Convex Analysis for Meta Analysis

Martin Adamčík

2017

Final Report for the Project

Convex Analysis for Meta Analysis
(research grant approved on 20th January 2016)

Asst. Prof. Dr. Martin Adameik
MSME, Department of Management
January 6, 2017

Contents

1 Introduction 5
1.1 Background 5

2 Literature Review 11
2.1 Venous thromboembolism and cancer 11

3 Research Methodology 15
3.1 Unexplained heterogeneity 15
3.2 Probabilistic propositional logic 24
3.2.1 Basic notions 24
3.2.2 Linear entropy operator 27
3.3 Guide to representation 29
3.3.1 How to select propositional variables 30
3.3.2 How to assign probabilities consistently 32
3.3.3 How to assign weighting 44
3.4 Roundabout computation 45
3.4.1 Algorithm 45
3.4.2 Implementation 47

4 Result and Discussion 51
4.1 Result 51
4.2 Discussion 52

5 Conclusion and Recommendation 55

Bibliography 57

ASSUMPTION UNIVERSITY OF THAILAND
SINCE1969
มหาวิทยาลัยอัสสัมชัญ

Chapter 1

Introduction

1.1 Background

In many fields of science, and particularly in medicine, there is a growing expectation to investigate results of previous research either to provide a baseline for comparison or to justify a need for a further empirical study. This could be illustrated by the following advice given by the UK's National Health Service that `research should take place only when a systematic review of previous research has been carried out and the need for new research has been established' [1].

If the quality of the previous studies is not to be questioned, and if they address the same research problems in similar ways and possible inconsistencies between studies are attributable to sampling errors then fixed—effect meta—analysis is a powerful tool for such an investigation. Such meta—analysis typically applies statistical techniques that take into account the sample sizes of given studies and produces a weighted arithmetic mean of a particular proportion being investigated [2]. For example, this proportion could be the odds that a specific treatment will reduce the risk of developing a specific disease or a proportion of certain patients that develop a certain condition.

The advantage of meta—analysis is that on one hand it allows dealing with statistically expected inconsistencies among studies to increase the statistical power by working with a larger pooled sample and on the other hand it allows one to generalise the results to a larger population. This latter use is due to the fact that many studies focus on a narrow population but span a much larger population. Such use of meta—analysis was advocated in [3] although it is controversial according to some other authors [4].

Nevertheless, there seems to be a consensus that such meta—analysis is not appropriate when there is heterogeneity between studies explained by the variability in research methods or in investigated samples. This is often referred to as 'mixing apples with oranges'. Indeed, if one study is about young patients and another about elderly patients then everyone in the medical field would take a combination of such studies' findings with extreme caution. As pointed out

in [4], even more striking would be a combination of weights of small children and dogs with large fish and cats even if all four samples appeared to be homogeneous; i.e., with variations as if they were taken randomly from the same population.

The question we ask in this paper is different and fairly specific. Is there a way to analyse studies that do not appear heterogeneous in terms of research problems and ways they investigated them but despite that studies show it statistically beyond any doubt? And of course we do not mean the obvious answer postulating the need to identify the actual source, which is indeed the right way to approach the problem if it is feasible because it explains it all and every researcher should do this in the first place. We have in mind the situation where despite an effort we have not clearly identified the source of heterogeneity. In this paper we call this unexplained heterogeneity.

Certainly statistical analysis that fail to account for heterogeneity (i.e., fixed—effect meta—analysis) is of no help since something else is going on other than that each study just took a sample from the same population. And although some statistical methods accounting for heterogeneity have been developed, in this paper we suggest a non-statistical technique, notably employing mathematical logic, which could be perhaps categorised as a version of the recently developed Bayesian approach to meta—analysis [5], but unlike the more general Bayesian approach it will be strictly applicable only to the restricted setting of our question. Although it is quite restricted, the technique still has appealing advantages of the Bayesian approach:

1. It deals with heterogeneity between studies.

2. It is able to combine related studies that do not necessarily measure the same proportions.

Similarly as in the case of the Bayesian approach to meta—analysis, given that the danger of combining unrelated studies has been avoided, we have the freedom to include into the analysis a whole range of studies. This can be quite convenient as the number of studies investigating a rare disease could be rather small for a meaningful statistical meta—analysis. The inability to deal with complex knowledge is one of the classical criticisms of meta-analysis [2]. To illustrate, when researchers wish to combine related studies into a single narrative they could end up with the following scenario. Say that one study investigates the effectiveness of a routine screening in discovering cancer and the other study adds to the same routine screening a CT scan. Although they measure effectively different proportions, the first the proportion of patients where the routine screening discovered cancer and the second one the proportions where the routine screening with a CT scan discovered cancer, these two proportions are much related so it appears natural to try to combine them.

Combining related studies is of course performed in systematic reviews of literature all the time [2]. However, such a combination is subjective in nature and often comes with bias as was shown in [6], where a catchy title of a study was all that was needed to bias the analysis. In this paper, in order to

achieve a rigorous combination of related studies, we use probabilistic logic and information geometry to transform studies to non—empty closed convex sets in a probabilistic simplex. The reasoning behind the encoding selection used in this paper in lieu of the plethora of choices shall be clearly explained. This will not only be a theoretical construction; the encoding will be demonstrated on real studies concerning the incidence of diagnosis of cancer in patients with unprovoked venous thromboembolism. The encoding will be similar to that suggested in [7] but further elaborated. We aspire here to provide the reader with a comprehensive tutorial to the approach we are proposing and demonstrate it on a real application.

The subsequent combination of resulting non—empty closed convex sets in probabilistic simplex will be performed by a probabilistic merging operator called the linear entropy operator, which was first introduced in [8]. More significantly, the fact that this operator is appropriate for analysing related studies with unexplained heterogeneity was showed in [9]. The following are the assumptions necessary for the argument to work and under different assumptions another operator could be more optimal:

1. There is a huge population from which each study selected a large sample on which it measured some proportions.

2. Heterogeneity can be statistically detected from the measured proportions.

3. The studies appear to investigate the same population by rather similar methods; there is no obvious source of heterogeneity.

The technical result from [9] was obtained similarly as the maximum entropy inference process was justified in [10]. Informally, an entropic approach to a daunting problem is about ignoring irrelevant processes that make the problem difficult. In this particular setting the exact nature of the unknown processes that do exist and cause the observations reported in studies more different than statistically expected could be ignored and instead these processes are treated probabilistically. There could be so many of them that we can just assume that each can occur with equal probability (but this would be false if a source of heterogeneity was known). And although this argument works even in the case of homogeneity, a statistical fixed—effect meta—analysis produces in such a case much more powerful results [9]. So the entropic approach should not be used if a more powerful method is available; either statistical in the case of homogeneity, or if we are able to identify the source of heterogeneity.

This shares with the Bayesian approach to meta—analysis the following. In the Bayesian approach a prior distribution is chosen for a particular practical meta—analysis and a posterior distribution is computed based on some additional information. The prior distributions used in [9] is that 'all possible ways in which studies could have obtained their observations are equally probable' and 'every population proportion is a *priory* equally probable'. Using a specific modelling of observing proportions as drawing with replacement the so called 'number of possible states argument' is applied to determine overwhelmingly most probable

population proportions given observations of studies and these coincide with the result of the linear entropy operator mentioned above. See [9] for more details.

For a reader who feels that the paper [9] is too technical but who still wants to understand the linear entropy operator better we will later list several properties which are in the context of uncertain reasoning [11, 12] called principles and we will point out which of them are satisfied by this operator.

Since the use of the linear entropy operator in the setting outlined above has been already supported in [9], it is the encoding of studies by several non—empty closed convex sets in a probabilistic simplex and the interpretation of the non—empty closed convex set resulting from the operator that will be addressed in this paper. This paper focuses on plain, accessible language rather than relying on mathematics in order to be useful to a greater array of future researchers interested in combining studies with unexplained heterogeneity.

Finally, we should point out that there is a major computational issue with the linear entropy operator and in order to apply the proposed method some roundabout computation in convex optimisation will be performed at the end of the paper by a method from [13, 14]. With all these nuances our approach to meta—analysis with unexplained heterogeneity is schematically outlined in Figure 1.1.

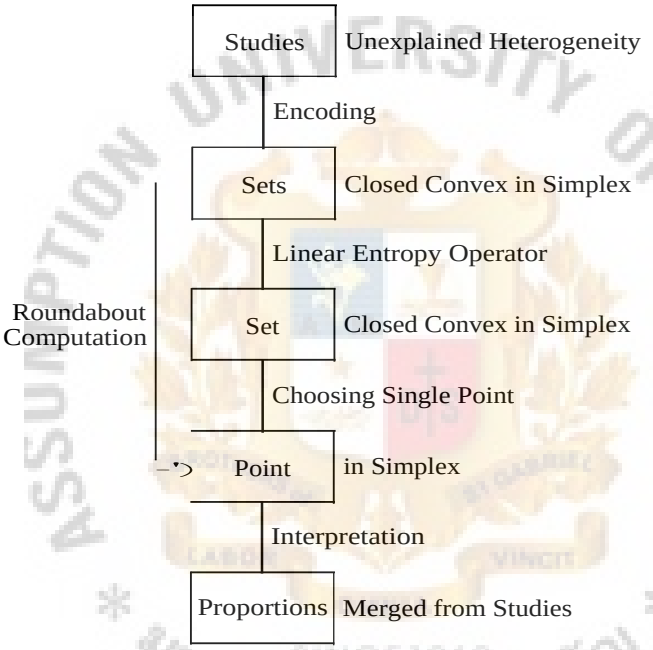

Figure 1.1: Our approach to meta—analysis with unexplained heterogeneity. We focus on blue arrows in this paper while red arrows were already developed in the literature.

ASSUMPTION UNIVERSITY OF THAILAND
SINCE1969

Chapter 2

Literature Review

2.1 Venous thromboembolism and cancer

Throughout this paper we explain the proposed approach to meta—analysis with unexplained heterogeneity on a particular medical problem concerning venous thromboembolism (VTE). VTE is a condition that includes either deep vein thrombosis (DVT) or pulmonary embolism (PE) or both. DVT is a formation of a blood clot in a vein. Such a clot often dislodges and travels through the heart into lungs. A sufficiently large clot can subsequently block a pulmonary vein causing PE and if left untreated it can lead to death [15]. In fact, VTE is the third most common cardiovascular illness [16].

Provoked venous thromboembolism is associated with prolonged rest or coagulation abnormalities and it could be induced by effort or a vein catheter [15]. Existing cancer, recent surgery and pregnancy are also risk factors [17]. When no such reason is apparent venous thromboembolism is referred to as unprovoked (in the literature alternatively the term 'idiopathic VTE' is used).

In 1872 Trousseau [18] observed that there is some sort of relationship between unprovoked VTE and previously undiagnosed (occult) cancer. Since then the presence of the relationship has been observed many times; a large meta—analysis [19] shows that 10% of patients diagnosed with unprovoked VTE are diagnosed with cancer within 12 months and the inclusion of extensive screening statistically significantly improves the detection of cancer from 49.4% of cancers detected with routine evaluation alone to 69.7% of cancers detected with extensive screening combined with routine evaluation. The exact nature of the relationship is still unknown and only recent studies started to cast some light on questions as to how patients with unprovoked VTE should be precisely screened for cancer when their condition becomes symptomatic [20, 21, 22, 23, 24, 25, 26, 27]. At the same time some of these studies reported a lower incidence of diagnosis of cancer in patients with unprovoked VTE than 10% and some of them also reported a low sensitivity of extensive screening. In this paper we are going to look at these studies and combine them. This selection of studies was

obtained by searching the PubMed Central database [28].

All the studies listed above are fairly similar in nature. Each is either a randomised or on observation study investigating adult patients with an unprovoked episode of VTE. This means that patients with the following risk factors were excluded: coagulation abnormalities (e.g., factor V Leiden mutation), recent surgery, prolonged rest, pregnancy or effort—induced episodes. There was however a variation on whether young adult patients were excluded (less than 40 years in [20], less than 25 years in [21] and less than 70 years in [25]). If a patient had no known cancer he or she was observed over a certain period of time to determine whether cancer occurs in his or her case. Studies that did not fit this template or were published before the year 2 000 were excluded.

It is important to note that the studies included in our analysis did not differentiate between different types of VTE; patients with DVT, PE or both conditions were eligible. This particularly means that DVTs in legs, which are much more common than DVTs in upper—extremity (such as in subclavian veins) [15], were much more represented in the studies. In fact, studies [23, 24] and [26] outright excluded upper—extremity DVTs. Studies were performed in western countries and no study selectively targeted a specific population in terms of life—style. Notably, studies included past and current smokers despite the fact that smoking is considered as a risk factor for both VTE [17] and cancer. Also, patients with previous history of VTE (unless recent) and family history of VTE were not excluded. Although most cancers were diagnosed in the older age groups no age differentiation was performed in any study.

The main objective of each study was to determine how the screening for cancer should be performed when unprovoked VTE is diagnosed. Since the one—year incidence of diagnosis of cancer in patients with unprovoked VTE at the level of 10% reported in [19] is much higher than in the normal population, naturally it seems appropriate to screen them. However, is screening considered as routine sufficient? Or do more invasive and expensive methods pay for themselves in terms of a much better detection rate?

It appears that the problem addressed is the same across the studies and that they investigated it in a fairly similar manner. Recall that this is a necessary assumption; otherwise, the method proposed here may not be used. (Needless to say, in such a case statistical meta—analysis would also be inapplicable.)

More explicitly, there was only a slight variation in the techniques used for routine screening across the studies. The techniques always included the following.

- Medical history, physical examination, complete blood count and chest x—rays.

Furthermore, creatinine levels, serum electrolytes and liver function testing were often included.

On the other hand, extensive screening techniques varied more significantly. They usually comprised a selection of the following.

- Chest computer tomography (CT) scan, abdomen CT scan, pelvis CT

scan, 18F—FDG—PET/CT whole body scan, ultrasound of abdomen and pelvis, mammography, tumor markers and prostate—specific antigen.

Before extensive screening techniques become part of a routine cancer screening for patients with unprovoked VTE we ought to make sure that the resources, psychological trauma and the danger of radiation induced cancer are all out—weighed by the number of lives saved.

