
Figure 2. Prostate cancer. Incidence rates by grade. SEER Program 1973-1995

were in decline, the fraction of cases with moderate differentiation continued to increase. This predominance of moderate differentiation during the PSA era suggests that PSA is largely unable to detect the large prevalent pool of well-differentiated prostate cancer. Low-grade misclassification of tumours by relatively inexperienced pathologists in the 1980s would have had a negligible effect on the overall observed PSA era changes in incidence rates by grade. Only 8% of the increase in incidence of the PSA era were due to well-differentiated tumours, with roughly 1/3 of those classified as well differentiated. Even if all low-grade cancers were incorrectly graded in the 1980s the contribution to the overall increased incidence by well-differentiated tumours in the PSA era would have increased by only a few percentage points. While poorly differentiated cancers were detected in increased numbers during the PSA era the magnitude was small relative to moderately differentiated tumours. This is probably due primarily to the smaller prevalence of poorly differentiated tumours in the population.

These trends in prostate cancer grade are somewhat counter-intuitive with respect to length time issues. Typically during screening, cancers of even lower biological potential are detected due to repeat screening effects where slower growing cancers are more easily identified. As measured by grade no such effect is evident in SEER grade data. TURP diagnosed cancers of the pre-PSA era were commonly assigned well differentiated (typically designated by the stage A1 category). Such cancers

198

are uncommonly identified in the PSA era. PSA appears to do the ideal thing by identifying cancers of greater biological significance while concurrently identifying these cancers at time points when they are more amenable to therapy. It is important to state that grade is only an approximate or surrogate indicator of biological relevance or potential. It is highly likely that many cancers of the PSA era have been identified which have insufficient biological potential to require treatment.

Stage

Distant stage rates have declined by more than 50% in the PSA era (Figure 3) (Stephenson, 1998; Stanford et al., 1999). Distant disease stage rates have declined in all age categories. Node positive rates (stage D1) are also declining according to data from the Utah Cancer Registry (Mansfield et al., 1996). Concurrently, rates of local and regional stage have increased substantially. Unfortunately SEER abstraction rules make separating local and regional stages difficult. Nevertheless, rising SEER prostate cancer incidence rates taken together with falling rates of distant stage indicate the presence of substantial trends toward early diagnosis during the PSA era. Whether this documented stage migration results in beneficial lead time effects where early diagnosis leads to more effective treatment, or whether it results in lead time bias where no impact on mortality is observed remains to be resolved with future mortality data.

Declining distant prostate cancer stage rates may be due to a cull phenomenon where distant cases are either removed from the prevalent population, or, due to early detection do not reach distant stage at diagnosis, or both. It may also be possible that health care providers during the PSA era have become less inclined to screen elderly men who are more likely to have distant disease at diagnosis. However, data from the Utah SEER registry reveal that elderly men are more likely to undergo PSA testing than younger men (Mansfield et al., 1997). This indicates that disinclination to screen older men is not a significant factor in declining distant stage rates.

Declining distant stage rates are a necessary but insufficient condition to be certain of future declines in mortality. Nevertheless, the fall of distant stage rates suggest that future declines in prostate cancer mortality are quite likely.
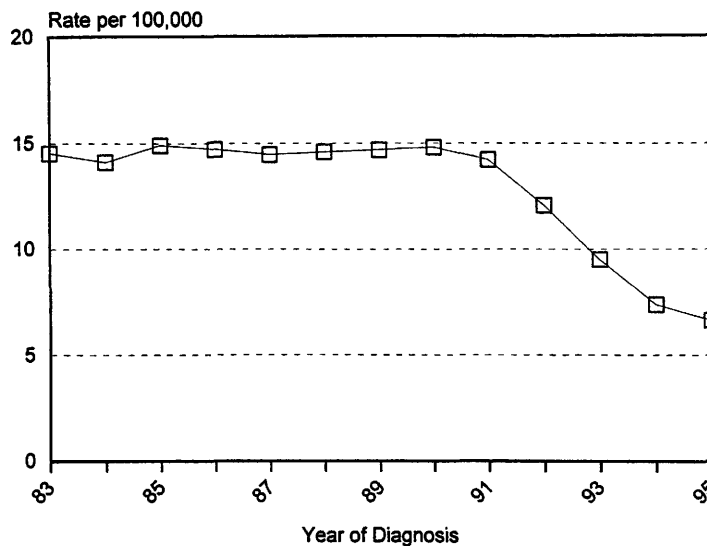
Figure 3. Prostate cancer. Incidence rates of distant stage. SEER Program 1983-1995

Age

Age at diagnosis has fallen sharply in the PSA era (Figure 4) (Stephenson, 1998; Stanford *et al.*, 1999). An abrupt fall in age began in 1990 and has continued through 1995, the last year for which information is available. Mean age at diagnosis has declined by 2.9 years from 71.9 to 69.0 years of age. As in the staging data presented above, declines in mean age at diagnosis reveal that a substantial lead-time effect has occurred in the United States during the PSA era. As can be seen in Figure 4, mean age at diagnosis is likely to continue to fall in coming years based of the steep slope of present mean age curves 1990 through 1995. In a large cohort of men with serially banked sera prior to the PSA era, retrospective analysis of PSA testing and prostate cancer rates established estimated lead time effects of approximately 4 to 5 years with use of PSA (Gann *et al.*, 1995).

200
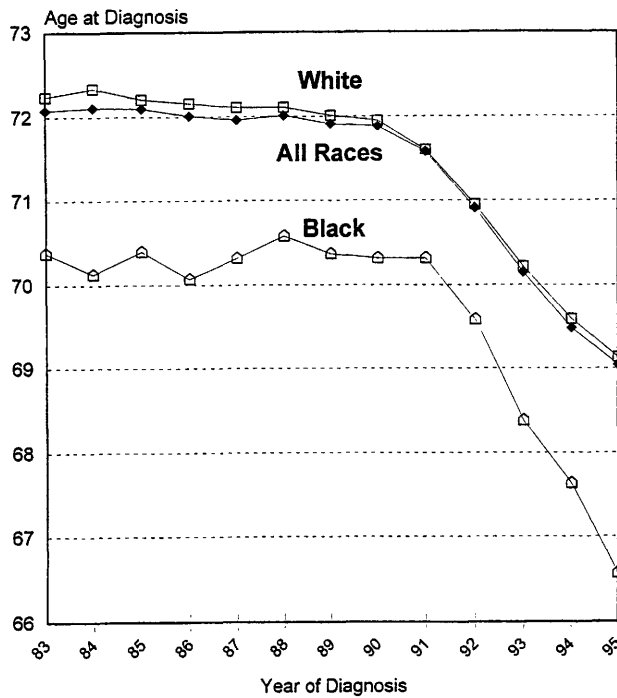
Figure 4. Prostate cancer. Mean age at diagnosis. SEER Program 1983-1995

Treatment

The majority of patients diagnosed in the PSA era received aggressive treatment (radical prostatectomy, radiation therapy) (Figure 5) (Stephenson, 1998; Stanford *et al.*, 1999). Radical prostatectomy became the most common type of treatment for local and regional prostate cancer in 1991. Figure 6 demonstrates how treatment choice is related to age in the US. Choice of radical prostatectomy is inversely related to age, while no treatment and hormone therapy are directly related to age. Choice of radiation therapy is biphasic with a peak in its use among men in their 70s. There is little evidence for a left or right shift in age of treatment choice when comparing years 1988-89 to 1993-94, suggesting that the decision tree for treatment as a function of age has not changed significantly during the PSA era in the US.

Rate per 100,000

Radical Prostatectomy

Radiation Therapy

No Treatment

Hormone Therapy

Other Treatments

60
50
40
30
20
10
0

88 89 90 91 92 93 94 95

Year of Diagnosis
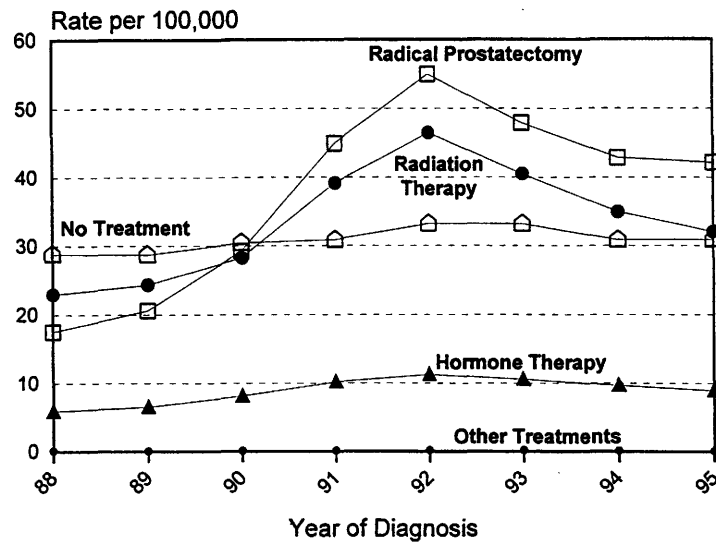
Figure 5. Prostate cancer. Incidence by treatment of localised and regional stages. SEER Program 1988-1994

Mortality

Age-adjusted US mortality rates peaked in 1991 and declined 7.3% by 1995 (Figure 7) (Stephenson, 1998; Stanford *et al.*, 1999). For men dying at ages of 75 or less, mortality rates declined by 15% during the same time interval. The magnitude of the decline in mortality is currently quite small. It is not expected that the large number of early stage prostate cancers which have been diagnosed and treated in the PSA era will have a substantial impact on mortality for several more years. It is clear that during the PSA era large numbers of men were diagnosed with prostate cancer and received treatment. It is unknown what fraction of these men were overtreated (therapy had no impact on outcome), or what fraction needed and benefited from therapy. The steepness of the decline in future mortality rates will be a measure of the relative number of cases from these two sub-populations.
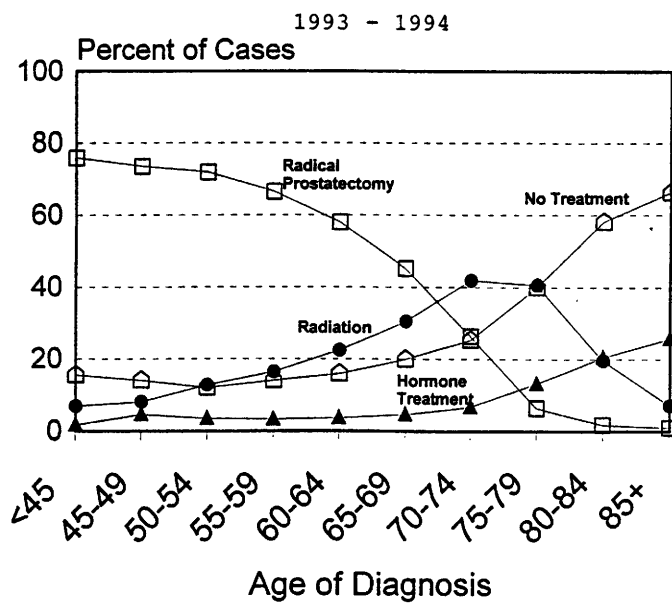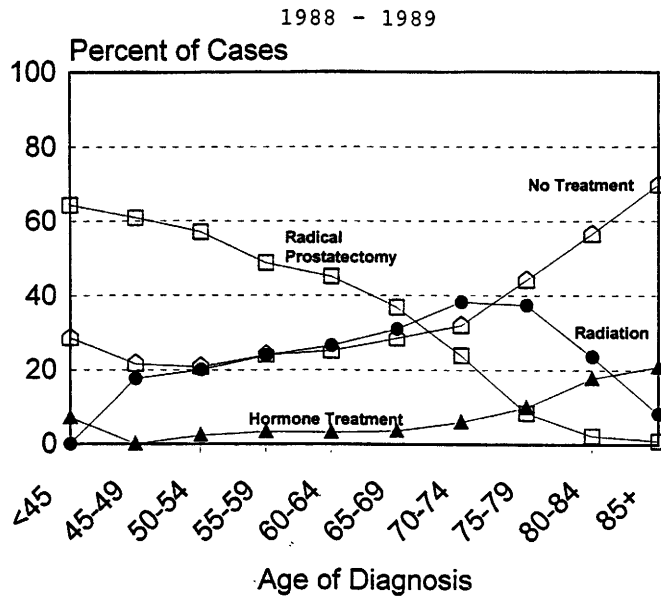
1988 - 1989

1993 - 1994

Figure 6. Prostate cancer. Distribution of cases by age and treatment for localised and regional stage. SEER Program 1988-1989 and 1993-1994
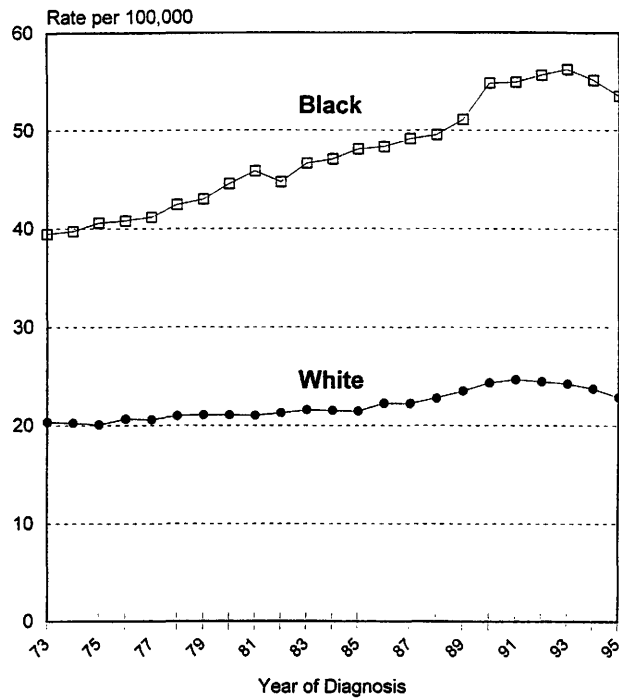
Figure 7. Prostate cancer. US mortality rates 1973-1995

Conclusions

Several important observations can be garnered from SEER prostate cancer data: (1) Prostate cancer incidence rapidly rose and fell during the PSA era (1988-1995) in the United States. The magnitude of those incidence changes is unprecedented in the history of oncology and is a reflection of enthusiastic use of PSA as a new detection method for prostate cancer in the United States. (2) Moderately differentiated tumours became the predominant grade in the PSA era. Fortuitously, this suggests that PSA is unable to detect the large prevalent pool of low-grade prostate cancer. (3) Distant stage rates fell by over 50%. This observation is a necessary but, by itself, insufficient precondition to observe certain declines in mortality rates in the future. (4) Age at diagnosis fell by more than three years. Taken together with stage migration effects, these age effects demonstrate a powerful lead time effect of PSA. (5) Large numbers of men with local and regional stage disease received aggressive therapy for prostate cancer in the US. (6) Mortality rates have fallen by only a small increment to date. Based on the

204

prolonged natural history of prostate cancer, any substantial effect of PSA detection on mortality is not yet expected.

These observations raise interesting questions regarding how to deal with problems of incomplete information and persisting uncertainty. In the case of PSA, enhanced detection of prostate cancer has been clearly demonstrated, but improved mortality outcomes are yet to be observed. How should individuals and societies approach such problems? Decision-making must often be made using inherently incomplete information in combination with the hopes and fears of the society's members. As examples, the societies of Sweden and the United States have arrived at distinctly different approaches to the problem of prostate cancer detection and treatment. While many argue that resolution of the uncertainty regarding PSA detection/screening can only be achieved by means of randomised screening trials, it is possible that convincing differences will be seen in population-based mortality data from regions with distinctly different management strategies. This is similar to the case with Pap smear testing where randomised screening trials were never performed, but convincing population-based mortality data were used to justify the use of Pap testing (Johannesson *et al.*, 1978; Läärä *et al.*, 1987; Benedet *et al.*, 1992; Sigurdsson, 1993). It appears likely that prostate cancer mortality data from these two countries and others will give clear answers to these troubling questions long before results from randomised screening trials are available.

References

Benedet JL, Anderson MB & Matisic JP (1992). A comprehensive program for cervical cancer detection and management. *Am J Obstet Gynecol* **166**:1254-1259

Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Juan JJ, Petros JA & Andriole GL (1991). Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N Engl J Med* **324**:1156-1161

Catalona WJ, Richie JP, Ahmann FR, Hudson MA, Scardino PT & Flanigan RC (1994). Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. *J Urol* **151**:1283-1290

Close DR, Kristal AR, Li S, Patterson RE & White E (1998). Associations of demographic and health-related characteristics with prostate cancer screening in Washington State. *Cancer Epidemiol Biomarkers Prev* **7**:627-630

Crawford ED & DeAntoni EP (1993). PSA as a screening test for prostate cancer. *Urol Clin North Am* **20**:637-646

Gann PH, Hennekens CH & Stampfer MJ (1995). A prospective evaluation of prostate-specific antigen for detection of prostate cancer. *JAMA* **273**:289-294

Johannesson G, Geirsson G & Day N (1978). The effect of mass screening in Iceland, 1965-1974, on the incidence and mortality of cervical carcinoma. *Int J Cancer* **21**:418-425

Läärä E, Day NE & Hakama M (1987). Trends in the mortality from cervical cancer in the Nordic countries: association with organized screening programmes. *Lancet* **1**:1247-1249

Mansfield JT, Lignel E, Andenoro JD, Spencer SL, Southwick AW & Stephenson RA (1997). The prevalence of PSA in the State of Utah. *J Urol* **157**:367A

Mansfield JT, Soper TH, James BC, Smart CR, Middleton RG & Stephenson RA (1996). Grade and stage migration in prostate cancer - pathologic findings in biopsy and radical prostatectomy specimens: population-based analysis from the Utah Cancer Registry. *J Urol* **155**:418A

Merrill RM, Feuer EJ, Warren JL, Schussler N & Stephenson RA (1999). The impact of transurethral resection of the prostate on population-based prostate cancer incidence. *Am J Epidemiol* **150**:848-860

Newman AJ, Graham MA, Carlton CE & Lieman S (1982). Incidental carcinoma of the prostate at the time of transurethral resection: importance of evaluating every chip. *J Urol* **128**:948-950

Potosky AL, Miller BA, Albertsen PC & Kramer BS (1995). The role of increasing detection in the rising incidence of prostate cancer. *JAMA* **273**:548-552

Sigurdsson K (1993). Effect of organized screening on the risk of cervical cancer: evaluation of screening activity in Iceland. *Int J Cancer* **54**:563-570

Stanford JL, Stephenson RA, Coyle LM, Cerhan J, Correa R, Eley JW, Gilliland F, Hankey B, Kolonel LN, Kosary, C, Ross R, Severson R & West D (1999). *Prostate Cancer Trends 1973-1995*, SEER Program (NIH Publication No. 99-4543), Bethesda, National Cancer Institute (See also www.seer.ims.nci.nih.gov)

Stephenson RA (1998). Population-based prostate cancer trends in the PSA era: data from the Surveillance, Epidemiology, and End Results (SEER) Program. *Monographs in Urology* **19**:1-19

Chapter 14. Evaluation of screening for prostate cancer

F.E. Alexander

Commentary

Stephenson acknowledges in his detailed evaluation of prostate cancer (PC) incidence and mortality trends from the US SEER database that the US prostate cancer experience is very different from the European: in Europe opinion is much less favourable towards PSA screening and PC mortality rates continue to increase. At present we have insufficient scientific evidence to determine whether these two are causally associated.

