國立臺灣大學公共衛生學院

流行病學與預防醫學研究所

碩士論文

Institute of Epidemiology and Preventive Medicine

College of Public Health

National Taiwan University

Master Thesis

評估大腸直腸癌族群篩檢過度偵測數量模型

Quantitative Models for Modelling Overdetection in Population-based Screening for Colorectal Cancer

廖翎均

Ling-Chun Liao

指導教授：陳秀熙 博士

Supervisor: Hsiu-Hsi Chen, Ph.D.

中華民國 107 年 8 月

Aug 2018

doi:10.6342/NTU201802429

國立臺灣大學碩士學位論文
口試委員會審定書

評估大腸直腸癌族群篩檢過度偵測數量模型

Quantitative Models for Modelling Overdetection in Population-based Screening for Colorectal Cancer

本論文係 廖翎均 君（學號 p05849005）在國立臺灣大學流行病學與預防醫學研究所完成之碩士學位論文，於民國 107 年 6 月 4 日承下列考試委員審查通過及口試及格，特此證明。

口試委員：陳秀熙（簽名）

（指導教授）

張淑惠

潘信良

誌謝

這本論文的完成，開啟了人生新的篇章。

回想這兩年的光陰，從一開始的痛苦到習慣，或許對於知識的浩瀚來說，學習很多卻似一無所獲，但能夠習得思考方向與探索的態度卻是含菁咀華。碩一下學期的篩檢統計課程讓從未接觸過相關議題的我，覺得很感興趣，尤其是關於造成偏誤的觀念，還有其中的研究提供的實證證據解決了以前在洗手間看到四癌篩檢的宣傳告示內容的好奇與困惑，原來其中充滿了道理。

口試時，被問到對於這篇論文主題的感覺，當時回答到一半已哽咽得說不出話來。至於當時沒有說完的內容是什麼呢? 透過研究與寫作的過程，試著去「思辨」的感受，從無限多個問號到瞬間灌頂的感動是前所未有的，從零開始一直到後來義無反顧的投入，希望能夠一直記得，並且持續努力去創造。從論文這裡出發去回頭看本來學習的領域，也獲得了新的想法，其中許多都是有連結的，所以從今以後還沒完成的事還有好多，雖然資歷較淺還需要更多經驗的累積與整合，這只是個起步，換一個視野看見的不足會讓人進步。

謝謝兩位口試委員張教授與潘醫師的建議，讓本論文的架構與論點更加完整。

謝謝陳教授與 533 研究室的老師們與學長學姊們，從不拒絕這段期間任何的問題，不是悉心回答，就是在適度引導到正確的問題的道路上。在課堂上與研究室的咖啡時光啟發了我對於研究的各種想像。陳教授的指導與邱老師的原型、嚴老師、許老師的從旁指引，並給予實際的建議，讓人觀摩了學者兼教育者的氣場、熱情與追求高效能的精神，也接納了我們所擁有的不一樣，很珍惜能有這樣的緣分。

另外感謝台大醫院新竹分院所有已離職、未離職的藥師先進，藥劑部主任、各單位的組長與同事們，尤其是化療藥局好夥伴，沒有你們的協助排假、職代，與精神支持，我可能就無法順利安心地完成學業與學位論文，即使是一週只有一天，也讓人加倍的感恩，因為在這樣的條件下擠出時間跟精神從來就不容易。

謝謝同為專班的六位好同學，在這段期間互相幫忙、互相鼓勵，如今能一起順利畢業了，而且是全部一起畢業，希望未來還能有很多機會相聚討論學習，因為我們真的不僅僅只是來拿個學位而已，「終身學習」對於我們來說並不是問題，就讓這一股改變從畢業後開始。

最後謝謝我親愛的父母、弟弟鏡文, 書庭以及家人支持我的每個選擇。

翎均 于 2018 年瓜月

doi:10.6342/NTU201802429

Contents

doi:10.6342/NTU201802429

doi:10.6342/NTU201802429

List of Figures

List of Tables

doi:10.6342/NTU201802429

Abstract

Background Overdetection resulting from population-based screening for common cancers has been debatable over the past decade, but the formal methodology for quantitative estimation of overdetection is lacking. Contextually, mass screening for breast cancer and prostate cancer has been well documented, but colorectal cancer with faecal occult blood (FOB) test and faecal immunological test (FIT) on this thorny issue has been scarcely highlighted.

Objective This thesis aims to develop various quantitative methods for overdetection and to apply these methods to colorectal cancer with either guaiac FOB test (gFOBT) or FIT to understand the mechanism of overdetection concerning the natural course of colorectal cancer.

Data sources We used data on two randomised controlled trials on gFOBT and Taiwan population-based service screening using FIT to assess the proposed methods to quantify overdetection in population-based cancer screening.

Method The graphic method was first proposed by comparing cumulative incidence curves of advanced and non-advanced cancers. There are two modelling approaches used for quantitative methods. First, we developed the standardized overdiagnosis ratio (SOR), and the expected (numerator) is simulated by a three-state stochastic process merely based on data from the screened group against the observed

doi:10.6342/NTU201802429

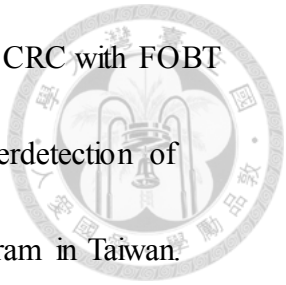
(denominator) either from the control group in the randomised controlled trial or the comparator from the pre-screening period in the service screening. The three state were, free of CRC, pre-clinical detectable phase, and the clinical phase in conjunction with the sensitivity of FOB test. Such a design was applied to two randomised controlled trials and one service screening program to estimate the proportion of overdetection of colorectal cancer resulting from FOB test and FIT by calculating (SOR-1)*100%.

The second proposed modelling approach is the development of a five-state Markov model in conjunction with SOR to quantify overdetection. Besides, to evaluate overdetection, the mechanism of overdetection through the pathway of progression to advanced pre-clinical detectable phase (PCDP) or the pathway to non-advanced clinical phase (CP) based on SOR was assessed in the light of the five-state Markov model.

Results Using the graphic method, the average of over-detection was 31.1% and 24.9% for the UK trial and the Denmark trial, respectively. However, the estimated proportion of over-diagnosis using the graphic method may be biased because the natural history of adenoma and high awareness in the routine clinical practice have not considered.

According to the proposed three-state disease progression model the 9%

doi:10.6342/NTU201802429



overdetection of CRC is expected in population-based screening for CRC with FOBT test based on the two randomised control trials in Europe and 7% overdetection of CRC were noted for FIT while using in a nationwide screening program in Taiwan. The estimated results derived from the five-state Markov model were 6.1% and 9.2% for the UK and Denmark trial, respectively. Indicated by SOR derived from the five-state Markov model, the important pathway for overdetection in population-based CRC screening is mainly the pathway from non-advanced PCDP to advanced PCDP. The pathway is supported by the evidence that SOR is deflated from the estimated results of 1.29 to 1.16 and 1.16 to 0.97 for the UK and Denmark trial, respectively.

Conclusions

This thesis systematically developed a series of statistical methods including the graphic method and the disease natural progression model for quantitative assessment of over-detection in colorectal cancer. The index of SOR was proposed to both assess the extent of over-detection and also to elucidate the mechanism of how overdetection affects the progression of colorectal cancer from PCDP to CP in the light of information on cancer stage.

Keywords: Colorectal cancer, population-based screening, overdetection, randomised controlled

doi:10.6342/NTU201802429

中文摘要

背景

對於目標癌症藉由不同的工具達到早期發現進而降低其死亡率為族群篩檢的主要目的，但此一過程中所衍生的過度偵測問題亦有許多的爭論。隨著族群篩檢對於不同癌症之廣泛運用，如何對篩檢中過度偵測進行量化評估對於篩檢之適切推行有迫切之重要性。目前在族群乳癌與攝護腺癌的篩檢已知存有過度偵測的情形，此兩種主要癌症之族群篩檢計畫之過度偵測亦在許多文獻中有所討論。而大腸直腸癌篩檢運用包含化學法糞便潛血檢查與免疫法糞便潛血檢查之族群篩檢計畫中的過度偵測則甚少被提及。

目的

本論文旨在發展一系列的方法對於族群篩檢中的過度偵測進行評估與量化。目前大腸直腸癌篩檢中的過度偵測在文獻上並未有許多探討，本研究將運用所發展的方法於以化學法以及免疫法為工具之大腸直腸癌族群篩檢。藉由此多種不同方式的評估方法結合大腸直腸癌疾病進展達到量化並評估大腸直腸癌篩檢中的過度偵測機轉。

資料來源

對於大腸直腸癌篩檢本研究運用兩個隨機分派試驗研究(化學法糞便潛血檢查)與台灣族群篩檢資料(免疫法糞便潛血檢查)對於其中之過度偵測進行量性評估與分析。

doi:10.6342/NTU201802429

研究方法

本研究發展三種對於族群篩檢中的過度偵測之量性評估方法。第一種方法為累積發生率曲線圖示法，透過比較隨機分派研究中篩檢組與未篩檢組之晚期癌症與非晚期癌症的累計發生率達到對於過度偵測之量性評估。第二類方法則運用數據模擬，藉由三階段隨機過程描述癌症之病程發展(無疾病期→無症狀可偵測期(臨床症狀前期)→臨床症狀期)達到估計疾病自然進展各階段之進展速度，同時對於篩檢工具敏感度加以考量，運用此實證評估結果推估得到篩檢邀請組中的預期個案數並且與篩檢組實際得到之觀察個案數相比得到過度偵測量性評估結果(標準過度偵測比率)。

本研究運用前述之方法於大腸直腸癌隨機分派研究資料，對於在篩檢計畫中之過度偵測進行量性評估。對於大腸直腸癌癌隨機分派研究則可藉由對照組作為比較基準評估過度偵測之情形；而台灣服務性族群篩檢之過度偵測評估，本研究則以篩檢前期之發生率資料作為比較基準，推估過度偵測比例(SOR，標準個案比之倒數-1)*100%。在前述對於族群篩檢中的過度偵測之了解下，本研究發展基於五階段大直直腸癌疾病進展模式之過度偵測評估方法，結合前述之標準化過度偵測率(SOR)達到量化過度偵測以及釐清其發生之機轉。

結果

運用累積發生率曲線圖示法於兩個大腸直腸癌隨機分派研究顯示英國與丹麥之研究中的過度偵測分別為 31.1%與 24.9%。但此一運用累積發生率之比較

doi:10.6342/NTU201802429

估計得到之數值由於並未將大腸直腸腺腫以及認知提高等因素納入考量，因此可能為具有偏誤之結果。

運用本研究提出之三階段馬可夫模型評估大腸直腸癌篩檢過度偵測之結果顯示，在兩個運用化學法糞便潛血檢查的歐洲隨機分派研究中過度偵測之情形接近 9%；而在台灣大型篩檢服務(運用免疫法糞便潛血檢查)則為 7%。進一步運用五階段大腸直腸癌疾病進展模式，英國與丹麥研究之過度偵測比率分別為 6.1%及 9.2%。運用本研究發展之標準化過度偵測率(SOR)對於兩個隨機分派研究評估之結果顯示，在排除過度偵測後，英國與丹麥之研究中 SOR 分別由 1.29 與 1.16 下降至 1.16 與 0.97。此一結果顯示，造成過度偵測之機轉主要源於大腸直腸癌疾病進展中，由早期臨床症前期轉移至晚期臨床症前期之疾病進展路徑所致。

結論

本論文系統性的發展出一系列量化估計過度偵測的統計方法，運用於評估大腸直腸癌族群篩檢中的過度偵測。藉由本研究發展之標準化過度偵測率(SOR)除了可達到對於過度偵測之量性評估外，亦進一步運用臨床症前期進展至臨床期之癌症期別變化達到對於大腸直腸癌篩檢中造成度偵測之機轉有所了解之目的。

關鍵字: 大腸直腸癌、族群篩檢、過度偵測、隨機分派

doi:10.6342/NTU201802429

Chapter 1: Introduction

Pros and cons of population-based screening

Cancer is the leading cause of death in the world. The burden of medical care for cancer is tremendous, particularly when it reaches the late stage. Population-based screening for early detection of screen-detectable cancer has been proposed to solve this issue. The benefits of population-based screening in the reduction of advanced cancer and cause-specific mortality have been well demonstrated in a series of common cancers including cervical cancer, breast cancer, colorectal cancer, oral cancer, and prostate cancer. (Sankaranarayanan, Ramadas et al. 2005, Koong, Yen et al. 2006, Tabar, Vitak et al. 2011, Schroder, Hugosson et al. 2014, Chiu, Chen et al. 2015, Yen, Auvinen et al. 2015, Auvinen, Moss et al. 2016, Yen, Tsau et al. 2016, Chuang, Su et al. 2017).While the effectiveness of mass screening for such screen-detectable cancers has been demonstrated their several cons that have been raised in recent years, including early investment of tremendous costs and unnecessary treatment for overdetection of early cancer that would not have been detected in the absence of population-based screening. The harm of population-based screening resulting from overdetection (or over-diagnosis) has been debated over the past decade. However, to estimate the quantity of overdetection is a thorny and challenging task.

doi:10.6342/NTU201802429

Overdetection in population-based screening

From the viewpoint of natural history, slow-growing cancer or non-progressive cancer detected via screening may not have progressed to advanced cancer and diagnosed with clinical symptoms or signs during life. Those cancers would not have existed in the absence of screening and are often defined as over-detected cases. From clinical aspect, there are several side-effects on overdetection. The following therapies and treatments of over-detected cases are unnecessary. Overtreatments and therapies not only cause harm to patients but also complicate the communication with healthcare givers and impose burdens on healthcare systems. From the aspect of health policy-makers, it is of necessity to provide quality assurance of a screening programme to control overdetection to relieve the anxiety of patients and also reduce the cost imposed on the limited resources allocated to health.

Lacking formal assessment of overdetection

Although overdetection is still debatable, there is lacking theoretically sound of methodology in the quantitative estimation of overdetection. The majority of methodologies in the evaluation of overdetection rely on data from a randomised controlled trial. The conventional epidemiological method for quantifying overdetection through the empirical data on randomised controlled trial is to make the comparison of overall cumulative incidence rate between the invited group and the

doi:10.6342/NTU201802429

uninvited group. The excess cumulative incidence is therefore attributable to over-detected cases. However, as demonstrated in the previous article, such an evaluation is also subject to the time of follow-up with forward time (lead-time), the time of advancing the date of diagnosis as a result of screening. To tackle this problem, the incorporation of cancer stage information would be conducive to clarifying the influence of forward time. This thesis would develop a graphic method, taking cancer stage information into account to cope with this problem. However, when data on randomised controlled trial are not available, the conventional method cannot be used. Besides, there are other issues of quantifying the proportion of overdetection when one evaluates the overdetection of service screening program. For example, the recently proposed evaluation of overdetection is prone to overestimation of overdiagnosis. The used of time series analysis on the basis of before and after study design in population-based screening program on breast cancer,reaching to 50% overdetection of breast cancer, due to the failure of considering forward time (lead-time) (Jørgensen, 2009) the time of advancing the date of diagnosis as a result of mammography. Moreover, estimating the overdetection in population-based service screening in the absence of the randomised controlled trial design would also be affected by the sensitivity of the screening program. To tackle these problems, one has to resort to the application of the stochastic process that takes into account forward

doi:10.6342/NTU201802429

time and sensitivity.