In this paper we will learn more about this interesting problem. We will see that despite similarities we can statistically detect heterogeneity in our selection of studies. We will also discover that they often measure different proportions but we will also explain how we can deal with such an issue using our representation in the form of non—empty closed convex sets in a probabilistic simplex. We will show the findings researchers established within individual studies and we will combine them using the method outlined in the introduction.

ASSUMPTION UNIVERSITY OF THAILAND
LABOR OMNIA VINCIT
SINCE1969
มหาวิทยาลัยอัสสัมชัญ

Chapter 3

Research Methodology

3.1 Unexplained heterogeneity

First, we are going to create a so called funnel plot for the incidence of diagnosis of cancer in patients with unprovoked VTE of the selection of studies [20, 21, 22, 23, 24, 25, 26, 27] we wish to combine against the sample size. An asymmetric funnel plot indicates that there is some sort of publication or selection bias [2], for example studies reporting negative results could tend to be unpublished. However, the funnel plot in Figure 3.1 does not appear to indicate any problem with our selection.

Second, we are going to create a Forest plot for the one—year incidence of diagnosis of cancer in patients with unprovoked VTE of our selection of studies, including the meta—analysis [19] and using a more recent paper ([29]) to identify additional details about the study [20]; see Figure 3.2. Note that the study [21] did not report the incidence of diagnosis of cancer in patients with unprovoked VTE so it is excluded from this Forest plot. A Forest plot is a graphical tool, which allows us to make initial analysis of studies; in particular in the following we try to see if the mean age of patients plays any role in the measured incidence of diagnosis of cancer in patients with unprovoked VTE. This was the motivation behind including one study that specifically targeted elderly patients [25] but in general this would be undesirable as a different mean age can be identified as a possible source of heterogeneity. It appears that mean age could play a small role in the reported incidence, although the studies investigating older patients have rather small sample sizes; see Figure 3.2.

In the Forest plots the pooled value was obtained as a weighted arithmetic mean of the reported incidences weighted by the corresponding sample sizes. Usually, weighting factors are not obtained directly from sample sizes. Instead, the inverse of the variance of a particular study is taken. Variance more accurately accounts for example differences in control and treatment groups when odds ratios are pooled but in our case we pool just the proportions so the inverse of the variance is proportional to the sample size. Furthermore, in the

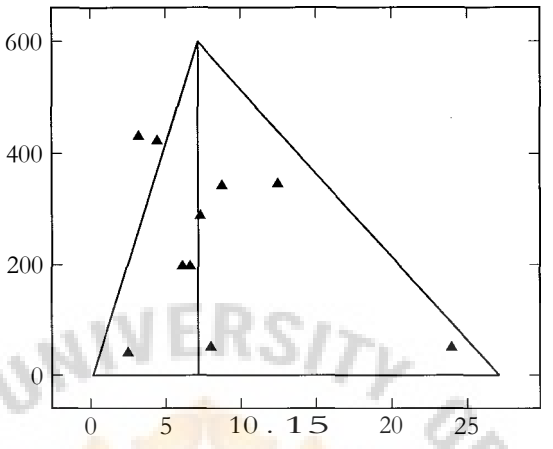

Figure 3.1: The funnel plot of the selection of studies we wish to combine. The x—axis denotes the proportion of cancers discovered over the period of each study. The y—axis denotes the sample size. The red line indicates the pooled incidence.

forest plots we use Clopper—Pearson confidence intervals since they are accurate confidence intervals for a random variable with a binomial distribution regardless of the sample size. In particular, in our following Forest plots (Figures 3.3 and 3.4) concerning the effectiveness of routine evaluation and the effectiveness of the technique combining routine evaluation with extensive screening we may not use the normal approximation to a binomial distribution since the number of patients with cancer in many studies is too small (less than 30).

According to the Forest plots, the pooled value of the one—year incidence of diagnosis of cancer in patients with unprovoked VTE (7.15%, 95% confidence interval 6.14 — 8.26) is lower than the one reported in the previous meta—analysis [19] (10%, 95% confidence interval 8.6 —11.3) and this difference is statistically significant as seen from the confidence intervals. On the other hand, the pooled value for sensitivity of routine evaluation (47.93%, 95% confidence interval 38.77-57.20) is similar to the one reported in the meta—analysis (49.4%, 95% confidence interval 40.2 — 58.5). Accordingly, the pooled value for sensitivity of the technique combining routine evaluation with extensive screening (75%, 95% confidence interval 66.27 — 82.45) is not statistically different from the one reported in the meta—analysis (69.7%, 95% confidence interval 61.1 — 77.8).

It appears that at this point we can just end our analysis. But ending the research at this point poses two issues:

1. Many studies did not report the proportions measured in the above mentioned Forest plots precisely. Notice empty spaces and that one study [21] is not included at all. Nevertheless, the studies measured similar propor-

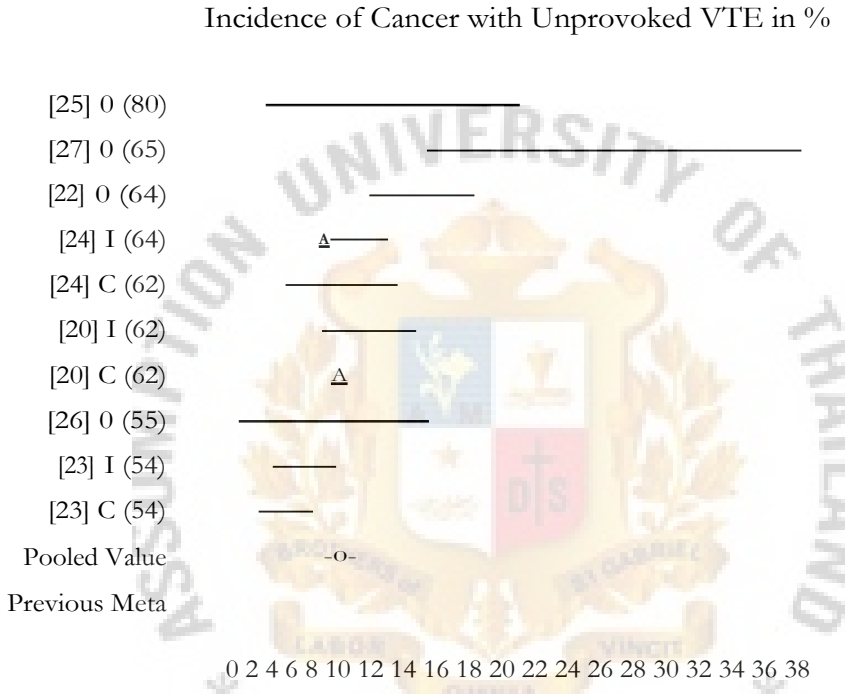

Figure 3.2: The Forest plot of the selection of studies in scope for combining. The x—axis denotes the percentage of people that had cancer discovered within a 12 months period (studies [20] and [24] had longer follow—up periods) after the occurrence of VTE. 95% Clopper—Pearson confidence intervals are used. On the y—axis the studies are arranged by the mean age of patients (the value in round brackets). The letter '0' indicates an observation study, the letter T indicates the incidence in the group that received the intervention (extensive screening) and the letter 'C' indicates the incidence in the control group.

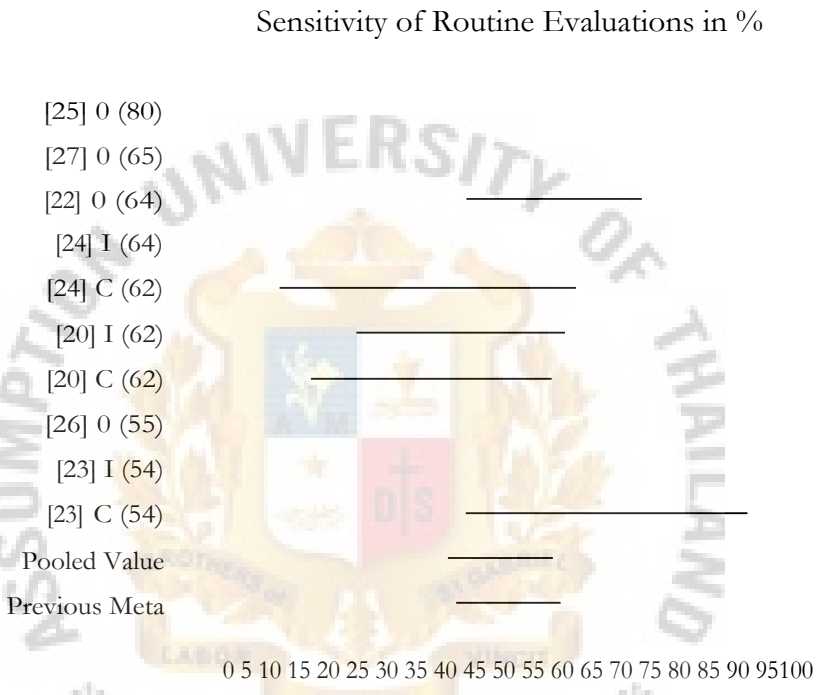

Figure 3.3: The Forest plot of the selection of studies in scope for combining. The x—axis denotes the proportion of cancers detected by routine evaluation when unprovoked VTE was diagnosed. 95% Clopper—Pearson confidence intervals are used. On the y—axis the studies are arranged by the mean age of patients (the value in round brackets). The letter '0' indicates an observation study, the letter 'I' indicates the incidence in the group that received the intervention (extensive screening) and the letter 'C' indicates the incidence in the control group. Not every study reported the proportion correctly, the sensitivity is missing for those studies.

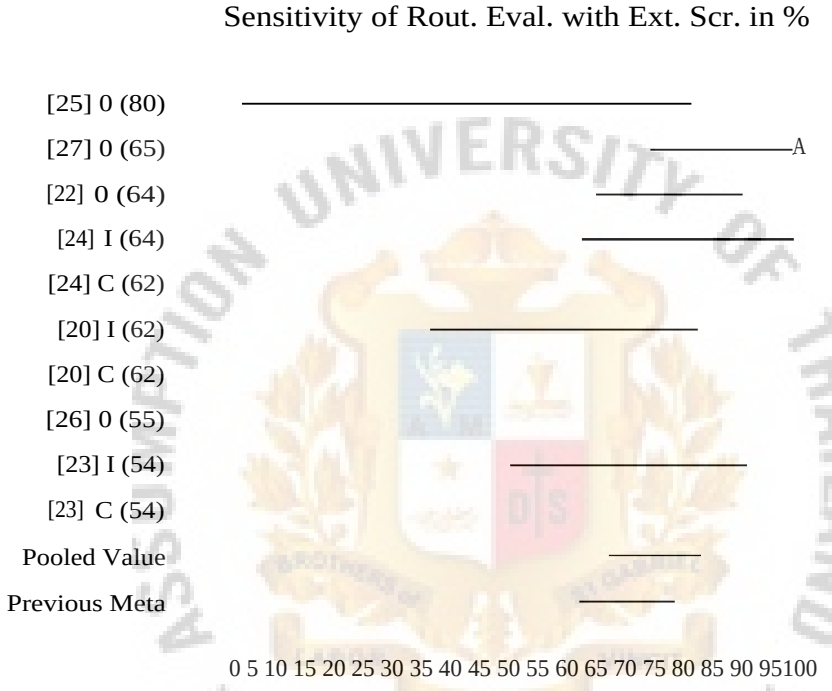

Figure 3.4: The Forest plot of the selection of studies in scope for combining. The x—axis denotes the proportion of cancers detected either by routine evaluation or by extensive screening when unprovoked VTE was diagnosed. 95% Clopper—Pearson confidence intervals are used. On the y—axis the studies are arranged by the mean age of patients (the value in round brackets). The letter '0' indicates an observation study, the letter T indicates the incidence in the group that received the intervention (extensive screening) and the letter `C' indicates the incidence in the control group. Not every study reported the proportion correctly, the sensitivity is missing for those studies.

tions. Can we somehow take advantage of the information they provide?

2. Heterogeneity between the studies can actually be statistically detected as we will shortly see. Is it right to just compute and interpret the pooled value (the weighted arithmetic mean of proportions reported in studies)?

Let us have a look at the reported incidence of diagnosis of cancer in patients with unprovoked VTE from Figure 3.2. Let us assume that the pooled proportion $p = 0.0715$ (7.15%) is the true proportion of patients that have cancer discovered in a 12 months period. If there are n patients in a sample then by the central limit theorem the number X of patients that have cancer discovered is a random variable and

$$\frac{X - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}} \approx Z$$

that is this fraction as a random variable has the standard normal distribution (this is indeed just an approximation and in general it works if $n \cdot p$ and $n \cdot (1-p)$ are both at least 5). Now, if p is obtained by taking m binomially distributed random variables $X_1, \ldots, X_m$ generating the same proportion by $p = \frac{X_1 + \cdots + X_m}{n}$, where $n_1 + \cdots + n_m = n$, then

$$\sum_{i=1}^{m} \frac{(X_i - n_i \cdot p)^2}{n_i \cdot p \cdot (1 - p)} \approx \chi^2_{m-1},$$

see [30]. One level of freedom (we sum only until $m - 1$) is lost by including the equation $p = \frac{X_1 + \cdots + X_m}{n}$. Finally, by observing

$$\frac{(X_i - n_i \cdot p)^2}{n_i \cdot p \cdot (1 - p)} = \frac{(X_i - n_i \cdot p)^2}{n_i \cdot p} + \frac{(n_i - X_i - n_i \cdot (1 - p))^2}{n_i \cdot (1 - p)},$$

and denoting the value of the random variable X_i respectively $n_i - X_i$ by O (`observed'), and denoting $n_i \cdot p$ respectively $n_i \cdot (1 - p)$ by $E$ (`expected') we obtain the familiar Pearson's chi—squared test [31] for homogeneity of binomially distributed $X_1, \ldots, X_m$ (i.e., the test whether they indeed generate the same proportion);

$$\sum \frac{(O - E)^2}{E} \approx \chi^2_{m-1}. \tag{3.1}$$

Now, if the studies have virtually taken samples from the same population then the numbers of patients $X_1, \ldots, X_m$ that have cancer discovered should be only a function of the sample size and the statistical error. If the value of the random variable in Formula 3.1 differs significantly from the expected value having the χ^2_{m-1} distribution then it is unlikely that the random variables generate the same proportion. This is what we call statistically detected heterogeneity. The reported proportions depend on something and we can ask what they depend on.