There are, potentially, three types of evidence on which a conclusion that PSA screening reduces PC mortality may be based: (I) randomised controlled trials (RCT); (II) other epidemiological studies – case-control and cohort studies and temporal and/or geographical comparisons; (III) the consensus of expert scientific opinion. These are not equivalent but represent a hierarchy with (I) being the gold standard.

I write as a European epidemiologist who has been actively involved in the era where evaluation of breast cancer screening by RCTs has pioneered research into cancer screening (Wald et al., 1993; Nyström et al., 1993; Alexander et al., 1999); this experience leads me to emphasise the necessity for the firmest evidence regarding PSA screening – that which comes from RCTs. On the basis of the evidence from RCTs, over 22 countries have organised programmes of mammographic screening (Shapiro et al., 1998). Several of these countries (e.g., Quinn and Allen, 1995; Smith et al., 1998) have reported reductions in breast cancer mortality but the authors are, justifiably, doubtful whether these can be attributed to the use of screening; for example, the reductions occur earlier than would be expected from an effect of screening as appears possible for the recent US reductions in prostate cancer mortality.

Substantial over-diagnosis from prostate cancer screening (i.e. men diagnosed and treated who, if unscreened, would have died of other causes before symptoms of PC emerged) is virtually certain from existing data (McGregor et al., 1998; Zappa et al., 1998; Alexander, 1997). Of 100 men with prostate cancer detected by screening up to age 70, only 16 would have died from PC before their 85th birthday (McGregor et al., 1998). It is clear that the human, social, psychological and economic costs of population-based screening programmes will be large. It is

essential that the <u>benefits</u> be quantified rigorously. RCTs, alone, can do this (Miller and Alexander, 1999).

Two RCTs of prostate cancer screening are in progress: the European Randomised Screening for Prostate Cancer (ERSPC) and the US Prostate, Colon, Lung and Ovary (PCLO) screening trial (Auvinen *et al.*, 1996; de Koning *et al.*, 2000). Together they have now randomised 160 000 men. The first analysis of ERSPC mortality data has been planned for 2008 when 120 000 men will have completed 10 years of follow-up. It is necessary to have this follow-up completed because PC often develops slowly; population-based series of locally confined disease have reported >75% survival at 10 years from diagnosis (Lu-Yao and Yao, 1997). The end-point for these trials, as for other trials of screening, is disease specific mortality. Since the expected number of prostate cancer deaths in the trials is small, accurate classification is essential. Reviewing all deaths is clearly impracticable; relying on death certificates alone may be unreliable. A sensible middle way involves cause-of-death review of all deaths, which may involve prostate cancer – those of known PC cases and those with PC on the death certificate. The assistance, here, of good quality cancer registration is invaluable since it allows accurate and unbiased ascertainment of prostate cancer cases in the two arms of the trial. Cancer registries are assisting ERSPC in several countries, including the Netherlands, Sweden and Finland.

RCTs of screening can, also, determine which 'process measures' are reliable early indicators of future mortality benefit. For breast cancer (Tabár *et al.*, 1985) cumulative incidence of 'distant' disease were less predictive than rates of disease of more modest stages ($\geq pT_2$ and/or pN_1). We do not yet know which benchmark will be appropriate for PC but the declining rates of distant stage disease (see Figure 3 in Chapter 13 of this volume) need not be as promising as first appears. Two things are, however, clear: firstly, Stephenson is absolutely correct in avoiding the temptation to base his arguments on increasing <u>proportions</u> of good prognosis cases and, secondly, ascertainment from cancer registries of cases in the trial populations is a prerequisite of scientific investigation of process measures.

Stephenson's data on grade (Figure 2) are particularly interesting. Typically, due to length biased sampling, initial (prevalence) screening detects large numbers (and high proportions) of cancers with markers of low biological aggressiveness (Tabár *et al.*, 1992a; Anderson *et al.*, 1991; Roberts *et al.*, 1987). Subsequent cancers arising between screens and detected at later screens are, when taken together, relatively free of length bias though their diagnosis has been moved forward in time. They have, therefore, favourable characteristics for the so-called 'chronological factors' (e.g., size and node status). Whether mammographic screening can be expected to lead to lower rates of high grade cancers has been controversial (Alexander *et al.*, 1997; Hakama *et al.*, 1995) but Tabár *et al.* (1992b) argue strongly that advancing the diagnosis should achieve this. Stephenson's data

suggest that PSA screening can advance the diagnosis to times when cancers are moderately but not well differentiated. He believes this to be 'ideal'; we shall in future know from the RCTs of screening what targets for grade should be applied in population screening programmes – if, that is, the trial results justify such programmes. Meanwhile, the key message to the scientific, public health and urological communities is that the ERSPC and PLCO trials are of enormous importance. It is not unreasonable to expect that, when their mortality results become available, analyses of their subsidiary data will enable future clinicians to have at their disposal scientifically valid criteria to select, for curative therapy, a minority of cancers detected by PSA screening. If so, the benefits (if any) demonstrated in the RCTs can be obtained at much reduced cost.

References

Alexander FE (1997). Rationale for trials of screening for prostate cancer. In: Schröder FH (eds). Recent advances in prostate cancer and BPH. In: *Proceedings of the IV Congress on Progress and Controversies in Oncological Urology* (PACIOU IV), held in Rotterdam, The Nederlands, April 1996. Lancaster, The Parthenon Publishing Group

Alexander FE, Anderson TJ & Hubbard AL (1997). Screening status in relation to biological and chronological characteristics of breast cancers: a cross sectional survey. *J Med Screen* **4**:152-157

Alexander FE, Anderson TJ, Brown HK, Forrest APM, Hepburn W, Kirkpatrick AE, Muir BB, Prescott RJ & Smith A (1999). 14 years of follow-up from the Edinburgh randomised trial of breast cancer screening. *Lancet* **353**:1903-1908

Anderson TJ, Lamb J, Donnan P, Alexander FE, Huggins A, Muir BB, Kirkpatrick AE, Chetty U, Hepburn W, Smith A, et al. (1991). Comparative Pathology of breast cancer in a randomised trial of screening. *Br J Cancer* **64**:108-113

Auvinen A, Rietbergen JBW, Denis LJ, Schröder FH & Prorok PC (International Prostate Cancer Screening Trial Evaluation Group) (1996). Prospective evaluation plan for randomized trials of prostate cancer screening. *J Med Screen* **3**:97-104

de Koning HJ, Auvinen A, Berenguer Sanchez F, Calais de Silva F, Ciatto S et al. (2000). Large-scale randomized prostate cancer screening trials; program performances in the ERSPC-and PLCO-trials (European Randomized Screening for Prostate Cancer-and Prostate, Lung, Colorectal and Ovary cancer trials). *J Natl Canc Inst* (in press)

Hakama M, Holli K, Isola J, Kallioniemi OP, Karkkainen A, Visakorpi T, Pukkala E, Saarenmaa I, Geiger U, Ikkala J, et al. (1995). Aggressiveness of screen detected breast cancers. *Lancet* **345**:221-224.

Lu-Yao GL &Yao SL (1997). Population-based study of long-term survival in patients with clinically localised prostate cancer. *Lancet* **349**:906-910

McGregor M, Hanley JA, Boivin JF & McLean RG (1998). Screening for prostate cancer: estimating the magnitude of over-detection. *Can Med Ass J* **159**:1368-1372.

Miller A & Alexander FE, on behalf of The International Prostate Screening Trial Evaluation Group (1999). Rationale for randomized trials of prostate cancer screening. *Eur J Cancer* **35**:262-271

Nyström L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Rydén S, Andersson I, Bjurstam N, Fagerberg G, Frisell J, et al. (1993). Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* **341**:973-978

Quinn ME & Allen E (1995). Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. *Br Med J* **311**:1391-1395

Roberts MM, Hawkins RA, Alexander FE, Anderson TJ & Steele RJ (1987). Oestrogen receptor activity in breast cancers detected at a prevalence screening examination. *Breast Cancer Res Treat* **10**:267-272

Shapiro S, Coleman EA, Broeders M, Codd M, de Koning H, Fracheboud J, Moss S, Paci E, Stachenko S & Ballard-Barbash R (1998). Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. *Int J Epidemiol* **27**:735-742

Smith CL, Kricker A & Armstrong BK (1998). Breast cancer mortality trends in Australia: 1921 to 1994. *Med J Aust* **168**:11-14

Tabár L, Fagerberg CJ, Gad A, Baldetrop L, Holmberg LH, Gröntoft O, Ljunquist U, Lundström B, Manson JC, Eklund G, et al. (1985). Reduction in mortality from breast cancer after mass screening with mammography. Randomized Trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet* **1**:829-832

Tabár L, Fagerberg G, Duffy SW, Day NE, Gad A & Gröntoft O (1992a). Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am* **30**:187-210

Tabár L, Fagerberg G, Day NE, Duffy SW & Kitchin RM (1992b). Breast cancer treatment and natural history: new insights from results of screening. *Lancet* **339**:412-414

Wald NJ, Chamberlain J & Hackshaw A, on behalf of the Evaluation Committee (1993). Report of the European Society for Mastology Breast Cancer Screening Evaluation Committee. *The Breast* 2:1-8

Zappa M, Ciatto S, Bonardi R & Mazzotto A (1998). Overdiagnosis of prostate carcinoma by screening: An estimate based on the results of the Florence Screening Pilot Study. *Ann Oncol* **9**:1292-1300

Chapter 15. Evaluation of screening for colorectal cancer

J. Faivre, A.M. Benhamiche and M.A. Tazi

Colorectal cancer fulfils the conditions required for mass screening. It is a major cause of morbidity and mortality in industrialised countries. However, it can be cured by the detection of early stage cancers and even prevented by the removal of adenomas. This situation has prompted considerable research efforts over the last 15 years to evaluate the ability of screening tests to decrease colorectal cancer mortality. Over the last five years more and more results have been published that suggest the benefits of screening. Cancer registries have been of great importance in planning and evaluating population-based studies. The purpose of this paper is to report their role in the planning, evaluation and monitoring of mass screening programmes for colorectal cancer.

Use of cancer registries in planning screening programmes

Descriptive epidemiological data provided by cancer registries proved useful in designing screening programmes for colorectal cancer. Data from cancer registries helped to indicate that colorectal cancers represented a major health problem and to focus attention on it. It is a major cause of illness in North America, Western Europe, Australia and New Zealand (Parkin et al., 1997). Japan is now also among the high-risk areas. Incidence data supplied by the European Network of Cancer Registries enable us to estimate the number of new cases per year at around 198000 in the 15 member states of the European Union in 1995 (EUCAN, 1999). Colorectal cancer is uncommon before the age of 50 (less than 6% of cases). Incidence increases rapidly from this age onwards. Thus the average risk population was defined as subjects over age 50. Considering these data and the results of screening trials it was recommended to implement screening in asymptomatic adults aged 50 and over by the European Group for Colorectal Cancer Screening (1999). Data from Cancer registries indicate that in Western Europe the risk of developing a colorectal cancer before the age of 74 is nearly 4.1% and 2.6% among males and females respectively (EUCAN, 1999). The cumulative risk of colorectal cancer in first degree relatives of a patient having a colorectal cancer before 50, or having two first-degree relatives affected, has been estimated to be over 10% (Benhamiche, 1998). On this basis the French consensus conference on colorectal cancer concluded that a screening colonoscopy was advisable for this population (Conférence de consensus, 1998). In the event of a

colorectal cancer after the age of 60 (with a risk slightly higher than in the general population) no advice was given about screening strategy, given the present state of knowledge.

Screening methods

Several tests and procedures have been proposed for the screening of colorectal cancer. The most commonly used is the faecal occult blood test. Most of these tests are guaiac-based tests, which are intended to detect the peroxidase-like activity of haemoglobin. The most extensively evaluated test is the Hemoccult II (Smith Kline Diagnostic, California). Two slides are prepared from three consecutive stool samples with or without dietary restrictions. This test is easy to perform, without great inconvenience to the individual, and is inexpensive. If any of the slides are positive, a complete colonoscopy is recommended. Sensitivity in detecting cancer with a non-rehydrated test and biennial screening in populations over 50 is situated between 50 and 60% for cancers (Young and St John, 1992) and between 20 and 30% for adenomas larger than 1 cm in diameter (Bertario *et al.*, 1988; Macrae and St John, 1982). The true positive rate is between 40 and 50%. Rehydratation increases sensitivity but also decreases specificity, so that the predictive accuracy of a positive test becomes very low (Young and St John, 1992).

More complex faecal occult blood tests, particularly immunochemical tests specific for human haemoglobin, have been developed. They are more sensitive, but their specificity at a population level is not well established. They are more expensive and not as suitable for a mass screening procedure.

Periodic sigmoidoscopy has been recommended by some organisations, whereas colonoscopy is rarely considered for individuals at average risk. The theoretical advantages of endoscopic screening include its high diagnostic sensitivity and specificity. However, it is unpleasant for the individual, has the risk of perforation, is expensive and compliance is not known. It does not fulfil the criteria generally required for a mass screening procedure. On-going studies in England and Italy will indicate the effectiveness of screening with sigmoidoscopy (Atkin *et al.*, 1993). Double contrast barium enema has the same drawbacks as endoscopy as well as lower sensitivity and specificity. The rest of this article will be confined to results from the faecal occult blood test.

Case-control studies

Six case-control studies have been conducted is order to estimate the efficacy of screening with faecal occult blood tests on colorectal cancer mortality. Three of them were population-based: one in Italy in the province of Florence (Zappa *et al.*, 1997), one in Germany, in Saarland (Wahrendorf *et al.*, 1993) and one in France in Burgundy (Faivre *et al.*, 1999a) (Table 1). Cases were less likely to have been screened than controls. Two studies indicate a significant protective effect for those screened within three years of case diagnosis compared with those not screened. No reduction in risk existed when considering longer screening intervals. Such findings lie behind current screening recommendations and suggest that the faecal occult blood test should be repeated at least every two years. The Italian study suggests a reduction in colorectal cancer mortality of 40% and the French study, a reduction of 33%. In the German study, in which results were reported by sex, mortality reduction was 57% for females and 8% for males for those asymptomatically screened 6-36 months before the reference date (Table 1). This difference in the protective effect of the faecal occult blood test between males and females was not reported in the other two studies. The reason for the discrepancy seen in Saarland has not been explained.

Table 1. Results of population based case-control studies

	Proportion screened		
	Cases	Controls	Odds Ratio
Florence	22%	29%	0.6 (0.4 - 0.9)
Burgundy	49%	61%	0.7 (0.5 - 0.9)
Saarland			
Males*	13%	14%	0.9 (0.5 - 1.6)
Females*	16%	29%	0.4 (0.3 - 0.7)

* Asymptomatically screened only

The main advantage of these population-based case-control studies, performed with the data from cancer registries, is the opportunity to include all deaths from colorectal cancer in a defined population and to hold information of similar quality on the history of participation in screening campaigns for both cases and controls.

Matching cases and controls by sex and birth year as well as by place of residence decreased the effect of socio-demographic and lifestyle risk factors that could have exaggerated the efficacy estimates. Prior screening can be a confounding factor in such studies, however before the studies began, no cancer screening programme existed in the study areas and faecal occult blood tests were not available to individuals. Cases and controls were unlikely to have been screened previously. Bias towards an apparent benefit of screening might operate when individuals with a lower risk of developing colorectal cancer are more likely to be screened. Data from the controlled part of the Burgundy study indicate little difference in the incidence of colorectal cancer between non-participants and the control group. It is thus unlikely that selection bias of this type accounts for a substantial part of the reported effect of screening. However, the difficulty in accounting for all relevant confounding factors could still limit the accuracy of case-control studies. We must not forget that as case-control studies compare participants with non-participants, they provide an indication of reduction in risk, which is independent of compliance rates. This means that the results are valid for 100% compliance.

Population-based controlled studies

There are four European controlled trials that compare colorectal cancer mortality between the study and the control group (Hardcastle et al., 1996; Kronborg et al., 1996; Kewenter et al., 1994; Tazi et al., 1997) (Table 2). Data from the cancer registries covering the area represented one source of information on diagnosed colorectal cancers. In Burgundy the researchers from the cancer registry were in charge of the design, set-up, collection of the data and analysis of the controlled trial.

The Funen, Nottingham and Gothenburg trials randomly allocated individuals or households identified from general practitioners records or population registers, to receive an invitation to participate in screening with Hemoccult, or to a control group. In the Burgundy trial, people from small administrative areas called «cantons» were allocated either to the screened or to the controlled population. The Minnesota study is not presented here as it was performed in volunteers.

Together with the efficacy of the screening test, compliance is a major determinant in the effectiveness of a screening programme. In Nordic countries and in England, the test is mailed with one or two reminders if required. In France this strategy resulted in a low compliance. It had to be combined with the active participation of primary care physicians who recommend the test to their patients. Compliance with the first screening campaign was 67% in Funen, 66% in Gothenburg, 53% in Burgundy and in Nottingham. The proportion of the population screened at least

216

once was 67%, 68%, 69% and 60% respectively. In total, 46% completed five screenings in Funen, 31% in Burgundy. In Nottingham 38% completed all screenings, i.e., three to six according to period of recruitment. Screening was limited to two rounds in Gothenburg.