Overdetection in population-based screening for colorectal cancer

While overdetection in population-based screening for cervical cancer, breast cancer, and prostate cancer has been evaluated by using the conventional method population-based screening for colorectal cancer with stool-based screening methods via guaic-Faecal Occult Blood Test(g-FOBT) or Faecal Immunological Test (FIT) has yet to be evaluated. In Taiwan, we have already launched a nationwide population-based screening for CRC with FIT, one of the largest service screening programs in the world, therefore, provides an opportunity to modelling overdetection of population-based service screening program.

Objective

By making use of two randomised controlled trials with g-FOBT in Denmark and UK and also one Taiwanese population-based service screening with FIT, this thesis aims to propose a series of statistical methods for modelling overdetection of population-based screening for colorectal cancer that has been never formally assessed quantitatively. The specific objectives are to

(1) Envisage a novel graphic method to estimate the quantity of overdetection with the consideration of using cancer stage information;

(2) Develop a standardised overdetection ratio (SOR) to quantify the proportion of

doi:10.6342/NTU201802429

overdetection based on a three-state Markov process;

(3) Develop a stage-adjusted overdetection ratio by using five-state Markov process to elucidate how overdetection affects the pathway leading to the progression of advanced colorectal cancer.

doi:10.6342/NTU201802429

Chapter 2. Literature Review

In this chapter, we first describe the situation of measuring overdetection in mass screening programme of breast cancer (2.1)and prostate cancer(2.2), the major two cancer which mass screening has been well documented, and the possible method for awaring or avoiding bias.

For screening programme of colorectal cancer (2.3) with guaic faecal occult blood test (g-FOBT) and faecal immunological test (FIT) on overdetection has been scarcely highlighted, we examine the evidence of screening tools relevant to study of a population-based screening program, for the detail to further dealing this issue.

2.1. Overdetection in screening programme for breast cancer

2.1.1 Evolution of population-based mammography screening

The issue of overdetection has drawn great attention among research people. In the era of population-based mammography screening, not only the benefit such as mortality reduction, but also the cost that incurred by overdiagnosis is to be considered. Although data derived from randomized controlled trial provide rich information for the estimation of overdetection, there are many countries where data on randomized controlled study design is not available. Furthermore, the implementation of mammography service screening involves large population and the

doi:10.6342/NTU201802429

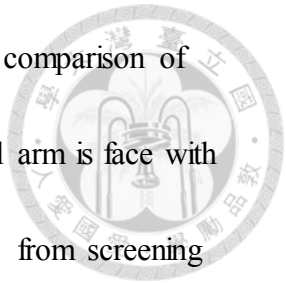
assessment in overdiagnosis is even more important during the trial period. Since the active case finding in screening programme will be discovered, not only those who eventually progress to clinical phase (CP) but also those will never be noticed if left untreated or diagnosed, namely overdetection or overdiagnosis. Assessing overdiagnosis based on the comparison between the trend of breast cancer incidence before and after the introduction of population-based screening programme is an intuitive approach. Jørgensen and Gøtzsche (2009) tempt to quantify overdiagnosis brought by population-based mammography screening based on such an intuition. The evaluated the extent of overdiagnosis by comparing the difference between the expected incidence of breast cancer derived from the era without mammography screening programme and the observed one of seven years after the introduction of population-based mammography screening programme. Based on before-after comparison, Jørgensen and Gøtzsche found that 52% (95% CI: 46-58%) of breast cancer cases identified in the area of mammography screening belong to the category of overdiagnosis. They concluded that the introduction of population-based mammography screening is responsible for the increasing incidence of breast cancer and one in three breast cancers detected in such a screening programme is due to overdiagnosis (Jørgensen and Gøtzsche, 2009). Although the basic methodological flaws are full with the article, we focus ourselves on the inappropriateness of making

doi:10.6342/NTU201802429

a comparison with biological aspects of breast cancer evolution associated with the implementation of population-based screening programme. These flaws such as using before-after design derived from the quasi-experimental research to analysing the data. They may neglect the lack of comparability between the two, thus misused the design that performed as in the randomised controlled trials.

As a fundamental biology, the occurrence of cancer starts from the molecular or cellular level. After the initiation of the dysplastic process, the occult lesion progresses and evolves into clinical overt disease. Between the initiation at the molecular level and clinical disease, there is some time period that the lesion can be detected by suitable tools such as mammography and hence the status of preclinical detectable phase, PCDP. In a screening-naïve population, the introduction of a screening programme (prevalent screen) will pick up cases that already entering the status of PCDP but not yet surface to clinical disease. Were it not for the active case finding of the screening programme, such cases remain occult and stay in the PCDP pool until the day that the lesion progresses and becomes a clinical overt disease. The soaring incidence during the early phase of the introduction of a screening programme is thus due to the discovery of cases in the PCDP pool, or the lead-time gain. The duration for the increasing incidence is closely related to the dwelling time of the target cancer to be screened. For breast cancer, a three-year period is well recognized.

doi:10.6342/NTU201802429



Event for data derived from randomized controlled group, the direct comparison of excess breast cancer cases between the screening arm and the control arm is face with the issue of imbalance time domain caused by the lead time inherited from screening activity. So, let alone the use of before-after design to assess the issue of overdiagnosis. However, in the setting of randomized controlled trial design, we have a good chance to have the estimate to quantify overdiagnosis as long as the follow-up period is sufficiently long not only to concur the aforementioned effect of lead time but also to allow for the non-advanced cancers to progress and surface to the clinical phase.

As a paradigm of evidence-based-policy making, the strategy of breast cancer prevention through population-based mammography screening has been adopted in western countries, including American, Canada, and most of Europe countries and also in Asian countries including Taiwan and Japan. Although the fruitful success in bringing down the breast cancer mortality through the widespread secondary measures targeting at early detection of breast neoplastic lesions together with the improvement in therapy and treatment for detecting breast cancer cases, the diversity of nature inherited within the characteristics of breast cancer render a small portion of breast cancer cases suffered from modest survival. This heterogeneity in biological characteristics also makes it sometimes difficult to stratify the population into the risk

doi:10.6342/NTU201802429

of the occurrence of breast neoplastic lesion. It is also of great interest to stratify the patients into the levels of prognostic risks for the implementation of effective treatment and therapy.

To detect this most common cancer of women in early stage, several modalities were used in strategies of breast cancer screening dependent on risk factors such as age, hormones, obesity and family history. These modalities including are clinical breast exam (CBE), imaging modalities such as ultrasound, mammography, digital breast tomosynthesis (DBT), and magnetic resonance imaging (MRI). Through the decades, mammography is used widely in detecting microcalcifications and remain the most important screening modality as the only one already proved its efficacy in mortality reduction (National Comprehensive Cancer Network 2017). Population-based randomized controlled trials for the evaluation the efficacy, including advanced cancer reduction and mortality reduction of mammography screening and the meta-analyses provide the base of evidence. Even though, the debate on adopting mammography in population-based breast cancer screening due to its harm mainly from overdetection draws attention in recent years. Arguable estimate of overdetection in mammography screening in publications varied from 0 % to 54 %, depending on the cohorts, intervention strategies, and measurement. Higher reports are mostly due to lack of adjustment of breast cancer risk or lead-time (Puliti, Duffy et al. 2012).

doi:10.6342/NTU201802429

2.1.2 Heterogeneity in the performance of population-based mammography screening

To clarify the recent debate on population-based mammography screening, Chen et al. developed a causal model to indicate the cause of heterogeneity across relative trials. They first did a systematic review included English articles between 1970 and 2015. Inclusion criteria of articles were randomised controlled design on breast cancer screening with mammography, an average-risk population with available tabular data, attendance rate, detection mode (e.g., interval cancers, screening-detected cancer, and cancer from non-attendee.), and reports of mortality and results of advanced breast cancer. Studies of High-risk women applied to mammography were excluded. Unlike the usual systematic review of RCTs, the number of participants in a trial for population-based screening was larger than drug clinical trial that score of weight was difficult to use, and publication bias would not be problematic because they were unlikely to be unreported. According to the above reason, the authors included all existing evidence on mammography screening for their study.

The degree of overdetection was one of the critical components that were described in the causal model, whereas the other two were attendance rate and the

doi:10.6342/NTU201802429

sensitivity using interval cancer incidence and expected incidence rate. Pooling data from nine trials, a meta-analysis was performed using Bayesian conjugated beta-binomial distribution. The random-effect model was used to capture the heterogeneity across trials, adjusted for year of conducting trials, screening interval, the logarithm of counts of interval cancer, and the population in different age group. Conducted the first trial was Canadian trial hold at 1963, while the latest UK trial was held in 1991, almost across 30 years.

They used "number of screens required for overdetection (NSO)" for one cancer to describe the absolute rate of overdetection by taking its inverse form. They estimated the possible range for overdetection and taking the average of the lower and upper estimation. Among nine trials, the estimated over-detected breast cancer rate was lowest in Gothenburg trial for 39 to 49 years old women (22 per 105), and highest in Canadian trial (167 per 100,000) with women aged 40 to 59 years. The average NSO ranged from 597 people in The Canadian National Breast Screening Study-1(CNBSS-1) to 4482 in Gothenburg trial, women age 40 to 59 years.
With the variance of strategy with heterogeneity between breast cancer screening trials, the robust estimation of overdetection in breast cancer should be therefore considerate these factors that play essential roles in the benefits or harms and find out the balance to make policy decisions.

doi:10.6342/NTU201802429

2.1.3 Overdetection and ductal carcinoma in situ (DCIS)

Another consideration of overdetection in breast cancer was the detection of ductal carcinoma in situ (DCIS), which increased with the introduction of mammography (IARC, 2002) and it now accounts for up to 20% of newly diagnosed breast cancer. The estimated rate of over-diagnosis varied in literature, because of the methods and protocols that include or exclude DCIS also affect the estimate. Experts may agree that a proportion of DCIS would not progress to invasive cancer without screening. However, the proportion is unknown, and this issue is more complicated with lead-time and the study design by the observed excess of participants and ethical issues as provided a closure screen at last in the control group. Yen and Tabar et al., use data from Swedish Two-County trial, population service screening programme from the UK(Chamberlain, Moss et al. 1993, Moss, Michel et al. 1995), New York (Dershaw, Loring et al. 1998), South Australia (Robinson, Crane et al. 1996), and the Netherlands (Fracheboud, de Koning et al. 2001) to access the over-diagnosis of DCIS in service screening. They propose a six-state continuous-time Markov model for capturing the disease progression, and they take progressive and non-progressive DCIS into account with the separated absorbing state. The study did not include Mammographically-undetectable DCIS. Non-progressive DCIS (DCIS0) will not

doi:10.6342/NTU201802429

progress to invasive cancer, which also remains unclear with its natural history, but progressive DCIS (DCIS1) will follow the route of the preclinical- detectable phase to invasive breast cancer.

In their study results, there was low proportion over-diagnosis consistence with randomised control trial of screening, 37 % (approximately 20% to 50%) detected DCIS cases were estimated to be non-invasive in their first screening, while in incidence screen is 4 %, the larger proportion of non-invasive cancer, the more severe problem of overdetection. (Yen, Tabar et al. 2003). In addition to the study performance and analytical approach method, the unclear natural progress of specific cancer would be another interesting issue in distribution of overdetection.

2.1.4 Assessing overdetection in breast cancer screening using data on randomized controlled trials

Cumulative incidence is one common expression of results in screening studies, and it can give us some information about overdetection. In the above systematic review with Bayesian meta-analysis of 9 population-based randomised controlled trials in breast cancer screening, Chen et al. compared the cumulative incidence by dividing breast cancer into advanced and non-advanced, both in invited group and uninvited group. They envisaged two different situations for high and low estimates,

doi:10.6342/NTU201802429

and further assessed the situations based on the incidence of advanced breast cancer. For high estimates, they assume in the enough observed time, all non-advanced cancer will progress to invasive cancer in non-screening cases; in low estimates, not all advanced breast cancer would progress from non-invasive cancer during the trial period (Chen, Yen et al. 2017). Then, they compared the cumulative incidence difference between advanced cancer in the invited group and overall non-advanced cancer for the low estimates of overdetection, and the cumulative incidence difference of non-advanced cancer in the invited group and overall advanced cancer was the high estimated. In their proposed model, the mean of advanced stage and non-advanced stage breast cancers in the study group and the control group per person-year were followed the Gamma distribution and then simulated to high and low estimates of overdetection. They consider the two extreme scenarios of estimation, and represented average estimates of overdetection.

The percentage of overdetection of mammography on breast cancer screening range from 0% to 28%, after adjusting for lead-time, it was noted that the proportion of overdetection depends on the lead-time gained by early detection and the follow-up time. The reported low estimates of overdetection were similar to the difference of cumulative incidence between two groups. They noted that the low estimated is often report as over-detect estimated that based on comparing two arms in RCT, and thus

doi:10.6342/NTU201802429

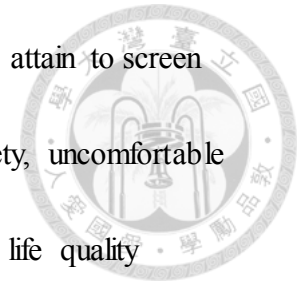
were not considered the influence of non-advanced cancer with longer lead-time. It would be interesting to apply their method to low and high estimates of overdetection in colorectal cancer service screening, but the method is dependent on randomised controlled design, and the lead-time and sojourn time should not be too long compared to the total follow-up time. The longer follow-up time after randomisation will wash out the excess of screen-detected cases that can be explained by lead-time, and more likely to be caused by overdetection.

2.2 Overdetection in prostate cancer screening programme

2.2.1 Debating of population-based screening of prostate cancer

Since the widespread application of prostate-specific antigen (PSA) as a screening tool for detecting prostate cancer, the benefit and drawbacks of this population-wide strategy have gained great attention. There were several population-based randomised controlled trials for evaluating the efficacy of the PSA-based screening strategy. However, these clinical trials gave controversy conclusions about benefits mainly due to the difference in the age group of the selected population, the proportion of contamination in the control group, and the PSA cutoff value for

doi:10.6342/NTU201802429



detecting positive cases. On the other hand, the cons were men who attain to screen may thus suffer from the undesired effect of screening such as anxiety, uncomfortable confirmatory process, over-aggressive treatment, and compromised life quality incurred by the complications (Donovan, Hamdy et al. 2016). Quality of life after overdetection of PCa screening and treatments also draws attention in the recent decade, Heijnsdijk and colleagues modelled QALYs loss caused by over-diagnosis based on the data derived from the ERSPC. The study demonstrated the efficacy of saving six cancer deaths and 73 life-year gained from the annual screening programme targeted at the man who aged between 69. The number need to screen and cases need to detect to prevent one cancer death in such a programme is 98 and 5, respectively (Heijnsdijk, Wever et al. 2012). Prostate cancer (PCa) is one of the most prevalent cancers in man in western countries. Therefore, it is reasonable to apply population-wide strategy to screening. On the contrary, the disease burden of the disease was low for the Asian population. This difference in disease burden further raised the concern on adapting the unified PSA-based screening programmes world-widely.