Study	Patients	Observed *(O)*	Expected *(E)*
[25] 0 (80)	with cancer	4	3.58
	without cancer	46	46.42
[27] 0 (65)	with cancer	12	3.58
	without cancer	38	46.42
[22] 0 (64)	with cancer	43	24.67
	without cancer	302	320.33
[24] I (64)	with cancer	12	14.09
	without cancer	185	182.91
[24] C (62)	with cancer	13	14.09
	without cancer	184	182.91
[20] I (62)	with cancer	30	24.45
	without cancer	312	317.55
[20] C (62)	with cancer	21	20.59
	without cancer	267	267.41
[26] 0 (55)	with cancer	1	2.86
	without cancer	39	37.14
[23] I (54)	with cancer	19	30.24
	without cancer	404	392.76
[23] C (54)	with cancer	14	30.82
	without cancer	417	400.18

Table 3.1: The number of patients with diagnosed cancer and without diagnosed cancer respectively. The expected value E is computed as the product of the sample size and the pooled proportion of patients with and without cancer respectively. On the y—axis the studies are arranged by the mean age of patients (the value in round brackets). The letter '0' indicates an observation study, the letter 'I' indicates the incidence in the group that received the intervention (extensive screening) and the letter 'C' indicates the incidence in the control group.

So what are the observed and expected values for the data reported in Figure 3.2? We explore them in Table 3.1.

Applying Formula 3.1 we establish that $x6 = \frac{(4-3.58)^2}{3.58} + \frac{(46-46.42)^2}{46.42} + \frac{(12-33.58)^2}{33.58}$ = 53.531, although a single study, namely [27], contributed by 21.33 to this value. A classical approach would be to remove this study from the meta—analysis. Statistically, in 90% of cases the value of A is below 13.36 should be there no heterogeneity. According to [2] 90% confidence level is usually used to determine this critical value 13.36 and the result above this critical point should invite caution as such a result indicates heterogeneity so studies should not be simply pooled.

Note that some expected values in Table 3.1 are below 5 which causes troubles with the application of the central limit theorem as explained above. Nevertheless, it is still applicable if the following conditions, originally attributed to [32], are satisfied.

1. No expected value is less than 1.

2. No more than 20% of the expected values are less than 5.

We can see that these conditions are satisfied in our case. Note that in the conditions above we are not concerned with the observed values; we are only interested in the expected values.

We also perform Pearson's chi—squared test for the reported sensitivity of routine evaluation (Figure 3.3) and the sensitivity of the technique combining routine evaluation with extensive screening (Figure 3.4). In the prior case the measured chi-squared value is xg = 8.38 while in 90% of cases this value is not higher than 7.78 should there be no heterogeneity. In the latter case the measured chi-squared value is xb = 15.11 while in 90% of cases this value is not higher than 9.24 should there be no heterogeneity. Although it must be noted that the number of studies considered and their respective sample sizes are worryingly low in these two tests.

In general, if the number of studies is low or the sample sizes are small then the chi—squared test is underpowered; in other words, it may not detect heterogeneity even if it is present. Despite this fact, we have detected heterogeneity.

The studies considered above investigated the same problem, the incidence of diagnosis of cancer in patients with unprovoked VTE, all using routine evaluation, extensive screening and a follow—up period. So there is no obvious source of the heterogeneity that was statistically detected above. Interestingly, in practice we have often statistical homogeneity but in fact studies are heterogeneous; recall the example combining small children and dogs with large fish and cats from the introduction. Here we have detected heterogeneity statistically but with no obvious explanation. Recall that this situation is referred to as unexplained heterogeneity. And although, digging further into studies, we uncover quite a few differences, these differences do not appear to explain the variability in reported proportions:

Although the study [25] excluded everyone younger than 70 years, looking at Figure 3.2 it can be seen that the variability in the mean age does not explain

heterogeneity in the measured incidence. Also, the fact that studies [20, 21] and [24] had longer follow—up periods should not significantly contribute to heterogeneity since the risk of being diagnosed with cancer falls to near normal levels a year after the diagnosis of VTE [33]. Even though studies [23, 24] and [26] excluded upper—extremity DVTs, these are only marginal to the overall number of VTEs [15]. Finally, there was large variability in what exactly a CT scan included, in particular it sometimes included and sometimes omitted pelvis, but the variability of differences means that it is difficult to attribute the specific heterogeneity we detected to it.

On the other hand, in particular the age and the choice of extensive screening techniques are certainly factors that contributed to heterogeneity. The simplification we adopt is that we treat these factors equally with any other possible factors contributing to heterogeneity. It would be much better to attribute clearly specific variability to detected heterogeneity, but since we are not able to do so then we argue that the high complexity of possible sources of heterogeneity should be approached probabilistically considering all possible situations equally. Precisely this was performed in [9] and the linear entropy operator emerged. So we are confident that applying it in this situation is justified.

In the literature the problem of unexplained heterogeneity is often approached statistically. Several existing methods are commonly referred to as random—effects meta—analysis [5]. Such analysis acknowledges that there are in fact differences between studies; for example, the incidence of diagnosis of cancer in patients with unprovoked VTE can depend on age and studies having differently aged patients naturally have different expected incidences. The reported proportions do not vary only due to chance but also due to the other factors. Naturally, the precise source of heterogeneity is not known; in general we just make sure that small studies are taken into account along with big studies; see [34]. This can be as extreme as virtually taking the un—weighted arithmetic mean of particular proportions reported in all studies.

However, this random—effects approach is sometimes criticised. As advocated in [34], large heterogeneity in no way indicates that smaller studies should be more trusted nor that there is a fault in bigger studies. The approach proposed later in this paper preserves the weights of big studies and thus it should not be susceptible to such a criticism. Furthermore, in [9] a specific model of meta—analysis with unexplained heterogeneity was created and it was shown that in that model the weighted arithmetic mean actually provides overwhelmingly the most probable estimate of the true proportion in the population that the studies investigated.

Furthermore, apart form statistically detected heterogeneity we have identified another problem. Proportions reported in many studies were related to the problem presented in this paper, but could not be incorporated in the analysis above, and we would not be able to use them also in random—effects meta—analysis. However, the following section explains a non—statistical technique that will permit taking them into consideration.

3.2 Probabilistic propositional logic

How would a logician approach the problem presented in the previous section? The author of this paper happens to be a logician so here is a method that we could perhaps call a logician's approach to meta—analysis with unexplained heterogeneity.

3.2.1 Basic notions

Our basic references for this framework are [11, 12, 35]. First, we need to build a propositional language in which we will talk about findings reported in studies. In general, this is a set L of a finite collection of variables, say

$$L =$$

For example, we can denote by a_1 the statement that a patient has cancer and by a_2 the statement that the patient has VTE. In order to combine variables into sentences we use several logical symbols: 'and' (A), 'or' (V) and 'negation' (--,) using the following recursive scheme: Every propositional variable is a sentence. If S is a sentence also is a sentence. For example, if S is a propositional variable a_1 then $= gal$ means that the patient does not have cancer. If S_i and S2 are sentences then also S_i A S2 and S_i V S2 are sentences. For example, if S_i denotes the sentence $_{gal}$ that the patient does not have cancer and S2 represents the sentence a2 that the patient has VTE the sentence S_i A S2 = $(—a_1)Aa2$ means that the patient does not have cancer and at the same time he or she has VTE. The sentence $a_2A(—a_1)$ has exactly the same meaning as $(¯a1)Aa2$; both 'A' and 'V' are commutative. The statement Si V $S1$ = $(¯a1)$ V al means that the patient either does not have cancer or the patient has it. Obviously, this sentence is always true. Most sentences however cannot be decided; they may be true for a particular patient but not necessarily.

To express the likelihood that a particular sentence is true for a random patient we assign to every sentence that can be constructed over our propositional language L a number between 0 and 1. 0 means that the sentence is false, 1 that the sentence is true and numbers rising from 0 to 1 represent our increasing confidence in validity of the sentence. Let us denote such a function

$$P : \text{sentences} \quad [0, 1] .$$

The function P may not be arbitrary and it must follow some probabilistic rules. There are many justifications as to why it is so but perhaps the most compelling are found in the Dutch book argument by de Finetti [36]. The rules are the following:

- If S is a true sentence then $P(S) = 1$,
- $P(—S) = 1 — P(S)$,
- $P(S_i \text{ V } S_2) = P(S_1) + P(S_2) — P(S_1 \text{ A } S2)$.

So in particular if S_i and S_2 are mutually exclusive (i.e., S_i A S_2 is false) then $P(S_i \vee S_2) = P(S_1) + P(S_2)$. We shall call P satisfying the rules above a probability function from now on.

There is another type of a sentence which we often need to consider: a conditional sentence. A conditional sentence S_i given S2 is a sentence where we consider validity of S_1 under the assumption that S_2 is true. We denote this conditional sentence S_i S2. The order of symbols matters; T is not commutative. For example, let a_1 denote the sentence that a patient has cancer and a2 the sentence that a CT scan detected a malignant tumor. Note that clearly a2 is false. Now for a random patient the sentence a_1 may or may not be true. But given a_2 we surely know that al is true. Hence $a_i|a_2$ is true and we would assign P(aiIa2) = 1. How to define this probability function in general is given by the so called Bayes formula (and it is due to mathematicians Thomas Bayes and Pierre—Simon Laplace):

$$P(S_1|S_2) = \frac{P(S_1 \wedge S_2)}{P(S_2)}$$

whenever $P(S_2) \neq 0$.

The logical and probabilistic rules above are valid in any situation. In particular, logically true sentences such as $(\neg a_1) \vee a_1$ are always assigned the value 1. If validity of a sentence is uncertain then we shall attempt to use the findings reported in a particular study to find the appropriate number in [0, 1] representing our confidence that it is valid for a random patient. For example, if 10% of patients with VTE have diagnosed cancer in a particular study and we define $L = \{a_1, a_2\}$, where a_1 stands for cancer being diagnosed and a_2 for VTE being present, then

$$P(a_1 \wedge a_2) = \frac{10}{100} = 0.1$$

shall represent our confidence in the sentence a_1 n a2 being true for a patient randomly selected from the population investigated in the study.

For every observation research or for every sample investigated in randomised studies we may consider the set of all probability functions that we can create from observations in the study in a given propositional language L. Different languages can lead to different sets and therefore we denote such a set by K^L and call it an L—knowledge base. In our example above we would write

$$KP = \{P(a_1 \wedge a_2) = 0.1\} \quad (3.2)$$

but this would not be possible if we had no propositional variables expressing the diagnosis of cancer and the presence of VTE.

However, even with a fixed language we still can create several different sets of statements. For example, we can replace the knowledge base ill Equation 3.2 with

$$KZ = \{P(\neg(a_1 \wedge a2)) = 0.9\}$$

but the intuitive meaning of the two knowledge bases Kf and la is the same. In other words, they both generate the same sets of possible probability functions P. More explicitly, this set of possible probability functions is a closed convex subset of the 2' — 1 dimensional probabilistic simplex where n is the number of propositional variables in L. (This is because the set of all logically and probabilistically possible probability functions P forms the whole probabilistic simplex which is closed and convex itself, and observations from studies in the form explained above are only linear constraints, which can restrict the simplex only in a closed and convex manner.) If two L—knowledge bases generate the same set of possible probability functions P then we call them equivalent. We shall be interested in knowledge bases only up to equivalence and we identify them with their respective closed convex sets in the probabilistic simplex. We shall use symbols WI', .1/174, and ViL, V4 to denote closed convex sets in the probabilistic simplex generated by L.

Now consider an L_1—knowledge base $K^{L}1$ generating a closed convex set $WL1$ where $L_1 =$, $a_n\}$ and expand our language by some additional propositional variables to form $L_2 = L_1 U \{b_1,$ $b_8\}$. If we disregard any implicit knowledge about the variables then $K^{L}1$ is also an L_2—knowledge base. Denote the closed convex set generated by K^{L1} in L_2 by W^{L2}. Now for every $P E W^{L2}$ let $P1L_1$ be the probability function from the simplex generated by L_1 that agrees with P on all sentences that can be created over L_1. Define

$$W^{L2}\,1L1 = \{PiLl: P \text{ E } W^{L2}\}.$$

Then clearly $W^{L}2$ $L, = W^{L1}$ (this can be shown using the disjunctive normal form theorem of propositional logic together with the rules that P obeys specified above).

In any practical problem it is undesirable to disregard all implicit knowledge we possess, in particular if it concerns dependence of propositional variables. For example, we may have two variables, one say a_1 representing general detection of cancer and the other say a2 representing detection by a CT scan. Clearly a2 implies a_1 as it cannot be that cancer is detected by a particular test but at the same time it is not detected in general. Such implicit knowledge always must be part of a knowledge base in which case the knowledge base may not be so easily transferable to an extended language as suggested above.

Now, let L1 = {al, a2} and $KL1 = \{P(a_i \text{ A } a2) = 0.2, P((\neg a1) \text{ A } a2) = 0.2, P(ai \text{ A } (—a_2)) = 0.3, P((\neg a1) \text{ A } (\neg a2)) = 0.3\}$. There is only one probability functions satisfying $KL1$ so the corresponding closed convex set WL1 has only one element. If ^{KL1} was constructed from observations in a particular study then the study provides full information regarding a_1 and a_2. However, the constraints formulated from observations in a particular study often only restrict the set of all possible probability functions, in particular when we consider many studies that make our propositional language rich. For example, if we add an extra variable a_3 to our language that is independent of a_1 and a_2 forming the language L_2 because another study talks about a_3 then we have no additional constraints to put to KL2 based only on the first study. Hence

K^{L2} does not specify for example the value of P(a3) at all and there are many different probability functions that could be possible and the corresponding closed convex set W^{L2} in the probabilistic simplex has many elements.

One needs to be careful not to include inconsistent constraints to a single knowledge base. For example, since sentences a_1 Aa2 and ($^{-}$2i) Aa2 are mutually exclusive there is no probability function P that satisfies $KL1 = \{P(a_i A a_2) = 0.7, P((—'a_1)$ A $a_2) = 0.81$. In other words, the corresponding set W^{L1} in the probabilistic simplex generated by L_1 is empty. This can happen for example if one mix observations from a control group with an intervention group in a randomised study.