Table 2. Population-based trials of Hemoccult screening for colorectal cancer

	Funen Denmark	Nottingham England	Gothenburg Sweden	Burgundy France
Study population	61 933 45-74 years	152 850 50-74 years	68 308 60-64 years	91 553 45-74 years
Screening test	Hemoccult unhydrated biennially	Hemoccult unhydrated biennially	Hemoccult most rehydrated 2 rounds	Hemoccult unhydrated biennially
Complete screening	67% did at least 1 screen, 46% completed 5 screens	60% did at least 1 screen, 38% completed all (3-6) screens	68% did at least 1 screen, 60% completed the 2 screening rounds	69% did at least 1 screen, 37% completed 5 screens
Positivity rate	1st screen 1.0%	1st screen 2.1%	1st screen 6.3%	1st screen 2.1%
Positivity predictive value for colorectal cancer	12.2%	11.5%	4.7%	11.4%
Proportion of colorectal cancer TNM Stage 1	Screen 22% Control 11%	Screen 20% Control 11%		Screen 29% Control 21%
Years of trial follow-up	10	median 7.8	median 8.3	9
Relative risk (95% CI) of colo-rectal cancer death with screening	0.82 (0.68 - 0.98)	0.85 (0.74 - 0.99)	0.88 (0.69 - 1.12)	0.86 (0.71 - 1.03)

The variation of positivity rates according to trials was related to the method of slide preparation. The positivity rate of the Hemoccult test in the initial screening was 1.1% in Funen where the test was performed with diet restriction, 2.1% in Nottingham and Burgundy, where there was no diet restriction, and 6.3% in Gothenburg where 80% of the tests were rehydrated. The proportion of positive tests in subsequent screenings was between 1% and 1.5% with the non-rehydrated Hemoccult test and 5.6% with a rehydrated test. The positive predictive value for the non-rehydrated test was about 10% for colorectal cancer and ranged between 30 and 40% for adenomas.

Cancer registries were useful in obtaining information on newly diagnosed cases of colorectal cancers in screening participants and in the control group. They were of particular importance to ensure complete ascertainment of incident colorectal cancers. In all four studies the proportion of Dukes A among cancers detected by screening was around 40%. The non-responders presented at a later stage than the controls. There was an intermediate situation for interval cancers, with a stage distribution between that of cases detected by screening and that of non-participants.

The shift towards a less advanced stage of the disease was maintained, when the test group as a whole (intention to screen) was compared to the control group. There was also a significant survival advantage for individuals in the screening group over those in the control group. But these data do not represent a sufficient argument in favour of the effectiveness of screening. There are a number of biases. Slow-growing cancers are more likely to be detected by screening (length bias). Screening hastens the diagnosis of incurable cancers, giving a longer lifespan to the disease without actually lengthening it (lead time bias) and subjects who participate in screening can be at lower risk (selection bias).

The effectiveness of the screening programme should be evaluated in terms of the number of cancer deaths prevented. The Funen trial reported an 18% (RR=0.82, 95% CI: 0.68-0.98) reduction in colorectal cancer mortality with a 10 year follow-up, the Nottingham study a 15% reduction (RR=0.85, 95% CI: 0.74-0.99) with a median follow-up of 7-8 years, the Burgundy study a 14% reduction (RR=0.86, CI: 0.71-1.03) with a 9 year follow-up trial and the Gothenburg study a 12% reduction (RR=0.88, 95% CI: 0.69-1.12) after 8 years of follow-up (Kronborg et al., 1996; Hardcastle et al., 1996; Faivre et al., 1999b; Towler et al., 1998). These four studies provide evidence that biennial screening with a Hemoccult test can reduce mortality from colorectal cancer. These findings can be extrapolated to the general population if the conditions of the screening programme active in these areas were to be reproduced. To be effective, a mass screening programme necessitates a rigid organisation with a call-recall system for individuals.

218

Monitoring of mass screening programmes

Screening for colorectal cancer by faecal occult blood tests began in Germany in 1977 as part of an annually offered cancer check-up. Because of confidentiality problems only a limited evaluation of this programme was performed (Gnauck, 1995). In particular, data from cancer registries could not be used.

Most available data concerning screening with Hemoccult II are provided by pilot studies in limited populations. This is because the scientific evidence for reduction in mortality from colorectal cancer by screening the stools for blood and performing colonoscopy to detect the source of bleeding, has only been provided recently. Cancer registries have been used in France to monitor screening programmes in Calvados (Normandy) and in Isère (Alps). In Calvados, the cancer registry was in charge of organising and evaluating the screening programme (Launoy et al., 1996). The Hemoccult test was offered to 170 000 subjects aged between 45 and 74. In case of a positive screening test, information on complementary investigations, as well as on treatment of diagnosed lesions were actively collected from GPs and specialists. The cancer registry routinely collected data on treatment and stage at diagnosis, and also provided data on interval cancers and on cancers in non-participants. These data were used to estimate the sensitivity of the Hemoccult test (Launoy et al., 1997). In Isère (Exbrayat et al., 1996) the screening programme was limited to 84 000 women between 50 and 69 (who were proposed a mammography, a Pap smear and Hemoccult test by their GP). The cancer registry data are used to identify interval cancers. It was recently decided to carry out two large pilot studies in the United Kingdom in order to verify the reproducibility of the Nottingham study. Other European countries will probably also implement such policy. Cancer registries will be of great importance in the evaluation of such programmes.

Conclusion

Data from cancer registries have been used to plan screening trials and to evaluate mass screening for colorectal cancer. They were used to provide data on colorectal cancers, both in case-control studies evaluating the efficacy of faecal occult blood tests and in controlled studies evaluating the effectiveness of such tests. The results of these studies will probably provide a basis to implement screening in larger populations. In the future, cancer registries will be important in monitoring these programmes.

References

Atkin W, Cuzik J, Nortover JMA & Whynes D (1993). Prevention of colorectal cancer by once-only sigmoidoscopy. *Lancet* **341**:736-740

Benhamiche AM (1998). Cancer du colon: épidemiologie descriptive et groupes à risque élevé. *Gastroenterol Clin Biol* **22**:S3-S11

Bertario L, Spinelli P, Gennari L, Sala P, Pizzetti P, Severini A, Cozzi G, Bellomi M & Berrino F (1988). Sensitivity of Hemmocult test for large bowel cancer in high-risk subjects. *Dig Dis Sci* **33**:609-613

Conférence de consensus (1998). Prévention, dépistage et prise en charge des cancers du côlon. *Gastroenterol Clin Biol* **22**:S275-S288

Ferlay J, Bray F, Sankila R & Parkin DM (1999). *EUCAN: Cancer Incidence, Mortality and Prevalence in the European Union, 1995*. Version 2.0. (IARC CancerBase No. 4). Lyon, IARC*Press*

European Group for Colorectal Cancer Screening (1999). Recommendation to include colorectal cancer screening in public health policy. *J Med Screen* **6**:80-81

Exbrayat C, Garnier A, Bolla M, Winckel P, Salicru B, Marron J, Faure H, Ménégoz F & Billette de Villemeur A (1996). Dépistage simultané des cancers du sein, du col, du colon et du rectum. L'expérience du département de l'Isère. *Bull Cancer* **83**: 641-648

Faivre J, Tazi MA, El Mrini T, Lejeune C, Benhamiche AM & Dassonvile F (1999a). Faecal occult blood screening and reduction of colorectal cancer mortality: a case-control study. *Br J Cancer* **79**:680-683

Faivre J, Tazi MA, Milan C, Lejeune C, Durand G & Lamour J (1999b). Controlled trial of faecal occult blood screening for colorectal cancer in Burgundy (France). Results of the first 9 years. *Gastroenterology* **116**:G1754

Gnauck R (1995). Screening for colon cancer in Germany. *Tumori* **80** (Suppl):30-37

Hardcastle JD, Chamberlain JO, Robinson MHE, Moss SM, Amar SS, Balfour TW, James PD & Mangham CM (1996). Randomized controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* **348**:1472-1477

Kewenter J, Brevenge H, Engaras B, Haglind E & Ahrén C (1994). Results of screening, rescreening, and follow-up in a prospective randomized study for detection of colorectal cancer by fecal occult blood testing. Results of 68,308 subjects. *Scand J Gastroenterol* **29**:468-473

Kronborg O, Fenger C, Olsen J, Jorgensen OD & Sondergaard O (1996). Randomized study as screening for colorectal cancer with faecal-occult-blood test. *Lancet* **348**:1467-1471

Launoy G, Herbert C, Vallée JP, Desoubeaux N, Reaud JM, Ollivier V, Bouvier V, Flachs A, Arsene D, Valla A & Gingoux M (1996). Le dépistage de masse du cancer colorectal en France. Expérience auprès de 165 000 personnes dans le Calvados. *Gastroenterol Clin Biol* **20**:228-236

Launoy G, Smith TC, Duffy SW & Bouvier V (1997). Colorectal cancer mass screening: estimation of fecal occult blood test sensitivity taking into account cancer mean sojourn time. *Int J Cancer* **73**:220-224

Macrae FA & St John DJ (1982). Relationship between patterns of bleeding and Hemoccult sensitivity in patients with colorectal cancer and adenomas. *Gastroenterology* **82**:891-898

Parkin DM, Whelan SL, Ferlay J, Raymond L & Young J (1997). *Cancer Incidence in Five Continents, Vol VII* (IARC Scientific Publications No. 143), Lyon, International Agency for Research on Cancer

Tazi MA, Faivre J, Dassonville F, Lamour J, Milan C & Durand G (1997). Participation in faecal occult blood screening for colorectal cancer in a well defined French population: results of five screening rounds from 1988 to 1996. *J Med Screen* **4**:147-151

Towler B, Irwig L, Glasziou P, Kewenter J & Silagy C (1998). A systematic review of the effects of screening for colorectal cancer using the faecal occult blood test, Hemoccult. *Br Med J* **317**:559-565

Wahrendorf J, Robra BP, Wiebelt H, Oberhausen R, Weiland M & Dhom G (1993). Effectiveness of colorectal cancer screening: results from a population-based case-control study in Saarland, Germany. *Eur J Cancer Prev* **2**:221-227

Young GP & St John DJ (1992). Faecal occult blood tests: choice, usage and clinical applications. *Clin Biochem Rev* **13**:161-167

Zappa M, Castiglione G, Grazzini G, Falini P, Giorgi D, Paci E & Ciatto S (1997). Effect of faecal occult blood testing on colorectal mortality: result of a population-based case-control study in district of Florence, Italy. *Int J Cancer* **73**:208-210

Chapter 16. Cancer registries in early detection of cutaneous malignant melanoma

J.W.W. Coebergh and H.J. van der Rhee

Introduction

Cutaneous malignant melanomas (CMM) account for 2 to 8% of all new malignant tumours in industrialised countries, but for only 1 to 2% of all cancer deaths. The lifetime cumulative risk of dying from CMM is less than 0.5% in most fair-skinned populations (Parkin et al., 1997). In the European Union about 31 000 new patients with CMM were diagnosed in 1995 and about 8000 deaths occurred (Ferlay et al., 1999). About 44 000 new patients are expected in the USA in 1999 and about 7000 deaths (Ries et al., 1999). Because CMM is a visible tumour, a variety of screening efforts (Elwood, 1994) has been undertaken during the 80's and 90's in response to the marked rise of incidence rates since the 60's (Coleman et al., 1993). Thanks to earlier detection, mortality has risen to a lesser extent, or even stabilised in middle-aged women and young adults in most countries (LaVecchia et al., 1999). Melanoma awareness has certainly increased in many countries and is linked with a more precise knowledge of the aetiology, largely intermittent UV radiation (Armstrong and Kricker, 1993).

The pressure to implement secondary and primary prevention has become considerable, not only due to the above mentioned scientific factors. There may be an emotional component in the medical reaction to the rising incidence, influenced by the experience of seeing young and middle-aged patients die from CMM. Cancer societies have also become more active in devising campaigns against avoidable dangers.

Cancer registries have played a vital role in alerting the medical profession and public to the melanoma epidemic. Until now, their data have mainly been used in describing the impressive changes in incidence, by subsite and sometimes according to thickness, in a variety of industrialised countries (Armstrong and Kricker, 1994) Some registries have also reported impressive changes in the incidence of the more frequent basal cell carcinoma (BCC) (Coebergh et al., 1995; Gallagher et al., 1990).

Early detection campaigns for CMM have largely been evaluated by monitoring of, for example, changes in incidence rate according to stage or thickness, sometimes also taking trends in mortality into account (MacKie et al., 1997; van der Rhee et

al., 1999). Occasionally have registries been used to identify the possible false negative or interval tumours (Rampen *et al.*, 1995), and few campaigns were continuous. If effects of screening had been assessed using mortality rates, they would probably have been small, or become manifest so much later that they could also be attributed to other influences. Assuming that no conclusive screening trials will be conducted in the near future, this chapter considers only the minimal requirements of the monitoring role of registries in support of medical and public health policies aimed at early detection and primary prevention.

Frequency: time trends and variation in mortality and incidence

Mortality rates from CMM have doubled or even tripled since the 50's (Coleman *et al.*, 1993; Armstrong and Kricker, 1994), although there is some overestimation due to under-registration and misclassification of skin tumours in the past. Incidence rates have also been rising dramatically since the 60's and 70's in the predominantly fairly skinned populations of Europe, at first in Northern and Central Europe, later in Eastern and Southern European countries. In Australia the most marked rises in incidence and mortality have occurred among people with recreational sun exposure, and among the sun-adoring Caucasian immigrants of Northwest European origin, especially when they migrated during childhood (Armstrong *et al.*, 1982). Marked rises in incidence (of 5 to 7% annually) have also been observed among whites in the USA (Glass and Hoover, 1989). These analyses also show that incidence figures started plateauing, and period-cohort analyses of incidence and mortality showed deceleration of the increases in generations born after 1950, sometimes even after 1920 (Chen *et al.*, 1994). Figures 1 and 2 show a worldwide overview of the incidence of CMM in the period 1988-92. The current south-north and east-west gradients (from low to high) in Europe will probably be reduced in the future, because rates are still rising in low-incidence populations, while they are tending to stabilise in higher risk countries. The absolute lifetime risk of CMM hardly exceeds 1% in most European countries and the risk of death remains below 0.3% (Ferlay *et al.*, 1999). Incidence curves usually start rising after the age of 20, but flatten beyond 50 to 60 years. In most countries incidence in females is some 25 to 50% higher than in males, whereas the differences in mortality are smaller or even reversed; this is largely due to a less favourable distribution of tumours with respect to subsite and stage in men (Streetly and Markowe, 1995).

Deaths from CMM comprise up to 2.5% of all mortality in Dutch women up to age 40, when overall death rates are very low, but this proportion declines with increasing age, in spite of a steep increase of the death rate (Figures 3-4).

224
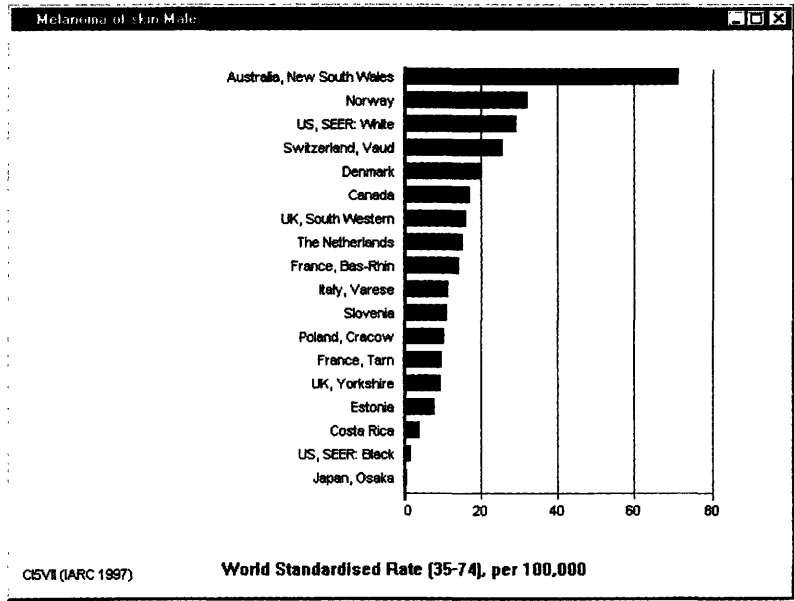
Figure 1. Worldwide age-adjusted incidence of cutaneous malignant melanoma in age-group 35-74 years, males, 1988-1992 (Parkin *et al.*, 1997)
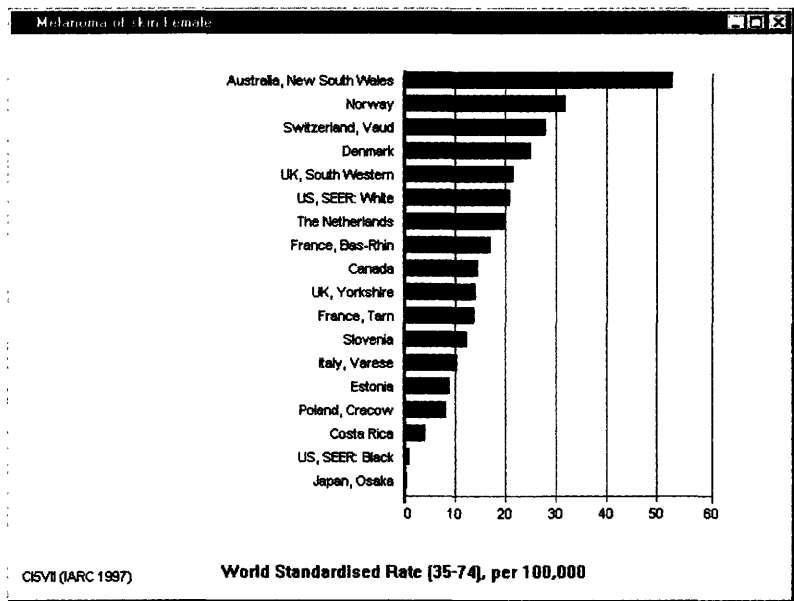

Figure 2. Worldwide age-adjusted incidence of cutaneous malignant melanoma in age-group 35-74 years, females, 1988-1992 (Parkin *et al.*, 1997)

Figure 3. Incidence of and mortality from cutaneous malignant melanoma in The Netherlands, 1991-1995. Rates per 100,000. (Source: Netherlands Cancer Registry and Statistics Netherlands)
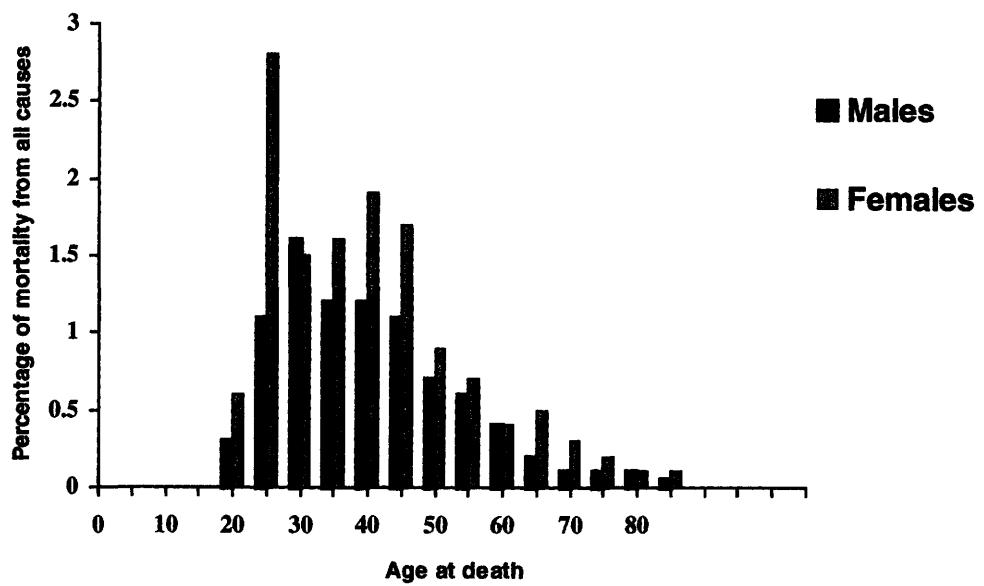

Figure 4. Mortality from cutaneous malignant melanoma (as percentage of all causes) by age in The Netherlands, 1991-1995 (Source: Statistics Netherlands)

It seems likely that increased awareness of melanoma leads to earlier detection. To assess its potential, one might first consider how often a general practitioner (GP) or dermatologist is confronted by a new patient with CMM.

