Review of sojourn time calculation models used in breast cancer screening

Shan Cheung, Jane L. Hutton, Julia A. Brettschneider

August 2017

Abstract

For decades, researchers have been estimating sojourn time of breast cancers. This is primarily for identifying a suitable round length in a new breast screening programme. In recent years, the aim of researches on sojourn time is to evaluate round lengths in existing breast screening programmes. Lead time is also a well studied topic, and it is used to adjust the survival time of screen-detected cancer patients in studies on the efficacy of screening programmes.

Because observing sojourn time and lead time is infeasible, mathematical models are built to perform estimations. However, the terminologies used in these models are often not well-defined. This can cause difficulties for researchers to learn and interpret their results. In this study, the models and the parameters were reviewed.

The definition of sojourn time is inconsistent, in which the end point of this time length is the clinically detectable time in some papers and the clinical detection time in others. The starting point of the lead time for screen-detected cancer is the diagnosis time, which is approximated by the screening date in lead time modelling. As for sensitivity, it has been wrongly defined as the sensitivity of mammogram in most studies of breast cancer sojourn time. Instead, it should be the sensitivity of the whole diagnostic process within the chosen screening programme.

Two main types of the stochastic process models were used for sojourn time estimation – the time recurrence models and Markov chain models. These models simplify the natural history of breast cancer and, as a result, they carry numerous assumptions. Especially, the progressive assumption is likely to be invalid. The commonly used estimation methods are maximum likelihood, least square approach and Bayesian inference. The study settings and cohorts often affect the model and the data. These subsequently affect the estimated results. Researcher should understand the disease nature, the study settings, the population cohorts and the data collection process in order to build an appropriate model for the study cohort.

1 Introduction

Cancer is a major disease in developed countries. In the UK, 352 thousand new cases of cancer were diagnosed in 2013 and there were 163 thousand cancer deaths in 2014 [3]. A population-based cancer screening programme is a preventive strategy intended to reduce cancer incidence and death, by identifying people who appear healthy but may be at increased risk of cancer. Currently, there are three population-based cancer screening programmes in the UK; they are the UK National Health Service (NHS) Breast, Cervical and Bowel Cancer Screening Programmes. In these screening programmes, eligible people, in a certain age range, are invited to screening repetitively. For example, in the UK NHS Breast Screening Programme (NHSBSP), women aged 50

to 70 are invited for breast mammography every 3 years.

One way to evaluate the efficacy of a screening programme is to measure the length of the survival time gained in patients diagnosed with cancer through screening compared to those diagnosed through other routes (i.e. clinical detection). As the diagnosis of a cancer through screening is usually earlier than if the cancer was diagnosed clinically, the survival time of the screening group will appear to be longer than that of the clinically diagnosed group. This happens even if the earlier screen detection has no impact on the progression of the disease. This phenomenon is called lead time bias [1, 12, 26, 30, 36, 40], and the lead time is the length of time between screen detection and hypothetical clinical detection if the patient had not attended screening [18]. The length of survival time gained in patients participating in a screening programme can be corrected by deducting the lead time.

The length of the survival time gained through screen detection depends on many factors, such as the overall health of the eligible population, the disease nature and the organisation of the screening programme. In fact, screening programmes are designed according to the specific disease nature for the eligible population. For example, a cancer's detectable pre-clinical period (DPCP) is the time period between a cancer being screen-detectable and the time that the cancer shows symptoms or is clinically diagnosed. It is the only period of time that screening can lead to a diagnosis of the cancer before clinical detection. In other words, screening has to happen within the DPCP of a cancer for the cancer to be diagnosed through screening. Because of this, a high frequency screening programme can detect more cancers before they are clinically diagnosed and can detect cancer earlier than a low frequency screening programme. However, there are cost and harm implications to high frequency screening which may outweigh the benefits of earlier diagnosis.

For many decades, researchers studied cancer sojourn time, which is the length of DPCP, in order to plan or evaluate the round length in screening programmes [39]. The sojourn time cannot be accurately observed unless a cancer patient is put under observation without treatment, and leaving a cancer, especially an invasive cancer, untreated can be dangerous. It is therefore not ethical to conduct a randomised controlled trial (RCT) to observe cancer growth during DPCP in people. Sojourn time should also not be observed from cancer patients who refused or are not suitable for treatment, as many of these patients may have other health problems and will not be representative of the eligible population. Modelling, using cancer data available from screening programme trials or existing screening programmes, is, therefore, an important technique used to estimate the sojourn time.

Although modelling is often used to estimate sojourn time and lead time, the terminologies used in the models that were developed are often not well-defined. The terminologies can also be ambiguous and inconsistent across academic papers and users. This can cause difficulties for interested parties, such as modellers and applied statisticians, when learning about these topics. The definitional and semantic inconsistencies can also lead to misinterpretation of estimated results and incomplete understanding of the restrictions of the models and results.

The objective of this study is to understand sojourn time modelling, and find out mean sojourn time of breast cancer in the UK population. This research report will explain the terminologies and discuss how the inconsistency in the definitions impacts on the formation of models and estimated results. One modelling method will be applied on the NHSBSP data, which have not been analysed before, to estimate the sojourn time of breast cancer in the UK.

2 Definitions

Breast cancer screening aims to detect breast cancer before it becomes clinically detectable in order to prevent or delay death from breast cancer and to minimise cancer treatment costs and harm. We consider most cancers grow from biological onset to having clinically detectable cancer symptoms in a cancer pathway. Figure 2.1 illustrates the time points of cancer development in a persons life. In sojourn time modelling, time points are often presented as the age of a person, instead of a date. For example, the time is zero when a person is born.

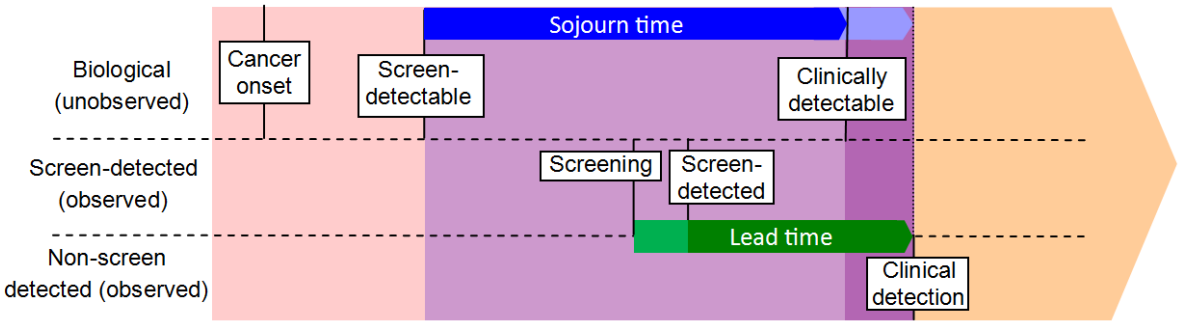

Figure 2.1: Unobserved biological cancer development time points and observed time points for screen-detected and non-screen detected cancers for patients who did not die before clinical detection. Three-stage model (explained in text): the pink area is the no disease stage; the purple area is the pre-clinical stage, which has two different end points depending on definition; the orange area is the clinical stage. The blue arrows represent sojourn time (explained in text). The green arrow represents lead time, which has two different starting points (explained in text). This shows the common scenario that screening happens between the screen-detectable time and the clinically detectable time, but screening can actually happen any time between the screen-detectable time and the clinical detection time.

This section will explain some terminologies often used in sojourn time and lead time modelling, including the cancer growth pathway above and the current breast screening pathway in the UK.