2.2.2 Disease progression of prostate cancer and screening strategy

The nature of disease progression of PCa further complicates the issue of whether PSA-based screening programme is feasible for being taken as a population-wide

doi:10.6342/NTU201802429

strategy. Based on the observation derived from PSA-based screening, most of the prostate neoplasms are localised and slow-growing lesions which take 5 to 12 years before they are surfacing to overt disease with the clinical symptom. Given that PCa occurred more often in the more elderly population, the probability of detecting indolent lesions that progress slowly or non-progressive disease is thus a concern of overdetection in population-wide PSA screening programme. A simulation study based on the European Randomized study of Screening for Prostate Cancer (ESRPC) data addressed one single PSA test at age 55, 75, and four years interval at age 55-67 male, the over-diagnosis was 27%, 56%, and 48%, respectively (Draisma, Boer et al. 2003). However, with variation across studies, we cannot give a simple conclusion to answer the complicated question of strategy in population-wide PCa screening, for evaluating the benefits and harms and get the advantages to outweigh the disadvantages. Lead-time bias and length bias both arise from sojourn time which represents the progression time of prostate cancer from asymptomatic to symptomatic, are often be ignored, as the similar example of one publication in breast cancer we mention on 2.1. The premise of the comparison between screen-detected and clinically-detected PCa was these biases should be first considered into adjustment, separating the excess cases arise from non-overdetection reason, then dealing with overdetection.

doi:10.6342/NTU201802429

2.2.3 Factors associated with overdetection in prostate cancer screening

After adjusting the bias to sojourn time for population-based screening RCT, It is also interesting to know the impact of screening policy on effects such as decrease of PCa mortality and their impact on overdetection at the different settings of incidence rate, screening tools and intervals for organised a tailored screening programme. In men aged 40 years with low PCa incidence, the population-wide screening is less cost-effective for the population. When it comes to the era of personalisation, we may curious about which strategy choice is the most utility. To access the impact of screening components such as the screening interval, target of age range at screening, attendance rate and contamination, for the efficacy of population-based screening at first, Wu et al build a seven-state Markov model and performed decision analyses with 25 years follow up in simulation, constructed with tumor characteristics, screening sensitivity and over-diagnosis. The model defined by three states included free of disease, preclinical stage, and clinical stage, and two tumour stages, advanced prostate cancer and early prostate cancer ,for prostate cancer death. Data cohort with age 55, 59, 63, and 67 years old in The Finnish screening trial were implemented to Markov decision analysis for comparing different screening policy, included no screening, single screening, and periodic screening in the 4-year interval. They compared the incidence of advanced prostate cancer without over-detected cases and

doi:10.6342/NTU201802429

the mortality of prostate cancer with the no-screening group for each strategy. They found 11.1% (95%CI 9.1-13.3%) reduction in the advanced tumour at the start age 55 to70 with 65% attendance rate and 20% contamination. The screening interval had more impact on the mortality reduction than the start screening age. Although the reduction of mortality was reduced when delayed 5 years of screening age by 9% in annual screening and 3% in biennial screening, the impact was negligible (Wu, Auvinen et al. 2012). Although the model can provide predictions of different strategies and the novel insight of personalised screen, conducting RCTs is still the most convincing way for making screening policy but unavoidable costly, therefore, modelling methods may be feasible.

2.2.4 Modelling approaches for overdetection in Prostate cancer screening

For policymakers and clinical health caregivers to evaluate the efficacy of mass screening of prostate-cancer or certain cancers, it is difficult to conduct a large RCT in Asia country as previous European and USA studies by the culture, economic environment, and political situation, and another concern is they are differ in population and cancer burden. As in most of the RCTs were designed to prove the efficacy for screening, factors that play a crucial role in affecting outcome were also be quantified as previous studies, based on pilot RCTs to determine to hold service

doi:10.6342/NTU201802429

screening program for a population. In public health aspects, service screening program was not a study, but a policy that would hold for years and provide services for many times that may further influence recently and future population health benefits or harms. As the evaluation of efficacy can be affected by many factors, there is no reason that the proportion of overdetection would remain the same in all situation of screening programme. Since few researcher designed RCTs for overdetection in the population-based setting and the disease burden are growing, the estimate of the proportion of overdetection could be a need to solve the question.

There were several modelling approaches used for adjusting biases when comparing the survival between groups in the screening programme. One of the methods based on continuous-time Markov process, in addition to the correction of lead-time bias and length bias, they also considered overdetetion cases. Wu at al proposed a Markov model of natural history for prognosis of prostate cancer, applied in comparing the survival in screening-detected cases (study group) and clinical detected case (control group) from Finnish data, also a part of ERSPC. Between January 1996 and January 1999, the trial randomly allocated 32000 men in the screening group and 48458 men in the control group. They invited participants in the study group to a screening with serum PSA cut off 3.0 ng/mL in the 4-year interval, the mean age at randomised was 58.7. Total 20,796 participants (587 in the screening

doi:10.6342/NTU201802429

group and 20209 in the control group) were followed from their first screen in 1996 until study end in 2005.

In their model, the definition of overdetection was differed from the traditional clinical definition that another cause of death occurring before clinical diagnosis of PCa is considered overdetection. The advantage of using non-conventional definition is they can separate three possible situations that could not overcome by using conventional definition. The three possibilities included non-susceptible to PCa, non-progressive PCa, and progressive PCa that died from another cause before PCa diagnosis. Data from population-based randomized controlled trial is possible to access by using parameters $\lambda 0(\cdot)$, $\lambda 1(\cdot)$, $u0(\cdot)$ and $u1(\cdot)$ to represent incidence rate, transition rate, and hazard rates of two different states of PCa, progressive and non-progressive, died from another cause of death.

They supposed two types of prostate cancer have different sojourn time from being detection to become symptom cancer, and actual survival of clinical detected PCa and screen-detected PCa without bias can be compared by separating two states of PCa. Cancer which has long sojourn time tend to be detected, by contrast, the interval cancer, which surfaces between screening intervals tend to have short sojourn time. They built a length-bias adjustment model according to the consideration to get unbiased estimated of transition rate($\lambda 0(\cdot)$, $\lambda 1(\cdot)$ and $\lambda 2(\cdot)$) in four state (Normal,

doi:10.6342/NTU201802429

Pre-clinical disease state, Clinical state, and death) and then extended it to mover-stayer model for estimating overdetection. They excluded following situations when comparing the survival, the non-progressive prostate cancer, called the "stayer" to capture the infinite time of never becoming progressive cancer, or never died from progressive cancer. Therefore the "stayer " did not need to estimate $\lambda 1(\cdot)$ and $\lambda 2(\cdot)$. The two parameters was corresponded to the lead-time and post lead-time, and they can be use in lead-time adjustment. For length-bias adjustment, $\lambda 1(\cdot)$ can also be used because lead-time distribution is derived from sojourn time distribution. By estimating the survival of mover after adjusting lead-time and length bias, they can further deal with overdetection, the major problem in PCa screening.

Wu et al. assumed there would be no gain in the survival of screening-detected prostate cancer if there is no overdetection. During the 9-year follow-up period, the hazard ratio of Prostate cancer death increased from 0.24 (95% CI: 0.16–0.35) to 1.03(95% CI: 0.79–1.33), after correcting for lead-time, length bias, and overdetection, and there was around 24% mortality reduction if no overdetection, but the calibration method of overdetection may need further verification with different and longer follow-up data. Insufficient follow-up time may lead to the distinction impossible between both stayer and mover with sojourn time longer than the follow-up time, and the screen-detected mover cases can be early treated may further

doi:10.6342/NTU201802429

contribute benefits of the screen and reverse the original conclusion. They defined the sojourn time of non-progressive PCa was infinite, and this means we can never see the detected non-progressive cancer becomes progress PCa. In such a case, the disease does not have the chance to result in the clinical symptom in the subjects were it not being detected in the active screening programme, and this made the evaluation more difficult in population service screening than in RCT (Wu, Auvinen et al. 2012).

Another method considered sensitivity, screening policies on age and interval as these factors varied across studies but few can deal with them to the issue of overdetection. Wu, Auvinen et al. taking sensitivity into account in the proposed multistate model depicting the natural progression of prostate cancer, and developed a quantitative measure for overdetection, named "number of screens for overdetection (NSO)" for estimating the absolute "risk of harm elicit by overdetection(RO)". In addition to the neoplastic lesion which will eventually progress and surfacing to clinical phase, the indolent lesions that are non-progressive and detected in screening programme are also incorporated into the proposed disease model. In the scenario of a screening programme for PCa, those recognised as normal cases included three components: the truly PCa free, false-negative-non-progressive PCa, and false-negative-progressive PCa. False-negative-progressive type attendees would be detected in subsequent rounds of screen or surface to clinical phase and being

doi:10.6342/NTU201802429

recognised as interval cancer before the next scheduled screening round. On the other hand, False-negative non-progressive type will not turn to clinical phase even with long-term follow-up. Due to the randomised controlled study design, the control group can provide information of progressive PCa, and make an unbiased estimation of the preclinical incidence rate of PCa, by assuming it equal to clinical incidence rate without overdetection. However, the method above requires several assumptions in their model construction. First, the time from PCa free to before screen-detectable phase was defined as finite. Those who died before entering preclinical-detectable phase would be considered as the loss to follow-up and non-susceptible subjects due to other cause of death. The second assumption was this noninformative censorship also had the potential to be detected as progressive PCa but died from competing causes of death. Third, the progress of advanced cancer was a hidden process that we cannot observe directly. They defined non-progressive PCa in the screen detect mode, while interval cancer, cancers in non-participants and control group were all progressive PCa. It is nonsense to pick up non-progressive PCa from progressive PCa based on the assumption of non-progressive PCa will never become progressive PCa in biology viewpoint. Another possible limitation of their model was the transition rates were assumed constant with time because of Markov property. As age change with time, sensitivity can also be different in each screen rounds.

doi:10.6342/NTU201802429

Comparing the result of estimated NSO in a single screen, started age and screening intervals differed in the degree of impact. First, NSO decreased with age at the first screen, demonstrated that the overdetection increase. NSO decreased from 104 to 48 at age 55 to 60, but the difference between age 60 and 65 was little. Second, longer screening intervals had the larger estimated NSO, it remarkably decreased from 8 years to 4 years, but this was not obvious when the interval was less than two years. Overall RO increases with increase screening round also implying that overdetection can occur in each round of screening. Comparing RO by the combination of starting screening age and screening interval, it decreases with the longer interval, and younger age before 60, indicating the chance that a person will be over-detected as PCa in the screening period. They thought this was related to the biological characteristic of long mean sojourn time of progressive PCa. The sensitivity for detecting the progressive and non-progressive type of PCa. As growing tumour produces more PSA and end up with clinical symptoms, sensitivity can also change during prevalence and subsequent screens, thus Wu et al., also applied different sensitivity estimates, and further considered it as an age-related factor, although the difference was not significant.

The study concluded that in the setting of Finish trial, the NSO was 29% (95 confidence interval (Cl) 18% to 48%) for screen men aged 55 to 67 years, equal to

doi:10.6342/NTU201802429

every 100 men screened, 3.4 men would be overdetection during three screen rounds. According to their discovery, NSO is related to age, screening interval, and the number of screening rounds (Wu, Auvinen et al. 2012), this epochal information that components in the strategy of screening had an impact on overdetection in population-based screening of PCa ,and could be possible quantified .

2.3 Overdetection in the screening programme for Colorectal cancer

Colorectal cancer, the third most common cancer, accounts for 694,000 deaths annually (Edwards, Ward et al. 2010). The adenoma-carcinoma pathway is considered responsible for the majority of colorectal cancer. Most of the colorectal cancers (CRCs) arose from small (<1.0cm) to large adenomatous polyps, which then takes 5 to10 years for this pre-cancer lesion to progress to early colorectal cancer (Winawer, Fletcher et al. 2003).

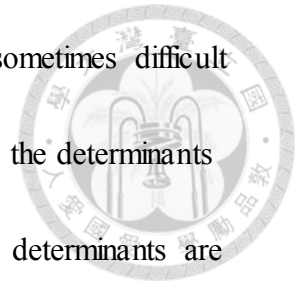
Among a variety of modalities, the stool-based tests focusing on the detection of bleeding phenotype have gained increasing attention for the role of secondary prevention of CRC at the population level (Shaukat, Mongin et al. 2013). The screening strategies using guaiac faecal occult blood test (gFOBT) have been proven to be effective in mortality reduction for CRC (Mandel, Bond et al. 1993, Winawer, Zauber et al. 1993, Hardcastle, Chamberlain et al. 1996, Steele, McClements et al. 2009). Following this success, faecal immunochemical test (FIT) was further

doi:10.6342/NTU201802429

developed in the recent decades and has demonstrated its effectiveness in several service screening programmes (Kronborg, Fenger et al. 1996, Chiu, Chen et al. 2015). Considering the efficacy of population-based colorectal cancer screening strategy using FIT as the tool, there is a lack of evidence in mortality reduction derived from the randomized controlled study. The endoscopic-based methods, like the sigmoidoscopy and colonoscopy, are two modalities to be considered. Sigmoidoscopy is limited in checking entire colon in screening. Colonoscopy is now considered the highest sensitive tools in detecting colorectal-cancer, and stand for the last station of stool-based screening to further confirm and remove the suspicious lesion, but overdetection remains uncertain (Lauby-Secretan, Vilahur et al. 2018). Furthermore, the corresponding issue on FOBT screening in the population-wide setting is almost no study mentioned.

Before we go to the discussion about how to deal with the overdetection in the screening programme of colorectal cancer , there are some questions come into our faces. First is the estimate of cost, involving time and budget, usually surrounded by the time we should follow and the number of people we have to invite. Second, the effectiveness evaluates of population-based cancer screening may be affected by attendance rate, compliance of follow-up procedure, contamination in the control group, and other situation we have described in 2.1 and 2.2 in details. Using a

doi:10.6342/NTU201802429



modelling approach not only because the benefits of screening are sometimes difficult to condition without RCT but also because we cannot easily change the determinants that have an influence on screening benefits in the real world. Some determinants are self-exists, such as the background incidence rate in the population, diseases, natural characteristic (the pattern of CRC progression). The attendance rate of invited population and compliance with procedure for diagnosis confirmation are other elements determine from a target population, they are elements that can improve by policy and designed strategy. Moreover, we will face the question of the calculation of the sample size, power, and even effectiveness, setting when we finally have a chance to design an executed RCT.

2.3.1 The modelling approach for assessing the effectiveness and overdiagnosis of colorectal cancer screening programme

Current studies have used a modelling approach for predicting the effectiveness of the screening programme, or even use the prediction to compute sample size required for mortality and other surrogate endpoints before planning stage of conducting an RCT. Because of the sojourn time of CRC is long, evaluating the effectiveness of CRC screening programme needs longer follow up time in RCTs, and therefore costly in the population-based setting for evaluating endpoints.

doi:10.6342/NTU201802429

Chiu and Malila proposed a five-state Markov model as an alternative way to predict the effectiveness of CRC screening. They analysed data in the Finnish population-based screening program using biennial faecal occult blood tests (FOBT) from 2004 to 2007, with participants aged 60 to 69, to estimate the natural history parameters and sensitivities simultaneously and then used them to predict the reduction of CRC mortality of six years and ten years through simulation. After excluding the diagnosed CRC cases before the screening, a total of 105,489 participants were randomly assigned to two groups (study group n=52,728, control group n=52,761). The five state model was expanded from three state model. They classified CRC into two major stages, non-localised and localised, non-localised cancer was the same meaning of non-advanced cancer and advanced stage of cancer without regional, and distant metastases or lymph node involvement. According to the two CRC stage have their own mean sojourn time and the inverse relationship between mean sojourn time and sensitivity, they considered the test sensitivity was also different in two CRC stage. They assume PCDP non-localized CRC progress from PCDP localised CRC. Based on the empirically observed data, expected number of each mode were calculated by applying the transition probabilities converted from transition rates using the Kolmogorov equation, including normal→PCDP localised CRC, PCDP localised CRC→PCDP non-localized CRC, PCDP localised

doi:10.6342/NTU201802429

CRC→Clinical localised CRC→Clinical non-localized CRC. As a result, preclinical incidence rate, sensitivity of PCDP localized CRC, and sensitivity of PCDP non-localized CRC were 0.0011 (0.00010,0.00127), 65.12% (27.05%, 77.05%), 73.70% (47.49%, 99.92%), respectively. By using Finnish data. The modelling pre-clinical incidence rate was slightly lower but closed to control group (0.00102), and overdiagnosis of CRC was 8.1% ((0.00111-0.00102) /0.00111).