In the Netherlands, with medium to high incidence rates of CMM, a GP (with an average practice of about 2500 people) is confronted with a new patient with CMM every 4-6 years and a dermatologist (one for every 50 000 inhabitants) once in 2-3 months (Coebergh et al., 1995). If the melanoma-predictive value of suspect, but unclear, skin signs varies from 2-10%, a Dutch GP would be confronted with a new patient every 2 months. If the predictive values lay between 5 and 10% for a dermatologist (who only receives referred patients), a new patient would attend for a diagnostic procedure every week. These frequencies would be modified in areas where (plastic) surgeons are active in skin cancer detection.

Concerning follow-up screening, an average GP in Holland would have about two prevalent patients (after removal of a CMM) in his practice, who may benefit from some surveillance (for another melanoma rather than for recurrence), whereas a dermatologist would have some 25 to 50 such (former?) patients, and gives attention to recurrence. The prevalence of dysplastic naevus syndrome, primarily in dermatologist practices, varies with definition and awareness. Such calculations (to be adapted by registries in every country) can also be made for BCC and SCC. The substantial prevalence of suspected lesions would at least underscore the need for sound dermatological training for GP's.

Stage distribution, survival and screening

Tumour stage, or thickness according to Breslow (Breslow, 1970) is the major determinant of recurrence and survival. CMM can be detected early at most sites, because they are readily visible. As a rule, patients with CMM of <1.5 mm thick have a very good prognosis, the cumulative relative 10-year survival rate being over 90%. Data on melanoma thickness are, however, rather sparse, since few registries have recorded this for more than 15 years (Coebergh et al., 1995; Roder, 1998; Thorn et al., 1994; Levi et al., 1998). In many studies stage-distribution is more favourable (thickness <1.5 mm) in women compared with men, in younger and middle-aged subjects, and in tumours on the extremities, rather than the trunk or the head and neck area.

Tumours have been increasingly detected at early stages since the 1970's, especially in European countries with good access to specialised care, leading to an impressive improvement of relative survival (Smith et al., 1998). As a consequence, the mortality/incidence ratio for melanoma has decreased in these countries and in the USA from 0.6 to 0.2, whereas unfavourable ratios still exist in eastern and southern Europe (Coleman et al., 1993; Armstrong and Kricker, 1994).

Beginning in the United States, early detection has been promoted by dermatologists and cancer societies, a pattern sometimes followed in Europe. There have been primarily exercises in raising awareness of the public and general practitioners, and there have been few attempts at systematic evaluation. In fact, detection rates for CMM appear to have increased, and stage distribution favourably influenced. Effects on population-based mortality are hard - if not impossible - to assess, because they would only become manifest after a long time. Marked changes are anyway unlikely, because the populations most at risk of fatal melanoma - such as men of middle and older age with the thicker melanomas - are often not reached by such campaigns. A shift in the classification of malignant behaviour may have occurred, although a systematic study in Europe did not provide any evidence of this during the 1970's and 80's (van der Esch et al., 1991).

In any case, mortality has not risen as much as incidence, and has even started to decrease in women after various campaigns in Scotland (MacKie et al., 1997) and in Connecticut (Khlat et al., 1992). However, the same thing happened in southern and western Holland without any or very limited screening effort (Coebergh et al., 1995; van der Rhee et al., 1999). The only documented, albeit still indirect, evidence of a positive effect of early detection comes from a skin self-examination (SSE) campaign in the state of Connecticut at the end of the 1980's. Although only practiced by 15% of the population, SSE occurred more frequently in patients with thin melanomas than in those with advanced or lethal lesions, the odds ratios for a protective effect being 0.7 to 0.5 (Berwick et al., 1996).

Role of the cancer registry in melanoma prevention

In developed countries, it is quite likely that cancer control policies will include primary and secondary prevention of CMM (Marks, 1995; Kroon et al., 1999). The latter will always include close surveillance of the small number of families with dysplastic naevi syndromes (DNS) (MacKie et al., 1993). With respect to promoting early detection among the general population, cancer registry reporting can support the setting of targets by indicating the room for improvement in the thickness distribution according to sex and age (Table 1). For example, a target statement can be formulated that more than 70 to 80% of the CMM's in male patients below 60 years must be <1.5 mm and 60 to 70 % in those over 60 years. In monitoring how closely such targets are approached, deficiences (such as subgroups which are insufficiently reached) can be identified. Of course, the targets must be realistic. Collection of data on tumour thickness may require retrospective reviews, with assistance of pathologists. Reports on survival can be particularly useful, particulary if the determinants can be explored using multivariate analysis (Levi et al., 1998).

Table 1. Function of cancer registry for melanoma prevention: intelligent monitoring

1. Public health trends of:

    - Mortality by sex and age-group
    - Incidence according to:
        - Sex and age-group (0-19, 20-39, 40-59, 60-69, 70+ yrs)
        - Histological type (SSM/nodular, lentigo maligna)
        - Stage, preferably Breslow thickness (<1.50, 1.51-3, 3+ mm)
        - Subsite (ICD-O, 4-digit)
        - Ratio in situ/invasive

2. Clinical evaluation of:

    - Relative survival according to:
        - Sex and age-group
        - Breslow thickness (in mm)
        - Subsite
    - Quality of care: Diagnostic process
        - Review of diagnosis
        - Reexcision rate
        - Referring and diagnosing specialty (according to subsite)

Summary

The epidemic of skin cancer is still continuing among fair-skinned populations in most industrialised countries, although it may be on the decline in some. With respect to prevention of CMM, public awareness is most important, especially if focussed on risk groups. For primary prevention, these are largely children and young adults and for secondary prevention, people at middle and old age. Continuous screening campaigns are unlikely to have much added value. Cancer registries have primarily a monitoring role in reporting of trends according to subsite and of thickness. Registries should try to support health policies aimed at controlling the CMM epidemic, which will also create opportunities to get involved in studies of aetiology and quality of care. Wherever possible, registries should report incidence trends of BCC and SCC by subsite.

References

Armstrong BK & Kricker A (1993). How much melanoma is caused by sun exposure. *Melanoma Res* 3:395-401

Armstrong BK & Kricker A (1994). Cutaneous melanoma. In: Trends in cancer incidence and mortality. In: Doll R, Fraumeni JF, Muir CS (eds) *Cancer Surveys* **19/20**:219-39

Armstrong BK, Holman CJ, Ford JM & Woodings TL (1982). Trends in melanoma incidence and mortality in Australia. In: Magnus K (ed) *Trends in Cancer Incidence: Causes and Practical Implications.* Washington, Hemisphere, pp. 399-417

Berwick M, Begg CB, Fine JA, Roush GC & Barnhill RL (1996). Screening for cutaneous melanoma by skin self-examination. *J Natl Cancer Inst* **88**:17-23

Breslow A (1970). Thickness, cross-sectional area and depth of invasion in the prognosis of cutaneous melanoma. *Ann Surg* **172**:902-08

Chen YT, Zheng T, Holford TT, Berwick M & Dubrow R (1994). Malignant melanoma incidence in Connecticut: time trends and age-period-cohort modelling by anatomic site. *Cancer Causes Control* **3**:341-350

Coebergh JWW (1995). Melanoma and non-melanoma skin cancer. In: Coebergh JWW, van der Heijden LH & Janssen-Heijnen MLG (eds) *Cancer Incidence and Survival in the Ssoutheast of the Netherlands, 1955-1994: A Report from the Eindhoven Cancer Registry.* Eindhoven, Integraal Kankercentrum Zuid, pp. 48-53

Coleman M, Estève J, Damiecki P, Arslan A & Renard H (eds) (1993). *Trends in Cancer Incidence and Mortality.* Melanoma1973-1987. (IARC Scientific Publication No. 121), Lyon, International Agency for Research on Cancer, pp. 379-411

Elwood JM (1994). Screening for melanoma and options for its evaluation. *J Med Screening* **1**:22-38

Ferlay J, Bray F, Sankila R & Parkin DM (1999). EUCAN: Cancer Incidence, Mortality and Prevalence in the European Union, 1995. Version 2.0. IARC CancerBase No. 4. Lyon, IARC*Press*

Gallagher RP, McLean DI, Yang CP, Ho V, Carruthers JA & Warshawski LM (1990). Trends in basal cell carcinoma, squamous cell carcinoma, and melanoma of the skin from 1973 through 1987. *J Am Acad Dermatol* **23**:413-21

230

Glass AG & Hoover RN (1989). The emerging epidemic of melanoma and squamous cell skin cancer in the USA. *JAMA* **262**:2097-2100

Khlat M, Vail A, Parkin DM & Green A (1992). Mortality from melanoma in migrants to Australia: variation by age and arrival and duration of stay. *Am J Epidemiol* **135**:1103-1113

Kroon BBR, Bergman W, Coebergh JWW & Ruiter DJ (eds.) (1999) Consensus on the management of malignant melanoma of the skin in The Netherlands. Dutch Melanoma Working Party. *Melanoma Res* **9**:207-212

La Vecchia C, Lucchini F, Negri E & Levi F (1999). Recent declines in wolrdwide mortality from cutaneous melanoma in youth and middle age. *Int J Cancer* **81**:62-66

Levi F, Randimbison L, La Vecchia C, Te VC & Franceschi S (1998). Prognostic factors for cutaneous malignant melanoma in Vaud, Switzerland. *Int J Cancer* **78**:315-319

MacKie RM, Henry P & Hole D (1993). Accelerated detection with prospective surveillance for cutaneous malignant melanoma in high-risk groups. *Lancet* **341**:1618-1620

MacKie RM, Hole D, Hunter JAA, Rankin R, Evans A, McLaren K, Fallowfield, M, Hutcheon A & Morris A (1997). Cutaneous malignant melanoma in Scotland: incidence, survival and mortality, 1979-94. *Br Med J* **315**:1117-1121

Marks R (1995). Skin cancer control in Australia: the balance between primary prevention and early detection. *Arch Dermatol* **131**:474-748

Parkin DM, Whelan S, Ferlay J, Raymond L & Young J (eds) (1997). *Cancer Incidence in Five Continents, Vol. VI* (IARC Scientific Publications No. 147), Lyon, International Agency for Research on Cancer, pp. 1004-1005

Rampen FHJ, Casparie–van Velzen IJA, Huystee van BEW, Kiemeney LALM & Schouten LJ (1995). False-negative findings in skin cancer and melanoma screening. *J Am Acad Dermatol* **33**:59-63

Ries LAG, Kosary CL, Hankey BF, Miller BA, Clegg L & Edwards BK (eds) (1999). *SEER Cancer Statistics Review, 1973-1996.* Bethesda, National Cancer Institute

Roder D (1998). *Incidence, Mortality and Survival, 1977 to 1997, Analysed by Type and Geographical Location.* Adelaide, South Australian Cancer Registry, pp. 125-131

Smith JAE, Whatley PM, Redburn JC, & the Eurocare Working Group (1998). Improving survival of melanoma patients in Europe since 1978. *Eur J Cancer* **34**:2197-2203

Streetly A & Markowe H (1995). Changing trends in the epidemiology of malignant melanoma: gender differences and their implications for public health. *Int J Epidemiol* **24**:897-907

Thorn M, Pontén J, Bergström R, Sparén P & Adami HO (1994). Clinical and histopathological predictors of survival in patients with malignant melanoma: a population-based study in Sweden. *J Natl Cancer Inst* **86**:761-769

van der Esch EP, Muir CS, Nectoux J, Macfarlane G, Maisonneuve P, Bharucha H, Briggs J, Cooke RA, Dempster AG, Essex WB, et al. (1991). Temporal change in diagnostic criteria as a cause of the increase of malignant melanoma is unlikely. *Int J Cancer* **47**:483-490

van der Rhee HJ, van der Spek-Keijser LM, van Westering R & Coebergh JW (1999). Increase in and stabilization of incidence and mortality of cutaneous malignant melanoma in Western Netherlands, 1980-95. *Br J Dermatol* **140**:463-467

Chapter 17. Lessons learned from neuroblastoma screening

R. Sankila

Introduction

The history of screening for neuroblastoma is a very good lesson for anyone involved in the planning, evaluation, or monitoring of screening programmes. As described below, general population-based cancer registries have had a limited role in the studies concerning the effects of screening for neuroblastoma. At the time the screening programmes were launched in Japan, there were rather few population-based cancer registries. Several studies in Europe and North America have relied upon data from hospital-based registries or collaborative groups, and some have used data from specialised childhood cancer registries.

Background

Neuroblastoma is the second most common solid tumour among children in Europe under 5 years of age, following astroglial brain tumours (Parkin et al., 1988). The prognosis is much better among children diagnosed at less than one year of age compared with older children, whose tumours more often are of advanced stage at diagnosis (Figure 1). About 90% of patients with neuroblastoma secrete elevated levels of catecholamine metabolites (vanillylmandelic acid, VMA, and/or homovanillic acid, HVA) in urine (Woods and Tuchman, 1987). The urine can, e.g., be easily blotted from the diaper and analysed later in a laboratory.

In the 1960s and 1970s, pioneering work was performed in Japan by Sawada, establishing the feasibility of early diagnosis of neuroblastoma by screening babies at the age of six months for elevated urinary levels of VMA/HVA (Sawada et al., 1971). In 1973, a mass-screening programme was started in Kyoto, Japan, and in 1985, it was extended to a nationwide programme throughout Japan (Sawada et al., 1991).

After a period of optimism, doubts began to rise (Miller et al., 1990, Murphy et al., 1991). Finally, in December 1998, the Consensus Conference on Neuroblastoma Screening concluded that neuroblastoma screening under seven months of age cannot be recommended as a public health policy, nor do new screening programmes in that age group appear to be justified on the basis of current evidence (Consensus Conference on Neuroblastoma Screening, 1999).

233
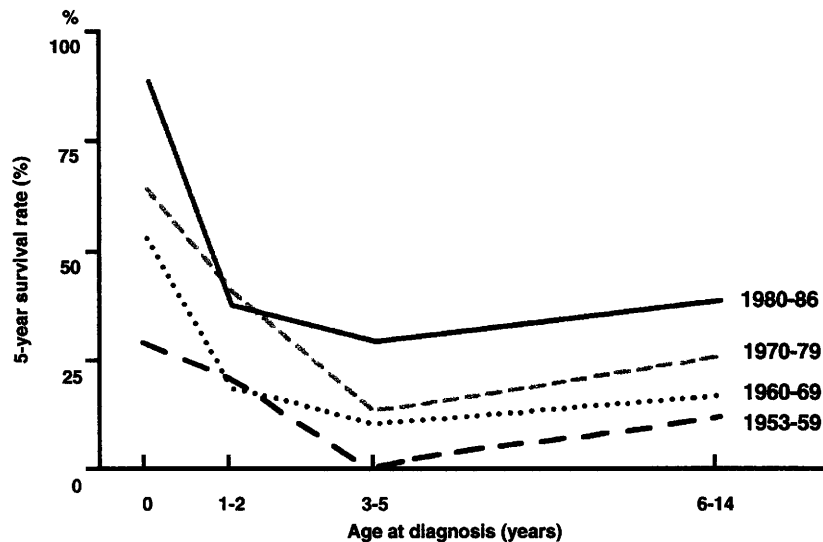
Figure 1. **5-year cumulative survival rates for 257 neuroblastoma patients (all stages) diagnosed in Finland in 1953-1986, by age and period of diagnosis** (reprinted from Sankila R & Hakama M, 1992. Survival trends for neuroblastoma patients in Finland: negative reflections on screening. Eur J Cancer 29A:122-123; with permission from Elsevier Science)

Prerequisites for screening

From reports written in English, one cannot find documentation on how systematic the planning process was, when the programme for screening children for neuroblastoma at 6 months of age was initiated in Kyoto in the early 1970s. It seems now, in retrospect, that our current understanding of many aspects of neuroblastoma is based largely on the results of research originating from these pioneering screening efforts. What is also evident, is that evaluating and monitoring the original Japanese screening exercises was very difficult, because there were no population-based cancer registries. These problems become obvious when reading the excellent review of all published Japanese studies (Parker and Powell, 1998).