2.1 Unobserved time points on cancer growth pathway

Biologically, cancer is caused by the increased growth of neoplastically transformed cells [9]. This group of cancer cells may be non-invasive and stay non-invasive, causing no harm to the person. The cells may be invasive from the outset, having the ability to spread to a vital organ and cause severe disease or death. Some non-invasive cells may change their behaviour and become invasive. At some time point in this development pathway (Figure 2.1), a cancer may enter the stage of being detectable by mammography. It may subsequently start to cause symptoms with sufficient impact for the person to seek medical advice before the cancer becomes severe enough to kill the person, or the person dies from another cause. The screen-detectable time begins when a cancer becomes detectable by a screening programme, and the clinically detectable time begins when cancer symptoms are physically noticeable to a person.

Although the screen-detectable time and clinically detectable time are well defined in theory, in reality the measurement of these two time points depends on additional factors. The screen-detectable time varies with breast cancer type, the performance of mammography machines and film readers, and the clinically detectable time is subjective to an individual's embodied experience of symptom development (i.e. from sensation to symptom) or which clinician examines the individual. The existence of these time points is revealed through the diagnosis of cancer and, therefore, if a cancer is not diagnosed, we do not know whether a person has cancer or not. The time points may not follow the order displayed in Figure 2.1. Sometimes, the clinically detectable time is before the screen-detectable time. For example, certain types of breast cancer (such as Paget's disease)

cause symptoms but are not detectable through mammography. This scenario is rare because most breast cancers as they grow bigger, are likely to contain microcalcifications which are visible on a mammogram. In rare occasions, tumours can regress or stop growing.

2.2 Breast cancer detection process

In the UK, the diagnostic process includes mammography, ultrasound, clinical examination, needle biopsies and, in rare cases, an operation [4]. The diagnostic pathway is illustrated in Figure 2.2. Each cycle starts with a woman with a breast cancer at any stage in the tumour growth time line (Figure 2.1) and ends with a definite result to the diagnosis: either diagnosing as a screen-detected cancer, a non-screen-detected cancer or no cancer. A woman may go through more than one cycle to have her cancer diagnosed.

A screen-detected cancer is defined as a breast cancer diagnosed through the NHSBSP (Box 1). Other breast cancer diagnosed through other pathways (Box 2) are recorded as a non-screen detected cancer in cancer registries in the UK. The following examples give a clear picture about how the classification links to the diagnostic pathway.

- **Screen-detected cancer:** Woman A had a cancer that was screen-detectable but not yet clinically detectable when entering the pathway. She attended her NHSBSP screening appointment (Box 1). Her mammogram was abnormal and she was recalled for assessment. The cancer was then diagnosed (Box 3).
- **Non-screen detectable cancer:** Woman B had a clinically detectable breast cancer. She went to see her GP because of a breast cancer symptom. She was referred to the symptomatic clinic (Box 2) and the cancer was diagnosed (Box 5).
- **Screen-detected cancer:** Woman C had a clinically detectable breast cancer. She attended her NHSBSP screening appointment (Box 1) and she told the clinician that she had symptoms of breast cancer. She was recalled for assessment and the cancer was then diagnosed (Box 3).
- **Non-screen detected cancer:** Woman D had a cancer that was screen-detectable but not yet clinically detectable. She made an appointment with the Mark & Spenser screening programme (Box 2) and the cancer was then diagnosed (Box 5).
- **Missed cancer; false negatives:** Woman E had a cancer that was screen-detectable but not yet clinically detectable when entering the pathway. She attended her NHSBSP screening appointment, but her cancer was missed (Box 4).
- **No cancer:** Woman F had a cancer that could not be detected because of its size (i.e. not yet screen-detectable). She attended her NHSBSP screening appointment (Box 1), but this cancer was not diagnosed (Box 4).

As demonstrated in the examples of Woman C and D, the classification of screen-detected and non-screen detected cancers is not influenced by whether or not a cancer has symptoms. A small proportion of the screen-detected breast cancer is clinically detectable at screening, and a small proportion of the non-screen detected breast cancer is not clinically detectable. Because by definition screen-detected cancer can only be diagnosed if it was not clinically detected, screen-detected cancers are likely to be asymptomatic. Fine details on the diagnostic practice for patients with symptoms and for those within the NHSBSP can be found in [44] and [25] respectively.

In terms of the tumour growth time line (Figure 2.1), the screening time refers to the time of the first mammogram taken when a woman attends a screening appointment. A screen-detected breast cancer is a breast cancer diagnosed through the NHSBSP after the cancer becomes screen-detectable and before clinical detection. Otherwise, it is a non-screen detected cancer.

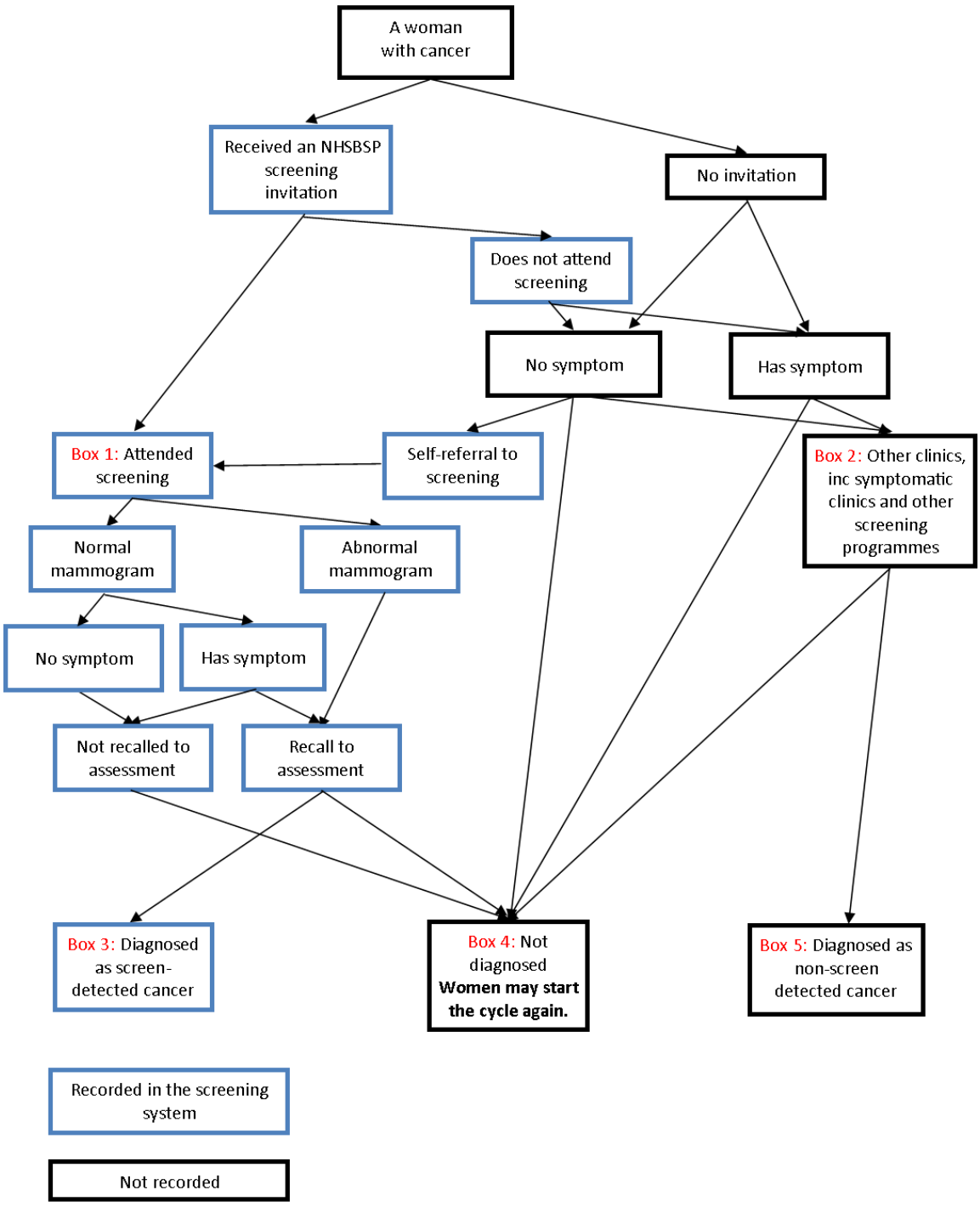

Figure 2.2: The UK diagnosis pathway for women with breast cancer.