During the estimation of transition parameters and sensitivity, they first estimated the sensitivity of FOBT that identifying localised CRC and got the incidence rate of overall CRC, because of the non-identifiable modelled by simultaneous estimation of all parameters. The incidence method was proposed by Day (Day 1985). Then they put the estimated sensitivity into likelihood function to calculate other transition rate and sensitivity of non-localized CRC. It was noted that the plugged in of sensitivity to calculate other parameters led to an iterative process, because the sensitivity for localised CRC depended on two transition rate ($\lambda 2$ and $\lambda 3$), and it will repeat and repeat until the two sensitivities and four transition rates finally converged. For our interests of overdetection, we focus on the trend of PCDP Localized CRC to become Clinical-Localized CRC or PCDP non-localized CRC, as we describe in 2.2.3, all estimation parameters including sensitivities were calculated based on empirical data, and varied by population characteristic, screening modalities,

doi:10.6342/NTU201802429

screening test-performance, and inter-screening intervals. The proportion of asymptomatic cases distributes in localized, and non-localized stage among all detected cancer cases might indicate the underlying cause of overdetection. (Chiu, Malila et al. 2017)

Chiu and colleagues also developed analytical decision model to solve the difficulties of RCT study plan for population-based screening for CRC cancer. A hypothetical cohort, with aged 45 to74 years, similar to two previous population-based RCTs, was simulated and randomly allocated into two arms in a stop-screen design. The study group had four rounds biannual FOB screening in study arm and six years study period while the control group was invited at the same time without offering the FOB test. Then they follow the CRC death of two groups for ten years since randomisation. They obtained the data from two CRC screening RCTs using FOB (Included Nottingham and Denmark trial with long-term follow-up instead of an ongoing Finnish randomised trial) to estimate the parameters governing the natural history. As the sensitivity reported was not the test sensitivity but program sensitivity, they used the Day method to convert, based on mean sojourn time with 3-year and 2-year screen interval. For predictive screening efficacy, mortality was the primary endpoint, and the advanced Duke's stage C or D (non-localized CRC) transition was the surrogate endpoint. Sensitivity analysis was also performed for another two

doi:10.6342/NTU201802429

associated elements, including FOB test attendance rate (50%, 60%, 70%, 80%, 90%, 100%) and colonoscopy compliance rate (80%, 90%, 100%) for three scenarios of pre-clinical CRC incidence (1, 1.5, 2 per 1000 years for low, intermediate, high, respectively). Then they calculated the power and sample size. According to their modelling methods, it not only provided a feasible way for determining sample size based on different endpoints and incidence settings but also provided a predictive method for screening effectiveness before holding large-scale RCT.

The credibility of parameters also had been tested. Chiu and colleagues also predicted the proportion of Dukes' A+B and Dukes' C+D, and they got 59.2 % and 40.8%, compared to 53.4% , 46.6% in the UK and 58.8%, 41.1% in Demark. When they set the incidence rate to 0.00102, which similar to the clinical incidence rate of Finland control group, they got expected number of CRC cases in Dukes' A+B and Dukes' C+D closed to the observed number based on 500,000 populations with age 60 to 69 years. Therefore, they pointed out the external validity of the parameters. (Y□H, Nea et al. 2011). Furthermore, this approach may have a possibility in the application of evaluating screening programs, for their feasible of control the different situation in a population-based screening programme.

doi:10.6342/NTU201802429

Chapter 3. Data Sources

3.1 Colorectal cancer screening data

3.1.1 Two randomized controlled trial (RCT) using FOB test

Data on both Denmark and UK RCTs for CRC screening using Hemoccult-II fecal occult blood test were accrued from two publications (Kronborg et al., 1996; Hardcastl e et al.,1996). The UK screening trial first recruited subjects aged 45-74 in Nottingham between February, 1981 and June 1983 (pilot study), and then the subjects of main study was enrolled between February 1985 and stopped screening in February, 1995. The Denmark trial recruited subjects aged 45-75 since August 1985 and stopped in August, 1995. The numbers of screen-detected cases by prevalent or subsequent repeated screening, interval cancer, and refuser were abstracted from two previous detailed reports. For UK trial, those who were FOB test positive but without further investigation by colonoscopy then surfaced to CRC were also obtained from study report. Aggregate data of two randomized control trials of screening with gFOBT are demonstrated in (**Table 5.3**) by detection modes.

doi:10.6342/NTU201802429

3.1.2 Taiwan CRC service screening using FIT

The complete data was from National Mass Screening Registry database, which included a nationwide colorectal cancer screening programme launched by Taiwan government. This national program offered biennial FIT for resident aged 50-69 years. During January 1, 2004, and December 31, 2009, a total of 1,160,895 participants (21% of the eligible population, 5,417,699) attended the FIT-based screening. Each patient used a single FIT kit (Eiken OC-SENSOR or Kyowa HM-JACK, deciding by each municipality) for screening. The collected sample was sent to qualified local laboratories for testing, with the cutoff points equal to 20 μg of haemoglobin per gram of stool (100ng/mL for Eiken and 8ng/mL for Kyowa). Subjects with the positive result (above the cutoff point) were referred to confirmatory examination using colonoscopy as the primary tool. The results of the confirmatory diagnosis of colorectal cancer following a positive FIT were reported to National Mass Screening Registry. Information on the status of colorectal cancer for attendees with the negative FIT result between successive screening rounds (interval colorectal cancer) and non-attendees (refuser) were ascertained from the Taiwan Cancer Registry. The Cancer Registry is a nationwide program for cancer incidence survey, with high coverage rate and acceptable validity.

doi:10.6342/NTU201802429

The comparator for assessing overdetection of Taiwan population-based colorectal cancer screening programme was derived from the historical control. Data on the occurrence of colorectal cancers were derived from the annual report of Taiwan cancer registry targeting at residents aged 50-69 years between the year of 2003 and 2004. The expected number of colorectal cancer cases during the screening period of 2004 to 2009 was then projected from the risk level of this pre-screening era with the consideration of increasing trend in incidence.

doi:10.6342/NTU201802429

Chapter 4. Statistical methods for the assessment of overdetection

4.1 Natural disease progression and overdetection

The phenomenon of overdetection can be defined by using the natural course of disease progression as depicted in Figure 4.1. Using the time of birth as the original point, PCDP occurs after age X. With the period Y of dwelling in the stat of PCDP, the disease of subject progress and then become an overt clinical disease of CP. Let the time of screening activity denoted by D. If the screen take place after time X and before X+Y (X<=D<X+Y) then the subject will be detected by the screening activity, and the detection mode will be defined the screening-detected case.

If screening takes place before X (D<X) and the disease progress rapidly, namely Y is short, the lesion may surface to clinical phase before he/she have the chance to attend next screening round. The detection mode of such a case will be interval cancer. Since the screening tool is not 100% accurate, the lesion at the phase of PCDP may be missed during screening round and then develop into the phase of CP.

In the scenario of overdetection, the subject is detected as having occult disease during screening activity. The neoplastic lesion has thus occurred before the time of screening (D>=X). However, the lesion will never progress into the state of CP, and the dwelling time Y becomes infinite (Y=∞). Were it not for screening; the lesion will

doi:10.6342/NTU201802429

never be noticed even with sufficiently long follow-up time. An infinite dwelling time Y is equivalent to have an infinitesimal progression rate for PCDP. Such kind of cases will be mixed up with screening-detected cases that will eventually surface to CP if not detected by screening. However, among the screening-detected cases, it is not possible to distinguish these two types of screening-detected cases. Neoplastic lesions with such a character will only be detected in the screening arm but not the control arm in which no screening activity was provided and thus a zero chance for the identification of cases with infinite Y. The comparison of screening arm and the control arm based on the randomised controlled design thus provide a solid ground to evaluate the extent of overdetection incurred by screening activity. However, this comparison must be based on the study with sufficiently long follow-up period. For the correlation between screening arm and control arm using study with short follow-up period, the lesion in its lead time period will be identified in the screening arm. However, such a case will not be discovered in the control arm hence an overestimate in the excess number of cases for the screening arm. In addition to use the comparison of cancer incidence between screening arm and control arm for assessing overdetection, the difference between the advanced and non-advanced cancer incidence provides more detailed information. The cancer stage here is distinguished merely as the non-advanced stage, and advanced stage, dependent on the cancer is in

doi:10.6342/NTU201802429

situ, localised growing or spreading to regional (which means nearby lymph nodes, tissue or organs) or even distance part of the body. If cancer did not spread to regional or distance will be classified as non-advanced cancer. we do not focus on more details of staging system on cancer report. As a biological phenomenon, the overdetection will be more likely to occur in cancers at non-advanced stage than the advanced stage. This concept is thus the basis for the development of methods for quantifying overdiagnosis in cancer screening.

4.2 Graphic method

The methods were developed by comparing cumulative incidence curves of advanced cancers and non-advanced cancers. Based on the study design of two arm randomized control trial, suppose we have invited group as study group (s) and uninvited group as control group(c), during the study period (t) ,the cumulative incidence (Cl) of advanced cancer in two groups can be show in the Figure 4.2

From the time since randomization to the timing of two curve separate is due to early detection in the screening activity, namely lead-time. The excess of advanced cancer compare control group to study group can be count as cumulative incidence gained by early detection arises from lead-time

$$\mathrm{CI}_{early} = CI^{C}_{adv} - CI^{S}_{adv} \tag{4.1}$$

When in the most ideally situation, the study period (t) is long enough, we can wait

doi:10.6342/NTU201802429

until the non-advanced cancer progress to advanced cancer (non-adv), at the end of study period there will be no more new case of advanced cancer arise from non-advanced cancer. If we defined the detection of non-advanced cancer in the study group is overdetection, the cumulative incidence of overdetection can be estimate by subtracting the non-advanced cumulative incidence from CI_{early}, and will be the high estimation (H) for detecting all non-advance cancer.

$$CI_{over}^{H} = CI_{non-adv}^{S} - \mathrm{CI}_{early} \tag{4.2}$$

However, the follow-up time is restricted that we usually have a proportion of non-advanced cancers in the study group would not deteriorate into advanced cancer before the study ends. Then we may further subtracted the counterpart in control group ($CI_{non-adv}^{C}$) from our equation (4.2) and get the low (L) estimation.

$$CI_{over}^{L} = \left(CI_{non-adv}^{S} - \mathrm{CI}_{early}\right) - CI_{non-adv}^{C} \tag{4.3}$$

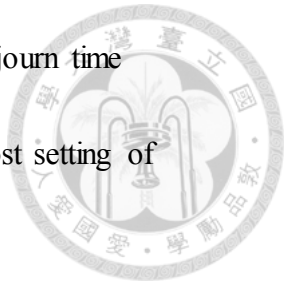
If see equation (4.3) more closely, CI_{over}^{L} is equivalent to the traditional excess incidence method

$$\begin{aligned} CI_{over}^{L} &= (CI_{non-adv}^{S} - [CI_{adv}^{C} - CI_{adv}^{S}]) - CI_{non-adv}^{C} \\ &= (CI_{adv}^{S} + CI_{non-adv}^{S}) - (CI_{adv}^{C} + CI_{non-adv}^{C}) \\ &= CI^{S} - CI^{C} \end{aligned} \tag{4.4}$$

This means the method a conservative approach depend on the length of the lead-time related to the study period. Equation (4.2) tend to get unbiased estimated if the lead-

doi:10.6342/NTU201802429

time is shorter, because lead-time distribution is derived from the sojourn time distribution. On the other hand, Equation (4.3) could be useful in most setting of randomized controlled trials.



It is also straightforward to take the average of the high and low estimates from two equations to estimate the proportion of over-detection, assuming an uniform distribution of sojourn time.

$$\mathrm{CI}_{Over} = \frac{CI^{H}_{over} + CI^{L}_{over}}{2} \quad (4.5)$$

doi:10.6342/NTU201802429

4.3 Modelling-based method

4.3.1 Model specification for disease progression

Three-state Markov model

Considering a three-state process depicting the progression of cancer from the state of free-of-disease to the disease status with overt clinical symptom, such an evolution can be depicted by continuous time Markov process. In the clinical scenario, the three state are free of cancer (state 0), preclinical detectable phase (PCDP) (state 1), and clinical phase (CP) (state 2). The state space (Ω) for such an disease evolution is thus written as $\Omega=\{0,1,2\}$. Let the random variable $X(t)$ denoted the state of certain subject of the eligible population of screening programme, $X(t)\in \Omega$. The observation on the history of cancer progression for such a subject with a total of r rounds of attended screening activity is thus written as the sequence of X(t), nemaly $\{X(t_0), X(t_1), \dots, X(t_r)\}$. The initial time, t_0, corresponds to the age of the attendee during the prevalent screening round.

The two rates dominate the development of CRC, incidence rate and progression rate, are the corresponding transition rates from the state 0 (free of CRC) to state 1 (PCDP) and state 1 (PCDP) to state 2 (CP), respectively, with the three-state Markov process underpinning. The transition rate matrix for the defined continuous-time three-state Markov process is thus written as follows.

doi:10.6342/NTU201802429

$$\mathbf{Q} = \begin{array}{l} \\ \text{Free of Cancer} \quad (0) \\ \text{PCDP Cancer} \quad (1) \\ \text{CP} \qquad\qquad (2) \end{array} \begin{array}{c} \begin{array}{ccc} \text{Free of Cancer}(0) & \text{PCDP}(1) & \text{CP}(2) \end{array} \\ \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 \\ 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 0 \end{pmatrix} \end{array} \quad (4.6)$$

The zero element for the transition from free-of-cancer to CP is due to the nature of CRC evolution during a infinitesimal time epoch can occurred only between the adjacent states. The zero element for the transition between PCDP to free-of-cancer is due to the progressive nature of cancer. For the purpose of assessing nature progression of CRC, CP is defined as the absorbing state of the Markov process which corresponds to the biological characteristics of zero probability for spontaneous recovery it means people with disease come into state 2 would never recover which results in the row of zeros for the initial state of CP. The progression model for cancer is depicted by Figure 4.3. Let the corresponding transition probability matrix given the follow-up period *t* for the defined three-state Markov model denoted by **P(*t*)**, which can be written as follows.

$$\begin{array}{l} \text{Free of Cancer} \quad (0) \\ \text{PCDP} \qquad\quad (1) \\ \text{CP} \qquad\quad (2) \end{array} \begin{array}{c} \begin{array}{ccc} \text{Free of Cancer}(0) & \text{PCDP}(1) & \text{CP}(2) \end{array} \\ \begin{pmatrix} P_{00}(t) & P_{01}(t) & P_{02}(t) \\ 0 & P_{11}(t) & P_{12}(t) \\ 0 & 0 & 1 \end{pmatrix} \end{array} \quad (4.7)$$

By using the forward Kolmogorov equation **P**'(t)=**P**(t)**Q**(t) (Cox & Miller, 1977), the element of the transition probabilities matrix, **P**(*t*), based the transition intensity matrix, **Q**, for the three-state disease progression model can be derived as follows.