The first prerequisite for any screening programme is some evidence of potential benefit from screening. Conclusive evidence can ultimately only be acquired through a well-designed randomised trial. Since no such trial was conducted in Japan, the potential benefits were based on clinical knowledge of the disease. In a very simple model of the disease, it was observed that neuroblastoma was more

234

common among children less than one year of age than among older children. Further, children diagnosed before the age of one year had proportionally more localised disease, and their overall survival rates were much better than those of older children (Figure 2). On the other hand, children diagnosed after one year of age had often stage 3 or 4 disease, and their prognosis was dismal (Tables 1 and 2).
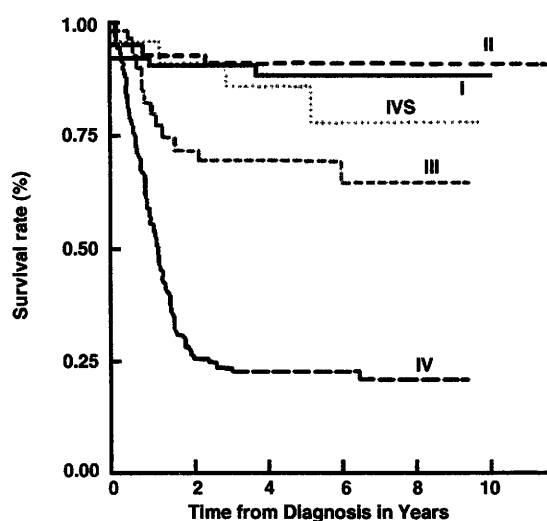
Figure 2. Survival since diagnosis by Evans stage of neuroblastoma (from Bernstein *et al.* , 1992. A population-based study of neuroblastoma incidence, survival, and mortality in North America. J Clin Oncol 10:323-329; by permission of Lippincott Williams & Wilkins)

It seemed obvious, that there might be a time window during which children with subclinical neuroblastoma could be diagnosed and treated curatively before their tumours advanced leading to inevitable death of the patients. However, there are two exceptional biological aspects complicating our understanding of the natural history of neuroblastoma. First, it is evident that neuroblastoma-like tissue is often found in organs of foetuses and new-borns (Acharya *et al.*, 1997). Furthermore, there is a special type of metastasised neuroblastoma (Stage 4S), which regresses spontaneously, matures into non-malignant ganglioneuromas, or is cured with minimal treatment, with very high survival rates (Brodeur *et al.*, 1988, Figure 2).

The second prerequisite is a test that could be used for mass screening. In the case of neuroblastoma, a test was available, since the tumours excrete catecholamine metabolites, which can be detected as elevated levels in urine. However, some 10% of the patients are non-secretors at time of clinical diagnosis. Testing was first started with a qualitative spot test for VMA, with further improvements in testing

235

with enzymatic immunoassay or, quantitatively, with high-performance liquid chromatography, and by including HMA and creatinine concentrations in the analysis package. Finally, a method using stable isotope dilution gas chromatography-mass spectrometry provided the most sensitive and precise analysis (especially) for control samples from children whose first samples had proven positive.

Table 1. Stage distribution (%) of all neuroblastoma cases (N=452) in Germany in 1987-1991, by age (modified from Schilling *et al.*,1998. German neuroblastoma mass screening study at 12 months of age: statistical aspects and preliminary results. Med Pediatr Oncol 31:435-441; reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.)

	Age		
Stage	< 1 year	1 year and above	Total
I-III/1-3	63%	37%	48%
IV/4	15%	64%	42%
IVS/4S	21%	-	9%

False-positive results

After the initiation of the mass screening programmes at age of 6 months in Japan, very promising results were published. The test was feasible, it was easy to perform by the mothers (just requiring a drop of urine in filter paper to be mailed to the laboratory), and generally, the compliance was very good. The test-positive children were retested, and, following clinical examination, neuroblastomas were diagnosed and treated. The stage distribution of the tumours was very favourable, and survival of the patients was excellent.

Table 2. 5-year survival rates (%) of all neuroblastoma cases (N=452) in Germany in 1987-1991, by stage and age (modified from Schilling *et al.,*1998. German neuroblastoma mass screening study at 12 months of age: statistical aspects and preliminary results. Med Pediatr Oncol 31:435-441; reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.)

	Age		
Stage	< 1 year	1 year and above	Total
I-III/1-3	95%	74%	85%
IV/4	69%	22%	28%
IVS/4S	81%	-	81%
Total	82%	38%	56%

The monitoring of screening programmes in Japan was difficult, because there were no population-based cancer registries and, at the beginning, the results of screening were published for each prefecture separately, often based on small numbers of patients from different time periods. The screening programmes were started without considering the need to obtain data on clear-cut end-points, and there was no reliable population-based information on the incidence of and mortality from neuroblastoma in Japan. As a consequence, later attempts to draw conclusions based on the original reports are challenging and imprecise. However, it soon became obvious that the detection rate in the screened population was much higher than would have been expected based on incidence data with no mass screening. With improvements in the laboratory methods, the detection rates increased from 5.1 to 24 per 100 000. The incidence rate among those who were intended to be screened, but were not, was about 50% lower. The incidence in Sapporo City among the screened population doubled. Similar results have been seen elsewhere in Japan, and in the largest screening trial in Northern America (Bessho, 1996; Nishi *et al.,* 1997; Woods *et al.,* 1998).

There were children with false-positive test results (between 28 and 53 per 100 000 screened children). For the lucky ones, after non-invasive clinical examinations at a hospital and repeated control urine samples (often for a period of several months), no other signs or symptoms of the disease were found. If, however, a clinical

tumour was revealed, at least a diagnostic operative procedure was performed. It is only in recent years that a "wait-and-see" approach has become the choice for patients less than one year of age with screen detected tumours, but with no apparent signs of aggressiveness, and with a potential for regression or maturation (Ladenstein et al., 1998).

The possibility of actually diagnosing something that looks like a neuroblastoma, but would never have become a clinical tumour, is further supported by the extremely good prognosis of patients with screen detected neuroblastoma. Of the more than 800 cases detected by screening in Japan, only 10 have died (Parker and Powell, 1998).

Missing evidence of benefit

The early reports describing the effects of neuroblastoma screening in Japan were based on such process indicators as sensitivity and specificity of the test, shifts in age and stage distributions, and the improvement in the survival rates. The end-points that should have been provided, but never were, included a decline in the incidence of (advanced stage) neuroblastoma after the first year of life, and crucially, a decline in the overall and age-specific mortality from neuroblastoma. Without complete catchment of all neuroblastoma patients and individual follow-up, any incidence, mortality, and survival calculations are inaccurate. Thus, the experiment in Japan never had any chance of proving the effect of screening for neuroblastoma at age of 6 months as a public health intervention. It is unfortunate that the unreliable reporting of the test-negative cases and their incomplete individual follow-up, hamper any further attempts to learn from the Japanese experience.

Based on the results of screening trials outside of Japan, it has become evident that mass screening of children at 6 months of age finds children with tumours that resemble clinical neuroblastomas. However, these tumours have favourable stage and other biological features. In recent years, it has become possible to "wait-and-see" with these patients, and regression of tumours has been observed. On the other hand, patients that were not detected by the screening programme, but were later clinically diagnosed with neuroblastoma, often have advanced stage and other biological markers of unfavourable prognosis (for review, see Brodeur et al., 1998). Further, there is no evidence of a decrease in the incidence of neuroblastoma after the age of one year, nor of a decrease in overall neuroblastoma mortality (Woods et al., 1997).

238

Conclusions

In its consensus statement in December 1998, the Jury found no support for mass screening for neuroblastoma as a public health policy at any age (Consensus conference on Neuroblastoma Screening, 1998). The results of current screening trials must be awaited before any further guidelines can be considered on screening for neuroblastoma at a later age than at 6 months. It has been proposed that screening at a later age might decrease the number of false positive cases, and increase the efficacy of detecting older patients with unfavourable biological markers and poor prognosis (Kerbl et al., 1997, Schilling et al., 1998). However, it is not clear whether these tumours can actually be found by screening at any earlier stage, nor if they can be curatively treated. It is also possible that the on-going trials may not be able to answer these questions due to limitations in their size and design.

A small decrease in mortality can not be ruled out, but even if it existed, the hideous adverse effects of mass screening for neuroblastoma at 6 months of age outweigh any benefits. It is apparent that the early experiments have provided a lot of data for further research on the natural history of neuroblastoma, and opened many unanswered questions regarding the consequences of screening. However, with current knowledge, it is far less obvious that all necessary precautions were taken into account to minimise potential harm, and, e.g., to provide the parents with adequate, neutral, and understandable informed consent forms. Serious consideration must be given to the ethical issues related to any screening programme involving healthy children. It is likely, that parents will be willing to enrol their children in screening programmes offering potential benefits. However, the enthusiasm of the screeners may limit their willingness to provide critical information on the adverse effects and false negative results. In the case of screening for neuroblastoma, it is still unclear how much morbidity - physical or psychological - is produced as a consequence of false positive test results. Nor has it been possible to estimate the number of healthy children (with spontaneously regressing subclinical neuroblastoma tissue), who have been exposed to invasive diagnostic and therapeutic procedures, some with fatal outcome.

As cancer registries become involved in evaluating and monitoring screening programmes, emphasis must be given to collecting data on side effects. Any screening trial or programme must be considered as an entity that not only includes assessing the biological and classic public health outcomes, but also issues related to ethics, quality of life, health economics, and keeping the public up-to-date with relevant information.

One of the major lessons learned from the neuroblastoma experience is that any cancer registry should be involved as early as possible in the planning phase of a screening programme, in order to provide a clear picture of how and what data can

239

be expected to be collected. Further, providing background data for power calculations and other estimates of the expected numbers and rates should be the responsibility of the cancer registry as well as - based on these figures - the neutral and sensible interpretation of the feasibility of any planned programme.

References

Acharya S, Jayabose S, Kogan SJ, Tugal O, Beneck D, Leslie D & Slim M (1997). Prenatally diagnosed neuroblastoma. *Cancer* **80**:304-310

Bernstein ML, Leclerc JM, Bunin G, Brisson L, Robison L, Shuster J, Byrne T, Gregory D, Hill G, Dougherty G, et al. (1992). A population-based study of neuroblastoma incidence, survival, and mortality in North America. *J Clin Oncol* **10**:323-329

Bessho F (1996). Effects of mass screening on age-specific incidence of neuroblastoma. *Int J Cancer* **67**:520-522

Brodeur GM, Seeger RC, Barrett A, Berthold F, Castleberry RP, D'Angio G, De Bernardi B, Evans AE, Favrot M, Freeman AI, et al. (1988). International criteria for diagnosis, staging, and response to treatment in patients with neuroblastoma. *J Clin Oncol* **6**:1874-1881

Brodeur GM, Ambros PF & Favrot MC (1998). Biological aspects of neuroblastoma screening. *Med Pediatr Oncol* **31**:394-400

Consensus Conference on Neuroblastoma Screening, December 8-10, 1998, IARC, Lyon, France. Preliminary conclusions. Association pour le Dépistage du Neuroblastome, Internal Report, Lyon, 1999

Kerbl R, Urban CE, Ladenstein R, Ambros IM, Spuller E, Mutz I, Amann G, Kovar H, Gadner H & Ambros PF (1997). Neuroblastoma screening in infants postponed after the sixth month of age: a trial to reduce "overdiagnosis" and to detect cases with "unfavorable" biologic features. *Med Pediatr Oncol* **29**:1-10

Ladenstein R, Matthay K, Berthold F, Kogner P, Pearson ADJ, Frappaz D, De Bernardi B, Yamamoto K & Hartman O (1998). What can we expect from neuroblastoma screening? Clinicians point of view. *Med Pediatr Oncol* **31**:408-418.

Miller AB, Chamberlain J, Day NE, Hakama M & Prorok PC (1990). Report on a workshop on the UICC project on evaluation of screening for cancer. *Int J Cancer* **46**:761-769

Murphy SB. Cohn SL, Craft AW, Woods WG, Sawada T, Castleberry RP, Levy HL, Prorok PC & Hammond GD (1991). Do children benefit from mass screening for neuroblastoma? Consensus statement from the American Cancer Society Workshop on Neuroblastoma Screening. *Lancet* **337**:344-346

Nishi M, Miyake H, Takeda T, Hanai J, Kikuchi Y & Takasugi N (1997). Mass screening for neuroblastoma and mortality in birth cohorts. *Int J Cancer* **71**:552-555

Parker L & Powell J (1998). Screening for neuroblastoma in infants younger than 1 year of age: Review of the first 30 years. *Med Pediatr Oncol* **31**:455-469.

Parkin DM, Stiller CA, Draper GJ, Bieber CA, Terracini B & Young JL (eds) (1988). *International Incidence of Childhood Cancer* (IARC Scientific Publications No. 87), Lyon, International Agency for Research on Cancer

Sankila R & Hakama M (1992). Survival trends for neuroblastoma patients in Finland: negative reflections on screening. *Eur J Cancer* **29A**:122-123

Sawada T, Imafuku S, Takada H, Kim PY & Inui A (1971). Early biochemical diagnosis of neuroblastoma - from screening to diagnosis (in Japanese). *Jpn J Cancer Clin* **17**:727-730

Sawada T, Sugimoto T, Kawakatsu H, Matsumura T & Matsuda Y (1991). Mass screening for neuroblastoma in Japan. *Pediatr Hematol Oncol* **8**:93-109

Schilling FH, Spix C, Berthold F, Erttmann R, Hero B, Michaelis J, Sander J, Tafese T & Treuner J (1998). German neuroblastoma mass screening study at 12 months of age: statistical aspects and preliminary results. *Med Pediatr Oncol* **31**:435-441

Woods WG & Tuchman M (1987). Neuroblastoma: the case for screening infants in North America. *Pediatrics* **79**:869-873

Woods WG, Tuchman M, Robison LL, Bernstein M, Leclerc JM, Brisson LC, Brossard J, Hill G, Shuster J, Luepker R, Byrne T, Weitzman S, Bunin G, Lemieux B & Brodeur GM (1997). Screening for neuroblastoma is ineffective in reducing the incidence of unfavourable advanced stage disease in older children. *Eur J Cancer* **33**:2106-2112

Woods WG, Tuchman M, Bernstein M & Lemieux B (1998). Screening infants for neuroblastoma does not reduce the incidence of poor prognosis disease. *Med Pediatr Oncol* **31**:450-454

Chapter 18. Cancer registries and genetic screening

L. Tryggvadóttir

Definitions

Germline mutations - Mutations that have arisen in germ cells. They segregate in families and are carried by all cells of the individual who inherits the mutation.

Somatic mutations - Spontaneous mutations that arise in somatic cells, such as tumour cells. They are therefore confined to the tissue of their origin and not transmitted to the descendants.

Genetic testing - There is more than one definition of genetic testing. In this text it means testing individuals for germline mutations. Almost any kind of body tissue can be used for this purpose. Usually, DNA from lymphocytes is used. Another definition of genetic testing, not applied here, includes testing individuals for somatic mutations. The DNA must in this case be extracted from the tissue where the mutations are supposed to have arisen.

Predictive genetic testing - Testing unaffected individuals for germline mutations that are known to increase the risk of a disease. This can be done decades before the onset of disease.

First degree relatives - Relatives that can be expected to share 50% of their genes with each other. Those are siblings, parents and offspring.

Familial cancer - Increased risk of cancer in the family. Definitions vary, but most demand at least two cases that are first degree relatives to each other. Often young age at diagnosis is included in the definition. Familial cancer is not necessarily due to inherited factors. Other explanations can be chance or some environmental factors shared by family members.

Hereditary cancer - Increased risk of cancer in the family, due to predisposing genetic alterations (germline mutations), usually in tumour suppressor genes. Hereditary cancer is a subset of familial cancers.

Proband – The first individual in a family in whom a familial syndrome has been manifest, e.g. a person diagnosed with FAP after presenting with colorectal symptoms, thus leading to the surveillance of his first degree relatives.

243

Genetic screening for cancer – Testing for inherited mutations in cancer related genes. To date, the target population consists of families that fulfill certain criteria, e.g. first-degree relatives of a proband.

Family cancer screening – A programme of screening for defined precursor lesions or early stages of cancer in members of families with increased risk of cancer, e.g. a programme consisting of colonoscopy every other year for young first degree relatives of FAP patients. In recent years, genetic screening is increasingly being included in family screening programmes.

244

Introduction

In recent years the rapid development of molecular genetics has opened possibilities for genetic screening for cancer. New information on high risk mutations is accumulating fast and this development is likely to be even further stimulated by a recently started programme of The National Cancer Institute, which aims to define all genes that are relevant to all cancer (NCI, Cancer Genome Anatomy Project). As hereditary cancer is a rare phenomenon, any genetic screening programme is destined to have very limited, if any, public health impact on a population level. On individual and family levels the benefits from such programmes may be substantial. Thus, genetic screening for cancer cannot be recommended in the general population, and is currently confined to individuals with family history of certain cancers.

This chapter deals first with the benefits and hazards of using familial and genetic information for surveillance and preventive purposes while at the same time trying to protect individuals from possible harmful effects of disclosure of sensitive data. Secondly, there is an overview over the three most common and best studied cancer syndromes and finally, there is discussion of the role of cancer registries in screening programmes for inherited predisposition to cancer.

Potential health benefits of genetic testing and the rights to personal privacy

The ability to identify inherited genetic aberrations that are associated with a markedly increased risk of cancer offers possibilities of preventive measures and early treatment to individuals carrying certain inherited mutations. However, at the same time, the prospect of being able to do so has given rise to many ethical questions and concerns.

The main concerns relate to the familial nature of the data and the sensitive nature of information based on DNA research. The participation of one family member involves not only himself, but usually also demands the involvement of other members of the family. Furthermore, the results from one individual are indicative of the status of other close relatives. The genetic test may reveal an inherited mutation that is indicative of a very high risk of a particular disease, even decades before the onset of any symptoms, and this poses a psychological burden on the tested individual. On the other hand, the results from genetic tests will free 50% of those who are suspected of carrying a dominant mutation from their worries. It has been suggested that, for such important information to be sought and disclosed, a

245

process with three well-defined stages should be offered to family members. Those are: the information stage, the test stage and the stage of interpretation and support (Harper, 1997). Counselling should be offered not just to the proband seeking advice, but to the rest of the family.

A further complication with genetic testing is that the results are of potential financial interest for insurance firms, employers and private health care institutions. Disclosure of testing results may thus lead to discrimination against those tested. There is also the danger of stigmatisation of individuals or whole families if they are known to carry a detrimental mutation.

Genetic tests have already become commercially available for diseases for which there are still serious uncertainties of the efficacy of recommended surveillance and management procedures, e.g., for breast and ovarian cancer. Therefore results are urgently needed from high quality studies of the effects of the currently recommended procedures on survival and quality of life. This uncertainty calls for careful education and counselling of individuals with family history (Vasen, 1998) and for a conservative approach with respect to genetic tests.

The sensitive nature of genetic testing calls for strict precautions, and special rules have been formulated for the handling of data and for approaching relatives, the most important rule being that of asking for informed consent from the individuals being tested. In Britain, the "Advisory Commission on Genetic Testing" (Government response to the third report of the House of Common, 1996) and in the United States the "Task Force on Genetic Testing" (1995) have published extensive rules concerning genetic testing.

The American Society of Clinical Oncology (ASCO) (1996) review of the main benefits and hazards of genetic testing recommends rules for oncologists regarding this complicated task. The statement stresses the clinical importance of identifying high-risk families and offering genetic counselling. It acknowledges the need for performing genetic testing in the setting of long-term outcome studies with appropriate confidentiality. The role of informed consent is stressed, and the importance of providing adequate information to the patients on possible risks and benefits of the detection of a high-risk cancer gene and of the prevention modalities. Also, the importance of studying the psychological impact of genetic testing is pointed out. ASCO supports legislation and other efforts aimed at hindering that results of genetic testing being used in a discriminatory way by insurance companies or employers.