2.3 Date of diagnosis

The date of diagnosis of a breast cancer is defined as the earliest time when the cancer was histologically confirmed. This can be the core biopsy or cytology result report date if the biopsy confirmed the breast cancer,

the surgery date if it was confirmed after an operation took place, or the date of death if the cancer was first noticed in a post-mortem. This date is neither a specific time point in the patient's diagnosis pathway nor in the cancer growth pathway. For screen-detected cancers, the screen-detection time is the date of diagnosis of the screen-detected cancer. For non-screen detected cancers, the clinical detection time is the date of diagnosis of the non-screen detected cancer. The date of diagnosis is important to patients and doctors as it is a date when the patient is confirmed to have cancer and treatment can be planned. It is also the most common starting point used in measuring cancer survival time.

2.4 Hypothetical time points

For screen-detected cancers with symptoms, the clinically detectable time is before the screening time. For screen-detected cancers without symptoms, the clinically detectable time does not exist after successful treatment. For both cases, the clinical detection time does not exist. We define a hypothetical clinically detectable time and a hypothetical clinical detection time for these cancers.

2.5 Three-stage process

A simple three-stage process has been proposed to model the biological growth of cancers [5, 6, 10, 16, 28, 29, 31, 39, 45, 48]. The three-stage process transits from a no disease stage to the pre-clinical stage and to the clinical stage (Figure 2.1). The no disease stage refers to the time before a cancer becomes screen-detectable (pink area). The pre-clinical stage is also called detectable pre-clinical period (DPCP). It has two definitions in academic papers: (1) when a cancer is asymptomatic (light purple area) [5, 17, 26] and (2) when a cancer can be diagnosed as a screen-detected cancer (light and dark purple area together) [14, 15, 48]. This will be further discussed in Section 3. The beginning of the clinical stage depends on the end point of the pre-clinical stage. If the definition of the pre-clinical stage is the asymptomatic period, then the clinical stage starts at the clinically detectable time (i.e. when a person has any breast cancer symptom). If the pre-clinical stage ends at the clinical detection time, the clinical stage refers to the stage when a cancer is no longer available for diagnosis through screening (orange area) [5, 48].

2.6 False negativity and sensitivity

A negative screen is defined as a mammogram without a visible abnormality in any part of the breast. In breast screening programmes, a false negative cancer refers to a screen-detectable cancer which was not diagnosed after the woman has been through the whole diagnostic process, which may be mammography alone or may include clinical examination, further imaging, ultrasound and needle biopsies. Mammography is the front line diagnostic method in the NHSBSP. It indicates the need for further investigations but does not necessarily diagnose a breast cancer. More specifically, a screen-detectable cancer becomes a false negative cancer because of a negative screen at the site where the cancer is located, or because it was not diagnosed during the further investigations following a positive screen. False negative rate is the probability to obtain a negative diagnosis for a woman with cancer.

The sensitivity of the diagnostic process is the probability of a cancer being detected by the diagnostic process. In breast screening, the sensitivity depends on many factors, such as the size and type of tumour, the density of the breast and the skill level of the technicians and clinicians [19, 27]. Further details about how false negative rate and sensitivity are defined in the sojourn time modelling publications are discussed in Section 3.2.

2.7 False positivity and specificity

The specificity of the diagnostic process is the probability of a negative diagnosis for women without cancer. In the context of a breast screening programme, that means the proportion of women not diagnosed by the screening programme within the cohort of all women without cancer participating in the screening programme. False positive cases are women who are wrongly diagnosed as having cancer after they have been through the diagnostic process. The false positive rate is the probability to obtain a positive diagnosis for a woman without cancer. This is 1 minus the specificity. There are fewer than 20 false positive cases per year [24].

False positive in mammography refers to the situation when abnormality is seen in a mammogram but the diagnosis is not a cancer. In NHSBSP, only one in three women who were recalled for further investigation after mammography examination is diagnosed with breast cancer. This is because mammography does not diagnose a cancer. Instead, the diagnosis of breast cancer is the first histological confirmation via a needle biopsy or a surgical operation in an assessment. False positive cases and specificity are not considered in most sojourn time models.

2.8 Interval cancers

In the NHSBSP, an interval cancer is a type of non-screen detected primary cancers diagnosed between two scheduled screenings. Interval cancers consist of false negative cancers and of cancers that have become screen-detectable after a negative screen at the same site as the interval cancer. In the diagnostic pathway (Figure 2.2), women with an interval cancer went through more than one cycle of diagnostic process. The following three examples illustrate the diagnostic pathway of women with the two types of interval cancers:

- **Previously false negative:** Woman G had a screen-detectable but not yet clinical detectable cancer when she attended her NHSBSP screening appointment (Box 1). Her cancer was missed (Box 4). Six months later, she found a lump (i.e. re-start the cycle) and went to her GP. She was referred to the symptomatic clinic (Box 2) and the cancer was diagnosed (Box 5).
- **Previously true negative:** Woman H had no cancer or a cancer which was not yet screen-detectable when she attended her NHSBSP screening appointment (Box 1). Her cancer was not diagnosed (Box 4). Six months later, she found a lump (i.e. re-start the cycle) and went to her GP. She was referred to the symptomatic clinic (Box 2) and the cancer was diagnosed (Box 5).
- **Previously false negatives:** Woman I had a screen-detectable but not yet clinical detectable cancer when she attended her first NHSBSP screening appointment (Box 1). Her cancer was missed (Box 4). Three years later, she attended her second NHSBSP screening appointment (Box 1) but her cancer was missed again (Box 4). Six months later, she found a lump and went to her GP. She was referred to the symptomatic clinic (Box 2) and the cancer was diagnosed (Box 5).

All three women attended an NHSBSP screening appointment and the results from this diagnostic process were negative. They then diagnosed with a non-screen detected cancer in a later diagnostic process. The difference on the classification is based on whether the cancer is screen-detectable at the time of the NHSBSP screen. A non-screen detected cancer is an interval cancer when it fits the following criteria:

- The woman must have attended at least one NHSBSP screening appointment.

- An interval cancer must be a primary breast cancers. Recurrences of previous primary breast cancers are not counted as interval cancers, but subsequent primary breast cancers in patients with a previous breast cancer are counted.
- A non-screen detected cancer diagnosed in the follow-up period, which starts from the date of the last screening examination. For example, the follow-up period of Woman I starts from the second attended screening appointment. The end point varies with patient age at diagnosis and regional data collection arrangements. For example, for a woman who attended only one screening appointment but no subsequent screening, the follow-up period for an interval cancer is the screening round length.

In the UK, all interval cancers must have a multi-disciplinary review [22]. During this process, some of the interval cancers are found to be false negatives at the woman's last screen.

3 Review of the use of definitions

In sojourn time and lead time modelling, sojourn time, sensitivity and lead time are the three main parameters for estimation. In most past papers, sojourn time is defined inconsistently [14, 15, 45, 49], sensitivity is often defined incorrectly [12, 41], and lead time is approximated using data available for sojourn time estimation [34, 49]. This section will discuss these issues in the definitions of the three parameters.

3.1 Sojourn time

The duration of the pre-clinical stage, or the length of the detectable pre-clinical period (DPCP), of a cancer in a three-stage process is the sojourn time of the cancer. The definition of sojourn time in the existing literature can be briefly split into two types:

1. the time length from screen-detectable time to clinically detectable time [5, 17, 26, 28, 29, 32, 45, 49]
2. the time length from screen-detectable time to clinical detection time [14, 15, 48].