Apart form dependence of selected propositional variables there is another implicit knowledge we would like to incorporate to every knowledge base: every sentence that is not logically false has some probability to be realised. Therefore, we add to every knowledge base linear constraints stipulating that P must assign to every logically satisfiable sentence a value bigger than some small positive constant, say $\frac{1}{100\,000}$. In our context this represents that everything that is logically possible should be manifested at least in one patient out of one hundred thousand. We shall often assume that such constraints are in place.

3.2.2 Linear entropy operator

Given a study, if we manage to identify a suitable propositional language L and find relevant and consistent constraints on possible probability functions P using observations from the study and implicit knowledge concerning propositional variables then we in fact represent the study by a non—empty closed convex set in the probabilistic simplex determined by L. The question as to how we should determine L and the constraints merits detailed investigation and it is the main objective of this paper. We will address it in detail in the following section. In this section we show how we can use this representation to combine the observations reported in studies.

For the purpose of combining non—empty closed convex sets in probabilistic simplex, the linear entropy operator denoted eKL was introduced in [8] (but only in the case when the weighting is uniform) and justified as optimal for combining studies with unexplained heterogeneity in [9] (this time for any non—zero weighting) under the assumptions that were specified in the introduction. However, do note that under different assumptions another operator might be more suitable and, additionally, whether these assumption are fully satisfiable in practice is questionable.

Now, formally, our preferred operator eKL takes an m—tuple of closed convex sets WI', , W_1f, and a non—zero weighting A = (A_i A_m), Al > 0, A_m > 0 and $E_i^m{}_i$ ai = 1, and maps them to a single non—empty closed convex set in the probabilistic simplex determined by L. We denote this set by

$$_{Q}\mathrm{KL}\ W\mathrm{L} \qquad _{w,7\!\!7_n}$$

In our context each WI' is determined by either an observation study or a

sample investigated in a randomised study and the corresponding weight A_i is defined by

$$Ai = \frac{\text{sample size for } W_i^L}{\text{pooled sample size}}$$

An argument supporting that the weighting should be determined in this way was given in [9].

$00^{\text{l}-\text{L}}$, Wif.',) is a set of possible probability functions P. Under assumptions employed in [9] each P in this set is overwhelmingly more likely to correspond to true probabilities on sentences over L in the population from which the studies have taken samples than all possible probability functions outside this set given the observations from studies are expressed by WP, $_{,}W^{L}$ with the corresponding weights. However, the argument employed in [9] does not imply whether some probability functions in q^{L}(WP, , W4) are more likely than others. Fortunately, the resulting set of probability functions often contains only a single P in which case there is no need to sort this ambiguity out. Nevertheless, the failure of the linear entropy operator to provide us with a single probability function in general can be viewed as a major philosophical and practical flaw of this operator which will need to be seriously addressed in Section 3.4 to compute its result.

Another problem with the linear entropy operator is that it cannot satisfy any reasonable adaptation of the general locality principle as defined in [12] due to [12, Example 1]. In that paper an attractive alternative to the linear entropy operator was investigated which satisfies the general locality principle.

On the other hand, the linear entropy operator satisfies a string of other principles and many such principles were extensively discussed in [35]. Here we list only several principles that will turn out somewhat useful in this paper:

- (Strong Consistency Principle) If $\mathrm{fl}_{\,1} \mathrm{WL}\ 0$ then

$$\Theta^{\mathrm{KL}}_{\vec{\lambda}}(W_1^L, \ldots, \quad =_{\mathrm{i}=,}^{\mathrm{n}}$$

In other words, if studies are jointly consistent then every P that studies agree on is also produced by the linear entropy operator. In particular, $\mathrm{etP},(\mathrm{w}^{\mathrm{L}}) =$ This property in the case when the weighting is uniform is proved in [8] but it is fairly easy to modify the proof so that it works for any non—zero weighting.

- (Language Invariance Principle) Let L_1 and $L2$ be propositional languages such that $L_1 \subset L2$ and all the variables are independent. Consider WP^{1}, $_{,}$ Then

Note that we already know that WI' = $\mathbf{WL}^1$, 1 < i < *m,* by the definition. This principle means that if we extend our propositional language by new variables but do not supply any new constraints (this includes also constraints postulating that every sentence that is not logically false has some probability to be realised) then the probability functions P produced by the linear entropy operator in the simplex generated by $L2$ agrees with the probability functions produced by this operator in the simplex generated by L_1 on all sentences formed only over L_1. In other words, we can freely extend our language with independent variables without affecting the result of combining studies by the linear entropy operator.

- (Consistent Irrelevant Information Principle) Let L_1 and $L2$ be distinct propositional languages such that all the variables are independent. Consider W1L' $W4^{P}$ and V_i^{L2}, , $\mathbf{V4}^{2}$ such that $\mathrm{nm}_1 v_{\mathbf{I}}^{L2}$ $\mathbf{O}$. Then

$$\mathbf{ef}^{kL}(\mathrm{wi}^{r,uL2}{}_{\mathrm{nvi}}{}^{i\text{„}uL2},\ldots,{}_{\mathrm{wni}}{}^{I,luL2}{}_{\mathrm{ny}}{}^{k,uL2})\mathbf{L}, =$$

This principle is stronger than the language invariance principle. It says that we can even supply new constraints with new variables as long as the constraints are jointly consistent across all studies. A simple example of such constraints are those postulating that every sentence expressible in $L2$ (but not jointly in L_1 and L_2) that is not logically false has some probability to be realised.

The two properties above in the case when the weighting is uniform are proved in [35, Lemma 4.2.8 and Corollary 4.2.9] but it is fairly easy to modify the proof so that it works also for any non—zero weighting.

The principles above will prove to be useful in the following section where we show how we should choose a propositional language and knowledge bases for our problem of combining studies concerning the incidence of diagnosis of cancer in patients with unprovoked VTE. The principles will help us to determine what matters and what does not when constructing them. Should the technical notation of this section appear impenetrable for the reader we believe that it could be fully understood upon reading the following section.

3.3 Guide to representation

The following should be considered only as a guide, one of many possible ways of representing real studies in a mathematical model of closed convex sets in a probabilistic simplex. It is not possible to prove that a particular representation of reality is 'correct'. Instead, we need to proceed back and forth between the model and reality trying to improve our model in each stage.

While examining the validity of the representation neither of the following will be addressed:

1. The stability of the representation (e.g., if a small change in constraints can cause big change in the results).

2. Whether different researchers can apply it consistently.

On the other hand, although we do not investigate this rigorously, the presented guide should be more transparent than subjective systematic reviews of conflicting studies.

3.3.1 How to select propositional variables

Our first task when representing studies by closed convex sets in a probabilistic simplex is to construct a propositional language that will allow us to talk about relevant findings. Looking at Section 2.1 we quickly identify three main things we would like to be able to express in the case of a random patient with unprovoked VTE:

1. Cancer manifested within 12 months (CA).
2. Cancer detected during routine evaluation when unprovoked VTE was diagnosed (RE').
3. Cancer detected during extensive screening when unprovoked VTE was diagnosed (ES').

These variables are however not independent. Indeed, if cancer is detected during initial routine evaluation or extensive screening then cancer was surely manifested within 12 months so propositional sentences CA) A RE' and CA) A ES' are false. Recall that in such a case we should assign $P((\text{—}'\ \text{CA}) \wedge \text{RE}') = 0$ and $P((\text{—}'\ \text{CA}) \wedge \text{ES}') = 0$ and place these constraints to every knowledge base over a language containing the variables CA, RE' and ES'. However, this would cause some technical difficulties; for example, we would violate the assumptions adopted in [9], where an argument supporting the linear entropy operator was given, so we instead add the following constraints:

1. $P((\text{—}\ \text{CA}) \wedge \text{REc}) = \frac{8}{000}$ and
2. $P((\text{—}'\ \text{CA}) \wedge \text{ES}^c) = \frac{8}{000}$

Given our sample sizes, this is not going to influence the interpretation of results in any way. In general, any positive number less than one can be used but practically it really needs to be significantly smaller that the proportions that can be reported in studies. The particular selection above is chosen for technical reasons discussed in Section 3.4.

Looking at the studies [21, 23, 24, 25] the propositional variables above seem sufficient. However, the studies [20, 22, 26, 27] differentiate between routine evaluation and extensive screening giving suspicious results and cancer being actually detected. In other words, only some of these suspicious findings are find to be actual cancers but if cancer is not detected then the patient still can manifest cancer within a 12 month period. So we are compelled to add the following two variables to make the most of the studies:

1. Routine evaluation yielded suspicious results in respect to cancer when unprovoked VTE was diagnosed (RES).

2. Extensive screening yielded suspicious results in respect to cancer when unprovoked VTE was diagnosed (ESS).

Again, we have problems with dependence so we add the following constraints to every knowledge base:

1. $P((\text{-}\cdot \text{RE}') \text{ A } \text{RE}^{c}) = {}_{100000}$ and
2. $P((\text{---}.\text{ES}^{s}) \text{ A } \text{ESc}) = {}_{100^{s}000}$.

This is because it is not possible to detect cancer using for example routine evaluation if it does not yield suspicious results in the first place.

Now, a more thorough examination of the studies shows us that the nature of extensive screening quite varied across the studies. This could compel us to add new variables to differentiate between CT scans and other tests such as ultrasound and tumor markers. Checking further details, we see differences in how suspicious results were treated, follow—up organised and so on and so forth.

In practice we may sometimes eventually conclude that every study is different and tailor propositional variables L_i specifically for each study 'V which would result in a language $L = L_1 U \ldots U Ln_{,,,}$ $_{Lit}$ n $L_{i2} = 0$, and individual studies being then represented by closed convex sets WI' , , Wk— respectively. Looking at the consistent irrelevant information principle from Section 3.2.2 we can see that in such an extreme case the linear entropy operator performs no actual merging; if we restrict the language into the one used in a single study L_i then we obtain precisely the closed convex set that was used to encode it; i.e., $\mathbf{W}_{\cdot}$ In our case we would like to add the implicit knowledge that every logically satisfiable sentence over whole L has some probability to be realised. In such a case the consistent irrelevant information is not applicable but the strong consistency principle will give us a similar result because clearly studies cannot disagree if their constraints are formulated in different languages.

So, it is necessary to stop somewhere in the process of adding variables. Otherwise no merging is possible. This is, however, not something intrinsic to meta—analysis but a trait of probability in general. Consider that one suffers from an episode of unprovoked VTE. What is his or her chance that cancer will be diagnosed within one year? Studies suggest 10%. But he or she is young and studies also suggest that younger people have much lower chance. So it must be less than 10%, right? Well, the incidence is lower, but this was obtained from a smaller sample. And that person had also a history of cancer in family and has been exposed to considerable amount of dust and smoke. But, on the other hand, he or she does sport and eat vegan and so on and so forth and ultimately there is no sample to establish the chance that he or she develops cancer because probability cannot be assigned to an individual.

We need an abstract population where we generalise from individuals. But if we generalise too much then we may end up 'mixing apples and oranges' as combining studies with explained heterogeneity was often described in the

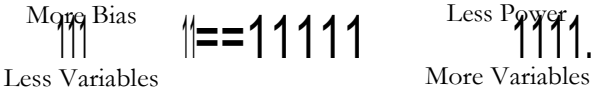

Figure 3.5: The illustration of the effect that the propositional language has on the bias and the power of a merging process.

literature. Our population here are studies and we need to abstract from details. Our studies appear sufficiently similar; extensive screening was after all treated as it is named in every study, something that should ultimately detect cancer if it is present, so in this particular case we feel comfortable ignoring some differences. But this is something that must be decided on a case—to—case basis. And this is not an easy task. We paraphrase Paris [37] who made a similar remark in a different context:

> `Most of us would surely prefer modes of reasoning which we could follow blindly without being required to make much effort, ideally no effort at all. Unfortunately, this guide to meta—analysis with unexplained heterogeneity is not one to follow blindly; it requires substantial effort to get meaningful results.'

To conclude, we trade precision of our propositional language for abstraction and power. Too fine language gives little power with not much merging performed and too rough language creates bias and wrong conclusions; see Figure 3.5 for an illustration.

Does our selection of propositional variables

$$L = \{\text{CA, REc, ESc, RE', ES'}\}$$

create bias? Of course, it has to be there. But in our particular selection of studies including variables concerning differences in extensive screening, the mean age, the length of the follow—up period and the exclusion of upper—extremity DVTs discussed in Section 3.1 would leave us with little power. Considering the treatment of extensive screening across studies as explained in Section 2.1 and looking at the analysis of variability identified across studies in Section 3.1 we are confident that the bias introduced by selecting the above propositional language is not going to significantly distort merging process. In practice it is necessary to perform such an analysis when constructing the propositional language.

3.3.2 How to assign probabilities consistently

In the previous section we have created and justified the propositional language $L = \{\text{CA, REe, ESC, RE', ESS}\}$. In this section we will explain in detail how we can assign to studies L—knowledge bases (see Section 3.2.1 for the definition of a knowledge base) by interpreting observations reported in studies as constraints

on possible probability functions P. There is no unique way of doing this; observations from studies can be often interpreted in many ways. But the more effort we put into it, the better the correspondence between our model in the form of knowledge bases and reality.

Now, notice that there are two types of studies:

1. observation; these report observations on a single sample and in our case they are [22, 25, 26, 27], and

2. randomised; these consider two samples: a control sample and an intervention sample and in our case they are [20, 21, 23, 24].

As we will shortly see, each type needs a separate treatment. In the following section we start with observation studies.

Probability and conditional probability

Let us have a look at the observation study [26].

> "Our final study population consisted of 40 patients who completed initial screening, FDG—PET/CT imaging, and follow—up. [...] During the initial screening evaluation, 16 patients (40%) had > 1 positive clinical finding concerning for underlying malignancy. [...] Twenty—five patients (62.5%) had > 1 abnormality on FDG—PET/CT suspicious for malignancy, [...] Of the 16 patients with positive clinical findings [...], 13 (81.2%) had one or more abnormalities on FDG—PET/CT requiring additional evaluation, [...], only one patient in this cohort was diagnosed with cancer (cancer incidence 2.5%, 95% CI: 0.6%-12.9%)."