doi:10.6342/NTU201802429

The probability of staying in the state of free-of-disease and that of transition from the state of free-of-disease to PCDP and CP, namely $P_{00}(t)$,$P_{01}(t)$, and $P_{02}(t)$ are thus written as

$$P_{00}(t) = e^{-\lambda_1 t} \tag{4.8}$$

$$P_{01}(t) = \frac{\lambda_1(e^{-\lambda_1 t} - e^{-\lambda_2 t})}{(\lambda_2 - \lambda_1)} \tag{4.9}$$

and

$$P_{02}(t) = 1 - \frac{\lambda_2 e^{-\lambda_1 t}}{\lambda_2 - \lambda_1} + \frac{\lambda_1 e^{-\lambda_2 t}}{\lambda_2 - \lambda_1} \quad , \tag{4.10}$$

respectively. For data on prevalent screening, t represents the age at attending screening and for data on subsequent screening, the time frame is the interscreening interval. Similarly, the probability of staying in the state of PCDP and that of progression from the state of PCDP to CP give the period of follow-up, t, namely $P_{11}(t)$ and $P_{12}(t)$ can be derived as follows.

$$P_{11}(t) = e^{-\lambda_2 t} \tag{4.11}$$

$$P_{12}(t) = 1 - e^{-\lambda_2 t} \quad . \tag{4.12}$$

doi:10.6342/NTU201802429

Five-state Markov model

The three-state progression model of cancer can be extended into five-state Markov model by considering both the PCDP and CP as non-advanced and advanced status as depicted by Figure 4.4. The state of cancer progression is thus including free-of-disease (state 0), non-advanced PCDP (state 1), advanced PCDP (state 2), non-advanced CP (state 3), and advanced CP (state 4). The state space in the five-state model is thus specified as Ω={0 (free-of-disease,1 (non-advanced PCDP) ,2 (advanced PCDP) ,3 (non-advanced CP),4 (advanced CP)} The incidence rate is thus corresponding to λ_1 which dominate the occurrence of cancer of PCDP at non-advanced stage from the state of free-of-disease. For the disease status of non-advanced PCDP, there two paths. One is the progression to advance stage of PCDP and the other is the progression to non-advanced CP with the rate of λ_2 and λ_3, respectively. Following the state of advanced PCDP, the disease can only progress to advanced CP. The transition rate matrix for the five-state disease progression model is thus written as

$$\mathbf{Q} = \begin{array}{l} \\ Free-of-disease\ (0) \\ PCDP_{non-advance}\ (1) \\ PCDP_{advance}\ (2) \\ CP_{non-advance}\ (3) \\ CP_{advance}\ (4) \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & 0 \\ 0 & -(\lambda_2+\lambda_3) & \lambda_2 & \lambda_3 & 0 \\ 0 & 0 & 0 & -\lambda_4 & \lambda_4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}. \quad (4.13)$$

doi:10.6342/NTU201802429

Similar to the transition intensity matrix for three-state model of disease progression, the two zero rows for non-advanced CP and advanced CP is also due to the absorbing status of CP, and the other zero elements is due to the biological characteristics of cancer progression given the infinitesimal time period implied by the rate of continuous-time Markov model. Base on the intensity matrix of the five state Markov model, the corresponding transition probability matrix given time period *t* can be derived from the forward Kolmogorov equation and is written as follows.

$$\mathbf{P}(\mathrm{t}) = \begin{array}{l} \\ Free-of-disease \quad (0) \\ PCDP_{non-advance} \quad (1) \\ PCDP_{advance} \quad (2) \\ CP_{non-advance} \quad (3) \\ CP_{advance} \quad (4) \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} P_{00}(t) & P_{01}(t) & P_{02}(t) & P_{03}(t) & P_{04}(t) \\ 0 & P_{11}(t) & P_{12}(t) & P_{13}(t) & P_{14}(t) \\ 0 & 0 & P_{22}(t) & 0 & P_{24}(t) \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}. \tag{4.14}$$

Due to the absorbing nature of CP, advanced or non-advances, the probability of staying in these two state is for sure. Other zero elements is due to the progressive nature and also the biological plausibility in the transition between states.

4.3.2 Likelihood function for data by modes of detection

Following previous work, the likelihood functions for three-state Markov model and five-state Markov model for the observed data collected from the scenario of screening programme can be derived by using the product of appropriate elements of

doi:10.6342/NTU201802429

transition probability matrix depicted above (Chen et al,.1996). For the eligible subject of a screening program, the probability of having observation on the sequence of the disease evolution, $\Pr\{X(t_0), X(t_1), \ldots, X(t_r)\}$ can be reduced to

$$\Pr\{X(t_0)\} \times \Pr\{X(t_1)|\, X(t_0)\} \times \ldots \times \Pr\{X(t_r)|\, X(t_{r-1})\} \qquad (4.15)$$

due to Markov property. This is the product of a series of transition probabilities corresponding to appropriate current state given the state at last epoch of observation. The parameter of interest, namely the transition rate between states are in turn embedded within the specification of transition probabilities.

Using the three-state Markov model as example, the likelihood of observed mode within the scenario of disease screening can be specified as follows.

Prevalent screen (first screen)

$$\text{Free of CRC} = \frac{P_{00}(age)+P_{01}(age)\times(1-sensitivity)}{P_{00}(age)+P_{01}(age)} \qquad (4.16)$$

$$\text{Prevalent sreening detected CRC} = \frac{P_{01}(age)\times(sensitivity)}{P_{00}(age)+\ P_{01}(age)} \qquad (4.17)$$

where age represents age at attending the prevalent screening round. Note that in the scenario of disease screening, only those free of cancer are eligible for the enrollment, hence the observed data are actually truncate those with the occurrence of cancer before he/she have the chance to attend the screening programme. This truncation data

doi:10.6342/NTU201802429

property was dealt with by the using of $P_{00}(age)+P_{01}(age)$ as the truncation probability for date on prevalent screening round.

Since the screening tool is not perfect in detecting the target disease, such an measurement error can be incorporated into the likelihood of observed data with the multistate disease progression underpinning. For data on prevalent screening, those detected as free-of-disease are actually including the true negative (TN) and the false negative (FN) components, written as

$$\mathrm{TN} = \frac{P_{00}(age)}{P_{00}(age) + P_{01}(age) \times (1 - sensitivity)} \tag{4.18}$$

and

$$\mathrm{FN} = \frac{P_{01}(age) \times (1 - sensitivity)}{P_{00}(age) + P_{01}(age) \times (1 - sensitivity)} \ . \tag{4.19}$$

Subsequent screening (second screening or more)

$$\mathrm{P(Free\ of\ CRC)} = \mathrm{TN}\ \times [P_{00}(t) + (P_{01}(t) \times (1 - sensiticity))] \tag{4.20}$$

$$\mathrm{P(Subsequent\ screen\ detected\ CRC)} = \mathrm{TN}\ \times P_{01}(\mathrm{t}) \times (sensitivity) + \mathrm{FN} \times P_{11}(\mathrm{t}) \tag{4.21}$$

where *t* denotes interscreening interval, the time between two screening.

doi:10.6342/NTU201802429

Interval cancer

$$\mathrm{P(Interval\ cancer)} = [TN \times P_{02}(t)] + [FN \times P_{12}(t)] \quad (4.22)$$

Refuser

$$\mathrm{P(Free\ of\ CP)} = \frac{P_{00}(age)\times(P_{00}(x)+P_{01}(x))}{P_{00}(age)+ P_{01}(age)} + \frac{P_{01}(age)\times P_{11}(t)}{P_{00}(age)+ P_{01}(age)} \quad (4.23)$$

$$\mathrm{P(Clinical\ detected\ CRC)} = \frac{P_{00}(age)\times P_{02}(x)}{P_{00}(age)+ P_{01}(age)} + \frac{P_{01}(age)\times P_{12}(x)}{P_{00}(age)+ P_{01}(age)} \quad (4.24)$$

The estimation of intensity rates parameters was derived by using the non-linear method to estimate temporal natural course of disease, by applying the aggregate data from the two previous randomized controlled trials from Kronborg and Hardcastle's publication (Hardcastle, Chamberlain et al. 1996, Kronborg, Fenger et al. 1996). The transition probabilities based on three-state Markov model were derived from forward Kolmogorov method. Sensitivity was also taken as another parameter for estimation, simultaneously, with the two parameters of transition rates (Chen, Kuo et al. 2000, Chiu, Duffy et al. 2010).

doi:10.6342/NTU201802429

4.3.3 Model-based study design for assessing overdetection

4.3.3.1 Three-state Markov model

The three-state process of CRC natural progression was defined as free of CRC, preclinical-screen detectable phase (PCDP), and clinical phase (CP). Following the screening scenario of the three-state model proposed previously , the three-state natural history of CRC (CRC free → the PCDP CRC→the CP CRC) was used to estimate annual incidence rate (λ_1) (corresponding to CRC free → the CRC in the PCDP) and annual progression rate (λ_2) (corresponding to the CRC in the PCDP → the CRC in the CP). It is postulated that the estimate of annual incidence rate would be overestimated were non-progressive CRC over-detected. Figure 4.3 shows why annual incidence rate of entering the PCDP would be overestimated when only data on the invited arm was used. The true incidence rate of CRC is inflated to the rate estimate (λ_1) due to the inclusion of non-progressive CRC during screening intervention. Also note that annual transition rate from the PCDP to the CP (λ_2) is needed to be corrected estimate as a result of the sensitivity of FOB Test.

The assessment of overdetection for screening programme was based on the standardized overdetection ratio (SOR). For the derivation of SOR, we first estimate the predicted case number using the rate of disease progression derived by a three-state Markov model incorporation the sensitivity of FOB test based on the data on

doi:10.6342/NTU201802429

screening group. The expected cases was then derived from the transition probability for each type of disease progression based on estimated results derived from screening arm and the follow-up person years of the control arm. The SOR was then derived by dividing the expected frequency of colorectal cancer by the observed frequency of colorectal cancer for the control group either from the control group in the randomized controlled trial or the comparator from pre-screening period in the service screening. Such a design was applied to two randomized controlled trials and one service screening program to estimate the proportion of overdetection of colorectal cancer resulting from FOB test and FIT by calculating (SOR-1)*100%. Note that the SOR-1 represent the proportion of excessed frequency in cancer due to the active case finding form screening progremme compared with the control arm.

The procedure for the assessment of overdetection with the three-state disease progression design is summarized as follows.

1. Estimate the annual incidence rate (λ_1) and the mean sojourn time (MST) after correcting for false negative cases using the data on screened group only.
2. Apply the pre-clinical incidence rate and the MST to project the expected number of progressive CRC in the PCDP.
3. Following the transition probability matrix, estimate the expected number of CRC among the control group with population size (N) given age of entry to study (m)

doi:10.6342/NTU201802429

and the follow-up time year (t) since last negative screen. This expected number of CRC is equal to

$$C(t) = N \times \frac{P_{00}(m) \times P_{02}(t) + P_{01}(m) \times P_{12}(t)}{1 - P_{02}(m)} \quad (4.25),$$

and the notations are summaries as follows.

$P_{00}(m)$: the probability of surfacing to free of CRC before age(m)

$P_{01}(m)$: the probability of surfacing to pre-clinical detected phase before age(m)

$P_{02}(m)$: the probability of surfacing to clinical phase before age(m)

$P_{02}(t)$: the probability of surfacing to clinical phase before time(t) since last negative screen

$P_{12}(t)$: the probability of pre-clinical detected phase to clinical phase before time(t) since last negative screen

4. The standardized overdetection ratio, SOR, is derived by

$$SOR = \left(\frac{C(t)}{D(t)} - 1\right) \times 100\% \quad (4.26)$$

C(t): the expected number of CRC among the control group by follow-up time t

For the two randomised controlled trial, the observed number of CRC (D(t)) was derived from the control group by follow-up time t. For the Taiwan Nationwide Colorectal Cancer Screening Programme, the incidence rate of those aged 50-69 in 2003, namely the year before the initiation of the screening programme with the

doi:10.6342/NTU201802429

projection to the study period used as the denominator.

4.3.3.2 Five-state Markov model

Extended from the method in assessing overdetection in screening programme using three-state Markov process, two approaches were developed on the basis of the five-state model of disease progression. Following the biological scenario of disease progression, overdetection occurred mainly in non-advanced PCDP rather than the advanced PCDP. There will be no overdetection considering observed CP, both advanced and non-advanced one. Facilitate by the five-state Markov model, the evaluation of overdetection for screening programme can be focused on the non-advanced PCDP detected during screening activity.

Overdetection in predicted frequency of non-advanced PCDP

Under the scenario of disease screening programme, the expected frequency of non-advanced PCDP, advanced PCDP, non-advanced CP and advanced CP for the control group can be projected by using estimated results on the rate of disease progression in conjunction with the person-year under observation of the control group. Similar to the rationale of assessing overdetection using three-state Markov model, the assessment can be done by using five-state Markov model. To test the hypothesis of the existence of overdetection is equivalent to test whether the expected proportion of non-advanced cancer derived from screening arm is higher than the

doi:10.6342/NTU201802429

observed frequency of non-advanced cancer in the control arm. This hypothesis is written as follows

$$\frac{Pred\,(N_{non-adv})}{Pred(N_{adv})+Pred(N_{non-adv})} > \frac{N_{non-adv}}{N_{adv}+N_{non-adv}} \quad (4.27)$$

The observed frequencies, N_{adv}, $N_{\text{non-adv}}$, and the corresponding proportion of non-advanced cancer is derived from the observed data in control arm. The predicted frequency for the non-advanced cancer and advanced cancer is derived from the transition probability estimated from the screening arm only. For the scenario in which overdetection was brought by the active detection in screening programme, the predicted frequency of cancers, especially the non-advanced one, will outweigh for the that observed from the control arm.

The mechanism of overdetection using five-state Markov process

Based on the five-state model, the index for overdetection cane be represented by the proportion of the two progression rates of non-advanced PCDP, λ_2 and λ_3. The use of five-state Markov process enables one to elucidate the relative contribution of overdetection between non-advanced CP and advanced PCDP by using the competing ratio of the progression rate from non-advanced PCDP to advanced PCDP in comparison with that from non-advanced PCDP to non-advanced CP (λ_2/λ_3). The relative competing ratio between that in the absence of overdetection and that in the presence of overdeteciton gives an indication of relative contribution of the influence

doi:10.6342/NTU201802429

of overdetection.

doi:10.6342/NTU201802429

Chapter 5. Results

5.1 Results of graphic methods

The diagrams of cumulated incidence for overall CRC, non-advanced CRC, and advanced CRC projected form the two trials of UK and Denmark are presented in **Figure 5.1** and **Figure 5.2** respectively. In the UK trial, the cumulative incidence of CRC was higher than that of control arm (**Figure 5.1 (a)**) whereas that for the Denmark trial was similar between the two groups (**Figure 5.2 (a)**). Considering the cumulative incidence of non-advanced CRC, both trials shows a higher cumulative incidence for the screening arm (**Figure 5.1 (b)** and **Figure 5.2 (b)**). Although it is generally accepted that the higher cumulative incidence in non-advanced lesion for the screen arm compared with the control arm resulted from the efficacy of early detection, some of the non-advanced lesions may be due to over-detection. The cumulative incidence of advanced CRC was lower for the screening arm in both trials (**Figure 5.1 (c)** and **Figure 5.2 (c)**). The summary of the cumulative incidences by the group of screening arm and control arm is provided in **Table 5.1**. Following the curve method proposed in section 4.1, the lower estimate for the overdiagnosis proportion is the difference between of cumulative incidence of overall CRC, CI^{L}_{Over}, which gives the value of 9.4% and -0.4% for the UK and the Denmark trial, respectively. The reasons of seeing only slightly increasing incidecne in both trials may be mainly due

doi:10.6342/NTU201802429

to early detection and removal of adenoma that leads to the reduction in the incidence of CRC in the subsequent screens. The cumulative incidence of early detection, CI_L, is the difference between that of advanced CRC, $CI_{adv}^{C} - CI_{adv}^{S}$, which gives the value of 61.3 and 130.6 per 100,000 for the UK and the Denmark trial, respectively. The upper limit of the over-detection proportion, based on the equation, $CI_{Over}^{H} = CI_{Non-adv}^{S} - CI_L$, was estimated as 52.8% and 50.2% for the UK and Denmark trial. This suggest high awareness of early detection of CRC through these two g-FOBT programs. The average of over-detection was 31.1% and 24.9%. The results of graphic methods show the estimated proportion of over-diagnosis may be biased as the natural history of adenoma and high awareness in the routine clinical practice may make this method inappropriate.