Some papers do not define sojourn time [6, 10, 31] or do not specify the end point of sojourn time [16, 39]. The authors may not realise there is a difference between clinically detectable time and clinical detection time, or the difference has been ignored. Definition (1) solely depends on the biology of the disease. It is a simpler definition than definition (2) because it covers only the asymptomatic period. Definition (2) covers the asymptomatic period (i.e. from screen-detectable time to clinically detectable time) and a part of the symptomatic period (i.e. from clinically detectable time to clinical detection). The advantage of this definition is that it fits better with the screening data used in the model discussed in Section 4, because diagnosis through screening can take place anytime between a cancer's screen-detectable time and its clinical detection time. Specifically, a cancer detected by screening between clinically detectable time and the clinical detection time is still classified as a screen-detected cancer. The screen-detected cancer and interval cancer data in the estimation of sojourn time reflect the clinical detection time, but not the clinically detectable time. Hence, the estimated mean sojourn time refers to the time length from screen-detectable time to clinical detection, regardless of the original definition.

The inconsistency of these definitions is at the end points – clinically detectable time and clinical detection time. The size of the difference depends on the culture in the population, waiting time in the health system and the disease nature. A population that is more receptive to seeing doctors, shorter waiting times in primary and secondary care, and more noticeable symptoms will result a smaller inconsistency. For example, the

inconsistency is probably larger for bowel cancer in men compared to breast cancer in women. This is because men are more likely to wait longer before going to see a doctor than women, and because some bowel cancer symptoms, such as constipation, are also the symptoms of minor illnesses which are not alarming to patients. The clinically detectable time is difficult to measure because most patients do not remember the date of their first symptom(s). Researchers should beware of these inconsistencies in sojourn time definition and investigate the impact of the inconsistencies on their application.

3.2 Sensitivity

Sensitivity is often estimated simultaneously with sojourn time [5, 10, 12, 14, 26, 28, 31, 39, 41, 45, 48], because breast cancer detection through screening depends on the sensitivity of all the diagnostic procedures used during the diagnostic process. With the exception of Taghipour et al. [38] and Shows et al. [33], authors have wrongly defined sensitivity as the sensitivity of the mammogram alone. Instead, the sensitivity should refer to the sensitivity of all the diagnostic procedures. The interval cancer data consist of false negatives, which are cancers missed during the whole diagnostic process, not just by mammography (Section 2.6). For estimations using interval cancers, the results do not give an estimate of the sensitivity of mammography alone, but rather the sensitivity of the diagnostic process. Although the definition is technically wrong, it is believed that the difference between the sensitivity of mammography and the sensitivity of the whole diagnostic process is minimal these days when the sensitivity of digital mammography is high.

Defining sensitivity as the sensitivity of the front line diagnostic method is appropriate in other diseases where the diagnosis is confirmed by the screening result. For example, in the abdominal aortic aneurysm screening programme, an aneurysm is diagnosed by a scan of an aorta without going through further investigations. In this case, false negatives are the aneurysms missed by the aorta scan and the sensitivity would correctly be referred to the sensitivity of the aorta scan. Therefore, researchers should understand the diagnostic process and the data used in their estimations in order to define sensitivity correctly.

3.3 Lead time

Lead time is the time length by which the diagnosis is advanced by screening [18, 34, 40, 46, 49]. For non-screen detected cancers, the lead time is zero, because these patients do not benefit from screening. For screen-detected cancers, lead time is the time length from screen-detection (i.e. the diagnosis date of a screen-detected cancer) to the hypothetical time of clinical detection. This definition fits with cancer survival analyses, because the date of diagnosis is commonly used as a starting point in these analyses. However, the screen-detected cancer and interval cancer data, used to estimate lead time and sojourn time, are usually based on the screening date rather than screen detection date. This approximation results in the lead-time being slightly over-estimated, but the difference is usually small. For example, the time between screening and screen detection for breast cancers in 2013/14 in the UK has a median of 17 days.

There is a simple relationship between sojourn time and lead time in screen-detected cancers. Let s denote sojourn time, which is defined as the time length from screen-detectable time, t_s, to clinical detection time, t_c. Let l denote lead time, which is defined as the time length from screen-detection time, t_1, to clinical detection time, t_c. Hence,

$$s = t_c - t_s \text{ and } l = t_c - t_1. \tag{1}$$

Therefore, the relationship between sojourn time and lead time can be written as

$$s = l + t_1 - t_s. \tag{2}$$

More detailed lead time modelling can be found in these published papers: [15, 26, 29, 39, 46, 48, 49]. From here, this document will focus on sojourn time.

4 Review of models

One of the methods to estimate sojourn time is to build stochastic process models using data available from RCTs or screening programmes [10, 14, 39, 45]. Some of these models can be used to estimate the sojourn time of cancers in the screening arm of RCTs without data from the control arm, and they have also been applied to studies on population screening programmes [26, 42]. This section summarise the stochastic process models used in sojourn time estimation and discuss the issues in using these models.

4.1 Stochastic process models for sojourn time estimation

Two types of stochastic process models have been used in sojourn time modelling: recurrence time models [16, 29, 31, 39, 45, 49] and Markov chain models [5, 6, 10, 11, 14].

The recurrence time models describe screening times in a repetitive screening programme as random events relative to the transition times in the three-stage process. The screening examinations sample a distribution of sojourn time in a way that a cancer can be detected if a woman attends a screening appointment when her cancer is in its pre-clinical stage. A basic recurrence time model [39] involves the sensitivity (parameter), the screening time (variable), the probability of a tumour entering its pre-clinical stage in a time certain period (constant) and the sojourn time function (parameter(s)) where the sojourn times of cancers are assumed to follow a statistical distribution.

A Markov chain is a stochastic process with the property that, given the present state, the future is independent of the past. This can be interpreted as the memoryless property, which implies that the amount of time a cancer spends in each state before proceeding to the next state (i.e. the holding time) is exponential. In breast cancer progression modelling, the states are given by disease states such as in the three-stage process [5, 6, 10]. In Markov chain models, the holding time can be used to derive the transition probability, which is the probability of a process currently in one state to transition into another state within a certain time period, for example one year [10]. The transition probabilities are then used to formulate the probability of diagnosing a cancer at each screen and the probability of diagnosing an interval cancer, say, in each month [10] after screen.

Duffy et al. [11] has developed a Markov chain model with 5 states that capture the nodal status of cancer. Gunsoy et al. [14] has developed a 10-state Markov chain model which aims at estimating the extent of over-diagnosis. Prevost et al. [28] uses a recurrence time model for the interval cancer component and Markov chain model for the screen-detected cancer component for colorectal cancer.

4.2 Distribution of sojourn time

Sojourn time varies from tumour to tumour. The majority of studies assumed that the sojourn time distribution is exponential, because Walter and Day [39] found this to be a good fit after comparing exponential, lognormal and an empirical step functions using the US Health Insurance Plan study data. Exponential distribution is a mathematically simple distribution to work with and it is the only distribution used as the holding time in Markov chain models because of its memoryless property. More flexible distributions, e.g. the log-logistic distribution [45], have been used as the sojourn time distribution in recurrence time models.

4.3 Sensitivity as a parameter

The sensitivity of the diagnosis process determines the probability of a screen-detectable cancer being diagnosed. It affects the incidence of screen-detected breast cancers and interval cancers, so it is usually built into the model and estimated with the sojourn time parameters simultaneously. However, some authors [10, 17, 26] assume sensitivity is 100% when estimating sojourn time. Most older models assume sensitivity is constant with womens age, whereas Wu et la. [45] assumes it depends on age. The definition of sensitivity was discussed in Section 3.2.

4.4 Incidence rate of cancer onset

The incidence rate of cancer onset is defined as the relative frequency of women with cancer entering the pre-clinical stage during a given time interval in the population of all women at risk of breast cancer at the beginning of the time interval. It is also called the probability of a transition from the no disease stage to the pre-clinical stage [45, 48], the incidence rate of pre-clinical disease [8, 28] or the underlying incidence [10]. It is treated as a constant [39] or a parameter [10, 11] in the stochastic process models for sojourn time estimation. This probability is unobserved and it is estimated or approximated in past studies [16, 26, 28, 39, 42]. However, previous papers seldom talk about why and how the incidence rate of cancer onset is estimated or approximated and the impact of this probability on the estimated results. These will be discussed in Section 6.1.