Candidates for genetic screening for cancer

Family screening programs that include genetic testing have to fulfil the same criteria as other screening programs. The number of commercially available genetic tests is rapidly increasing, and this too calls for high-quality research into the effects of presumed, but not proven, benefits of the recommendations given. Some of these are quite dramatic, such as the prophylactic removal of the breasts of women carrying mutations in BRCA1/2 genes. It is not likely that randomised trials can be used to evaluate genetic screening and surveillance programs, but evaluation is possible by comparing the experience of groups of potential carriers of high risk mutations that have participated in screening and surveillance programs, with the experience of groups that have not participated.

In the report from ASCO (1996) three categories of tests are defined for cancer predisposition testing. One of those categories consists of tests where the medical benefits are still not apparent, e.g. tests for mutations in the gene associated with Ataxia-Telangiectasia. The other two categories are described below. However, it must be borne in mind that new tests will become available, and regrouping of tests will occur, along with advances in molecular biology and according to results from research of the effects of screening.

The first category consists of tests for well-defined syndromes for which either a positive or negative result will change medical care. These are Familial Adenomatous Polyposis (FAP), Multiple Endocrine Neoplasia 2a and 2b, Retinoblastoma and Von Hippel-Lindau Syndrome.

The second category consists of tests for hereditary syndromes, where the medical benefits of the identification of a mutation carrier are less well established. The syndromes concerned are Hereditary Non-Polyposis Colon Cancer (HNPCC), Hereditary Breast and Ovarian Cancer and Li-Fraumeni Syndrome. In the future, a prostate cancer syndrome might be included in this category (Grönberg *et al.,* 1997)

FAP, HNPCC and hereditary breast and ovarian cancer

For three of the above syndromes, extensive registration of families and clinical screening for precursor lesions or initial stages of the diseases have been ongoing for many years. These syndromes are FAP, HNPCC and Hereditary Breast and Ovarian Cancer. Family screening and counselling has been the traditional approach, and only recently has it become possible to offer genetic testing as a part of the programmes. The first indication of the localisation of a gene associated with one of the syndromes, the FAP syndrome, came in 1987 (Bodmer *et al.,* 1987). In

247

1991 the gene for this syndrome, the *APC* gene, was identified (Kinzler *et al.*, 1991). The first gene associated with HNPCC was localised and identified in 1993 (Peltomäki *et al.*, 1993; Fishel *et al.*, 1993; Leach *et al.*, 1993) and the *BRCA1* gene was identified in 1994 (Miki *et al.*, 1994). The family registries for FAP, HNPCC and Hereditary Breast and Ovarian Cancer have made it possible to identify individuals at high risk of those syndromes, as judged from their relationship with affected family members and to offer them screening and treatment. Also, these registries have proved to be a very important source of family information, essential for the linkage analyses that have led to the identification of the genes.

FAP

Familial adenomatous polyposis (FAP) was first described in 1882 (Cripps, 1882) and the risk of malignant transformation of the adenomas was first discussed in 1925 (Lockhart-Mummery, 1925). FAP is a rare dominantly inherited susceptibility to colorectal cancer (CRC), with an estimated incidence of between 1/5000 and 1/10 000. The disease is characterised by the formation of up to thousands of adenomatous polyps in the large bowel of affected individuals. Patients with polyps almost inevitably develop colorectal carcinoma (Järvinen, 1992).

Registration of FAP has now been ongoing for several decades and in many countries (Bülow *et al.*, 1993; Rhodes *et al.*, 1991). In 1985, national and regional FAP registers from all over the world formed the "Leeds Castle Polyposis Group" as an international research forum (Thomson, 1988). Since 1989 the European Commission has financially supported a concerted action on familial cancer with special emphasis on FAP (EuroFAP). The aim is to increase the identification and registration of FAP patients in all countries of the European Union and to facilitate molecular genetic research (Bülow *et al.*, 1993).

For individuals with a family history of FAP, the management offered has been regular bowel examination with colonoscopy every two years between the ages of 10 and 40, and then with less intensity up to the age of 60 (Bülow *et al.*, 1993). Prophylactic colectomy is the recommended treatment for newly diagnosed patients with polyps. Several FAP-registry based studies have reported beneficial effects of regular colonoscopy screening of family members at risk. FAP patients detected by screening of family members (call-up patients) have been shown to have a considerably lower incidence of colorectal carcinoma than patients presenting with symptoms (probands) (Bülow *et al.*, 1995; Järvinen *et al.*, 1992; Rhodes *et al.*, 1991; Vasen *et al.*, 1990). Survival after the diagnosis of FAP has been found to be considerably better for call-up patients than for probands, but lead-time bias will play a role here. Even though life expectancy has been found to be increased in call-up patients relative to probands (Järvinen, 1992; Nugent *et al.*,

248

1993), a comparison with the general population revealed a considerably increased risk of dying in the call-up patients (Nugent *et al.*, 1993). A study based on the population-based Finnish Cancer Registry in the years 1961 to 1990 indicated that the proportion of FAP among all cases of colorectal cancer in Finland had declined from 0.53% to 0.14%, which might be partly attributed to the preventive effects of family screening (Järvinen, 1992).

The *APC* gene, responsible for FAP, identified in 1991 (Kinzler *et al.*, 1991), is a tumour suppressor gene localised on chromosome 5q21. Spontaneous mutations in the gene appear to account for between a tenth and a third of patients, whereas the rest of the patients have inherited mutations. Heterogeneity between families has been manifest by a marked variation in age of onset of polyposis and the number and size of adenomas. Commercially available genotyping for FAP became available in 1994 and may become a regular part of family screening programs. The test has been reported to detect an *APC* gene mutation in approximately 80% of FAP families (Cromwell *et al.*, 1998), but it has nearly 100% sensitivity and specificity when the mutation in the particular pedigree is known.

HNPCC

Hereditary nonpolyposis colorectal cancer syndrome (HNPCC) or Lynch syndrome (Lynch *et al.*, 1988) was first described in 1913 (Warthin, 1913) and re-discovered in 1971 by Lynch and Krush (Lynch and Krush, 1971). It was estimated that HNPCC may account for between 1 and 5 percent of all cases of colorectal carcinoma, but the true prevalence may be lower (Aaltonen *et al.*, 1994; Marra *et al.*, 1995; Evans *et al.*, 1997; Fante *et al.*, 1997; Riegler *et al.*, 1999). HNPCC is characterised by an autosomal dominant mode of inheritance, early-onset colorectal cancers predominantly located in the proximal colon and the occurrence of various other cancers, such as cancer of the endometrium (Aarnio *et al.*, 1999).

The syndrome has since 1971 been registered and described in many countries. In 1990 an international group (ICG-HNPCC) was formed, aiming at collaborative studies on the prevalence, natural history and clinical expression of the HNPCC syndrome (Vasen *et al.*, 1991a). The group agreed on a set of minimum criteria for recruitment of patients to the collaborative studies, the so-called Amsterdam criteria (Vasen *et al.*, 1991b). The criteria refer to number of affected near relatives and young age at diagnosis.

Several genes have been identified, mutations in which predispose to HNPCC. They are *hMSH2, hMLH1, hPMS1, hPMS2* and *hMSH6/GTBP* (Miyaki *et al.*, 1997; Bronner *et al.*, 1994; Nicholaides *et al.*, 1994; Papadopoulos *et al.*, 1994; Fishel *et al.*, 1993; Leach *et al.*, 1993; Peltomäki *et al.*, 1993). The genes play a role in mismatch repair function, and the tumours are characterised by replication

errors (RER) (Brown *et al.*, 1998). Mutation carriers have an estimated lifetime risk of over 80 % of developing CRC and an increased risk of other cancers (Vasen *et al.*, 1996). However, this estimate is based on high-risk families. It remains to be studied whether the risk is lower in unselected mutation carriers (Aarnio *et al.*, 1999). To date, more than 120 germline mutations have been identified in the mismatch repair genes. A database of mutations was established in 1994 by ICG-HNPCC (Peltomäki and Vasen, 1997), aiming at facilitating research that can help decide on mutation detection strategies.

Clinical screening of patients fulfilling the Amsterdam criteria has been ongoing for several years, and more recently also the genetic screening of suspected carriers of mutations. Cancer screening among HNPCC family members is more demanding than for FAP. In the absence of clear preceding clinical markers (such as multiple polyps), full colonoscopies are needed and the screening has to go on for the rest of the life among the mutation carriers and in HNPCC-like families where mutations cannot be found (Syngal *et al.*, 1999; Järvinen *et al.*, 1995). It is therefore very important to investigate whether the screening has beneficial effects. This is best done by comparing tumour stages, incidence and especially mortality from CRC, between screened and unscreened HNPCC family members or suspected mutation carriers. A non-randomised Finnish prospective study compared two groups of asymptomatic at-risk members of 22 families with HNPCC (Järvinen *et al.*, 1995). One group of 133 study subjects had screening at 3-year intervals; the other group consisted of 118 controls without screening. CRC occurred in six (4.5%) of the study subjects and in 14 (11.9%) of the controls, a reduction of 62% in the screened group. In the screened group, the tumour stages were more favourable and there were no deaths caused by CRC, compared with five deaths among the 14 cases in the control group. The follow-up was partially based on data from the population-based Finnish Cancer Registry.

A cost-effectiveness analysis of colorectal cancer screening among HNPCC mutation carriers has been made (Vasen *et al.*, 1998a). This study used estimates of the lifetime risk of developing CRC among mutation carriers (Vasen *et al.*, 1996), and comparisons of survival rates between sporadic CRC patients and HNPCC patients (for most of whom mutation status was known). The other estimates used in the cost-effectiveness analysis were based on studying members of HNPCC families with unknown mutation status. The comparison involved the stage distribution between patients from HNPCC families who were under surveillance and who were not (Vasen *et al.*, 1995), CRC stage-specific relative survival rates (Sankila *et al.*, 1996), and the above estimates of the effectiveness of surveillance (Järvinen *et al.*, 1995). The results indicated that the surveillance of mutation carriers led to an increase of seven years in the average life expectancy. Also, the costs of surveillance appeared to be less than the costs of no CRC surveillance.

250

HNPCC is presently the cancer syndrome with the best information on the effectiveness of clinical family screening and genetic testing. However, more research is needed, especially on results from genetic screening and on how the screening and preventive measures affect the quality of life and psychological status of family members.

Hereditary breast and ovarian cancer

Research on familial breast cancer and collection of high-risk families has been going on for several decades and many studies have confirmed that familial breast cancer is characterised by young age at onset and increased risk of bilaterial cancer (Bishop 1992; Tulinius *et al.*, 1992). In 1989, the Breast Cancer Linkage Consortium was established. It is an international collaboration, initially aimed at conducting linkage analyses of breast cancer families with increased risk (Bishop, 1995).

Two breast cancer genes, *BRCA1* and *BRCA2* were located to chromosomes 17q21 and 13q12-q13 in 1994 and 1995, respectively (Tavtigian *et al.*, 1996; Wooster *et al.*, 1995; Miki *et al.*, 1994). The prevalence of mutations in these genes varies considerably between populations. In relatively inbred populations, such as Ashkenazi Jews and Icelanders, the prevalence in unselected breast cancer cases has been found to be 12% (mutations in *BRCA1* and *BRCA2*) and 8% (*BRCA2*) respectively (Warner *et al.*, 1999, Thorlacius *et al.*, 1998). Lower prevalence estimates have been reported in unselected patients in other populations such as from USA (North Carolina) and UK, or 3,3% (*BRCA1*) and 2-3% (*BRCA1* and *BRCA2*) respectively (Newman *et al.*, 1998; Peto *et al.*, 1999). However, the UK prevalence estimate may be somewhat too low, because it is partly based on the high penetrance (or cumulative risk) estimate of 84% by age 70 years in carriers of *BRCA2* mutations, derived from the Breast Cancer Linkage Consortium (BCLC), using high-risk families (Ford *et al.*, 1998). Estimates of cumulative risk by age 70 years are considerably lower when based on unselected breast cancer patients, or between 28% and 37% in carriers of *BRCA2* mutations (Warner *et al.*, 1999; Thorlacius *et al.*, 1998) and between 36% and 56% when *BRCA1* and *BRCA2* mutations are considered together (Hopper *et al.*, 1999; Fodor *et al.*, 1998; Struewing *et al.*, 1997). The risk appears to vary considerably between families, even if they carry the same mutation (Thorlacius *et al.*, 1996). This indicates other important determinants of risk in mutation carriers - inherited, environmental or both. Carriers of *BRCA1* mutations also have an increased risk of ovarian cancer (Miki *et al.*, 1994). Heterogeneity of risk has also been demonstrated for ovarian cancer (Claus and Schwartz, 1995).

To date, hundreds of different germline mutations have been reported in the *BRCA1/2* genes and genetic testing is now frequently included in family screening

251

for breast cancer. However, screening is not recommended for the general population. The benefits and costs of screening Ashkenazi Jewish women (who have a much higher prevalence of BRCA1 mutations than other European populations) were discussed in a recent paper (Grann et al., 1999), concluding that genetic screening in this population may prolong survival, but that the results need confirmation through prospective studies.

There are still many questions unanswered with respect to the surveillance and management of breast cancer families. A European collaborative group of clinicians specialising in familial breast cancer was formed in 1996. In a survey among 16 family cancer clinics participating in this collaboration, most centres recommend surveillance of the breasts if the lifetime risk exceeds 15-20% (Vasen et al., 1998b). The surveillance protocol that is generally advised comprises monthly breast self-examination, examination by a specialist every 6 months and annual mammography, all starting from an age between 25 and 35 years. Surveillance of the ovaries is recommended in BRCA1/2 gene mutation carriers. Prophylactic mastectomy is considered for proven mutation carriers in half of the centres. Most of those recommendations are based on experts' opinion, and studies of the efficacy are urgently needed.

The efficacy of prophylactic mastectomy has been evaluated by comparing the incidence of breast cancer in groups of women with family history of breast cancer who underwent bilateral mastectomy, with the expected incidence in the group. Results from a recent, well-conducted study indicated a significant reduction in the incidence of breast cancer after mastectomy (Hartmann et al., 1999). However, critics have pointed out that bias could account for some of the observed beneficial effects and that results from familial patients may not apply for mutation carriers. Furthermore, it must be considered that this disfiguring and potentially psychologically damaging operation on a large group of women is a high price paid for the prevention of relatively few cancers and fewer deaths (Eisen and Weber, 1999).

Large scale follow-up studies are needed, where groups of suspected or proven mutation carriers under surveillance are compared with similar groups without surveillance, with respect to tumour stage, incidence, mortality, stage specific relative survival and quality of life. As is apparent from the above-cited European survey, the clinical management of a woman with family history depends on her estimated lifetime risk. Therefore, precise risk estimates are critical. They may differ between families (Hopper et al., 1999; Thorlacius et al., 1996) and should therefore be viewed in the context of the extent of family history of the woman.

Role of cancer registries in genetic cancer screening programmes

The important task of evaluating the efficacy of family screening programmes, with or without genetic screening, is totally dependent on close co-operation of two kinds of registries. These are the cancer family registries and the cancer registries. This co-operation is also important for other aspects of family screening, such as recruitment of probands and confirmation of cancer diagnoses in family members.

Cancer family registries

As described above, a large number of family registries and other collections of cancer-prone families have been in existence for some time. These registries have collected information on probands and relatives, usually offering family screening, with a view to early diagnosis and prophylactic treatment. Genetic counselling has become an important part of their activities as well as active research into the efficacy of family screening, often in close collaboration with classical cancer registries. The information collected on high-risk families in the family registries formed the basis for the linkage analyses that led to the identification of the currently known cancer suppressor genes. Presently, genetic screening for cancer is not done in the general population, but is confined to individuals with a family history. There are two main reasons for this: first, the mutation detection rate can be expected to be extremely low in the general population. Second, because a large number of mutations are known for each syndrome, the interpretation of results from genetic testing must be done in the context of test results from other family members. The sensitivity and specificity of the genetic test are much lower if the mutation segregating in the family is not known.

Cancer registries

The ideal design for studying the efficacy of programmes of family screening and surveillance, also including genetic screening, would be the randomised controlled trial. Trials are not likely to be conducted for this purpose, because the estimated risk for individuals with a family history is too high, and because of the high degree of anxiety associated with belonging to a high-risk family. As pointed out by Vasen et al. (1995), the next best design is a comparison of the incidence and mortality between a screened group and a control group of family members who have not been under surveillance. Population-based cancer registries can be utilised for this purpose, because of access to information on the incidence and mortality in both groups. As summarised below, evaluating family screening programmes and genetic testing can be performed in collaboration with population-based cancer registries.

1. In a research setting, the identification of young cases of a particular cancer type from a cancer registry is a practical way of collecting data on putative probands into a family registry, especially if the study is aimed at being nationwide.

2. Confirmation of cancer diagnoses in relatives of the probands is much facilitated where there is access to a cancer registry covering the entire study population.

3. When it comes to the important task of evaluating the results of family and genetic screening for cancer, access to population-based cancer registries is crucial. The cancer registries can provide unbiased follow-up information when comparing the incidence, stage distribution, morphology, mortality and relative survival between screened and unscreened groups of individuals in the putative cancer families. The comparisons are the only sound method of evaluating the efficacy of current procedures, and they form a basis for cost benefit analyses.

4. Another important role for the cancer registries is in estimating the cancer risk associated with mutations in cancer suppressor genes. Unbiased estimates can only be derived from population-based cancer registries, along with information on unselected mutation carriers. The risk estimates are critical, because they form the basis for decision making of clinicians on an individual basis with respect to the choice of surveillance and management procedures. Some of the risk estimates currently used are based on high-risk families and are likely to apply only to a selected subgroup of mutation carriers. The risk estimates also form an important basis for estimates of costs and benefits of screening and surveillance procedures.

5. The cancer registries can further provide unbiased estimates of the proportion of all cancer of a particular type that is attributable to known mutations. Studies of changes in this proportion with time can be indicative of effectiveness of surveillance programmes in reducing the incidence of the particular inherited cancer.

6. Finally, cancer registries are often research institutions where state-of-the-art epidemiological methodology is in every day use. Thus, cancer registry personnel can have important roles in providing consultation on epidemiological methods, standardisation of definitions of cancer syndromes and co-ordinating research methods between institutions, to facilitate the combination of data for collaborative studies with high statistical power.