In the sample there were 40 patients of which 16 had suspicious routine evaluation. Therefore

$$\mathrm{P(RES)} = \frac{16}{40}$$

So in general we assign probabilities as proportions; we take the number of patients with a trait over the overall number of patients we have investigated for the trait. The addition 'we have investigated for the trait' is important. $P(ESS \text{ A RES})$ because the trait 'having extensive screening suspicious' was here reported only for those who had routine evaluation suspicious. The correct interpretation of the observation from the study is the following conditional probability (see Section 3.2.1):

$$\mathrm{P(ES's\ RE^S)} = \frac{13}{16}$$

In general if N is the sample size in the study and a and b are sentences then by the Bayes rule

$$P(a|b) \quad \frac{P(a \wedge b)}{P(b)} \quad \frac{N_{a \wedge b}}{N_b} \quad \frac{N_{a \wedge b}}{N_b}$$

where $N_{a \wedge b}$ is the number of patients for which both sentences a and b are true and N_b is the number of patients for which the sentence b is true. In our case, $N = 40$, a = ESS, b = RES, Nam = 13 (for 13 patients both screening methods produced suspicious results) and $N_b = 16$ (16 patients had routine evaluation suspicious).

Note that the study in fact reported that no patient had cancer without having it diagnosed by extensive screening. Since for technical reasons we do not wish to assign zero probabilities we instead include the following constraint to express that having cancer and not being diagnosed by extensive screening is unlikely:

$$P(CA \quad ES`)) = \frac{8}{100\,000}$$

Similarly, no cancer was diagnosed by routine evaluation, hence

$$P(CA \wedge RE^c) = \frac{8}{100\,000}$$

Encoding all observations expressible in our language L into constraints together with implicit knowledge mentioned in the previous section we obtain the following L—knowledge base:

$$K_L = \{P(RE_s) = \frac{16}{40}, P(ES' \mid RE') = \frac{13}{16}, P(CA \wedge (\neg ES^c)) = \frac{8}{100\,000}$$

$$P(CA \wedge RE_c) = \frac{8}{100\,000} \quad P(ES_s) = \frac{25}{40}, P(ES') = \frac{1}{40}, P(CA) = \frac{1}{10^{-}},$$

$$P((\neg RE') \wedge RE_c) = \frac{8}{100\,000}, P((\neg ES_s) \wedge ES_c) = \frac{8}{100\,000}'$$

$$P((\neg CA) \wedge RE_c) = \frac{8}{100\,000}, P((\neg CA) \wedge ES_c) = \frac{8}{100\,000}'$$

$$P(`\text{logically satisfiable sentence}') > \frac{1}{100\,000}\ I'$$

The first two lines above were obtained from observations, the third and fourth lines express our implicit knowledge about the dependence of the variables and the last line is to express that every logically satisfiable sentence that we can construct over L can be true at least for one patient out of one hundred thousand. Recall that this last requirement is only technical so that for example the justification from [9] for the linear entropy operator works.

Our final step in constructing a knowledge base is to check its consistency. Normally, observation obtained from a single sample should be consistent, but there always could be errors and distortions due to patients lost during the study

period (in particular if there is a lengthy follow—up period). Furthermore, our requirement that every sentence must be assigned a non—zero probability could cause further problems. In the case above

$$P(CA \quad ES^e)) + P(CA \text{ A } ESc) = P(CA) = \frac{1}{40}'$$

$$P((— CA) \text{ A } ESC) + P(CA \text{ A } ESC) = P(ES\ ) = \frac{1}{To} \text{ and}$$

$$P((— CA) \text{ A } ESC) = \frac{8}{100\ 000}$$

yield also

$$P(CA \text{ A}(— ESc)) = \frac{8}{100\ 000}$$

so everything appears consistent. Checking consistency for many constraints however may be difficult. This should be performed using a computational software as we used in Section 3.4.

Randomised studies

Unlike the observation study above, randomised studies report proportions observed on two different samples and these can be easily inconsistent with each other. Inconsistent constraints considered together yield a knowledge base to which corresponds an empty set in the probabilistic simplex given by L. Such contradictions are expected when there are several different sources of information (e.g., more studies) and the purpose of the linear entropy operator is to interpret such conflicting reports if each report is individually consistent. If all observations across all studies are consistent then by the strong consistency principle the linear entropy operator produces a set of probability functions that corresponds to a knowledge base containing all those observations. So we do have internal consistency of our methodology. On the other hand. the linear entropy operator is not able to produce any results if it is given an inconsistent knowledge base as one of the inputs. We therefore must express observations obtained from control and intervention samples as separate knowledge bases. We will demonstrate this on the study [21].

> "Apparently cancer—free patients with acute idiopathic venous thromboembolism were randomized to either the strategy of extensive screening for occult cancer or to no further testing. [...] Of 201 patients, 99 were allocated to the extensive screening group and 102 to the control group. In 13 (13.1%) patients, the extensive screening identified occult cancer. In the extensive screening group, a single (1.0%) malignancy became apparent during follow—up, whereas in the control group a total of 10 (9.8%) malignancies became symptomatic [...]"

Looking carefully on the study we realise that all included patients went through routine evaluation without being diagnosed with cancer. Note that this on the other hand does not mean that routine evaluation was never suspicious, only that cancer was not proved. Interpreting the quote from the study we have in mind that every observation is a conditional probability subject to —; REc.

Patients in the intervention sample were extensively screened and the following knowledge base could be constructed from the observations:

$$\mathcal{A}_{T11,1} = \{P(CA\ REc) = \frac{14}{99}, P(CA\ A\ ESc\ RE') = \frac{13}{}$$

$$P((\text{—}REs)\ A\ REc) = \frac{8}{100\ 000\ p} \quad \text{– ESs) A ES}^{C}) = \frac{}{1008000'}$$

$$P((\text{–}\ CA)\ A\ REc) = \frac{8}{100\ 000\ pi(} \quad CA)\ A\ ESC) = \frac{8}{100\ 000}$$

$$P(\text{`logically satisfiable sentence'}) > \frac{1}{100\ 000\ I\ \bullet}$$

Note that $P(\text{—}\ REc)$ was not reported hence we cannot determine unconditional P(CA REc)). This is the reason why this study was not included in the analysis performed in Section 3.1 and it is a demonstration of the main advantage of the presented method over statistical (not Bayesian) meta—analysis; the present method has an ability to deal with complex knowledge.

Now we represent observations from the control sample:

$$1([21],0 = IP(c\ A\ HREc) = \frac{10}{102'}$$

$$P((\text{—}\ RE') \quad RE^{c}) = \frac{8}{100\ 000\ '} \quad P((\text{—}\ ESs)\ A\ ESC) = \frac{8}{100\ 000}$$

$$P((\text{—}\ CA)\ A\ RE') \text{—} \frac{8}{100\ 000} \quad P((\text{—}\ CA)\ A\ ESc) = \frac{8}{100\ 000}$$

$$P(\text{`logically satisfiable sentence'}) > \frac{1}{100\ 000}$$

Notice that these two knowledge bases above are jointly inconsistent: Constraints $P(CA(\text{—}\ RE') = \frac{ib}{15}$, and $PICA\ RE`) = 99$ cannot possibly hold at the same time. We therefore always split intervention and control samples.

Another randomised study is the study [24]:

> "[...] leaving 197 in each group for the modified intention—to—test analysis. After initial screening assessment, cancer was diagnosed in 11 (5.6%) patients in the 18F—FDG PET/CT group and four (2.0%) patients in the limited screening group (absolute risk difference 3.6%, 95% CI 0.4 to 7.9; $p = 0.07$). [...] One (0.5%) occult malignancy was detected in 186 patients who had negative initial screening in the 18F—FDG PET/CT group, compared with nine (4.7%) in 193 patients in the limited screening group (absolute risk difference 4.1%, 95% CI 0.8 to 8.4, $p = 0.01$)."

Patients in the intervention sample obtained both routine evaluation and 18F—FDG PET/CT while patients in the control sample obtained only routine evaluation (in the paper the term limited screening is used instead). We construct a knowledge base for the intervention sample:

$$\mathcal{A}_{A1,1} = \{P(REc \text{ V } ESe) = \frac{11}{197}, P(CA \text{ I } REc) \text{ A } ESC)) = \frac{}{186}$$

$$P((\text{-} RES) \text{ A } RE') = \frac{8}{100\,000\,'} \quad P((\text{-} ESs) \text{ A } ESc) — \frac{8}{100\,000\,'}$$

$$P((\text{-} CA) \text{ A } RE') = \frac{8}{100\,000} \quad P((—' CA) \text{ A } ES') — \frac{8}{100\,000\,'}$$

$$P(\text{`logically satisfiable sentence'}) > \frac{1}{100\,000\,\text{I}\,'}$$

and for the control sample:

$$K[a4],c = \{P(RE') = \frac{4}{197}, P(CA \quad RE') = 9$$

$$P((— RE') \text{ A } REe) = \frac{8}{100\,000} \quad P((— ESs) \text{ A } ESc) = \frac{8}{100\,000\,'}$$

$$P((— CA) \text{ A } REc) = \frac{8}{100\,000} \quad P((— CA) \text{ A } ESe) = \frac{8}{100\,000}$$

$$P(\text{`logically satisfiable sentence'}) > \frac{1}{100\,000\,\text{I}\,'}$$

Note that P(CA REc)) = P(CA REc))•P(-' RE') = P(CA RE9)• (1 — P(REc)) = th • (1 — th) = The constraint P(CA RE')) = would be in $K[$241 0 redundant. We do not have to include redundant constraints in knowledge bases, but we may do it (although a computational software checking, for example, consistency can reward us with a lower computational time should we do not include such constraints).

Also P(CA A RE^c) = P(RE^c) —$P((\neg$ CA) A RE) = 4 97 so P(CA) = P(CA A(—' REe)) +P(CA A REc) = . Therefore, including the constraint P(CA) = would lead to an inconsistent knowledge base. Nevertheless, one can easily interpret the quote above using this very constraint and there would be nothing wrong with that if one also manages to reassign other probabilities so that the resulting knowledge base is consistent. Only our way of interpreting the observations gives a rather small chance of one to ten thousand to the event that a patient has cancer detected by routine evaluation while he or she actually does not have it.

Sensitivity, specificity and positive predictive value

In medicine there are three main ways how to explain the usefulness of a screening strategy. These are its sensitivity, specificity and positive predictive value. We have already worked with a notion of sensitivity in Figures 3.3 and 3.4. The sensitivity of a screening strategy specifies how many cancers were identified by the strategy relative to the overall number of cancers.

For example, taking 4'$_{41\ O}$ considered above the sensitivity of routine evaluation is

$$P(\text{REc A RE' ICA}) = \frac{P(\text{CA A(REc A REs))}}{P(\text{CA})}$$

$$\frac{P(\text{RE'}) - P(\text{REc RES})) - P((\text{-- CA}) \text{ A (RE' A REs))}}{P(\text{CA})}$$

$$\frac{\frac{4}{197} - \frac{8}{100\,000} - \frac{4}{100\,000}}{\frac{13}{197} - \frac{8}{100\,000}} \approx \frac{4}{13}.$$

Now, looking at the randomised study [20], which was in detail explained in [29], in the control sample 7 of 21 cancers were discovered by routine evaluation which gives the sensitivity of routine evaluation

$$P(\text{RE' A REs CA}) = \frac{}{21}$$

In the intervention sample 12 of 30 cancers were discovered by routine evaluation thus

$$P(\text{REc A RE' CA}) = 12$$

Routine evaluation jointly with extensive screening detected 18 out of 30 cancers:

$$P((\text{RE' A RE'}) \text{ V } (\text{ES' A ES'})\text{ICA}) = 18$$

Compare this with reports in Figures 3.3 and 3.4.

Now let us have a look at a report from [29] that is detailed in Table 3.2.

	Sensitivity	Specificity	Posit. Predict. Value
Routine Eval.	19	479	19 / 1 9
Extensive Scr.	18	199 / 284	91

Table 3.2: The sensitivity, the specificity and the positive predictive value of routine evaluation and extensive screening as reported in [29, Table 2].

The specificity expresses the proportions of non—suspicious results to the number of patients without cancer and the positive predictive value expresses

the proportion of actual cancers among only suspicious results; see Table 3.3 for an interpretation of these quantities in our language.

	Sensitivity	Specificity	Posit. Predict. Value
Routine Eval.	P(RE' A RE' CA)	P(-RE[8]1- CA)	P(CA A RE' RE')
Extensive Scr.	P(ES' A ESS CA)	P(ES' H CA)	P(CA A ES' I ESS)

Table 3.3: Our interpretation of the sensitivity, the specificity and the positive predictive value of routine evaluation and extensive screening using the propositional language L.

Sensitivity	Specificity	Posit. Predict. Value
P((RE' RE')V V(ES' A ES')1 CA)	P((-,RE') A ES')H CA)	P((CA A RE' A RE')V v(CA A ES' A ESs)1 RES v ESS)

Table 3.4: Our interpretation of the sensitivity, the specificity and the positive predictive value of routine evaluation combined with extensive screening using the propositional language L.