4.5 Links between data, parameters and constants

Screen-detected cancer and interval cancer data can be obtained from RCTs or existing screening programmes. Researchers formulate the probability of diagnosing a cancer through screening and a cancer between screening examinations as functions of sensitivity, mean sojourn time and, sometimes, the incidence rate of cancer onset [10, 14, 39, 45, 48]. The interval cancer functions include functions of interval cancers which entered their pre-clinical stage before and after the woman's last screen. Chen et al. [6] did not use interval cancer data in their model, so screen-detected cancers from multiple screening rounds were required to capture more than one observed time point. The constants in previous models were set according to the model structure and the data available in the study. For example, in the Markov chain model developed by Duffy et al. [10], the constants are the number of women screened and the length of the time interval (i.e. one month) in the follow-up period of interval cancers.

4.6 Estimation methods

There are three main estimation methods used to estimate sojourn time.

1. Maximum likelihood estimation [28, 31, 45, 48] – A likelihood function is built according to the relationship between the screen-detected cancer and interval cancer data in each screening round in the corresponding data set. For example, the number of screen-detected cancers and interval cancers is assumed to be multinomial [31, 45], which implies an independent relationship between the number of screen-detected cancers and interval cancers in the cohort. Precisely, it is a partial likelihood, because the full likelihood of the natural history of disease is difficult or impossible to compute. As the likelihood function is complex, iterative methods, such as Newton's method, are used to maximise the likelihood.

2. Least squares approach [10, 28, 39, 42] – The function is a Poisson regression equation that the observed numbers of women diagnosed with a screen-detected cancer (independent variable) are equal to the expected numbers of women diagnosed with a cancer through screening (dependent variable) and error. The expected numbers are derived from the probability of diagnosing a cancer through screening and the probability of diagnosing a cancer between screenings, which have sensitivity and mean sojourn time as the parameters for estimation.

3. Bayesian approach [14, 28, 45] – A random sample is generated from the joint posterior distribution of parameters for Bayesian inference. For example, Prevost et al. [28] think that the mean sojourn time is probably between 0.005 and 5 years for their data set. They use a gamma(0.001,0.001) distribution truncated outside the region of 0.005 to 5 as the prior of the mean sojourn time.

4.7 Assumptions

The three-stage process simplifies the natural history of breast cancer. The important assumption is that all cancers are progressive [5, 6, 10, 16, 28, 29, 31, 39, 45, 48]. This progressive assumption implies that once a cancer reaches its pre-clinical stage, it will sooner or later leave the pre-clinical stage and enter the clinical stage. Most cancers follow this process, except cancers that regress and cancers that will never become clinically detectable. This assumption affects how the incidence rate of cancer onset is estimated and this will be discussed in Section 6.1. It also ignores over-diagnosis, which will be discussed in Section **??**.

In the Markov chain models, the time-homogeneous assumption implies that the incidences of pre-clinical and clinical disease are constant over patient age. However, older women are more likely to have breast cancer than younger women and it is likely that breast cancer incidence goes up with women's age. Nevertheless, if a study focuses on a smaller older age group, say age 50 to 60, the memorylessness is a reasonable assumption.

In the Markov chain models, the memoryless property implies that the sojourn time is exponentially distributed, but this assumption is not required for the recurrence time models. However, modellers using the recurrence time models often assume that the sojourn time is exponentially distributed. This is because, after comparing a log-normal, an exponential distribution and an empirical step function using the US Health Insurance Plan data, Walter and Day [39] found that the sojourn time fits an exponential distribution best. However, researchers should beware that an exponential distributed sojourn time implies some cancers stay in the pre-clinical stage for an unrealistically long time, which can be beyond death.

The three-stage process is designed to fit the most common scenarios in data collection. Uncommon cases are often over-looked. To put this in context, the stochastic process models with an interval cancer component carry an assumption that all interval cancers are clinically detectable at the time of diagnosis, but, in fact, some interval cancers are still in their pre-clinical stage at the time of detection. In the UK, although the majority of interval cancers are cancers diagnosed clinically through symptomatic clinics, some are diagnosed through private screening programmes and some are found accidentally without symptoms through, for example an x-ray of the chest or through routine health checks. However, this does not necessarily imply that the number of interval cancers recorded is more than the true number of interval cancers. This is because some interval cancers are recorded as screen-detected cancers because of the early recall process. This will be discussed in Section 5.

There are assumptions related to the estimation methods mentioned in the last section. A common assumption is that the random variables of numbers of screen-detected cancer and interval cancer cases are independent and identically distributed among women. This assumption is probably true because there is no known confounding factor in the probability of diagnosing a cancer between women. For the Poisson regression in the least square approach, the error term is assumed to follow a Normal distribution with a mean of zero and a constant standard derivation. The errors should also be independent and identically distributed. The Bayesian approach assumes that the parameters vary according to a posterior distribution, instead of being fixed as in the maximum likelihood method.

4.8 Approximation of the incidence rate of cancer onset

The incidence rate of cancer onset is the transition rate from the no disease stage to the pre-clinical stage. In most studies, it is measured in a unit of one year. Under the progressive assumption, all cancers that entered the pre-clinical stage will exit the pre-clinical stage. Therefore, when the number of cancers in the pre-clinical stage is constant over time, the incidence rate of cancer onset in a year and the incidence rate of clinical disease in that year, assuming screening is unavailable, will converge. In sojourn time estimations, the incidence of clinical disease was used to approximate the incidence rate of cancer onset in some past studies [16, 31, 39]. In a RCT, the incidence of clinical disease for a certain time period in the absence of a screening programme is the observed relative frequency of women with cancer in the control arm. This is approximated as the incidence of cancer onset in sojourn time studies of the RCT. We should be mindful that the population in the control arm cannot fully represent the women who attended screening [30]. This is because the screen-detected cancer and interval cancer data for a sojourn time estimation come from women who attended screening, which is just a part of the screening arm in a RCT. When the uptake (i.e. participation rate) is low, the observed incidence of clinical disease in the control arm can be higher than that in the women who attended screening if their cancer was not diagnosed through screening. In order to make the estimation more precise, researchers should adjust the clinical incidence for the selection bias.

Other approaches were used to approximate the incidence rate of cancer onset. Walter and Day [39] approximated it by the incidence of clinical disease before the screening programme rolled out. They assumed that the incidence rate of cancer onset is constant over year and age. For example, the incidence rates are the same between the year 1970 and 2000, and they are the same for all ages in each year. This method is not suitable for studies with the UK data, because the incidence rate of cancer onset in the UK is likely to be increasing continuously (Section **??**). Two other studies [16, 31] considered a step function of age, with constant incidence of clinical disease within each 5-year age group. They suggested utilising the cancer incidence data from the control arm of RCTs to approximate the age dependent probabilities of entering the pre-clinical stage. Weedon-Fekjær et al. [42] used historical data with an added time trend and they acknowledged that the change in usage of hormone replacement treatment may affect the accuracy of their estimate. Wu et al. [45] assumed the incidence rate of cancer onset follows a log-normal distribution of patient age. They estimated the two parameters of the log-normal distribution, sojourn time and sensitivity simultaneously. The incidence rate of cancer onset varies with country, age and social-economic level of the patients in the study cohort. Researcher should review the previous incidence rates in their study in order to choose or develop an appropriate estimation method.