5.2 Results from disease progression model

5.2.1 Estimated results based on three-state Markov model

5.2.1.1 CRC progression rate based on three-state Markov model

The estimated results based on the three-state Markov model for CRC progression from two randomized controlled trials using gFOBT as the tool and the Taiwan nationwide service screening using faecal immunochemical test (FIT) were listed in **Table 5.2**. The annual incidence rate (CRC free→PCDP CRC) is higher in

doi:10.6342/NTU201802429

the UK trial and Demark trial than that derived from Taiwan nationwide service screening programme (0.00147 and 0.00172 vs 0.00096 (per person-year)), so as the progression rate (0.3475 and 0.4433 VS 0.1858(per year)). The estimated results on the sensitivity for the detection of CRC at preclinical stage using gFOBT was around 53% (53.4%, 95% CI: 34.3-69.6% for UK trial and 52.05%, 95% CI: 35.4-68.6% for the Denmark trial), a figure lower than that of FIT (82.23%, 95% CI: 46.8-96.0%).

Table 5.3 demonstrated that there is lacking of statistical significance for the comparison between the expected and the observed number of each detection mode. These findings suggest that the model is adequate for temporal natural history of colorectal cancer from CRC free(0), through the PCDP(1), and finally to the CP(2) with adjustment for sensitivity of FOB Test.

5.2.1.2 Overdetection of CRC screening based on three-state Markov model

Take one of the data retrieved from Nottingham trial as an example, there are 74,998 subjects in the control group, by using transition probability estimated from incidence rate and progress rate, (Probability value, **Table 5.4**) the expected CRC number can be calculate as below

$$C(t)=N\times(P_{00}(m)\times P_{02}(t)+P_{01}(m)\times P_{12}(t))/(1-P_{02}(m))$$

$$=74{,}998\times(0.9142\times 0.0084+0.0039\times 0.9479)/(1-0.0819)$$

doi:10.6342/NTU201802429

=931.26

(SOR-1)*100% represents the extra proportion of screening detected cases, and these cases cannot be discover in the absence of screening .To continue our example above, the observed CRC number (D(t)) is 856, therefore we have inverse standardized overdiagnosis ratio(SOR) can be wrote as (C(t)/D(t)), and

(SOR-1) *100%= (931.26-856)/856=8.79%

By using the three-state Markov model, the expected number of CRC were 931.26, 528.06, and 3656.63 in UK trial, Demark trial and Taiwan service screening programme, respectively (Expected CRC, C(t), **Table 5.4**). In comparison of the observed CRC cases in gFOBT control group in Europe, the percentage of overdetection were 8.79% (95% CI: 8.29-9.65%) and 9.33% (8.81-10.20%) for the UK and Denmark trial, respectively, it was slightly lower for Taiwan service screening programme using FIT as tool (7.05%, **Table 5.4**).

5.2.2 Results based on five-state Markov model

Table 5.5 lists the incidence rate of colorectal cancer, the three progression rate (from non-advanceed PCDP to advanced PCDP, λ_2; from non-advanced PCDP to non-advanced CP, λ_3; from advanced PCDP to advanced CP, λ_4) along with the observed frequency of CRC by the status of advance and non-advance for the two

doi:10.6342/NTU201802429

randomized controlled trial using gFBOT of the screening tool. The incidence rates of CRC for the control arm representing the background incidence of the study population were also provided in **Table 5.5**. The follow-up period for the UK trial and the Denmark trial was 7.9 and 9.1 year, respectively. The estimated results on the incidence rate was 0.00146 and 0.00158 for the UK trial and the Denmark trial respectively. The estimated results on the proportion and SOR based on five-state Markov model for the two randomized controlled trial are listed in **Table 5.6**. Based on the estimated results on CRC progression for the non-advanced CRC, the rate ratio (λ_2/λ_3) was estimated as 1.29 and 1.16 for the UK and Denmark trial, respectively (**Table 5.6**). Note this ratio is based on the data with overdetection CRC.

By using the estimated result for disease progression the percentage of overdetection derived from the proportion of expected frequency of colorectal cancer cases compared with that observed form the control arm is 6.1% and 9.2% for the UK and Denmark trial, respectively. After exclude the influence of overdetection, the ratio of progression rate for non-advanced PCDP was decreased for both trial. The ratio of the progression rate for non-advanced PCDP after taking into account overdiagnosis was estimated as 1.16 and 0.97 for the UK and Denmark trial, respectively. Based on the ratio of progression rate with and without overdiagnosis, the standardized overdiagnosis ratio, SOR, was estimated as 1.10 and 1.19 for the UK and Denmark

doi:10.6342/NTU201802429

trial, respectively.

5.2 Computer simulation for assessing influential factors on standardized overdetection ratio (SOR)

To evaluate the impact of incidence rate, progression rate, follow-up time, and sensitivity of the SOR, a computer simulation was performed. Using the base case estimate with the incidence rate of 147 per 100,000, disease progression rate of 0.35, sensitivity of 53.4%, and the follow-up period of 7.9 years, the impact of these components on the estimated results of SOR was evaluated using a series of values.

The results on computer simulation were listed in **Table 5.7**. As expected, the SOR will be underestimated for the value of incidence rate lower than that of control arm, namely underestimated incidence rate. Compared with the base case of 147 per 100,000 incidence and the 8.8% SOR, the 130 per 100,000 incidence gives the SOR of -3.7%. In contrary, a high incidence of 160 per 100,000 gives the estimated SOR of 18.4%. The estimated results on SOR were robust to the change in progression rate. For the progression rate ranged from 0.1 to 1, the estimates results on SOR were close to the base case of 8.8%. Considering the effect of follow-up time, the longer the follow-up time, the lower the SOR. The sensitivity also correlated positively with lower SOR.

doi:10.6342/NTU201802429

Chapter 6. Discussion

We proposed a series of novel statistical methods, which have been never addressed, for estimating the influence of overdetection with an illustration of population-based screening for colorectal cancer with fecal immunological test (FIT). The proposed methodologies include a new graphic method taking information on cancer stage (advanced and non-advanced cancer) into account, standardized overdiagnosis ratio (SOR) using a three-state Markov process, and stage-standardized overdiagnosis ratio using five-state Markov process. The merit and weakness of these methodology are discussed as follows.

6.1 Graphic method

This method is the extension of the conventional method for the comparison of overall cumulative incidence between the invited group and the uninvited group. There is no excess of screen-detected cancer after reaching catch-up time (two curves would converge to the same curve) if there are no over-detected cancers in the result of screening. When overdetection is present, the extra screen-detected cancers after comparing two curves are deemed over-detected cases. However, such a conventional method would mix up with catch-up time that is associated with lead-time and the

doi:10.6342/NTU201802429

follow-up time in the presence of overdetection. To deal with this issue, one can extend this traditional method by considering cancer stage information. We, therefore, proposed a novel graphic method. It is very interesting to note that when cancer stage information is taken into account, the lower and upper limit of excess proportion of overdetection are available. The basic principle is that we first subtract the extra advanced cases from the control group in order to to make allowance for lead-time, the time taken for the occult transition from non-advanced status to advanced status. Screening leads to early detection of CRC (non-advanced) for arresting the progression of these non-advanced cases turn into advanced cases as seen in the control group. Such a transition is not seen for over-detected non-advanced CRC. If study period is long enough for such an occult transition in controlled group and there is lacking of awareness of detecting non-advanced cases. This is upper limit of overdetection if we assume all over-detected cases arise from non-advanced cancers. When we further subtract non-advanced cancer from the control group, it is very interesting to see the equivalence of this graphic method with the conventional method. Such an estimate is regarded the lowest estimate of overdetection. Accordingly, using this graphic method provides a new insight into upper and lower limit of overdetection. The other merit is that graphic method is distribution-free and dispenses with the distribution of sojourn time used in modelling method. However,

doi:10.6342/NTU201802429

the weakness of using graphic method is the failure of getting a better understanding of the mechanism of overdetection. In addition, the estimated result on overdetection proportion based on the graphic method for population-based colorectal cancer screening may be appropriate for cancer without premalignant lesion such as breast cancer, this is assumption can generally be supported given a sufficient long follow-up period. However, for cancer with premalignant lesion, this assumption may not hold and the estimated results may be biased, such as cervical cancer and colorectal cancer. Other factors that had impact on the cumulative incidence also cannot be seen directly. We also observed the negative value in our result in the lower limit (-0.4 in Demark) and developed possible speculation according to equation 4.3

$$CI^{L}_{over} = \left(CI^{S}_{non-adv} - \mathrm{CI}_{early}\right) - CI^{C}_{non-adv} \quad (4.3)$$

When $CI^{C}_{non-adv}$ is larger than $\left(CI^{S}_{non-adv} - \mathrm{CI}_{early}\right)$, CI^{L}_{over} can be negative. In other situation, we may figure CI_{early} (equal to $CI^{C}_{adv} - CI^{S}_{adv}$) is larger than $CI^{S}_{non-adv}$, which means there are many advanced-cancer in the control group (CI^{C}_{adv}). In the control group, patients with cancer were diagnosed after symptoms appear in clinical condition, the excess advanced case in control group made us further speculate under-detection or nonperformance may exist, dependent on healthcare system or others, but this suspicion needs further research and verification.

doi:10.6342/NTU201802429

6.2 Overdetection with modelling approach

6.2.1 Three-state Markov process

The alternative method to model overdetection is the use of a modelling approach. The basic idea is that we first used only data from the invited group to estimate pre-clinical incidence rate, sojourn time, and sensitivity. We expected that if there is overdetection, the pre-clinical incidence is expected to be higher than that in the absence of screening (control group). The higher pe-clinical incidence would further affect the mean sojourn time, the inverse of annual progression rate, and also sensitivity. The application of transition probabilities encoded with these parameters to the data of the control group would estimate the expected cases in the presence of overdetection. We estimated the proportion of overdetection by calculating the ratio of the expected to the observed from the control group subtracting from one. This index is called standardised overdetection ratio (SOR). Note that this modelling approach not only provides a quantitative estimate of overdetection but also enables one to elucidate how incidence rate, progression rate, and sensitivity affect overdetection resulting from population-based screening. Such influences are not possible to be investigated by using the graphic method as indicated.

doi:10.6342/NTU201802429

6.2.2 Five-state Markov process

The extension of three-state Markov process to five-state Markov process enables one to elucidate further and decipher the mechanism of how overdetection affects the progression of colorectal cancer from PCDP to CP in the light of information on cancer stage (non-advanced versus advanced state). According to Figure 4.3, the benefit of screening on early detection of CRC in reducing mortality is to arrest the progression from non-advanced to advanced state in PCDP rather than focus on the transition from non-advanced PCDP to non-advanced CP. It is therefore postulated to assess whether the influence of overdetection on the pathway from non-advanced PCDP to advanced PCDP is more significant than the pathway from non-advanced PCDP to non-advanced CP.

The results from Table 5.4 confirm this postulate as SOR is deflated from the estimated results of 1.29 to 1.16 and 1.16 to 0.97 for the UK and Denmark trial, respectively.

6.2.3 Influence of sensitivity

Estimating the proportion of over-detected CRC cases is less straightforward as many factors may affect the results. In addition to the overestimation of pre-clinical rate, a long sojourn time (MST) may also lead to more over-detected cases. As the

doi:10.6342/NTU201802429

sensitivity has inverse relationship with the MST, it also affects the estimate of overdetection. Since the two parameters, sensitivity and overdetection, cannot be directly observed, we applied a three-state natural history model making allowance for sensitivity to estimate both parameters in conjunction with pre-clinical incidence rate by using data from screen-detected data and interval cancer. We then estimated the parameters of transition rates governing the temporal disease natural history model. These transition estimates yield the expected CRC cases by treating the over-detected CRC as progressive CRC. The underlying assumption is that over-detected cases would not have surfaced to clinical phase had not screening been applied. The sojourn time is therefore infinity after the lesion entered the PCDP. In the control group, these over-detected cases would not be observed in the absence of screening. This provides the basis for estimating the proportion of over-detected cancers. One of the novelties of the application of our model that this is the first time to estimate the proportion of over-detected CRC with adjustment for sensitivity.

6.2.4 Implications for over-detection in economic evaluation

Screening for colorectal cancer may result in over-detected CRC cases despite the benefit of detecting early-stage colorectal cancers. By using a population-based randomised controlled trial, We may over-estimated the expected CRC cases using the data from the screening arm due to treating non-progressive CRC in the PCDP as

doi:10.6342/NTU201802429

if they were progressive CRC. Subtracting the expected CRC cases from the observed CRC cases estimated the proportion of over-detected CRC attributable to screening. Quantifying the proportion of overdetection of CRC not only indicates the extent of harm of the application of FOBT but also offer an estimate of unnecessary treatment, therapies and cost involved with these overdetection cases. This information will particularly contribute to the cost-utility analysis of mass screening when taking into account this over-detection to calculating the quality-adjusted life years.

6.2.5 Limitations

The first limitation is that our proposed modelling approach cannot estimate the over-detection of adenoma as the natural history of CRC has not included the disease progression of adenoma. According to the natural history of CRC, the incidence rate might be reduced due to polypectomy. Regarding the publication by Scholefield et al. based on UK trial after the 20-year follow-up, the CRC mortality still showed 18% reduction, but there was no difference in CRC incidence rates between intervention and control trials. This result does not indicate there was no overdetection in UK trial using FOB test for CRC screening when using empirical data on incidence as indicated above because the incidence would be affected by polypectomy. However, comparing the detection of adenoma between two groups, we highly suspected the awareness, and clinical accessibility increased by year, especially for subjects in

doi:10.6342/NTU201802429

control. Those reasons are, first, the increasing small adenoma (<10mm) detection proportion in control group which demonstrated on paper in 1996 and 2012. The small adenoma proportions of intervention and control arms were 25% (253/1001) vs 35% (129/370) in 1996 and 38% (868/2291) vs 40% (601/1484) in 2012 respectively. Another, the findings of total adenomas increased in control arm. The intervention/control ratio of adenoma was 2.71-time (1001/370) and 1.54-time (2291/1484) respectively. So, that phenomenon showed the incidence reduction was diminished by more detection of polyp in control which renders the evaluation of over-detection become complex as seen using the graphic method. The increase detection of the polyp during the recent 20 years becomes a glowing subject in the future.

The second limitation is that estimating the proportion of overdetection needs to rely on information from the control group. In service screening program, there is no unselected control group without screening. In such circumstance, we have no choice but to use data on prevalence screening to estimate the sensitivity and the MST. We then applied these two parameters to project the expected interval cancers, one of the clinically-detected cancers that would not contain over-detected cases, to compare the observed cases. In this circumstance, the sensitivity should be taken into account as interval cancers consist of both false negative cases and cases that entered the PCDP

doi:10.6342/NTU201802429

after the first screen. The concern arises as to whether the identifiability problem of estimating parameters is encountered due to insufficient on data. This becomes the important issue of ongoing research on evaluation of overdetection in population-based service screening for CRC with FOBT.