There are two main problems with translating Table 3.2 into constraints from Table 3.3. The first problem is that the reports concerning routine evaluation are from the sample that combines both control and intervention groups. The second problem is that the reports regarding extensive screening are made after excluding patients having cancer discovered during routine evaluation. A more appropriate translation of the results are therefore the following constraints:

$$\mathrm{P}(\mathrm{ESc\ A\ } ESs \mathrm{\ ICA\ A(-\ REe))} = \tau_{3},\ P(\text{-}ESs1(\text{-}\ \mathrm{CA})\ \mathrm{A}\ (\ \text{-}\mathrm{RE}^{\circ})) = \frac{199}{284}$$

$$\mathrm{P(CA\ A\ ESC\ I\ ESS\ A(\neg\ REc))} = \frac{6}{91}$$

for the intervention sample and

$$\mathrm{P(REc\ A\ RE'\ ICA)} = \frac{7}{21'}\ \mathrm{P(CA\ A\ RE`\ I\ REs)} = \frac{}{62}$$

for the control sample (note that we are not able to determine any constraints for the specificity from the reported observations). Obviously, we have not obtained the latter constraints for the sensitivity and the positive predictive value from Table 3.2; we have obtained them from the following:

> "During the total study period a malignancy occurred in 21 of the 288 patients undergoing limited screening (7.3%), vs 30 out of 342 extensively screened patients (8.8%) [. ..1"

Furthermore, in [29, Table 1] it was reported that 62 out of 288 routine evaluations were suspicious. Finally, we form the following knowledge base.

$$K^{L}_{[201} = \{\text{P(RE' A RES ICA)} = \frac{7}{21}\ \text{P(CA A RE' I RE')} = \frac{}{2}$$

$$\text{P(CA)} = \frac{21}{288}\text{P(RE}^{5}\text{)} = \frac{62}{288}$$

$$\text{P((-r RE') A RE')} = \frac{8}{100\,000}\ P((\text{- ESs) A ES')} = \frac{8}{100\,000'}$$

$$P((\text{--' CA) A RE')} = \frac{8}{100\,000}\ P((\text{-, CA) A ES')} = \frac{8}{100\,000}$$

$$\text{P(`logically satisfiable sentence')} > \frac{1}{100\,000\ \}}$$

In [29, Table 1] it was also reported that 57 out of 342 routine evaluations were suspicious in the intervention sample. Finally, we form the following knowledge base.

$$K^{L}_{[201,I}\{\text{P(ES' A ESs ICA A(-' REc))} = \frac{}{18}$$

$$\text{P(CA A ES' 1ES' A(-RED))}\qquad \text{P(REc A RE' ICA)} = \frac{12}{}$$

$$\text{P(CA A RE' I RE')} = \frac{12}{77}, \text{P(CA)} = \frac{30}{42\ '}\ \text{P(RES)} = \frac{57}{342}$$

$$\text{P((REs) A REc)} = \frac{8}{100\,000}\ P((\text{- ESs) A ESe)} = \frac{8}{100\,000'}$$

$$P((\text{-, CA) A RE')} = \frac{8}{100\,000}\ P((\text{¯ CA) A ES}^{c}\text{)} = \frac{8}{100\,000}$$

$$\text{P(`logically satisfiable sentence')} \quad \frac{1}{100\,000}$$

12 patients in the intervention sample had cancer discovered during routine evaluation. Considering that 342 - 12 = 330 patients were supposed to be extensively screened the constraint for specificity $P($- ES' CA) A (-' REc)) = $\frac{}{284}$ but also some other constraints that we can construct from the reports in the paper such as P(ESs REc) = $\frac{91}{302}$ and P(ESc H REc) = $\frac{}{1}$g7, do not appear to make sense and in fact they are inconsistent with the knowledge base above. These value are due to the fact that only 302 patients actually went for extensive screening procedures and the corresponding proportions were reported in the study. In practice we shall always expect that patients are lost in particular during follow-up and we investigate how to address such an issue in the following section.

Danger of Inconsistencies	Less Power
Keeping constraints together	Splitting contraints

Figure 3.6: The illustration of the effect that splitting constraints into more knowledge bases has on the power of a representation.

Adjusting constraints

If there are only a few patients lost during the study period (i.e., there are only few such patients in respect to the overall sample size) then we may safely ignore such lose and we in fact ignored this in all previously mentioned studies but [20]. However, 28 lost patients in [20] appear significant in a sample of 330 patients.

The down-side of adjusting constraints for missing patients is the danger of creating inconsistencies so one would perhaps suggest splitting one knowledge base into two. After all, we always split the control and intervention samples for the very same reason. We need to however realise that splitting constraints into several knowledge bases is not desirable at all as it decreases their power. It introduces ambiguity where it was not in the first place. See Figure 3.6 for an illustration.

In the following we adjust the randomised study [23] for missing patients:

> "A total of 14 patients (3.9%; 95% CI, 1.9 to 5.4) in the limited-screening group [431 patients] and 19 patients (4.5%; 95% CI, 2.9 to 6.9) in the limited-screening-plus-CT group [423 patients] received a diagnosis of occult cancer [in the interval between randomisation and the 1-year follow-up]. [...] In the primary outcome analysis, 4 of 14 occult cancers (29%; 95% CI, 8 to 58) were missed by the limited screening strategy (i.e., cancer was diagnosed after the screening strategy had deemed the patient as being free from cancer and before the end of the 1-year follow-up period), whereas 5 of 19 occult cancers (26%; 95% CI, 9 to 51) were missed by the strategy of limited screening plus CT."

In the control sample only 4 patients did not receive any actual screening and more significantly 17 patients discontinued follow-up. We consider the amount of missing patents rather insignificant and therefore we encode the above by the following knowledge base.

$$K^{L}_{[2}31,0 = \{P(CA) = \frac{14}{431}, P(Rff\ A\ RE'\ I\ CA) = \frac{10}{T4},$$

$$P((\text{- REs})\ A\ REc) = \frac{8}{100\,000}, \text{-ESs})\ A\ ESe) \text{—} \frac{8}{10\,000'}$$

$$P((\text{- CA})\ A\ RE') = \frac{8}{100\,000}, \text{CA})\ A\ ES") = \frac{8}{10\,000'}$$

$$\text{P(`logically satisfiable sentence')} > \frac{1}{100\,000}$$

On the other hand, 33 patients in the intervention sample did not receive the envisaged screening and 15 were lost during follow—up. Having 33 patients who missed the screening appears more significant and we adopt the following.

$$1\text{Q}_{314} = P(\text{CA}) \quad \frac{19}{423}, \text{P(REc V ES')} = \frac{14}{423 - 33'}$$

$$P((—, \text{RE')} \text{ ARE')} - \frac{8}{100\,000} \quad P((\neg \text{ ESs) A ESc)} = \frac{8}{100\,000'}$$

$$P((\text{--}' \text{ CA) A RE')} — \frac{8}{100\,000} \quad \text{CA) A ES')} = \frac{8}{100\,000'}$$

$$\text{P(`logically satisfiable sentence')} \quad \frac{1}{100\,000}$$

Since we do not allow zero probabilities we actually do not allow to assign the probability 1 to any sentence that can be false. In the following observation study [27] the sensitivity of extensive screening was reported as 1; however, we need to adjust this value slightly down for a technical reason of avoiding zeros.

> "Fifty patients were included. [...] In 22 (44%) patients, PET—CT showed increased FDG—uptake suspicious for malignancy [.. .1. After additional procedures, malignancy was confirmed in 54.5% ($1^2/_22$) of cases. [...] In 45.5% (10/22) of cases, additional evaluation [...] discarded the presence of an occult cancer. [...] Twenty—eight patients (56%) had a negative PET—CT at VTE diagnosis. Five patients were lost to follow—up. Among the 23 other patients, none had developed malignancy at the end of follow—up."

The sensitivity of extensive screening P(ESc A ESS ICA) is reported as iz but this would mean that P((ESC A ESS) A CA) = P(CA) which in turn gives P((—.(ES'AESs))ACA) = 0. Therefore, in order to create a consistent knowledge base we represent this observation as P(CA) — P((ES' A ESS) A CA) = lool000

$$\text{If}(_{271,0} = \{\text{P(CA)} — \text{P((ES' A ES') A CA)} = \frac{12}{100000}'$$

$$\text{P(CA A ES' I ES')} = \frac{12}{22} \text{ P(ESS)} = \frac{22}{}$$

$$P((\text{--}' \text{RES) A RE')} = \frac{8}{100\,000} \quad P((— \text{ ES}^{\text{S}}) \text{ A ES')} = \frac{8}{100\,000}$$

$$P((\text{--}' \text{ CA) A RE')} = \frac{8}{100\,000}' P((— \text{ CA) A ES')} = \frac{8}{100\,000}$$

$$\text{P(`logically satisfiable sentence')} > \frac{1}{100\,000}$$

The knowledge base above gives us the following statement about the sensitivity of extensive screening:

$$P(ESc\ A\ ESS\ CA) = \frac{P((ES'\ A\ ESs)\ A\ CA)}{P(CA)} \quad \frac{P(CA)}{P(CA)} \quad \frac{12}{100000}$$

Note that adding P(CA) = 0, would lead to an inconsistent knowledge base: P(CA A ES' ESS) = 1 and P(ESS) = 0 gives P(CA A(ESc A ESS)) = 0 while P(CA) > P((CA A ES') n ESS)).

Conclusion

In Section 3.3.2 we have learnt couple of useful guidelines:

- Distinguish between probability and conditional probability.
- Split constraints in randomised studies to those concerning the intervention sample and those concerning the control sample.
- Do not split constraints into more knowledge bases if not necessary.
- Interpret the sensitivity, the specificity and the positive predictive value of a screening method.
- Read through a whole paper not just through its abstract.

We apply them to represent the two remaining observation studies [22, 25] considered in this paper.

First, the study [25] observed the following:

> "Fifty patients [...] were included. One patient was diagnosed [with cancer] at inclusion and cancers were found in three other patients during the follow—up period [...]."

Our interpretation is the knowledge base

$$= \{P(CA) = \quad P(REc\ V\ ES') = \frac{}{573},$$

$$P((\text{—}'\ RES)\ A\ RE') = \frac{8}{100\,000} \qquad ESS)\ A\ ES') = \frac{8}{100\,000}$$

$$P((\text{—}'\ CA)\ A\ RE') = \frac{8}{100\,000},\ P((\text{---}'\ CA)\ A\ ES') = \frac{8}{100\,000}$$

$$P('Iogically\ \text{satisfiable sentence}') \ ?\ \frac{1}{100\,000}$$

Second, the study [22] observed the following:

> " [...] 345 patients considered to have idiopathic VTE. [...] Ninety—two of [patients with suspicious findings on routine evaluation] had an idiopathic thrombosis. [...] In patients with idiopathic venous thromboembolism [the frequencies of malignancy not detected by routine evaluation] were 3.14% (six of 191) [in patients who were younger than 70 years] and 9.30% (12 of 129) [in patients who were above 70 years] [...]."

Furthermore, [22, Table 3] reported that there were 9 patients whose cancer was missed during both screenings but they were diagnosed before the end of the follow—up period.

Note that we had to look at the actual paper and not just at the abstract since they investigated both patients with unprovoked VTE and patients with provoked VTE but the abstract reports only proportions on the combined population.

We interpret the above by

$$= \{P(RES) = \frac{92}{145}, P(CA \text{ A } RE' \text{ I } RE') = \frac{25}{}$$

$$P(CA) = 945' \; P(CA \text{ A } ES' \text{ H } RE') = \frac{}{320'}$$

$$P((— RE') \text{ A } RED) = \frac{8}{100\,000'} \quad P((\text{-} ESs) \text{ A } ES^c) = \frac{8}{100\,000'}$$

$$P((\text{--} CA) \text{ A } RE') = \frac{8}{100\,000} \quad P((—r CA) \text{ A } ES') = \frac{8}{100\,000'}$$

$$P(\text{`logically satisfiable sentence'}) > \frac{}{100^1000 \text{ I}}$$

Note that the above gives that the sensitivity of routine evaluation is

$$P(REc \text{ A } RE' \text{ I } CA) = \frac{P((RE' \text{ A } RE') \text{ A } CA)}{P(CA)}$$

$$\frac{P(CA \text{ A } RE' \mid REs) \cdot P(RES)}{P(CA)} \quad \frac{E \cdot a}{\frac{43}{345}} \quad \frac{25}{43}$$

This concludes our construction of knowledge bases and in the next section we will show how to assign a weighting to knowledge bases.

3.3.3 How to assign weighting

In the previous section we have created the following knowledge bases, each created from a sample investigated in a study (the corresponding sample size is in the parentheses): $K_{[201/}$ (342), K(2'0],c, (288), $K_{[z1],1}$(99), $K_{[\text{-/<-}\ 1]\ c;}$ (102), $K_{[2.2]o}$ (345), $KEL_{,31,i}$ (423), $K_{,31\ o}$ (431), $K_{,41.i}$ (197), $If(_{241,0}$ (197), $KI_{[251,0}$ (50), $./<_{51\ o}$ (40), $./f_{71\ o}$ (50). As we have explained in Section 3.2.1 a non—empty closed convex set in a probabilistic simplex corresponds to every consistent knowledge base. Let us denote by WI' the set corresponding to the sample 'i'. Given several non—empty closed convex sets, the linear entropy operator was justified for the purpose of combining them if we deal with unexplained heterogeneity under specific assumptions; see Section 3.2.2. The resulting closed convex set is denoted by $\text{-'A}\ {}^{t}w_{[20}{}_{]}{}_{I'}\cdot\ .{}^{5}w^{L}{}_{[27LO}$), where the weighting A is given by

$$A_i = \frac{\text{sample size for WL}}{\text{pooled sample size}}$$

For example,

$$A[2(1j -= \frac{\text{sample size for } \underline{WL2011}}{\text{pooled sample size}} \quad 2\,564'$$

Therefore, in the merging process, closed convex sets obtained from larger samples have also bigger weights in respect to closed convex sets obtained from smaller samples. Working with the weights obtained from sample sizes in the context of unexplained heterogeneity was supported in [9]. Apart from the corresponding sample size we do not otherwise judge the quality of a study. This could be perhaps considered as a weakness of the presented representation as our only choice here is to not include flawed studies at all.

Furthermore, the closed convex sets were obtained from knowledge bases and constraints in any given knowledge base do not appear equally strong. For example, the sensitivity of routine evaluation was obtained only by looking at the patients having cancer and these comprised only a small fraction of the overall sample size. This is not problematic for the result obtained in [9] due to the assumption that a sufficiently large sample was observed. Unfortunately, it is questionable whether this is the case for some studies included here. In general, analysis proposed here should be performed only if all constraints included in knowledge bases were observed on large samples. In the other case a different method for combining them could be more appropriate.

3.4 Roundabout computation

3.4.1 Algorithm

In the previous section, we have identified 12 knowledge bases each of which determines a non—empty closed convex set of L—probability functions. We denote those $W^L_{[2OLP}$ $WL\ wL\ wL\ wL\ wL\ wL\ wL\ wL$ $[20],C^5$ " $[211,1^5$ " $[21],C^5$ " $[221,05$ " $[23],I^5$ " $[23],C^5$ " $[24],P$ " $[24],C^7$ '' $[251,0'$ W[8] o' '' $[271]O$ in this order. As we have mentioned previously, the use of the linear entropy operator for meta—analysis with unexplained heterogeneity was justified and this operator was shown to have nice properties. Thus far, we have not explicitly defined this operator, but we will do so in this section as we need such a definition to actually compute the result.