4.9 How death is accounted for in sojourn time models?

Publications on sojourn time does not discuss about death, but the progressive assumption suggests that if a woman dies before her cancer reaches its clinical detection time, the cancer at its pre-clinical stage is assumed to be diagnosed at death. In practice, a breast cancer may not be found at death, because most deaths do not go through an autopsy. However, we can think of it in another way, namely that the pre-clinical stage is a stage when the cancer can be picked up by screening and the clinical stage is a stage when the woman with cancer is no longer available for screening. When a woman dies, her cancer will enter the latter stage, because she will cease to participate in screening. One reason to study sojourn time is to evaluate the round length in a population screening programme. From the screening programme point of view, the round length is for the screening population as a whole which includes women who die before their cancer is clinically detected. Hence, assuming a cancer is clinically diagnosed at death is actually fit for purpose in sojourn time estimations.

The three-stage process has no separate stage for death, but death can be a separate stage in a multi-stage model [14]. In studies on screening programmes, some authors [10, 28] account for death by using the number of person years at risk of having cancer at the time of hypothetical clinical detection to calculate the expected number of hypothetical clinical detections in the screening population. By definition, the women at risk of having cancer at the time of hypothetical clinical detection are alive at this time.

In sojourn time estimations, death is not considered in the distribution function of sojourn time. Most studies assume that sojourn time is exponentially distributed, which implies that the sojourn time is a positive value without an upper bound. This is contradictive to the progressive assumption, which implies the sojourn time cannot be infinite. A more recent model [14] has death as a separate absorbing stage and the sojourn times for cancers in patients who are alive and in those who die before a cancer's clinical detection are calculated separately. This is probably a clearer way to model death. However, for the purpose of evaluating round length in a screening programme, an overall sojourn time for all cancers taking account of the deaths in the screening population is preferred, rather than two separately estimated sojourn times.

4.10 How do the models handle multiple cancers?

A patient can have multiple primary breast cancers and multiple recurrences of breast cancer. There is no national recording of the incidence of recurrent breast cancer in England, but it is known that 4% of the women diagnosed with a primary breast cancer in 2008 had a previous primary breast cancer within the previous 22 years [43]. Some models [10, 17] are built to apply on data without previous breast cancers, and others [14, 31, 39, 45] have no restriction on this. There is a lack of information on how the multiple breast cancers are handled in sojourn time models. Most RCTs were performed on women who did not have a breast cancer before entering the trial, but it is unknown how the other type of multiple breast cancer were recorded in the trials' dataset. Nevertheless, if first primary breast cancer data are used in the estimation, the outputs (e.g. mean sojourn time and sensitivity) will also refer only to first primary breast cancers. In practice, the NHSBSP is interested in the sojourn time for all breast cancers in the screening cohort. This is because women are invited for breast screening regardless of having had a previous breast cancer, with the exception of women who are receiving cancer treatment and women who are ceased from being invited (such as women who had all their breast tissue removed at a double mastectomy).

5 Review of data sets

It is essential to understand the data before constructing a model. In sojourn time and lead time estimations, the models are built based on the data available in RCTs or screening programmes. The sojourn time definition varies in studies of the RCTs. Moreover, the RCTs are different in study settings. As a result, the data structures of them are different. This consequently affects the formulation of a model and the estimated results. This section will discuss the definition of sojourn time used in the RCTs' data sets, and how the study settings may change the model and the estimated results.

In past studies, various sojourn time estimation methods have been applied on data sets of RCTs. The majority of the models define sojourn time as the time length from screen-detectable time to clinically detectable time (Table 1). As mentioned in Section 3.1, this is a convenient definition, but the estimated sojourn time produced from these data sets fits into definition (2) instead.

A study by Shen and Zhen [32] presented a few factors that may influence estimated results. Using one estimation method on the screening arm of six RCTs, Shen and Zhen [32] found that the estimated mean sojourn times for

Data sets	Definition (1)	Definition (2)	No definition	Does not mention the end point
HIP, US	[26, 32, 33, 45]		[31]	[16, 39]
Two-county	[5, 26]	[48]	[6, 10]	
Malmö	[32]			
Stockholm	[32]			
Edinburgh	[32, 33]	[1]		
Canada 1	[7, 32, 33]		[31]	
Canada 2	[7, 32, 33]			

Table 1: Sojourn time definition used in studies with RCTs. Definition (1) : the time length from screen-detectable time to clinically detectable time. Definbition (2) : the time length from screen-detectable time to clinical detection time.

cancer had a high variability from 1.9 to 5.5 years. Such a variation may have resulted from differences in the cohort characteristics, study settings (Table 2) and confounding effects of hormonal replacement drugs used in the female population before and during the study and follow-up periods. Moreover, the specific randomisation method and inclusion criteria may also cause selection bias [13] and affect estimated sojourn time. For example, women in the control group of the Canada 1 and 2 trials had a physical examination for breast cancer; whereas women in the control group of the other trials had no active clinician involvement in detecting breast cancer. Consequently, these issues limit the generalisability of the estimate results.

In data sets from different parts of the world, the estimated mean sojourn times for breast cancer varies between countries. This may be because the development of a cancer is determined not only by biological processes but also by human intervention. At present, the majority of the sojourn time models have been developed based on data from the US [16, 18, 26, 31, 39, 45]. There are also models based on data from other countries, such as Taiwan [6], the Netherlands [34], Sweden [5, 6, 10, 26, 48], Canada [31], Norway [42] and Italy [26]. For the UK, Gunsoy et al. [14] developed an extensive Markov chain model to estimate mean sojourn time based women aged 40-49 years from the UK age trial data set. Yen et al. [47] analysed UK screen-detected cancer data from the early 1990s but did not use interval cancer data.

Study	Year Begun	Age at Entry	Screen Rounds	Screening Interval (years)	Modality	Sample Size	
						Study	Control
HIP	1963	40-64	4	1	2-MM & PE	30,239	30,756
Sweden							
Two-County	1977	40-74	2	2 (< 50) 2.8 (50+)	1-MM	78,085	56,782
Malmö	1976	45-69	2	1.8	2-MM	21,088	21,195
Gothenburg	1983	39-49	5	1.5	2-MM	11,724	14,217
Stockholm	1981	40-64	2	2.3	1-MM	39,164	19,943
Edinburgh	1976	45-64	6	1	PE	23,226	21,904
			3	2	2-MM (first)		
Canada 1	1980	40-49	4	1	2-MM & PE	25,214	25,216
Canada 2	1980	50-59	4	1	2-MM & PE	19,711	19,694

Abbreviations: 1-MM, single-view mammogram; 2-MM, two-views mammogram; PE, physical examination.

Table 2: Background information of eight breast cancer randomized controlled trials – US Health Insurance Plan (HIP), Swedish two-county trial, Swedish Malmö, Swedish Gothenburg, Swedish Stockholm, UK Edinburgh, Canadian 1 and Canadian 2 trials [32]

The estimated sojourn time may also be affected by the amount of data included. The minimum data items needed to estimate sojourn time, sensitivity or lead time are the number of women screened and the number of women diagnosed with cancer at multiple screening examinations. In order to make the estimate more

precise, most researchers also use the number of women diagnosed with an interval cancer between screening examinations. Screen-detected cancer data can be obtained from any RCT or existing screening programme, but interval cancer data may not be available easily, because they rely on rigorous extensive follow-up or good quality cancer registration records which can be linked to the records for women who attended screening. If the last screening date before an interval cancer diagnosis and the diagnosis date of the first interval cancer since the screening examination are available, researchers can calculate the number of women with an interval cancer diagnosed in defined time periods, for example on a monthly basis, after the last screen [10]. This can further improve the precision of the estimation. An example of this application is provided in Section **??**. While some earlier models require data from both the screening and control arms of an RCT [15, 49], most further developed models only need data from the screening arm, which means that the model is applicable to existing screening programmes.