References

Aaltonen LA, Sankila R, Mecklin JP, Järvinen H, Pukkala E, Peltomäki P & de la Chapelle A (1994). A novel approach to estimate the proportion of hereditary nonpolyposis colorectal cancer of total colorectal cancer burden. *Cancer Detect Prev* **18**:57-63

Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, Peltomäki P, Mecklin JP & Järvinen HJ (1999). Cancer risk in mutation carriers of DNA-mismatch-repair genes. *Int J Cancer* **81**:214-218

Bishop DT (1992). Family history of breast cancer: how important is it? *J Med Genet* **29**:152-153

Bishop DT (1995). Breast Cancer Linkage Consortium. The GAW9 Breast Cancer Linkage Data Set. *Genet Epidemiol* **12**:837-840

Bodmer WF, Bailey CJ, Bodmer J, Bussey HJ, Ellis A, Gorman P, Lucibello FC, Murday VA, Rider SH, Scambler P, et al. (1987). Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* **328**:614-616

Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A, et al. (1994). Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368**:258-261

Brown SR, Finan PJ, Cawkwell L, Quirke P & Bishop DT (1998). Frequency of replication errors in colorectal cancer and their association with family history. *Gut* **43**:553-557

Bülow S, Burn J, Neale K, Northover J & Vasen H (1993). The establishment of a polyposis register. *Int J Colorect Dis* **8**:34-38

Bülow S, Bülow C, Faurschou Nielsen T, Karlsen L & Moesgaard F (1995). Centralised registration, prophylactic examination, and treatment results in improved prognosis in familial adenomatous polyposis. Results from the Danish Polyposis Register. *Scand J Gastroenterol* **30**:989-993

Claus EB & Schwartz PE (1995). Familial ovarian cancer. Update and clinical applications. *Cancer* **76**:1998-2003

Cripps WH (1882). Two cases of disseminated polyps of the rectum. *Trans Path Soc Lond* **33**:165-168

Cromwell DM, Moore RD, Brensinger JD, Petersen GM, Bass EB & Giardiello FM (1998). Cost analysis of alternative approaches to colorectal screening in familial adenomatous polyposis. *Gastroenterology* **114**:893-901

Eisen A & Weber BL (1999). Prophylactic mastectomy - the price of fear. *N Engl J Med* **340**:137-138

Evans DG, Walsh S, Jeacock J, Robinson C, Hadfield L, Davies DR & Kingston R (1997). Incidence of hereditary non-polyposis colorectal cancer in a population-based study of 1137 consecutive cases of colorectal cancer. *Br J Surg* **84**:1281-1285

Fante R, Benatti P, di Gregorio C, De Pietri S, Pedroni M, Tamassia MG, Percesepe A, Rossi G, Losi L, Roncucci L & Ponz de Leon M (1997). Colorectal carcinoma in different age groups: a population-based investigation. *Am J Gastroenterol* **92**:1505-1509

Fishel R, Lescoe MK, Rao MRS, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R, et al. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colorectal cancer. *Cell* **75**:1027-1038

Fodor FH, Weston A, Bleiweiss IJ, McCurdy LD, Walsh MM, Tartter PI, Brower ST & Eng CM (1998). Frequency and carrier risk associated with common BRCA1 and BRCA2 mutations in Ashkenazi Jewish breast cancer patients. *Am J Hum Genet* **63**:45-51

Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, Sobol H, Teare MD, Struewing J, Arason A, Scherneck S, Peto J, Rebbeck TR, Tonin P, Neuhausen S, Barkardottir R, Eyfjord J, Lynch H, Ponder BA, Gayther SA, Zelada-Hedman M, & the Breast Cancer Linkage Consortium (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *Am J Hum Genet* **62**:676-689

Government response to the third report of the House of Commons Select Committee on Science and Technology, 1994-95 session. January 1996

Grann VR, Whang W, Jacobson JS, Heitjan DF, Antman KH & Neugut AI (1999). Benefits and costs of screening Ashkenazi Jewish women for BRCA1 and BRCA2. *J Clin Oncol* **17**:494-500

Grönberg H, Isaacs SD, Smith JR, Carpten JD, Bova GS, Freije D, Xu J, Meyers DA, Collins FS, Trent JM, Walsh PC & Isaacs WB (1997). Characteristics of prostate cancer in families potentially linked to the hereditary prostate cancer 1 (HPC1) locus. *JAMA* **278**:1251-1255

Harper PS (1997). What do we mean by genetic testing? *J Med Genet* 34:749-752

Hartmann LC, Schaid DJ, Woods JE, Crotty TP, Myers JL, Arnold PG, Petty PM, Sellers TA, Johnson JL, McDonnell SK, Frost MH & Jenkins RB (1999). Efficacy of bilateral prophylactic mastectomy in women with a family history of breast cancer. *N Engl J Med* 340:77-84

Hopper JL, Southey MC, Dite GS, Jolley DJ, Giles GG, McCRedie MR, Easton DF, Venter DJ, & the Australian Breast Cancer Family Study (1999). Population-based estimate of the average age-specific cumulative risk of breast cancer for a defined set of protein-truncating mutations in BRCA1 and BRCA2. *Cancer Epidemiol Biomarkers Prev* 8:741-747

Järvinen HJ (1992). Epidemiology of familial adenomatous polyposis in Finland: impact of family screening on the colorectal cancer rate and survival. *Gut* 33:357-360

Järvinen HJ, Mecklin J-P & Sistonen P (1995). Screening reduces colorectal cancer rate in families with hereditary nonpolyposis colorectal cancer. *Gastroenterol* 108:1405-1411

Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hedge P, McKechnie D, et al. (1991). Identification of FAP locus genes from chromosome 5q21. *Science* 253:661-665

Leach FS, Nicholaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomäki P, Sistonen P, Aaltonen LA, Nyström-Lahti M, et al. (1993). Mutations of a MutS homolog in hereditary non-polyposis colorectal cancer. *Cell* 75:1215-1225

Lockhart-Mummery P (1925). Cancer and heredity. *Lancet* i:427-429

Lynch HT (1981). Genetics and breast cancer. New York: Van Nostrand Reinhold Company

Lynch HT & Krush AJ (1971). Cancer family "G" revisited: 1895-1970. *Cancer* 27:1505-1511

Lynch HT, Lanspa SJ, Boman BM, Smyrk T, Watson P, Lynch JF, Lynch PM, Cristofaro G, Bufo P, Tauro AV, et al. (1988). Hereditary nonpolyposis colorectal cancer. Lynch syndromes I and II. *Gastroenterol Clin North Am* 17:679-712

Marra G & Boland CR (1995). Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. *J Natl Cancer Inst* 87:1114-1125

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**:66-71

Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, Igari T, Koike M, Chiba M, Mori T, et al. (1997). Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nature Genet* **17**:271-272

National Cancer Institute, National Center for Biotechnology Information. Cancer Genome Anatomy Project. Available at: http://www3.ncbi.nlm.nih.gov/ncicgap and http://www/ncbi.nlm.gov/CGAP

Newman B, Mu H, Butler LM, Millikan RC, Moorman PG & King MC (1998). Frequency of breast cancer attributable to BRCA1 in a population-based series of American women. *JAMA* **279**:915-921

Nicolaides NC, Papadopoulos N, Wei YF, Carter KC, Ruben SM, Rosen SA, Haseltine WA, Fleischmann RD, Fraser CM, et al. (1994). Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* **371**:75-80

Nugent KP, Spigelman AD & Phillips RK (1993). Life expectancy after colectomy and ileorectal anastomosis for familial adenomatous polyposis. *Dis Colon Rectum* **36**:1059-1062

Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen SA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD, et al. (1994). Mutation of a mutL homolog in hereditary colon cancer. *Science* **263**:1625-1629

Peltomäki P, Vasen H & the International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer: Database and Results of a Collaborative Study (1997). *Gastroenterol* **113**:1146-1158

Peltomäki P, Aaltonen LA, Sistonen P, Pylkkänen L, Mecklin JP, Järvinen H, Green JS, Jass JR, Weber JL, Leach FS, Petersen GM, Hamilton SR, de la Chapelle A & Vogelstein B (1993). Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **260**:810-812

Peto J, Collins N, Barfoot R, Seal S, Warren W, Rahman N. Easton DF, Evans C, Deacon J & Stratton MR (1999). Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst* **91**:943-949

Rhodes M. Chapman PD, Burn J & Gunn A (1991). Role of a regional register for familial adenomatous polyposis: experience in the Northern Region. *Br J Surg* **78**:451-452

258

Riegler G, Savastano A, Selvaggi F, Ciociano R, Martino R, Riccio G, Iorio R, Ponti G, Carratu R, Borgheresi P, de Filippo G, Rossi GB, Tempesta AM, de Palma GD, Catanzano C, Russo P, Bianco MA, Piscitelli A, di Carlo V, Baldi V, Avagliano P, Guardascione F, Petrelli G, di Giorgio P, Beatrice M, et al. (1999). Prevalence of HNPCC in a series of consecutive patients on the first endoscopic diagnosis of colorectal cancer: a multicenter study. The Italian Collaborative Group. *Endoscopy* **31**:337-341

Sankila R, Aaltonen LA, Järvinen HJ & Mecklin JP (1996). Better survival rates in patients with MLH1-associated hereditary colorectal cancer. *Gastroenterology* **110**:682-687

Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, Timmerman MM, Brody LC & Tucker MA (1997). The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N Engl J Med* **336**:1401-1408

Syngal S, Fox EA, Li C, Dovidio M, Eng C, Kolodner RD & Garber JE (1999). Interpretation of genetic test results for hereditary nonpolyposis colorectal cancer: implications for clinical predisposition testing. *JAMA* **282**:247-253

Task Force of Genetic Testing. Interim Principles. Baltimore, USA, 1995. Available at: http://ww2.med.jhu.edu/tfgtelsi/contents.html

Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, Merajver S, Thorlacius S, Offit K, Stoppa-Lyonnet D, Belanger C, Bell R, Berry S, Bogden R, Chen Q, Davis T, Dumont M, Frye C, Hattier T, Jammulapati S, Janecki T, Jiang P, Kehrer R, Leblanc JF, Goldgar DE, et al. (1996). The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. *Nat Genet* **12**:333-337

The American Society of Clinical Oncology: Genetic Testing for Cancer Susceptibility (1996). *J Clin Oncol* **14**:1730-1736

Thomson JP (1988). Leeds Castle Polyposis Group Meeting. *Dis Colon Rectum* **31**:613-616

Thorlacius S, Ólafsdóttir GH, Tryggvadóttir L, Neuhausen S, Jonasson JG, Tavtigian SV, Tulinius H, Ogmundsdóttir HM & Eyfjord JE (1996). A single BRCA2 mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. *Nature Genet* **13**:117-119

Thorlacius S, Struewing JP, Hartge P, Ólafsdóttir GH, Sigvaldason H, Tryggvadóttir L, Wacholder S, Tulinius H, Eyfjord JE, et al. (1998). Population-based study of risk of breast cancer in carriers of BRCA2 mutation. *Lancet* **352**:1337-1339

Tulinius H, Sigvaldason H, Ólafsdóttir G & Tryggvadóttir L (1992). Epidemiology of breast cancer in families in Iceland. *J Med Genet* **29**:158-164

Vasen HF, Griffioen G, Offerhaus GJ, Den Hartog Jager FC, Van Leeuwen-Cornelisse IS, Meera Khan P, Lamers CB & Van Slooten EA (1990). The value of screening and central registration of families with familial adenomatous polyposis. A study of 82 families in The Netherlands. *Dis Colon Rectum* **33**:227-230

Vasen HF, Mecklin JP, Meera Khan P & Lynch HT (1991a). The international collaborative group on hereditary non-polyposis colorectal cancer (ICG-HNPCC). *Dis Colon Rect* **34**:424-425

Vasen HF, Mecklin JP, Meera Khan P & Lynch HT (1991b). Hereditary non-polyposis colorectal cancer. *Lancet* **338**:877

Vasen HF, Taal BG, Nagengast FM, Griffioen G, Menko FH, Kleibeuker JH, Offerhaus GJ & Meera Khan P (1995). Hereditary nonpolyposis colorectal cancer: results of long-term surveillance in 50 families. *Eur J Cancer* **31**:1145-1148

Vasen HF, Wunen JT, Menko FH, Kleibeuker JH, Taal BG, Griffioen G, Nagengast FM, Meijers-Heijboer EH, Bertario L, Varesco L, Bisgaard ML, Mohr J, Fodde R & Khan PM (1996). Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. *Gastroenterol* **110**:1020-1027

Vasen HF, van Ballegooijen M, Buskens E, Kleibeuker JK, Taal BG, Griffioen G, Nagengast FM, Menko FH & Meera Khan P (1998a). A cost-effectiveness analysis of colorectal screening of hereditary nonpolyposis colorectal carcinoma gene carriers. *Cancer* **82**:1632-1637

Vasen HF, Haites NE, Evans DG, Steel CM, Moller P, Hodgson S, Eccles D, Morrison P, Stoppa Lyonet D, Chang-Claude J & Caligo M (1998b). Current policies for surveillance and management in women at risk of breast and ovarian cancer: a survey among 16 European family cancer clinics. European Familial Breast Cancer Collaborative Group. *Eur J Cancer* **34**:1922-1926

Warner E, Foulkes W, Goodwin P, Meschino W, Blondal J, Paterson C, Ozcelik H, Goss P, Allingham-Hawkins D, Hamel N, Di Prospero L, Contiga V, Serruya C, Klein M, Moslehi R, Honeyford J, Liede A, Glendon G, Brunet JS & Narod S (1999). Prevalence and penetrance of BRCA1 and BRCA2 gene mutations in unselected Ashkenazi Jewish women with breast cancer. *J Natl Cancer Inst* 91:1241-1247

Warthin AS (1913). Heredity with reference to carcinoma. *Arch Int Med* **12**:546-555

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C & Micklem G (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**:789-792

Chapter 19. Conclusions and recommendations

R. Sankila, L.J. Schouten and D. M. Parkin

Cancer registries play a key role in the process of evaluating and monitoring mass screening programmes. By providing timely analyses of the trends in incidence and tumour characteristics, and through linking the cancer registry database to that of the mass screening programme, they can contribute to a variety of analyses of the screening programmes.

By studying trends in the incidence of the target cancer, a cancer registry can provide insight into the effects of mass screening. A good example is the decrease in the incidence of cervical carcinoma in Western European countries. It is very likely that screening contributed greatly to this development (see Chapter 6 in this book). Studies in Scandinavia showed that the decrease in incidence of this tumour varied widely between countries and that the incidence decreased most rapidly in countries with well-organised mass screening programmes (see Chapters 5 and 6 in this volume). At the start of a new screening programme, it is possible to measure the effect of the screening by studying regional differences. Hakama (see Chapter 2 in this volume) demonstrated how to evaluate the effects when a breast-screening programme was introduced. Trends in incidence and survival have also been studied in Finland based on data which were obtained at different locations and from women of varying ages who had been invited for their first screening at different times.

The case of screening for neuroblastoma has shown how much of a disadvantage it can be not to have a cancer registry. Screening was introduced in some parts of Japan, but because there were no cancer registries in the regions concerned, it was not clear how large its (probably negative) impact was. It seems likely that a large number of abnormalities were detected and treated, which would have otherwise regressed spontaneously (see Chapter 17 in this volume). Indeed, in the era of evidence-based medicine, no screening programme should be launched without intensive collaboration with the population-based cancer registry.

Variables

The value of a registry in the evaluation of screening is greatly enhanced if it is able to collect data on the stage at diagnosis, and possibly also on other relevant indicators, such as tumour size, grade and treatment. On the basis of data from the American SEER registry, Stephenson (see Chapter 13 in this volume) showed that

263

there had been a huge increase in the registrations of prostate cancer, followed by a subsequent decrease. This trend in America was caused by large-scale opportunistic screening using the PSA test. It also became clear how the stage distribution of the prostate cancer cases registered differed before and during the 'PSA era'.

To be able to follow trends in tumour stage, a cancer registry must have complete and reliable data on this item (see Chapters 4 and 9 in this volume). This is a problem for many cancer registries, but it is hoped that new recommendations from the ENCR for recording stage in terms of 'binary TNM' will improve the situation.

In order to be able to better estimate the influence of such trends in diagnosis, it is valuable to separate cancers detected by screening examinations, and those found by other means. Ideally, this can be done by linking data from the cancer registry with the databank of the mass screening programme. If this is not possible, registries could utilise a variable 'method of detection' to capture this information (see Chapter 4 in this volume). However, it is not clear how reproducible the results would be using such a variable. Further, this item does require further definition, depending on the reason why data are being collected. For instance, the category of patients who present for a screening examination because of non-specific symptoms pose a considerable classification problem. If a tumour is detected, should it be considered that it was screen-detected or that it was detected because of the symptoms?

Indicators of effect

Since early detection of invasive cancers aims to increase the effectiveness of treatment, the most objective indicator of success is a reduction in deaths (other benefits, such as improved quality of life, are more difficult to quantify). For screening programmes, which aim to detect (and ensure treatment of) pre-invasive cancers, a reduction in the incidence of invasive disease is the target, as for cancer of the cervix, and, in part, for colon cancer (see Chapter 15 in this volume) and oral cancers. Generally, however, when detecting early invasive cancer, the idea is to improve the effectiveness of treatment, and hence reduce mortality. Naturally, survival is increased by early detection, which is a necessary condition for lower mortality, but not a guarantee (see Chapter 3 in this volume). In any case, survival data will rarely be available before mortality rates. Improved quality of life between diagnosis and death might also be considered a legitimate endpoint for evaluation and one that could be investigated through cancer registries.

It should be noted that incidence of cancer will not be reduced by screening for early cancers, but it may be increased. This effect will certainly be seen after screening is introduced and asymptomatic prevalent cases are discovered. If earlier

264

diagnosis were the only consequence of screening, then the final, cumulative incidence should be unaffected. However, the possibility that cancers are being detected which, in absence of screening, would never have surfaced clinically (overdiagnosis) implies that incidence, especially that of small cancers, may be permanently higher in screened than in unscreened populations (see Chapters 9 and 13 in this volume). It is, therefore, more useful to monitor the incidence of large and metastasised cancers, since a decreased incidence of these cancers is a necessary condition for a lowered death rate, although, again, not a guarantee (see Chapter 2 in this volume). When studying trends by stage, it is important to use incidence rates instead of proportions. If screening causes a sharp increase in the diagnosis of small tumours, there will be a decrease in the percentage of large tumours, even though there has been no actual decrease in the number of large tumours.