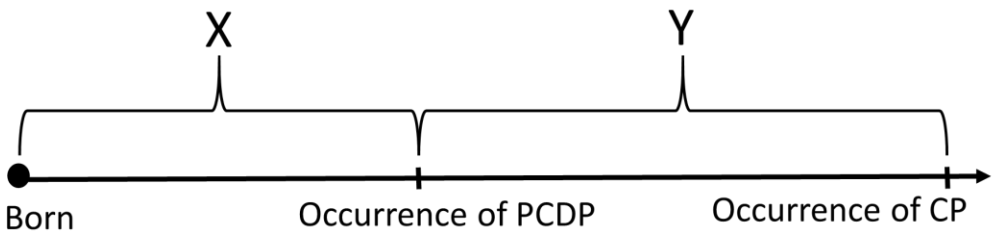
Conclusion

This thesis systematically developed a series of statistical methods, including the graphic method and the disease natural progression model for quantitative assessment of over-deteciton of colorectal cancer screening. The index of SOR was proposed to both assess the extent of over-detection and also to elucidate the mechanism of how overdection affects the progression of colorectal cancer from PCDP to CP in the light of information on cancer stage.

doi:10.6342/NTU201802429

Figures

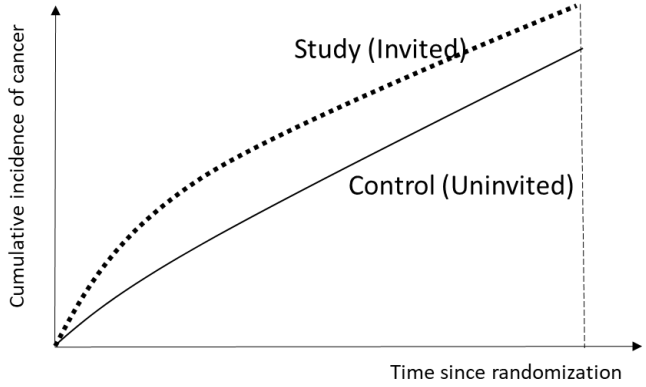
Figure 4.1 Timeline of cancer progression



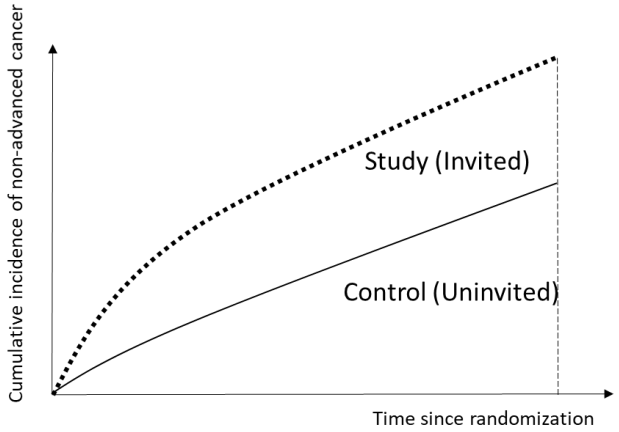
doi:10.6342/NTU201802429

Figure 4.2 The hypothetical illustration on the cumulative incidence by invitation groups.

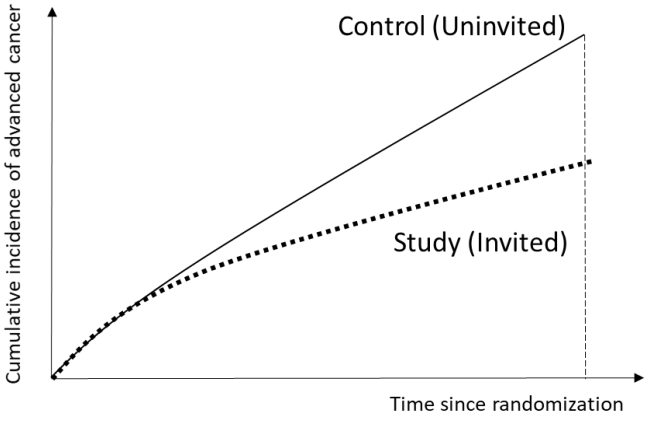
(a) Cumulative incidence of cancer by screening groups



(b) Cumulative incidence of non-advanced cancer by screening groups



(c) Cumulative incidence of advanced cancer by screening groups

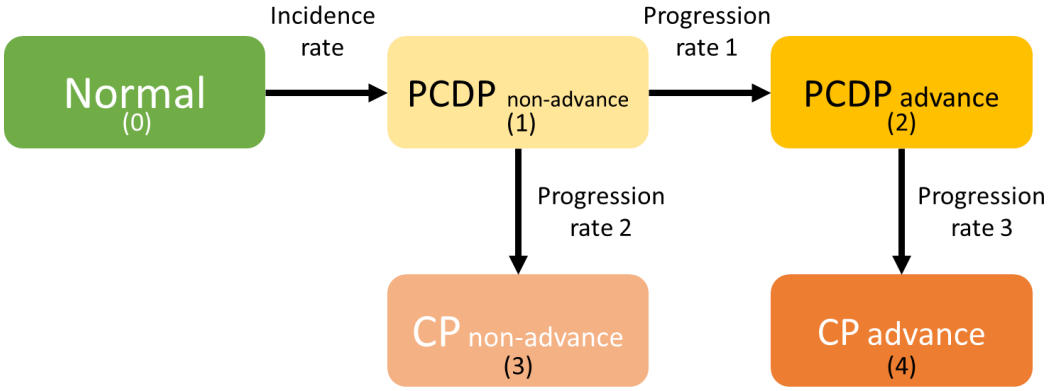


doi:10.6342/NTU201802429

Figure 4.3 Three-state Markov model for cancer progression

doi:10.6342/NTU201802429

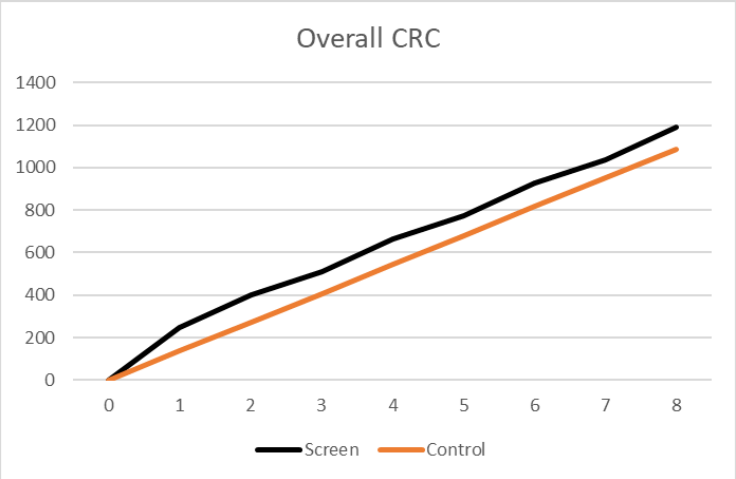
Figure 4.4 Five-state Markov model for cancer progression



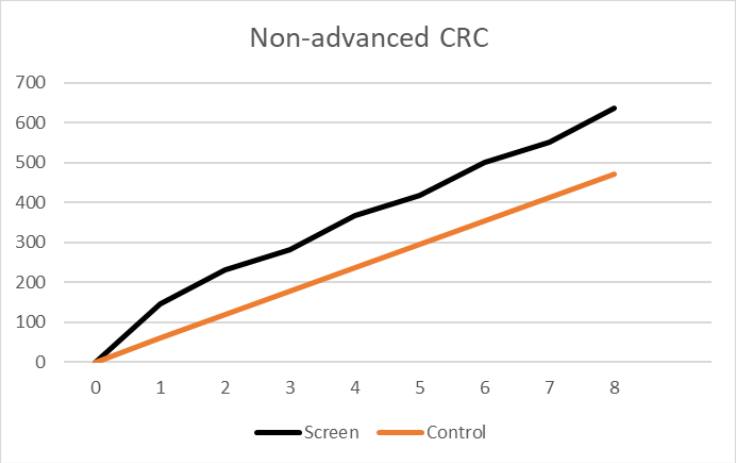
doi:10.6342/NTU201802429

Figure 5.1 Cumulated incidence of overall CRC, non-advanced CRC, and advanced CRC for UK trial using empirical data.

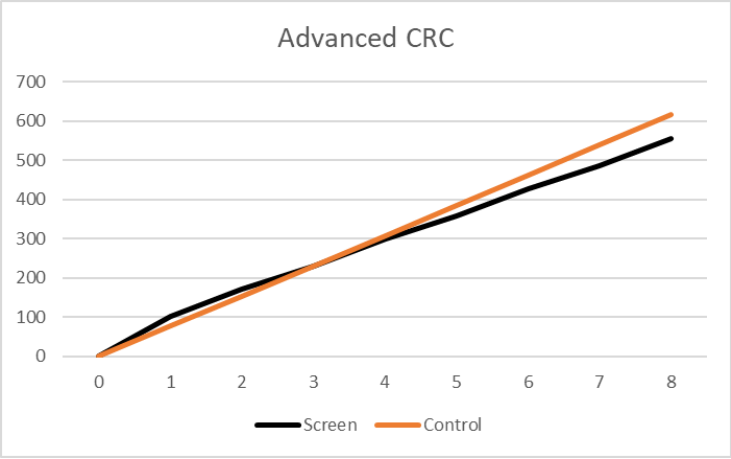
(a) Cumulative incidence of overall CRC



(b) Cumulative incidence of non-advanced CRC



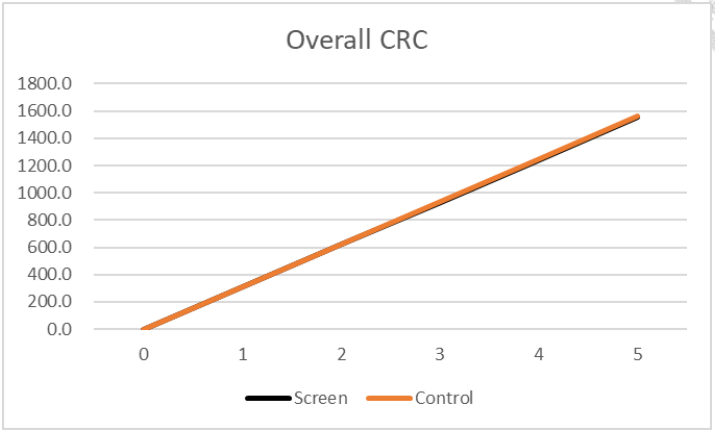
(c) Cumulative incidence of advanced CRC



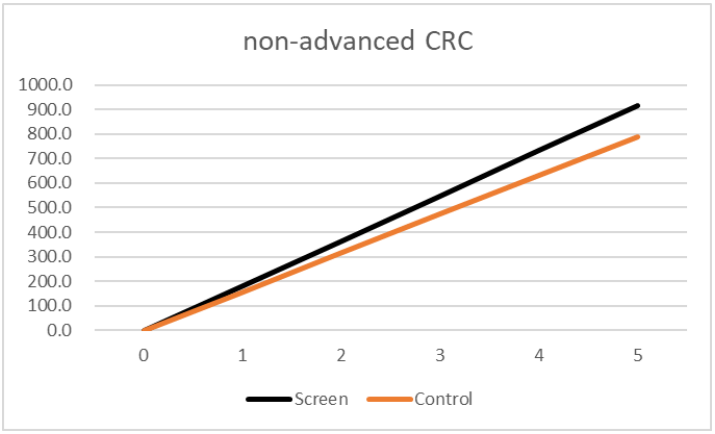
doi:10.6342/NTU201802429

Figure 5.2 Cumulated incidence of overall CRC, non-advanced CRC, and advanced CRC for Denmark trial using empirical data.

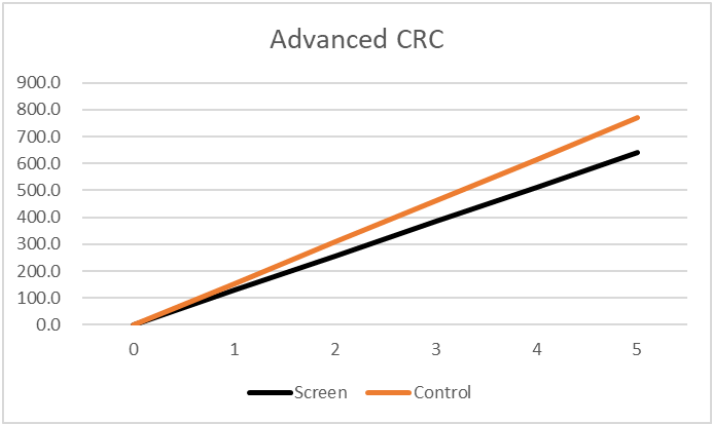
(a) Cumulative incidence of overall CRC



(b) Cumulative incidence of non-advanced CRC



(c) Cumulative incidence of advanced CRC



doi:10.6342/NTU201802429

Tables

Table 5.1 Estimated results on the percentage of overdetection for the UK and Denmark trail on colorectal cancer screening

Indicator		Group	UK	Denmark
Cumulative incidence	Overall	Screen	1190.7	1553.3
		Control	1088.0	1559.8
	Non-advanced	Screen	636.0	914.3
		Control	472.0	790.1
	Advanced	Screen	554.7	639.0
		Control	616.0	769.6
	Early detection		61.3	130.6
Percentage of overdetection	Lower limit		9.4%	-0.4%
	Upper limit		52.8%	50.2%
Average			31.1%	24.9%

doi:10.6342/NTU201802429

Table 5.2 Estimated results on the transition rates of CRC based on three-state Markov model

Parameters	UK (Nottingham)	Demark (Funen)	Taiwan
λ_1(Normal→PCDP) (per person-year)	0.00147 (0.00136, 0.00159)	0.00172 (0.00155, 0.00189)	0.00096(0.00085, 0.00107)
λ_2 (PCDP →Clinical) (per year)	0.3475 (0.2437, 0.4513)	0.4433 (0.3226, 0.5639)	0.1858(0.0488, 0.7068)
Sensitivity of PCDP CRC detection	53.40% (34.26%, 69.55%)	52.05% (35.43%, 68.56%)	82.23% (46.82%, 96.05%)
Mean sojourn time of PCDP CRC ($1/\lambda_2$)	2.88 (2.22, 4.10)	2.26 (1.77, 3.10)	5.28(1.41, 20.48)

CRC: colorectal cancer

PCDP: pre-clinical detectable phase

doi:10.6342/NTU201802429

Table 5.3 Information of modes and model fitting of two randomized controlled trials of CRC screening

Parameter	Status	Nottingham in UK		Funen in Demark	
Screening finding by round		Observed	Expected	Observed	Expected
Prevalent screening	Normal	44733	44735.32	20635	20630.24
	CRC	104	101.58	37	41.76
Interval cancer	CRC	164	140.04	148	147.23
Positive but without confirmation (first round)	CRC	28	27.76	----	----
Positive but without confirmation (repeated round)	CRC	57	66.03	----	----
Repeated screening	Normal	88008	87977.51	66025	66014.12
	CRC	132	109.69	83	67.74
Refuser	Normal	30015	30014.06	9895	9895.61
	CRC	400	400.94	195	194.29

doi:10.6342/NTU201802429

Table 5.4 Estimated results on standardized overdetection ratio (SOR) based on the three-state Markov model and expected and observed frequencies of colorectal cancer for control group.