The recording of multiple breast cancers may change the estimated results and the lack of information on how they are recorded can lead to misinterpretation. The existence of multiple primary breast cancers and recurrences is often overlooked by researchers. None of the research papers mention how these cancers are recorded in their data sets. Moreover, the words "cases" and "women" are used interchangeably with "cancer" in some RCTs, making it difficult for the reader to tell whether it is the number of cancer patients or the number of cancers in patients that are used in the analyses [2, 20, 21, 35, 37]. To give a specific example, in the National Breast Screening System (NBSS) of the NHSBSP, while both the number of "women with screen-detected cancer(s)" and of "screen-detected cancers" are recorded, it is usually the former but not the latter that is reported. Furthermore, while "cancers" and "screen-detected cancers" include multiple primary breast cancers and recurrences, "interval cancers" include multiple primary breast cancers but not recurrences. Therefore, researchers should minimise ambiguity and eliminate potential sources of misinterpretation in the wording for the data record, and they should either include or exclude multiple cancers for all patients involved in the data set, in order to obtain more precision in the sojourn time estimation.

In estimations based on the maximum likelihood method, the joint likelihood equation of the probabilities of diagnosing a cancer through screening and the probabilities of diagnosing a cancer between screenings is a reflection of the relationship between the cancers diagnosed through screening and the interval cancers recorded in a data set. For example, the likelihood equation published by Shen and Zhen [31] is based on multinomial mutually exclusive random outcomes of having either a screen-detected cancer, an interval cancer or no cancer for each woman in their study cohort. A data set of first primary breast cancers would be suitable for this likelihood. Practically, to gather a set of first primaries, the follow-up of cancer for a woman ends after she is diagnosed with a breast cancer. In a situation where a data set includes additional cancers diagnosed after the first diagnosis, the multinomial likelihood method would not be suitable. Unfortunately, the follow-up processes in the RCTs are not documented fully in academic papers (Table 3). The follow-up process and recording of cancers in the studies should be understood in order to build a correct likelihood equation for estimation.

The study design can affect the accuracy of the data, and this can lead to inaccurate estimation of sojourn time. For example, in the UK, the observed number of interval cancers is likely to be lower than expected because of three reasons: assessment, early recall practice and adjuvant therapy treatment. Firstly, an assessment is a series of diagnostic procedures, such as physical examination, further imaging and needle biopsy. After an initial mammography, a woman may be recalled for assessment to further investigate and to confirm the existence of breast cancer. The length of this investigation period ranges from a couple weeks to a few months. Secondly, a woman whose diagnostic process reveal early signs of cancer is usually invited for an early screening six months after the last screen. This scenario is called "early recall". Interval cancers developed during diagnostic or early recall periods are recorded as screen-detected cancers instead of interval cancers. RCTs and screening

Data sets	Q1	Q2	Q3	Q4	Q5
HIP trial	Yes	No	Unknown	12 months from the women's last screening	Histologically confirmed breast cancers detected between screenings
Swedish 2 county trial	Yes	Yes	Unknown	Average interval was 33 months for women 50 years of age and older and 24 months for women under 50 years of age.	Breast carcinomas diagnosed between screening examinations
Edinburgh	Yes	Yes	Yes	36 - 42 months	Cancers detected between the last trial screening and the first routine screening
Canada 1 trial	Yes	Un-known	Interval cancer must not have a cancer diagnosed in the previous screen.	12 months	Cancer occurred less than 12 months after a screening examination that did not result in recommendation for diagnostic evaluation.
NHSBSP	No	Yes	Include second primaries but not recurrence	Most women were followed up for around 36 months of the last screen.	Cancers clinically diagnosed from last screen to the next offered screening appointment or the end of follow-up whichever the earliest.

Table 3: Data recordings in RCTs and NHSBSP.

Q1: Are patients with previous breast cancers excluded from the study cohort?
Q2: Will a patient diagnosed with a breast cancer at screening be invited again for screening?
Q3: Do interval cancers include recurrence and second primaries?
Q4: What is the follow-up period of interval cancers?
Q5: What is the definition of interval cancers?

programmes with annual screening (e.g. Canada 1 and 2 trials and the part of the American breast cancer early detection program) are worst hit by these issues, because a significant part of the 1-year follow-up period for interval cancers are took up by the assessment and/or early recall period. Finally, adjuvant therapy treatment (e.g. endocrine therapy) given after a diagnosis of breast cancer suppresses the growth of breast cancers and patients can be treated for 5 years or more. The risk of developing an interval cancer for these women is lower than those who do not have adjuvant therapy treatment. Researchers should beware of the practices in the study population which could affect the accuracy of the data.

6 Summary

In this study, previous studies on sojourn time and lead time models have been reviewed. The parameters in these models were defined inconsistently and incorrectly. Sojourn time was defined in two ways with different end points – clinically detectable time and clinical detection time. In most studies on breast cancers, sensitivity was wrongly defined as the sensitivity of mammography, instead of the sensitivity of the whole diagnostic process used in a screening programme. The starting point of lead time, which is the diagnosis date of a screen-detected cancer, was usually approximated by the screening date. The effect of these on estimated results depends on the disease nature, culture in the population, waiting time in the health system and the diagnosis process.

Most sojourn time models assume a simple three-stage stochastic process of cancer growth, in which cancers progress from no disease stage to pre-clinical disease stage and then to clinical disease stage. It is assumed that all cancers will eventually be clinically diagnosed. Two types of stochastic models were used to estimate sojourn

time: the recurrence time model and the Markov chain model. Most publications assume that sojourn time is exponentially distributed.

Most data sets used in the previous sojourn time studies are from randomized control trials (RCTs) and some are from existing screening programmes. RCTs have more reliable information on the incidence of clinical disease from the control arm. However, most RCTs were carried out before the 1980s. It is unknown whether or not there is any change to the cancer sojourn time since then. In contract, using data sets from an existing programme can produce a more up-to-date estimate of cancer sojourn time. However, the incidence of clinical disease should be estimated with care.

References

[1] F. E. Alexander. Estimation of sojourn time distributions and false negative rates in screening programs which use two modalities. *Statistics In Medicine*, 8(6):743–756, 1989.

[2] F. E. Alexander, T. J. Anderson, H. K. Brown, A. P. Forrest, W. Hepburn, A. E. Kirkpatrick, C. McDonald, B. B. Muir, R. J. Prescott, and S. M. Shepherd. The Edinburgh randomised trial of breast cancer screenng: results after 10 years of follow-up. *British Journal of Cancer*, 70(3):542548, 1994.

[3] Cancer Research UK. Cancer statistics for the UK. Online, 2013 and 2014. http://www.cancerresearchuk.org/health-professional/cancer-statistics.

[4] Cancer Research UK. Breast cancer screening in the UK. Online, cited October 2016. http://www.cancerresearchuk.org/about-cancer/type/breast-cancer/about/screening/who-is-screened-for-breast-cancer.

[5] H. H. Chen, S. Duffy, and L. Tabár. A markov chain method to estimate the tumour progression rate from preclinical to clinical phase, sensitivity and positive predictive value for mammography in breast cancer screening. *The Statistician*, 45(3):307–317, 1996.

[6] H. H. Chen, H. S. Kuo, M. F. Yen, M. S. Lai, L. Tabár, and S. W. Duffy. Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics*, 56(1):167–172, 2000.

[7] Y. Chen, G. Brock, and D. Wu. Estimating key parameters in periodic breast cancer screening – Application to the Canadian National Breast Screening Study data. *Cancer Epidemiology*, 34(4):429–433, August 2010.

[8] N. Day and S. D. Walter. Simplified models of screening for chronic disease: Estimation procedures from mass screening programmes. *Biometrics*, 40:1–14, 1984.

[9] V. DeVita, S. Hellman, and S. Rosenberg. *Cancer principles and practice of oncology*, print 2, pages 102–103. Lippincott Williams and Wilkins, 6 edition, 2001.

[10] S. W. Duffy, H. H. Chen, L. Tabar, and N. E. Day. Estimation of mean sojourn time in breast cancer screening using a markov chain model of both entry to and exit from the preclinical detectable phase. *Statistics in Medicine*, 14:1531–1543, 1995.