First, by the disjunctive normal form theorem, the value of a probability function on any sentence is determined by its values on special atomic sentences [11]. There are two to the number of variables in L of such sentences so in our case we have $2^5 = 32$ atomic sentences. If we list them in some fixed order, say $a_1, \ldots, a_{32}$, then we may identify an L—probability function P with the vector of its values on these atomic sentences: $(P(a_i), \ldots, P(a_{32}))$. Note that $P(a_i) \in [0, 1]$, $1 < j < 32$, by the definition and $\sum_{i=1} P(a_i) = 1$ follows from

the probabilistic rules that P satisfies. So the set of all L—probability functions is the 31 dimensional simplex.

Now consider two L—probability functions P and Q in our simplex. The Kullback—Leibler divergence (an asymmetric distance) from P to Q is the following real number

$$KL(Q \,||\, P) := \sum_{j=1}^{32} Q(a_j) \log \frac{Q(a_j)}{P(a_j)}.$$

What the linear entropy operator here in effect does is to create a set of L—probability functions P that minimise the following weighted sum of Kullback—Leibler divergences

$$\sum_{i=1}^{12} KL(Q^{(i)} \,||\, P)$$

subject to $Q^{(i)} \in W[2a],I, \ldots, n(12)$ ω_L " [27] 0 where the weighting factors

$$A, — \frac{\text{sample size for WI'}}{\text{pooled sample size}}$$

were identified in Section 3.3.3. We denote this set by $(\omega^L_{[2\,01,1^7} \ldots, W[27],0)$.

However, no method that could feasibly determine this set explicitly is known to us. On the other hand, it is only a simple generalisation of the famous alternating projective procedure due to Csiszar and TusnLly [38] performed in the literature many times and with various degrees of generality [13, 14, 39, 40] that the following procedure converges to a point in $001^{(L}_{A\,x}(w_{[20]\,I^7} \bullet " \ w_{[27]\,0})$:

Let P_o be an arbitrary L-probability function with no zero coordinates. We define recursively a sequence $\{Pk\}ZL_0$ by the following:

1. set $Q^{(i)} = \arg\min_{c2(i)\,Ewi} KL(Q^{(i)}|A)$ and
2. define $Pk\pm i(a_i) = \_\_Ai \bullet Q^{(i)}(\%)$ for all $1 < j < 32$.

Now,

$$\lim_{k\to\infty} Pk \in "A \setminus {}^{w\,L}_{[20],I7\cdots,\ {}^{L}_{v\,[27],}})$$

provides us with a means to compute a point inside the set; unfortunately, no explicit characterisation of such a point is known to us.

Fortunately, there is a roundabout way. Let $1 > r > 0$ be a fixed real number. Let .// be an arbitrary L-probability function with no zero coordinates. We define recursively a sequence $\{P1,\bullet\}Z_{=O}$ by the following:

1. set $Q^{(i)} = \arg\min_{Q(i)\,E\,WI.} {}^{KL(Q(i)}{}_{Pic.)}$ and
2. define $PT_{+,}(a_j) = r \bullet 3 (1 — r) \bullet E_{i-1}^{12} A, \bullet Q(i)(a_j)$ for all $1 < j < 32$.

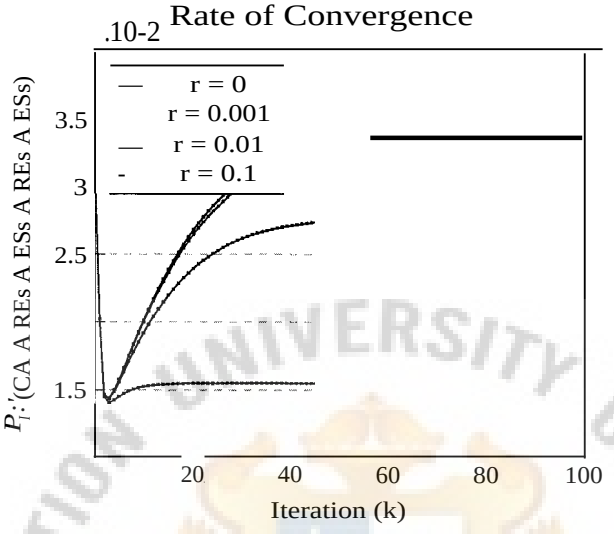

Figure 3.7: The behaviour of TY_{ir}(CA A RE' A ESC A RES A ESS) for several fixed r and as we change $k = 0, \ldots, 99$.

What we have in effect done above was to bias the alternating projective procedure towards the uniform probability function and we may decrease the level of bias by lowering the value of the parameter r. In particular $P_i^{\circ} = Pk$. Now, it is a consequence of a modification of the chairman theorem due to Wilmers [12] performed in [13, 14] that

$$\lim_{r \to 0} \lim_{k \to \infty} Pk = \text{cmoc}^{\text{iclICL}}_{k \to A'}(TAT L_{[20]}, \ldots, w L_{[271,0)})$$

see [11] for a definition of the CM' operator, which selects a unique probability function in a closed convex set of probability functions.

In general it does not hold that $\lim_{r \to 0} \lim_{k \to \infty} Pk = \lim_{k \to \infty} \lim_{r \to 0} Pk$ (see [13, 14]) but as seen in Figure 3.7 this is true in our situation. A possible and likely explanation is that the set $9^{1.1}(w_{[20],1}, \ldots, w_{[27]\ 0})$ in fact contains only a single probability function. We will therefore simply use the alternating projective procedure to compute a probability function P to represent our selection of studies.

3.4.2 Implementation

The algorithm specified in Section 3.4.1 was implemented by using the tools specified in Table 3.5. In this section we explore some technical issues that we have encountered during the implementation.

	Tool
Processor	Intel Core i5-4690 CPU only one core (thread) was used
Compiler	Microsoft Visual Studio 2015 Community Version 14.0.25123.00 Update 2
Programming Language	C++
Optimisation Package	Alglib 3.10.0, Free Edition Licence: GPL 2+, Reference: [41]

Table 3.5: Tools used to perform the computation.

First, the only computationally complex task in the algorithm specified in Section 3.4.1 is to compute the unique $Q^{(i)} = \arg\min_{Q(\cdot)_{Ew}i,} KL(Q(i)1/A)$ for every sample 'i'. Since the Kullback—Leibler divergence is a strictly convex function in its first argument and each set W_i^L is determined by several linear equalities and inequalities this is a problem of linear convex optimisation. Alglib, more precisely its sub—package Minbleic, employs the gradient of the Kullback—Leibler divergence (in our case analytically obtained) to minimise this divergence and it handles equality constraints by linear projections and inequality constraints by the method of activation and deactivation (in our case we do not have any genuine inequality constraints; we have only boundary conditions). More details can be found in [41].

Since for technical reasons, we do not allow to assign zero probabilities to sentences (otherwise the algorithm from Section 3.4.1 would not work), our variables were bounded by small constants. This in effect could cause some of our variables to vary in magnitudes that are much smaller than those of other variables. This is a major problem for optimisation. There are two methods to remedy this:

1. We may select constraints in such a way that all variables varying in small magnitudes are in fact fixed to unique numbers.

2. We may inform Minbleic how each variable should be scaled.

Both these methods were employed here.

Minbleic automatically checks whether the constraints in every knowledge base are consistent. We have obtained no indication of inconsistency.

There are several stopping conditions available in Minbleic. We have tested that EpsG (gradient change) smaller than 0.000 000 01 does not lead to any improvement in minimising the divergence. Also EpsX (step change) smaller than 0.000 000 1 does not lead to any improvement.

Second, the question how many iterations for the parameter k are needed was necessary to be answered. From Figures 3.8 and 3.9 it is apparent that preforming 1 000 iterations is sufficient. These computations took 359 seconds and 243 seconds respectively.

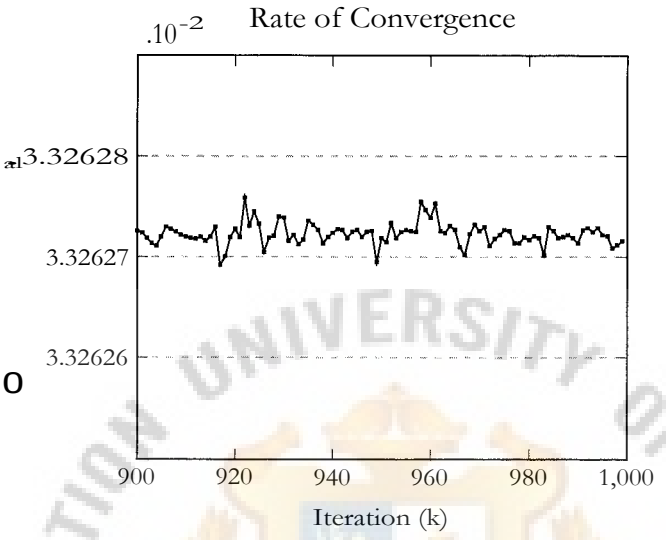

Figure 3.8: The behaviour of $P_{\mathbf{r}}$(CA A REc A ESc A RE' A ESs) for fixed $\mathbf{r}$ = 0.001 and as we change k = 900, ... , 999.

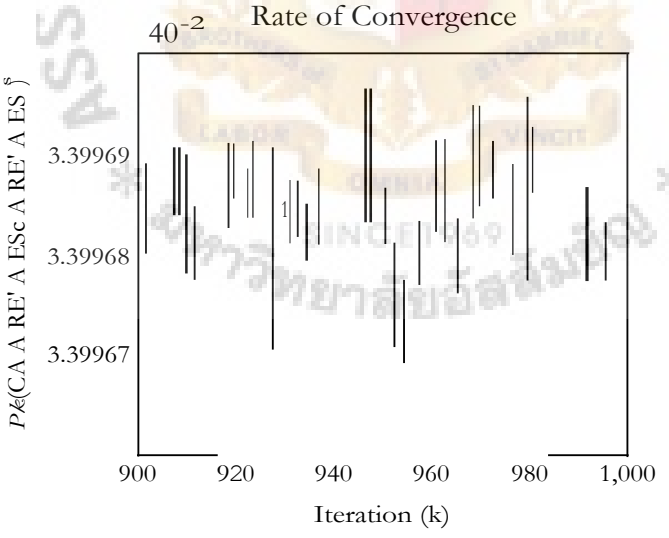

Figure 3.9: The behaviour of P_k(CA A REc A ESe A RE' A ESS) and as we change k = 900, ... , 999.

ASSUMPTION UNIVERSITY OF THAILAND
SINCE1969

Chapter 4

Result and Discussion

4.1 Result

In this paper we have developed a special non—statistical technique for meta—analysis with unexplained heterogeneity and we have applied it to the problem of determining the incidence of diagnosis of cancer in patients with unprovoked venous thromboembolism (VTE); see Section 2.1 for a detailed explanation of this condition. The new method was needed as traditional fixed—effect meta—analysis could not be applied due statistically detected heterogeneity; see Section 3.1. Furthermore, any non—Bayesian meta—analysis would not be able to deal with complex knowledge. In Section 3.3 we have managed to interpret complex findings reported in the studies [20, 21, 22, 23, 24, 25, 26, 27]. For comparison, statistical techniques allow only observations that were made in Section 3.1. The initial stage of our method offers much flexibility in interpreting the findings and further improvements are certainly possible but the subsequent application of the linear entropy operator was justified in [9] as optimal under assumptions specified in the introduction. One of the assumptions however states that the reported proportions were observed on sufficiently large samples. The pooled sample size of the selection of studies considered here is 2 564 patients which appears sufficient; however, many proportions were reported on rather small samples as noted in Section 3.3.3. The following results should therefore be treated as our best guess based on the information on hand:

The one—year incidence of diagnosis of cancer in patients with unprovoked VTE is 7.78%. This is significantly lower than 10% (95% confidence interval 8.6 — 11.3) reported in the previous meta—analysis [19]; see Figure 4.1 for comparison. Note that our method does not produce confidence intervals. This is because if all assumptions mentioned in the introduction are satisfied (although never the case in practice) then our method employing the linear entropy operator yields the overwhelmingly the most likely proportions; see [9]. Although our result suggests that the incidence is lower than 10% it is still much higher than in the normal population so it still appears reasonable to screen for cancer

in patients when they are diagnosed with unprovoked VTE. We will analyse the reliability of this result in the following section.

Which screening method for cancer should be used when unprovoked VTE is diagnosed can be determined by Table 4.1 where the sensitivity, the specificity and the positive predictive value (see Tables 3.3 and 3.4 for explicit definitions) of routine evaluation, extensive screening and combined screening are reported; see Section 2.1 for a detailed description of these screening strategies but do note that these are abstracted notions and the definition of the strategies varied from study to study. Our results should be therefore considered as a general guidance whether routine methods are sufficient or if including more extensive methods is warranted.

Despite the impression made by some individual studies included in our analysis that the sensitivity of 69.7% (95% confidence interval 61.1-77.8) for screening combining routine evaluation with extensive screening reported in the meta—analysis [19] is too high, our result that the sensitivity of combined screening is 77.76% indicates even higher sensitivity that the one established in [19].

Our results also indicate that extensive screening has much better detection rate than routine evaluation which detects only 43.77% of cancers. This roughly supports the original findings of [19] that routine evaluation has the sensitivity of 49.4% (95% confidence interval 40.2-58.5).

	Sensitivity	Specificity	Posit. Predict. Value
Routine Eval.	43.77%	81.55%	15.63%
Extensive Scr.	77.72%	68.45%	17.19%
Combined Scr.	77.78%	63.29%	15.15%

Table 4.1: Our results based on a new non—statistical analysis of several studies concerning effectiveness of different screening strategies for cancer upon diagnosis of unprovoked VTE.

4.2 Discussion

In this section we investigate how reliable our results are, namely whether exclusion of a single study from our collection [20, 21, 22, 23, 24, 25, 26, 27] has a significant effect on the results reported in the previous section:

The incidence of diagnosis of cancer in patients with unprovoked VTE computed in such a manner varied from 6.97% to 9.79%. This means that we cannot reliably claim that the incidence is in fact lower than the one reported in the previous meta—analysis [19]. Such a claim would depend on the inclusion of the study [23] and it is worth noting that this study investigated patients with the lowest average age among all studies considered here.