To clarify the effects of screening on mortality, the cancer registry data can be combined with the data from the death certificates (see Chapter 2 in this volume). Often the cancer registry data are more detailed regarding the cancer diagnosis than reports available to the clinician by the time he writes the death certificate. For example, many cervical cancer deaths have been allocated to 'uterus unspecified' in the death certificates. Further, only cancer registries can link mortality data to the respective dates of diagnoses. This is necessary in evaluation screening programmes, as they can only affect the mortality of cases diagnosed after the launch of the programme.

A cancer registry becomes even more useful if its database can be linked with those from mass screening programmes on a record level. However, this linkage is only possible if both registries have sufficient identification data or have a unique, identical and permanent patient number (see Chapters 4 and 8 in this volume). Such linkage provides a great deal of information about the impact of mass screening (see Chapter 9 in this volume). Trends can be studied in the number of screen-detected malignancies (if the screening organisation is able to retrieve this information), the number of interval cancers and the number of malignancies in people who do not participate in the screening or who do not belong to the target group. A complete population-based cancer registry usually forms the major means of identifying the interval cancers. The number of interval cancers is considered to be one of the most important early indicators of the quality of a screening programme. By providing feedback for the screening organisation about 'missed' cases, it is possible to adjust the detection criteria and referral pattern.

If the cancer registry has data on treatment, then this information can be useful for cost-benefit analyses on the screening. In case of neuroblastoma screening, the full cost of the screening could be assessed by including estimates of the unnecessary follow-up and treatments. Further, PSA screening undoubtedly causes unnecessary treatment, which, however, might be judged financially acceptable if there is

readiness to pay for the unknown, but potential, benefit and for the treatment of the evident side-effects.

Collaboration with screening organisations

Evaluation of mass screening can be one important reason to start and maintain a population-based cancer registry. Cancer registries should therefore be prepared to contribute to the evaluation of mass screening. One condition, however, is that the cancer registry data are complete, accurate and valid and that the data are rapidly available (see Chapters 4 and 7 in this volume). As a general rule, but only as far as feasible, a cancer registry should be willing to collect data that enable worthwhile evaluation of mass screening.

The cancer registry should be in contact with the organisation planning to start a screening programme. It is important to start the collaboration as early as possible during the planning phase. Often at the beginning, there is little understanding of what the evaluation and monitoring of such a programme requires. It is therefore necessary to explain the available data sources as well as the feasibility of collecting extra data items. In this way, the registry personnel can avoid future confusion when unrealistic expectations regarding the possibilities of evaluation may arise. It is also important to consult the screening organisation regarding the way their database is handled. Due to their experience in processing and linking databases, the cancer registries should be the best sources of advice for this purpose.

The design and analytical methodology used in any evaluation or monitoring exercise must always be decided after considering the unique aspects of each setting. The evaluation of the screening programme may require specially designed studies, which may or may not be incorporated in the day-to-day operations of the screening programme and those of the cancer registry. On the other hand, the periodical monitoring of the programme should be organised as a simple routine. The structure of the screening database, the delays in the data flow and in the linkages with other registers, e.g. the population register, will be the limiting factors in the monitoring process. These issues may not always be clear to the screening organisers, clinicians and public health professionals. Thus, the cancer registry should keep the other parties involved fully informed about the feasibility and requirements concerning the monitoring process.

In the planning phase, the cancer registry may have to consider adapting new methods of data collection and follow-up to fulfil the expectations on timeliness. Further, the monitoring of the screening programme may require extra data items to be collected. If judged feasible, some of these items may be added to the cancer registry database, e.g. method of detection. However, a separate screening registry

266

dataset may be more feasible as it may contain extra items necessary for the screening programme only, e.g., invitation data, technical data concerning the test and its results as well as data concerning re-screening and diagnostic procedures. Finally, once the programme is running, the monitoring process itself may reveal that modifications to the data collection procedures, data flow, and items collected may become necessary. Indeed, these special operations may be focused on the target cancer only, e.g., the data processing of breast cancer cases could be given special priority to provide rapid accumulation of data needed for monitoring a mammographic screening programme.

Venta • Salg • Verkauf • Πωλήσεις • Sales • Vente • Vendita • Verkoop • Venda • Myynti • Försäljning
http://eur-op.eu.int/general/en/s-ad.htm

BELGIQUE/BELGIË

Jean De Lannoy
Avenue du Roi 202/Koningslaan 202
B-1190 Bruxelles/Brussel
Tél. (32-2) 538 43 08
Fax (32-2) 538 08 41
E-mail: jean.de.lannoy@infoboard.be
URL: http://www.jean-de-lannoy.be

**La librairie européenne/
De Europese Boekhandel**
Rue de la Loi 244/Wetstraat 244
B-1040 Bruxelles/Brussel
Tél. (32-2) 295 26 39
Fax (32-2) 735 08 60
E-mail: mail@libeurop.be
URL: http://www.libeurop.be

Moniteur belge/Belgisch Staatsblad
Rue de Louvain 40-42/Leuvenseweg 40-42
B-1000 Bruxelles/Brussel
Tél. (32-2) 552 22 11
Fax (32-2) 511 01 84
E-mail: eusales@just.fgov.be

DANMARK

J. H. Schultz Information A/S
Herstedvang 12
DK-2620 Albertslund
Tlf. (45) 43 63 23 00
Fax (45) 43 63 19 69
E-mail: schultz@schultz.dk
URL: http://www.schultz.dk

DEUTSCHLAND

Bundesanzeiger Verlag GmbH
Vertriebsabteilung
Amsterdamer Straße 192
D-50735 Köln
Tel. (49-221) 97 66 80
Fax (49-221) 97 66 82 78
E-Mail: vertrieb@bundesanzeiger.de
URL: http://www.bundesanzeiger.de

ΕΛΛΑΔΑ/GREECE

G. C. Eleftheroudakis SA
International Bookstore
Panepistimiou 17
GR-10564 Athina
Tel. (30-1) 331 41 80/1/2/3/4/5
Fax (30-1) 323 98 21
E-mail: elebooks@netor.gr
URL: elebooks@hellasnet.gr

ESPAÑA

Boletín Oficial del Estado
Trafalgar, 27
E-28071 Madrid
Tel. (34) 915 38 21 11 (libros)
913 84 17 15 (suscripción)
Fax (34) 915 38 21 21 (libros),
913 84 17 14 (suscripción)
E-mail: clientes@com.boe.es
URL: http://www.boe.es

Mundi Prensa Libros, SA
Castelló, 37
E-28001 Madrid
Tel. (34) 914 36 37 00
Fax (34) 915 75 39 98
E-mail: libreria@mundiprensa.es
URL: http://www.mundiprensa.com

FRANCE

Journal officiel
Service des publications des CE
26, rue Desaix
F-75727 Paris Cedex 15
Tél. (33) 140 58 77 31
Fax (33) 140 58 77 00
E-mail: europublications@journal-officiel.gouv.fr
URL: http://www.journal-officiel.gouv.fr

IRELAND

Alan Hanna's Bookshop
270 Lower Rathmines Road
Dublin 6
Tel. (353-1) 496 73 98
Fax (353-1) 496 02 28
E-mail: hannas@iol.ie

ITALIA

Licosa SpA
Via Duca di Calabria, 1/1
Casella postale 552
I-50125 Firenze
Tel. (39) 055 64 83 1
Fax (39) 055 64 12 57
E-mail: licosa@licosa.com
URL: http://www.licosa.com

LUXEMBOURG

Messageries du livre SARL
5, rue Raiffeisen
L-2411 Luxembourg
Tél. (352) 40 10 20
Fax (352) 49 06 61
E-mail: mail@mdl.lu
URL: http://www.mdl.lu

NEDERLAND

SDU Servicecentrum Uitgevers
Christoffel Plantijnstraat 2
Postbus 20014
2500 EA Den Haag
Tel. (31-70) 378 98 80
Fax (31-70) 378 97 83
E-mail: sdu@sdu.nl
URL: http://www.sdu.nl

ÖSTERREICH

**Manz'sche Verlags- und
Universitätsbuchhandlung GmbH**
Kohlmarkt 16
A-1014 Wien
Tel. (43-1) 53 16 11 00
Fax (43-1) 53 16 11 67
E-Mail: manz@schwinge.at
URL: http://www.manz.at

PORTUGAL

Distribuidora de Livros Bertrand Ld.ª
Grupo Bertrand, SA
Rua das Terras dos Vales, 4-A
Apartado 60037
P-2700 Amadora
Tel. (351) 214 95 87 87
Fax (351) 214 96 02 55
E-mail: dlb@ip.pt

Imprensa Nacional-Casa da Moeda, SA
Sector de Publicações Oficiais
Rua da Escola Politécnica, 135
P-1250-100 Lisboa Codex
Tel. (351) 213 94 57 00
Fax (351) 213 94 57 50
E-mail: spoce@incm.pt
URL: http://www.incm.pt

SUOMI/FINLAND

**Akateeminen Kirjakauppa/
Akademiska Bokhandeln**
Keskuskatu 1/Centralgatan 1
PL/PB 128
FIN-00101 Helsinki/Helsingfors
P./tfn (358-9) 121 44 18
F./fax (358-9) 121 44 35
Sähköposti: sps@akateeminen.com
URL: http://www.akateeminen.com

SVERIGE

BTJ AB
Traktorvägen 11-13
S-221 82 Lund
Tlf. (46-46) 18 00 00
Fax (46-46) 30 79 47
E-post: btjeu-pub@btj.se
URL: http://www.btj.se

UNITED KINGDOM

The Stationery Office Ltd
Customer Services
PO Box 29
Norwich NR3 1GN
Tel. (44) 870 60 05-522
Fax (44) 870 60 05-533
E-mail: book.orders@theso.co.uk
URL: http://www.itsofficial.net

ÍSLAND

Bokabud Larusar Blöndal
Skólavördustig, 2
IS-101 Reykjavik
Tel. (354) 552 55 40
Fax (354) 552 55 60
E-mail: bokabud@simnet.is

NORGE

Swets Blackwell AS
Østenjoveien 18
Boks 6512 Etterstad
N-0606 Oslo
Tel. (47) 22 97 45 00
Fax (47) 22 97 45 45
E-mail: info@no.swetsblackwell.com

SCHWEIZ/SUISSE/SVIZZERA

Euro Info Center Schweiz
c/o OSEC
Stampfenbachstraße 85
PF 492
CH-8035 Zürich
Tel. (41-1) 365 53 15
Fax (41-1) 365 54 11
E-mail: eics@osec.ch
URL: http://www.osec.ch/eics

BĂLGARIJA

Europress Euromedia Ltd
59, blvd Vitosha
BG-1000 Sofia
Tel. (359-2) 980 37 66
Fax (359-2) 980 42 30
E-mail: Milena@mbox.cit.bg
URL: http://www.europress.bg

ČESKÁ REPUBLIKA

ÚVIS
odd. Publikaci
Havelkova 22
CZ-130 00 Praha 3
Tel. (420-2) 22 72 07 34
Fax (420-2) 22 71 57 38
URL: http://www.uvis.cz

CYPRUS

Cyprus Chamber of Commerce and Industry
PO Box 21455
CY-1509 Nicosia
Tel. (357-2) 88 97 52
Fax (357-2) 66 10 44
E-mail: demetrap@ccci.org.cy

EESTI

Eesti Kaubandus-Tööstuskoda
(Estonian Chamber of Commerce and Industry)
Toom-Kooli 17
EE-10130 Tallinn
Tel. (372) 646 02 44
Fax (372) 646 02 45
E-mail: einfo@koda.ee
URL: http://www.koda.ee

HRVATSKA

Mediatrade Ltd
Pavla Hatza 1
HR-10000 Zagreb
Tel. (385-1) 481 94 11
Fax (385-1) 481 94 11

MAGYARORSZÁG

Euro Info Service
Szt. István krt.12
II emelet 1/A
PO Box 1039
H-1137 Budapest
Tel. (36-1) 329 21 70
Fax (36-1) 349 20 53
E-mail: euroinfo@euroinfo.hu
URL: http://www.euroinfo.hu

MALTA

Miller Distributors Ltd
Malta International Airport
PO Box 25
Luqa LQA 05
Tel. (356) 66 44 88
Fax (356) 67 67 99
E-mail: gwirth@usa.net

POLSKA

Ars Polona
Krakowskie Przedmiescie 7
Skr. pocztowa 1001
PL-00-950 Warszawa
Tel. (48-22) 826 12 01
Fax (48-22) 826 62 40
E-mail: books119@arspolona.com.pl

ROMÂNIA

Euromedia
Str.Dionisie Lupu nr. 65, sector 1
RO-70184 Bucuresti
Tel. (40-1) 315 44 03
Fax (40-1) 312 96 46
E-mail: euromedia@mailcity.com

SLOVAKIA

Centrum VTI SR
Nám. Slobody, 19
SK-81223 Bratislava
Tel. (421-7) 54 41 83 64
Fax (421-7) 54 41 83 64
E-mail: europ@tbb1.sltk.stuba.sk
URL: http://www.sltk.stuba.sk

SLOVENIJA

Gospodarski Vestnik
Dunajska cesta 5
SLO-1000 Ljubljana
Tel. (386) 613 09 16 40
Fax (386) 613 09 16 45
E-mail: europ@gvestnik.si
URL: http://www.gvestnik.si

TÜRKIYE

Dünya Infotel AS
100, Yil Mahallessi 34440
TR-80050 Bagcilar-Istanbul
Tel. (90-212) 629 46 89
Fax (90-212) 629 46 27
E-mail: infotel@dunya-gazete.com.tr

ARGENTINA

World Publications SA
Av. Cordoba 1877
C1120 AAA Buenos Aires
Tel. (54-11) 48 15 81 56
Fax (54-11) 48 15 81 56
E-mail: wpbooks@infovia.com.ar
URL: http://www.wpbooks.com.ar

AUSTRALIA

Hunter Publications
PO Box 404
Abbotsford, Victoria 3067
Tel. (61-3) 94 17 53 61
Fax (61-3) 94 19 71 54
E-mail: jpdavies@ozemail.com.au

BRESIL

Livraria Camões
Rua Bittencourt da Silva, 12 C
CEP
20043-900 Rio de Janeiro
Tel. (55-21) 262 47 76
Fax (55-21) 262 47 76
E-mail: livraria.camoes@incm.com.br
URL: http://www.incm.com.br

CANADA

Les éditions La Liberté Inc.
3020, chemin Sainte-Foy
Sainte-Foy, Québec G1X 3V6
Tel. (1-418) 658 37 63
Fax (1-800) 567 54 49
E-mail: liberte@mediom.qc.ca

Renouf Publishing Co. Ltd
5369 Chemin Canotek Road, Unit 1
Ottawa, Ontario K1J 9J3
Tel. (1-613) 745 26 65
Fax (1-613) 745 76 60
E-mail: order.dept@renoufbooks.com
URL: http://www.renoufbooks.com

EGYPT

The Middle East Observer
41 Sherif Street
Cairo
Tel. (20-2) 392 69 19
Fax (20-2) 393 97 32
E-mail: inquiry@meobserver.com
URL: http://www.meobserver.com.eg

INDIA

EBIC India
3rd Floor, Y. B. Chavan Centre
Gen. J. Bhosale Marg.
Mumbai 400 021
Tel. (91-22) 282 60 64
Fax (91-22) 285 45 64
E-mail: ebicindia@vsnl.com
URL: http://www.ebicindia.com

JAPAN

PSI-Japan
Asahi Sanbancho Plaza #206
7-1 Sanbancho, Chiyoda-ku
Tokyo 102
Tel. (81-3) 32 34 69 21
Fax (81-3) 32 34 69 15
E-mail: books@psi-japan.co.jp
URL: http://www.psi-japan.co.jp

MALAYSIA

EBIC Malaysia
Suite 45.02, Level 45
Plaza MBf (Letter Box 45)
8 Jalan Yap Kwan Seng
50450 Kuala Lumpur
Tel. (60-3) 21 62 92 98
Fax (60-3) 21 62 61 98
E-mail: ebic@tm.net.my

MÉXICO

Mundi Prensa México, SA de CV
Río Pánuco, 141
Colonia Cuauhtémoc
MX-06500 México, DF
Tel. (52-5) 533 56 58
Fax (52-5) 514 67 99
E-mail: 101545.2361@compuserve.com

PHILIPPINES

EBIC Philippines
19th Floor, PS Bank Tower
Sen. Gil J. Puyat Ave. cor. Tindalo St.
Makati City
Metro Manilla
Tel. (63-2) 759 66 80
Fax (63-2) 759 66 90
E-mail: eccpcom@globe.com.ph
URL: http://www.eccp.com

SOUTH AFRICA

Eurochamber of Commerce in South Africa
PO Box 781738
2146 Sandton
Tel. (27-11) 884 39 52
Fax (27-11) 883 55 73
E-mail: info@eurochamber.co.za

SOUTH KOREA

**The European Union Chamber of
Commerce in Korea**
5th Fl, The Shilla Hotel
202, Jangchung-dong 2 Ga, Chung-ku
Seoul 100-392
Tel. (82-2) 22 53-5631/4
Fax (82-2) 22 53-5635/6
E-mail: eucck@eucck.org
URL: http://www.eucck.org

SRI LANKA

EBIC Sri Lanka
Trans Asia Hotel
115 Sir Chittampalam
A. Gardiner Mawatha
Colombo 2
Tel. (94-1) 074 71 50 78
Fax (94-1) 44 87 79
E-mail: ebicsl@slnet.ik

UNITED STATES OF AMERICA

Bernan Associates
4611-F Assembly Drive
Lanham MD 20706-4391
Tel. (1-800) 274 44 47 (toll free telephone)
Fax (1-800) 865 34 50 (toll free fax)
E-mail: query@bernan.com
URL: http://www.bernan.com

**ANDERE LÄNDER/OTHER COUNTRIES/
AUTRES PAYS**

**Bitte wenden Sie sich an ein Büro Ihrer
Wahl/Please contact the sales office of
your choice/Veuillez vous adresser au
bureau de vente de votre choix**
Office for Official Publications of the European
Communities
2, rue Mercier
L-2985 Luxembourg
Tel. (352) 29 29-42455
Fax (352) 29 29-42758
E-mail: info-info-opoce@cec.eu.int
URL: http://eur-op.eu.int

1/2001

5

These publications include the annual reports produced by the Employment and Social Affairs DG, such as 'Employment in Europe' and 'Equal opportunities for women and men in the European Union', and other key documents on the subjects of 'Employment and European Social Fund', 'Equality between women and men', 'Health and safety at work', 'Social security and social integration', 'Industrial relations and industrial change', 'Fundamental rights and anti-discrimination'.

Please send all orders to the sales office of your choice (see addresses overleaf or http://eur-op.eu.int).