[11] S. W. Duffy, N. Day, L. Tabár, H. H. Chen, and T. Smith. Markov model of breast tumor progression: some age-specific results. *J Natl Cancer Inst Monogr*, 22:93–97, 1997.

[12] R. Etzioni and Y. Shen. Estimating asymptomatic duration in cancer: the aids connection. *Statistics in Medicine*, 16:627–644, 1997.

[13] P. Gøtzsche and M. Nielsen. *Screening for breast cancer with mammography*, volume 1: CD001877. Cochrane Database Systemic Review, 2011.

[14] N. B. Gunsoy, M. Garcia-Closas, and S. M. Moss. Modelling the overdiagnosis of breast cancer due to mammography screening in women aged 40 to 49 in the United Kingdom. *Breast Cancer Research*, 14:R152, 2012.

[15] G. B. Hutchison and S. Shapiro. Lead time gained by diagnostic screening for breast cancer. *Journal of National Cancer Institute*, 41(3):665–681, 1968.

[16] S. Lee and M. Zelen. Scheduling periodic examinations for the early detection of disease: applications to breast cancer. *Journal of the American Statistical Association*, 93(444):1271–1281, 1998.

[17] T. Louis, A. Albert, and S. Heghinian. Screening for the early detection of cancer III. Estimation of disease natural history. *Mathematical Biosciences*, 40(1-2):111–144, 1978.

[18] J. D. Mahnken, W. Chan, J. D. H. Freeman, and J. L. Freeman. Reducing the effects of lead-time bias, length bias and over-detection in evaluating screening mammography: a censored bivariate data approach. *Statistical Methods in Medical Research*, 17:643–663, 2008.

[19] J. Michaelson, S. Satija, R. Moore, G. Weber, E. Halpern, A. Garland, D. B. Kopans, and K. Hughes. Estimates of sizes at which breast cancers become detectable on mammographic and clinical grounds. *Journal of Womens Imaging*, 5(1):3–10, 2003.

[20] A. B. Miller, C. J. Baines, T. To, and C. Wall. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *Canadian Medical Association Journal*, 140(10):1459–1476, 1992.

[21] A. B. Miller, C. J. Baines, T. To, and C. Wall. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *Canadian Medical Association Journal*, 140(10):1477–1488, 1992.

[22] National Health Service Breast Screening Programme. *Quality assurance guidelines for breast cancer screening radiology*, pages 14–15. NHSBSP Publication, 2 edition, 2011.

[23] NCIN National Cancer Intelligence Network. Cancer incidence by deprivation, England 1995-2004, 2009.

[24] NHSBSP and ABS. An audit of screen-detected breast cancers for the year of screening April 2011 to March 2012. Annual report, May 2013.

[25] Non-operative diagnosis subgroup of the National coordinating group for breast screening pathology, editor. *Guidelines for non-opeartive diagnostic procedures and reporting in breast cancer screening*. Number 50. NHS Cancer Screening Programmes, June 2001.

[26] E. Paci and S. W. Duffy. Modelling the analysis of breast cancer screening programmes: Sensitivity, lead time and predictive value in the florence district programme (1975-1986). *International Journal of Epidemiology*, 20(4):852–855, 1991.

[27] P. L. Porter, A. Y. El-Bastawissi, M. T. Mandelson, M. G. Lin, N. Khalid, E. A. Watney, L. Cousens, D. White, S. Taplin, and E. White. Breast tumor characteristics as predictors of mammographic detection: Comparison of interval and screen-detected cancers. *J Natl Cancer Inst*, 91(23):2020–2028, 1999.

[28] T. C. Prevost, G. Launoy, S. W. Duffy, and H. H. Chen. Estimating sensitivity and sojourn time in screening for colorectal cancer a comparison of statistical approaches. *American Journal of Epidemiology*, 148(6):609–619, 1998.

[29] P. C. Prorok. Bounded recurrence times and lead time in the design of a repetitive screening programme. *Journal of Applied Probability*, 19:10–19, 1982.

[30] S. Shapiro. Periodic Screening for Breast Cancer: The HIP randomized controlled trial. *Journal of the National Cancer Institute Monographs*, 22:27–30, 1997.

[31] Y. Shen and M. Zelen. Parametric estimation procedures for screening programmes: Stable and nonstable disease models for multimodality case finding. *Biometrika*, 56(3):503–515, 1999.

[32] Y. Shen and M. Zelen. Screening sensitivity and sojourn time from breast cancer early detection clinical trials: Mammograms and physical examinations. *Journal of Clinical Oncology*, 19:3490–3499, 2001.

[33] J. Shows and D. Wu. Inference for the lead time in breast cancer screening trials under a stable disease model. *Cancers*, 3:2131–2140, 2011.

[34] H. Straatman, P. G. M. Peer, and A. L. M. Verbeek. Estimating lead time and sensitivity in a screening program without estimating the incidence in the screened group. *Biometrics*, 53:217–229, 1997.

[35] L. Tabár, G. Fagerberg, N. E. Day, and L. Holmberg. 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. *British Journal of Cancer*, 55:547–551, 1987.

[36] L. Tabár, G. Fagerberg, S. W. Duffy, and N. E. Day. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *Journal of Epidemiology and Community Health*, 43:107–114, 1989.

[37] L. Tabár, G. Fagerberg, S. W. Duffy, N. E. Day, A. Gad, and O. Gröntoft. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiologic Clinics of North America*, 30(1):187–210, January 1992.

[38] S. Taghipour, D. Banjevic, A. B. Miller, N. Montgomery, A. K. S. Jardine, and B. J. Harvey. Parameter estimates for invasive breast cancer progression in the Canadian National Breast Screening Study. *British Journal of Cancer*, 108:542–548, 2013.

[39] S. D. Walter and N. E. Day. Estimation of the duration of a pre-clinical disease state using screening data. *American Journal of Epidemiology*, 118(6):865–885, 1983.

[40] S. D. Walter and L. W. Stitt. Evaluating the survival of cancer cases detected by screening. *Statistics in Medicine*, 6(8):885–900, December 1987.

[41] H. Weedon-Fekjær, B. H. Lindqvist, L. J. Vatten, O. O. Aalen, and S. Tretli. Breast cancer tumor growth estimated through mammography screening data. *Breast Cancer Research*, 10(3):R41, 2008.

[42] H. Weedon-Fekjær, L. J. Vatten, O. O. Aalen, B. Lindqvist, and S. Tretli. Estimating mean sojourn time and screening test sensitivity in breast cancer mammography screening: new results. *Journal of Medical Screening*, 12:172–178, 2005.

[43] West Midlands Knowledge and Intelligence Team. *The third all breast cancer report*. Public Health England and the National Cancer Intelligence Network Breast Site Specific Clinical Reference Group, 2015.

[44] A. M. Willett, M. J. Michell, and M. J. R. Lee, editors. *Best practice diagnostic guidelines for patients presenting with breast symptoms*. Department of Health, November 2010.

[45] D. Wu, G. L. Rosner, and L. D. Broemeling. MLE and bayesian inference of age-dependent sensitivity and transition probability in periodic screening. *Biometrics*, 61(4):1056–1063, December 2005.

[46] D. Wu, G. L. Rosner, and L. D. Broemeling. Bayesian inference for the lead time in periodic cancer screening. *Biometrics*, 63(3):873–880, September 2007.

[47] M. F. Yen, L. Tab´r, B. Vitak, R. A. Smith, H. H. Chen, and S. W. Duffy. Quantifying the potential problem of overdiagnosis of ductal carcinoma in-situ in breast cancer screening. *European Journal of Cancer*, 39(12):1746–1754, 2003.

[48] M. Zelen. Optimal scheduling of examinations for the early detection of disease. *Biometrika*, 80(2):279–293, 1993.

[49] M. Zelen and M. Feinleib. On the theory of screening for chronic diseases. *Biomatrika*, 56:601–614, 1969.