Study	Probability	Value	Expected CRC, C(t)	Observed CRC, D(t)	SOR-1 (%) (95% CI)
Nottingham, UK	N (control group population)	74,998	931.26	856	8.79%(8.28,9.65)
	$P_{00}(m)$	0.9142			
	$P_{02}(t)$	0.0084			
	$P_{01}(m)$	0.0039			
	$P_{12}(t)$	0.9479			
	$P_{02}(m)$	0.0819			
Funen, Denmark	N (control group population)	30,966	528.06	483	9.33%(8.81,10.20)
	$P_{00}(m)$	0.8988			
	$P_{02}(t)$	0.0133			
	$P_{01}(m)$	0.0035			
	$P_{12}(t)$	0.9881			
	$P_{02}(m)$	0.0976			
Taiwan	N (national 50-69 population)	3,811,011	3656.63	3416	7.05%(6.56,7.89)
	$P_{00}(m)$	0.9440			
	$P_{02}(t)$	0.0001			
	$P_{01}(m)$	0.0044			
	$P_{12}(t)$	0.1857			
	$P_{02}(m)$	0.0515			

doi:10.6342/NTU201802429

Table 5.5 Observed frequency of CRC of control arm and estimated results on the transition rate for colorectal cancer based on five-state Markov model of CRC progression for the two randomised controlled trials

		Nottinghan(UK)	Funen(Demark)
Control arm	**Total**	74998	30966
	Normal	74142	30483
	Non-advanced CRC (D(t))	395	245
	Advanced CRC	461	238
Incidence of CRC in Control arm		0.00144	0.00172
Follow up (years)		7.9	9.1
Parameters estimate from Screening arm	**λ1 (Normal→PCDP Non-advanced CRC)**	0.00146	0.00158
	λ2 (PCDP Non-advanced CRC→PCDP advanced CRC)	0.2754	0.3247
	λ3 (PCDP Non-advanced CRC → clinical Non-advanced CRC)	0.2142	0.2801
	λ4 (PCDP advanced CRC → clinical advanced CRC)	0.7627	0.6478

doi:10.6342/NTU201802429

Table 5.6 Estimated results on the proportion and standardized overdetection ratio based on five-state Markov model for the two randomized controlled trials.

Index	State	Nottinghan(UK)	Funen(Demark)
Prediction number	PCDP non-advanced CRC	217.9	86.9
	PCDP advanced CRC	78.8	43.5
	Clinical non-advanced CRC	370.8	222.4
	Clinical advanced CRC	477.6	258.5
Proportion of overdetection		6.1%	9.2%
Ratio of progression rate for non-advanced PCDP (λ_2/λ_3)	With overdetection	1.29	1.16
	Without overdetection	1.16	0.97
Standardized Overdetection Ratio (SOR)		1.10	1.19

doi:10.6342/NTU201802429

Table 5.7 Computer simulation for the effect of disease transition rate, follow-up time, and sensitivity on estimated results of over-diagnosis percentage

* Base case estimate: Incidence rate: 147 per 100,000; progression rate: 0.35; Sensitivity: 53.4%; C(t): 931.26; D(t): 856; SOR 8.79%

Parameters	Value	**% Over-diagnosis(SOR)**	C(t) (expected)	D(t) (observed)
Lambda1	0.0013	-3.72%	824.2	856
	0.0014	3.64%	887.2	856
	0.00147*	8.79%*	931.3	856
	0.0015	11.00%	950.2	856
	0.0016	18.35%	1013.1	856
Lambda2	0.1	8.62%	929.8	856
	0.25	8.79%	931.3	856
	0.35*	8.79%*	931.3	856
	0.5	8.79%	931.3	856
	1	8.79%	862.0	856
Follow-up Time	3	9.28%	330.0	302
	6	9.03%	658.6	604
	8.5*	8.78%	931.2	856 *
	12	8.56%	1311.4	1208
	20	7.89%	2172.8	2014
Sensitivity	0.4	12.47%	962.7	856
	0.5	10.26%	943.9	856
	0.6	7.32%	918.7	856
	0.8	2.91%	880.9	856

doi:10.6342/NTU201802429

References

Auvinen, A., S. M. Moss, T. L. Tammela, K. Taari, M. J. Roobol, F. H. Schroder, C. H. Bangma, S. Carlsson, G. Aus, M. Zappa, D. Puliti, L. J. Denis, V. Nelen, M. Kwiatkowski, M. Randazzo, A. Paez, M. Lujan and J. Hugosson (2016). "Absolute Effect of Prostate Cancer Screening: Balance of Benefits and Harms by Center within the European Randomized Study of Prostate Cancer Screening." Clin Cancer Res 22(1): 243-249.

Chamberlain, J., S. M. Moss, A. E. Kirkpatrick, M. Michell and L. Johns (1993). "National Health Service breast screening programme results for 1991-2." Bmj **307**(6900): 353-356.

Chen, T. H., A. M. Yen, J. C. Fann, P. Gordon, S. L. Chen, S. Y. Chiu, C. Y. Hsu, K. J. Chang, W. C. Lee, K. G. Yeoh, H. Saito, S. Promthet, C. Hamashima, A. Maidin, F. Robinson and L. Z. Zhao (2017). "Clarifying the debate on population-based screening for breast cancer with mammography: A systematic review of randomized controlled trials on mammography with Bayesian meta-analysis and causal model." Medicine (Baltimore) **96**(3): e5684.

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

Chiu, H. M., S. L. Chen, A. M. Yen, S. Y. Chiu, J. C. Fann, Y. C. Lee, S. L. Pan, M. S. Wu, C. S. Liao, H. H. Chen, S. L. Koong and S. T. Chiou (2015). "Effectiveness of fecal immunochemical testing in reducing colorectal cancer mortality from the One Million Taiwanese Screening Program." Cancer **121**(18): 3221-3229.

Chiu, S. Y., S. Duffy, A. M. Yen, L. Tabar, R. A. Smith and H. H. Chen (2010). "Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening." Cancer Epidemiol Biomarkers Prev **19**(5): 1219-1228.

Chiu, S. Y.-H., N. Malila, A. M.-F. Yen, S. L.-S. Chen, J. C.-Y. Fann and M. Hakama (2017). "Predicting the effectiveness of the Finnish population-based colorectal cancer screening programme." Journal of Medical Screening **24**(4): 182-188

doi:10.6342/NTU201802429

Chuang, S. L., W. W. Su, S. L. Chen, A. M. Yen, C. P. Wang, J. C. Fann, S. Y. Chiu, Y. C. Lee, H. M. Chiu, D. C. Chang, Y. Y. Jou, C. Y. Wu, H. H. Chen, M. K. Chen and S. T. Chiou (2017). "Population-based screening program for reducing oral cancer mortality in 2,334,299 Taiwanese cigarette smokers and/or betel quid chewers." Cancer **123**(9): 1597-1609.

Day, N. E. (1985). "Estimating the sensitivity of a screening test." Journal of Epidemiology and Community Health **39**(4): 364-366.

Dershaw, D. D., L. Loring, S. J. True, D. C. Momrow, L. Liberman and E. A. Morris (1998). "A comparison of screening mammography results from programs for women of different socioeconomic status." Cancer **82**(9): 1692-1697.

Edwards, B. K., E. Ward, B. A. Kohler, C. Eheman, A. G. Zauber, R. N. Anderson, A. Jemal, M. J. Schymura, I. Lansdorp-Vogelaar, L. C. Seeff, M. van Ballegooijen, S. L. Goede and L. A. Ries (2010). "Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates." Cancer **116**(3): 544-573.

Fracheboud, J., H. J. de Koning, R. Boer, J. H. Groenewoud, A. L. Verbeek, M. J. Broeders, B. M. van Ineveld, J. H. Hendriks, A. E. de Bruyn, R. Holland and P. J. van der Maas (2001). "Nationwide breast cancer screening programme fully implemented in The Netherlands." Breast **10**(1): 6-11.

Hardcastle, J. D., J. O. Chamberlain, M. H. Robinson, S. M. Moss, S. S. Amar, T. W. Balfour, P. D. James and C. M. Mangham (1996). "Randomised controlled trial of faecal-occult-blood screening for colorectal cancer." Lancet **348**(9040): 1472-1477.
Jørgensen, K. J. and P. C. Gøtzsche (2009). "Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends." BMJ **339**.

Koong, S. L., A. M. Yen and T. H. Chen (2006). "Efficacy and cost-effectiveness of nationwide cervical cancer screening in Taiwan." J Med Screen 13 Suppl 1: S44-47.
Kronborg, O., C. Fenger, J. Olsen, O. D. Jorgensen and O. Sondergaard (1996). "Randomised study of screening for colorectal cancer with faecal-occult-blood test." Lancet **348**(9040): 1467-1471.

doi:10.6342/NTU201802429

Lauby-Secretan, B., N. Vilahur, F. Bianchini, N. Guha and K. Straif (2018). "The IARC Perspective on Colorectal Cancer Screening." New England Journal of Medicine **378**(18): 1734-1740.

Moss, S. M., M. Michel, J. Patnick, L. Johns, R. Blanks and J. Chamberlain (1995). "Results from the NHS breast screening programme 1990-1993." J Med Screen **2**(4): 186-190.

Bevers, T. B., B. O. Anderson, E. Bonaccio, S. Buys, M. B. Daly, P. J. Dempsey, W. B. Farrar, I. Fleming, J. E. Garber, R. E. Harris, A. S. Heerdt, M. Helvie, J. G. Huff, N. Khakpour, S. A. Khan, H. Krontiras, G. Lyman, E. Rafferty, S. Shaw, M. L. Smith, T. N. Tsangaris, C. Williams, T. Yankeelov and N. National Comprehensive Cancer (2009). "NCCN clinical practice guidelines in oncology: breast cancer screening and diagnosis." J Natl Compr Canc Netw **7**(10): 1060-1096.

Puliti, D., S. W. Duffy, G. Miccinesi, H. De Koning, E. Lynge, M. Zappa and E. Paci (2012). "Overdetection in Mammographic Screening for Breast Cancer in Europe: A Literature Review." Journal of Medical Screening **19**(1_suppl): 42-56.

Robinson, J. I., C. E. Crane, J. M. King, D. I. Scarce and C. E. Hoffmann (1996). "The South Australian Breast X-Ray Service: results from a statewide mammographic screening programme." Br J Cancer **73**(6): 837-842.

Sankaranarayanan, R., K. Ramadas, G. Thomas, R. Muwonge, S. Thara, B. Mathew and B. Rajan (2005). "Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial." Lancet **365**(9475): 1927-1933.

Schroder, F. H., J. Hugosson, M. J. Roobol, T. L. Tammela, M. Zappa, V. Nelen, M. Kwiatkowski, M. Lujan, L. Maattanen, H. Lilja, L. J. Denis, F. Recker, A. Paez, C. H. Bangma, S. Carlsson, D. Puliti, A. Villers, X. Rebillard, M. Hakama, U. H. Stenman, P. Kujala, K. Taari, G. Aus, A. Huber, T. H. van der Kwast, R. H. van Schaik, H. J. de Koning, S. M. Moss and A. Auvinen (2014). "Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up." Lancet **384**(9959): 2027-2035.

Shaukat, A., S. J. Mongin, M. S. Geisser, F. A. Lederle, J. H. Bond, J. S. Mandel and T. R. Church (2013). "Long-term mortality after screening for colorectal cancer." N

doi:10.6342/NTU201802429

Engl J Med **369**(12): 1106-1114.

Tabar, L., B. Vitak, T. H. Chen, A. M. Yen, A. Cohen, T. Tot, S. Y. Chiu, S. L. Chen, J. C. Fann, J. Rosell, H. Fohlin, R. A. Smith and S. W. Duffy (2011). "Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades." Radiology **260**(3): 658-663.

Tabar, L., B. Vitak, T. H. Chen, A. M. Yen, A. Cohen, T. Tot, S. Y. Chiu, S. L. Chen, J. C. Fann, J. Rosell, H. Fohlin, R. A. Smith and S. W. Duffy (2011). "Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades." Radiology **260**(3): 658-663.

Winawer, S., R. Fletcher, D. Rex, J. Bond, R. Burt, J. Ferrucci, T. Ganiats, T. Levin, S. Woolf, D. Johnson, L. Kirk, S. Litin and C. Simmang (2003). "Colorectal cancer screening and surveillance: clinical guidelines and rationale-Update based on new evidence." Gastroenterology **124**(2): 544-560.

Wu, G. H., A. Auvinen, A. M. Yen, M. Hakama, S. D. Walter and H. H. Chen (2012). "A stochastic model for survival of early prostate cancer with adjustments for leadtime, length bias, and Overdetection." Biom J **54**(1): 20-44.

Wu, G. H., A. Auvinen, A. M. Yen, M. Hakama, T. L. Tammela, U. H. Stenman, P. Kujala, M. Ruutu and H. H. Chen (2012). "The impact of interscreening interval and age on prostate cancer screening with prostate-specific antigen." Eur Urol **61**(5): 1011-1018.

Wu, G. H., A. Auvinen, L. Maattanen, T. L. Tammela, U. H. Stenman, M. Hakama, A. M. Yen and H. H. Chen (2012). "Number of screens for Overdetection as an indicator of absolute risk of Overdetection in prostate cancer screening." Int J Cancer **131**(6): 1367-1375.

Yen, A. M., A. Auvinen, J. Schleutker, Y. Y. Wu, J. C. Fann, T. Tammela, S. L. Chen, S. Y. Chiu and H. H. Chen (2015). "Prostate cancer screening using risk stratification based on a multi-state model of genetic variants." Prostate **75**(8): 825-835.

Yen, A. M., A. Auvinen, J. Schleutker, Y. Y. Wu, J. C. Fann, T. Tammela, S. L. Chen, S. Y. Chiu and H. H. Chen (2015). "Prostate cancer screening using risk stratification based on a multi-state model of genetic variants." Prostate **75**(8): 825-835.

doi:10.6342/NTU201802429

Yen, A. M., H. S. Tsau, J. C. Fann, S. L. Chen, S. Y. Chiu, Y. C. Lee, S. L. Pan, H. M. Chiu, W. H. Kuo, K. J. Chang, Y. Y. Wu, S. L. Chuang, C. Y. Hsu, D. C. Chang, S. L. Koong, C. Y. Wu, S. L. Chia, M. J. Chen, H. H. Chen and S. T. Chiou (2016). "Population-Based Breast Cancer Screening With Risk-Based and Universal Mammography Screening Compared With Clinical Breast Examination: A Propensity Score Analysis of 1429890 Taiwanese Women." JAMA Oncol **2**(7): 915-921.

Yen, A. M., H. S. Tsau, J. C. Fann, S. L. Chen, S. Y. Chiu, Y. C. Lee, S. L. Pan, H. M. Chiu, W. H. Kuo, K. J. Chang, Y. Y. Wu, S. L. Chuang, C. Y. Hsu, D. C. Chang, S. L. Koong, C. Y. Wu, S. L. Chia, M. J. Chen, H. H. Chen and S. T. Chiou (2016). "Population-Based Breast Cancer Screening With Risk-Based and Universal Mammography Screening Compared With Clinical Breast Examination: A Propensity Score Analysis of 1429890 Taiwanese Women." JAMA Oncol **2**(7): 915-921.

Yen, M. F., L. Tabar, B. Vitak, R. A. Smith, H. H. Chen and S. W. Duffy (2003). "Quantifying the potential problem of Overdetection of ductal carcinoma in situ in breast cancer screening." Eur J Cancer **39**(12): 1746-1754.

Y□H, C. S., M. Nea, Y. A. M□F, A. Ahti, H. Matti and C. H□H (2011). "Analytical decision model for sample size and effectiveness projections for use in planning a population□based randomized controlled trial of colorectal cancer screening." Journal of Evaluation in Clinical Practice **17**(1): 123-129.

doi:10.6342/NTU201802429