The variability of our results concerning effectiveness of different screening strategies is in Table 4.2. The sensitivity, the specificity and the positive predic-

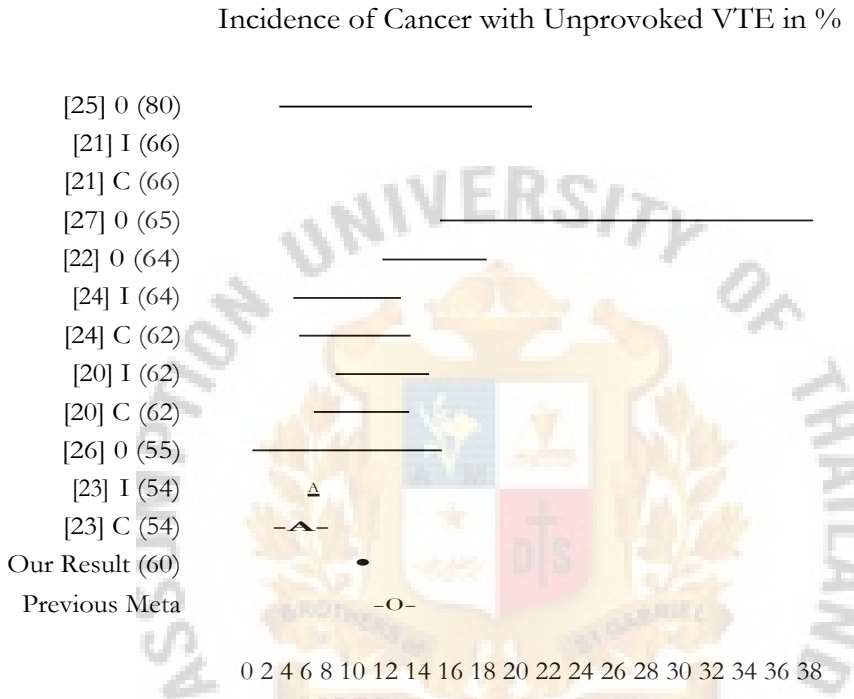

Figure 4.1: The Forest plot of our selection of studies (in blue), our interpretation of them (in red) and the previous meta—analysis [19] (in black). The x—axis denotes the percentage of people that had cancer discovered within a 12 months period (studies [20, 21] and [24] had a longer follow—up period) after the occurrence of VTE. Note that the study [21] did not report this proportion explicitly. This is the main reason why our result is different from the pooled value reported in Figure 3.2. 95% Clopper—Pearson confidence intervals are used but note that our non—statistical method does not produce confidence intervals. On the y—axis the studies are arranged by the mean age of patients (the value in round brackets). The letter '0' indicates an observation study, the letter 'I' indicates the incidence in the group that received the intervention (extensive screening) and the letter 'C' indicates the incidence in the control group.

tive value of routine evaluation vary symmetrically around our results 43.77%, 81.55% and 15.63% respectively in magnitudes that appear to indicate that our results are reliable.

	Sensitivity	Specificity	Posit. Predict. Value
Routine Eval.	36.59 — 49.61%	76.24 — 83.67%	12.69 — 18.07%
Extensive Ser.	47.11 — 83.20%	58.58 — 73.48%	10.22 — 20.89%
Combined Scr.	74.99 — 83.25%	52.54 — 64.07%	12.58 — 18.51%

Table 4.2: The variability of our results when a single study is removed concerning effectiveness of different screening strategies.

On the other hand the sensitivity, the specificity and the positive predictive value of extensive screening vary strongly towards lower values from established 77.72%, 68.45% and 17.19% respectively in rather high magnitudes. This variation was predominantly achieved by excluding a study that investigated mere 50 patients. In particular the sensitivity of extensive screening could be in fact as low as the sensitivity of routine evaluation alone. More research is needed to establish this value.

Finally, the sensitivity of combined screening of 77.78% appears reliable. Combined screening is much more sensitive than routine evaluation alone. The specificity of combined screening vary asymmetrically. This value could be smaller than established 63.29%. There does not appear to be a significant problem with the positive predictive value established as 15.15%.

Excluding a single study proved to be a useful tool and one would perhaps suggest to exclude more studies using some criteria in an attempt to perform subgroup—analysis. In general in meta—analysis this was shown to be a problematic strategy that reduces the power of the analysis and thus creates results only by chance [2]. More studies being included in the analysis here would be preferable.

Chapter 5

Conclusion and Recommendation

Our analysis indicates that 7.78% of patients with unprovoked venous thromboembolism are diagnosed with cancer within one year. This result is however sensitive to the average age of the patients included in the studies. We are not able to reliably claim that the incidence is lower than 10% as reported in [19]; this could be simply because our pooled sample has a lower average age (of 60 years).

Our results concerning routine evaluation and combined screening for cancer upon diagnosis of unprovoked VTE appear reliable and support the finding of the previous meta—analysis [19] that combined screening employing both routine and extensive screening techniques has a much better occult—cancer detection rate than routine evaluation alone. The difference in the detection rate according to our analysis is 77.78% versus 43.77%. Despite the fact that extensive screening is expensive, it could be traumatic for patients and there is also the danger of radiation induced cancer it may be worth to consider to employ it jointly with routine evaluation.

Finally, it is an open problem whether extensive screening without routine evaluation detects significantly more occult cancers in patients with unprovoked VTE than routine evaluation. Although combined screening has much better sensitivity, this could be because each technique detects cancer in different patients. The fact that extensive screening alone under—performs was indicated in some studies included in our analysis (e.g., [20]) nevertheless such findings are usually obscured as routine evaluation is performed first and patients having cancer discovered in this stage are removed. Further research on the sensitivity of extensive screening is needed.

ASSUMPTION UNIVERSITY OF THAILAND
SINCE1969

Bibliography

[1] NHS, Clinical trials and medical research — collecting the evidence, http://www.nhs.uk/Conditions/clinical-trials/Pages/puttingittogether.aspx, [Online; accessed 22-December-2015] (2015).

[2] M. Petticrew, H. Roberts, Systematic reviews in the social sciences: a practical guide, Blackwell Publishing Ltd, Oxford, 2006.

[3] Cochrane handbook for systematic reviews of interventions 4.2.6 [updated September 2006], in: J. P. T. Higgins, S. Green (Eds.), The Cochrane Library, Issue 4, John Wiley and Sons Ltd, 2006, pp. 1-257.

[4] D. Sharpe, Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away, Clinical Psychology Review 17 (1997) 881-901.

[5] Cochrane handbook for systematic reviews of interventions 5.1.0 [updated March 2011], in: J. P. T. Higgins, S. Green (Eds.), The Cochrane Collaboration, 2011.

[6] B. Bushman, G. Wells, Narrative impressions of literature: The availability bias and the corrective properties of meta-analytic approaches, Personality and Social Psychology Bulletin 27 (2001) 1123-1130.

[7] M. Adamdk, A common point of disjoint closed convex sets and a prototype application in medical reasoning, in: S. Dhompongsa, N. Petrot, S. Plubtieng, S. Suantai (Eds.), Proceedings of the 9th International Conference on Nonlinear Analysis and Convex Analysis, Yokohama publishers, 2016, pp. 1-18.

[8] G. M. Wilmers, M. Adamdk, Probabilistic merging operators, Logique et Analyse 228 (2014) 563-590.

[9] M. Adamdk, On the applicability of the 'number of possible states' argument in multi-expert reasoning, Journal of Applied Logic 19 (2016) 20-49. doi:10.1016/j.ja1.2016.10.001.

[10] J. B. Paris, A. Vencovska, On the applicability of maximum entropy to inexact reasoning, International Journal of Approximate Reasoning 3 (1989) 1-34.

[11] J. B. Paris, The uncertain reasoner companion, Cambridge University Press, Cambridge, 1994.

[12] G. M. Wilmers, A foundational approach to generalising the maximum entropy inference process to the multi—agent context, Entropy 17 (2015) 594-645.

[13] M. Adameik, The information geometry of Bregman divergences and some applications in multi—expert reasoning, Entropy 16 (2014) 6338-6381.

[14] M. Adamefk, Corrections on Adamefk, M. The information geometry of Bregman divergences and some applications in multi—expert reasoning, Entropy 2014, 16; 6338-6381 (March 2016). doi : 10 . 13140/RG . 2 . 1 . 2289 . 3201.

[15] H. V. Joffe, S. Z. Goldhaber, Upper—extremity deep vein thrombosis, Circulation 106 (2002) 1874-1880.

[16] S. Z. Goldhaber, Pulmonary embolism thrombolysis: A clarion call for international collaboration, Journal of the American College of Cardiology 19 (1992) 246-247.

[17] E. Previtali, P. Bucciarelli, S. M. Passamonti, I. Martinelli, Risk factors for venous and arterial thrombosis, Blood Transfus 9 (2011) 120-138.

[18] A. Trousseau, Phlegmasia alba dolens, Clinique Medicale de l'Hotel Dieu de Paris 3 (1872) 654-712.

[19] M. Carrier, G. Le Gal, P. S. Wells, D. Fergusson, T. Ramsay, M. A. Rodger, Systematic review: the Trousseau syndrome revisited: should we screen extensively for cancer in patients with venous thromboembolism?, Annals of Internal Medicine 149 (2008) 323-333.

[20] F. F. van Doormall, W. Terpstra, R. van der Griend, M. H. Prins, M. R. Nijziel, M. A. van de Ree, H. R. Buller, J. C. Dulith, A. Ten Cate-Hoek, S. M. van den Heiligenberg, J. van der Meer, J. M. M. B. Otten, Is extensive screening for cancer in idiopathic venous thromboembolism warranted?, Journal of Thrombosis and Haemostasis 9 (2011) 79-84.

[21] A. Piccioli, A. W. Lensing, M. H. Prins, A. Falanga, G. L. Scannapieco, M. Ieran, M. Cigolini, G. B. Ambrosio, M. Monreal, A. Girolami, P. Prandoni, Extensive screening for occult malignant disease in idiopathic venous thromboembolism: a prospective randomized clinical trial, Journal of Thrombosis and Haemostasis 2 (2004) 884-889.

[22] M. Monreal, A. W. Lensing, M. H. Prins, M. Bonet, J. Fernandez-Llamazares, J. Muchart, P. Prandoni, J. A. Jimenez, Sreening for occult cancer in patients with acute deep vein thrombosis or pulmonary embolism, Journal of Thrombosis and Haemostasis 2 (2004) 876-881.

[23] M. Carrier, A. Lazo-Langner, S. Shivakumar, V. Tagalakis, R. Zarychanski, S. Solymoss, N. Routhier, J. Douketis, K. Danovitch, A. *Y.* Lee, G. Le Gal, P. S. Wells, D. J. Corsi, T. Ramsay, D. Coyle, I. Chagnon, Z. Kassam, H. Tao, M. A. Rodger, Sreening for occult cancer in unprovoked venous thromboembolism, The New England Journal of Medicine 373 (2015) 697-704.

[24] P. Robin, P. Y. Le Roux, B. Planquette, S. Accassat, P. M. Roy, F. Couturaud, N. Ghazzar, N. Prevot-Bitot, 0. Couturier, A. Delluc, 0. Sanchez, B. Tardy, G. Le Gal, P. Y. Salaun, Limited screening with versus without 18—fluorodeoxyglucose PET/CT for occult malignancy in unprovoked venous thromboembolism: an open—label randomised controlled trial, The Lancet Oncology S1470-2045, e-publication ahead of print 2015 Dec 7.

[25] V. Rieu, S. Chanier, P. Philippe, M. Ruivard, Systematic screening for occult cancer in eldery patients with venous thromboembolism: a prospective study, Internal Medicine Journal 41 (2011) 769-775.

[26] M. T. Rondina, N. Wanner, R. C. Pendleton, L. W. Kraiss, R. Vinik, G. A. Zimmerman, M. Heilbrun, J. M. Hoffman, K. A. Morton, A pilot study utilizing whole body 18 F—FDG—PET/CT as a comprehensive screening strategy for occult malignancy in patients with unprovoked venous thromboembolism, Throbosis research 129 (2012) 22-27.

[27] M. Chauchard, K. Benali, T. Papo, K. Sacre, Positron emission tomography combined with computed tomography as a screening tool for occult malignancy in patients with unprovoked venous thromboembolism: an observational study, Medicine 93 (2014) 1-4.

[28] PubMed, National center for biotechnology information, U.S. national library of medicine, http: //www.ncbi .nlm.nih.gov/pubmed, [Online; accessed 8—January-2016].

[29] A. Kleinjan, F. F. van Doormall, M. Prins, H. R. Buller, J. M. M. B. Otten, Limitations of screening for occult cancer in patients with idiopathic venous thromboembolism, The Netherlands Journal of Medicine 70 (2012) 311-317.

[30] H. R. Gibson, B. L. Dillard, Elementary Statistics, Kendall Hunt Publishing, 2012.

[31] K. Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, Philosophical Magazine Series 5 (1900) 157-175.

[32] W. G. Cochran, The X^2 test of goodness of fit, The Annals of Mathematical Statistics 23 (1952) 315-345.

[33] H. T. Sorensen, L. Mellemkjwr, F. H. Steffensen, J. H. Olsen, G. Nielsen, The risk of a diagnosis of cancer after primary deep venous thrombosis or pulmonary embolism, The New England Journal of Medicine 338 (1998) 1169-1173.

[34] M. M. Al-khalaf, L. Thalib, S. A. R. Doi, Combining heterogenous studies using the random—effects model is a mistake and leads to inconclusive meta—analyses, Journal of Clinical Epidemiology 64 (2011) 119-123.

[35] M. Adam6Ik, Collective reasoning under uncertainty and inconsistency, Phd thesis, University of Manchester (2014).

[36] B. de Finetti, Sul significato soggettivo della probabilita, Fundamenta Mathematicae 17 (1931) 298-329.

[37] J. B. Paris, What you see is what you get, Entropy 16 (2014) 6186-6194.

[38] I. Csiszar, G. Tusnady, Informational geometry and alternating minimization procedures, Statistic and Decisions 1 (1984) 205-237.

[39] P. P. B. Eggermont, V. N. LaRiccia, On em-like algorithms for minimum distance estimation (1998).

[40] H. H. Bauschke, P. L. Combettes, D. Noll, Joint minimization with alternating bregman proximity operators, Pacific Jornal of Optimization 2 (2006) 401-424.

[41] S. A. Bochkanov, Alglib project, http //www alglib .neti, [Online; accessed 2—April-2016] (1999-2016).