Institute for Clinical Trials and Methodology
University College London

A Quantitative Process for Enhancing End of Phase 2 Drug Development Decisions

Submission for PhD by Antony James Sabin

I, Antony James Sabin, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis

Signed:

ABSTRACT

The objectives of the phase 2 stage in a drug development programme are to evaluate the safety and tolerability of different doses, select a promising dose range and look for early signs of activity. At the end of phase 2 (EOP2), a decision to initiate phase 3 studies involves the commitment of considerable resources and associated costs. This multi-factorial, critical business decision is made by balancing the current condition of a sponsor organisation's portfolio, the future cost of development, the competitive landscape, and the expected safety and efficacy benefits of a new therapy. The decision process needs to be efficient, effective and evidentially supported.

This PhD presents a practical, formalised and quantitative approach for enhancing the EOP2 decision process. This process ensures that a consistent and explicit evidence-based approach is used to inform decisions for new drug candidates. Broadly following this process will help statisticians increase their strategic influence in drug development programmes.

Embedded within the process is a predominantly bayesian approach to predicting the probability of efficacy success (PoS) in a future (frequentist) phase 3 programme at the end of phase 2. Also included are early predictions of the outcomes of indirect treatment comparisons and future treatment ranking between the investigational treatment and alternative treatments either already approved or still in development. Moreover, the incorporation of qualitative factors into the decision making process, and the implementation of a PoS

framework when limited or no prior clinical data is available is discussed. The entire process is illustrated using an example from the pancreatic cancer indication, with further supporting examples of predicting the PoS provided for gastric cancer, soft tissue sarcoma, non-small cell lung cancer and ovarian cancer. Additionally, the utility of the PoS framework in practice is highlighted through a review of the PoS for 63 completed phase 3 studies.

IMPACT

The research conducted for this PhD has led to the development of a practical, formalised and quantitative approach for enhancing the phase 2 decision process. This process aims to ensure that a consistent and explicit evidence-based approach is used to inform decisions for new drug candidates which, if broadly followed, will improve the quality of decisions made. The research has already had considerable impact within the pharmaceutical industry. The process developed received wide acceptance within Amgen Ltd and is now being followed for all drugs entering phase 2. Similar quantitative approaches are being followed across other companies, including AstraZeneca and GSK.

The EOP2 decision is a critical decision point during the drug development process. Incorrect decisions result in a waste of resources and have major impact on company finances, cost of medicines and share-holder confidence. From the patient perspective, incorrect ‘go’ decisions lead to large numbers of patients being treated with an ineffective therapy and potentially missing out on other better treatment options. Despite the criticality of this decision, the phase 3 success rate, particularly within oncology, has historically remained low. Enhancing the quality of the EOP2 decision is therefore particularly impactful.

The research performed starts by defining the key questions to be addressed when making an EOP2 decision and focuses on the systematic abstraction of relevant data from available literature. This is then used to develop a novel statistical modelling approach for predicting the probability of efficacy success in

phase 3. The model is also extended to compare and rank the expected efficacy of the investigational treatment to the current and other potential treatment options under development at the time when the new investigational product is expected to achieve market approval. This extension greatly complements the modelling of the PoS and further enhances the EOP2 decision. Such information could be used to stop development of an investigational drug prior to investing in a costly phase 3 trial when it is unlikely to beat competing treatments and become profitable, despite having a high PoS in a future comparative phase 3 trial against the currently available standard of care.

Additional research is also presented on the implementation of a PoS framework to enhance the decision to start phase 3 when limited or no prior clinical data is available; a situation that may arise as sponsors expand their medicines to include additional indications close to patent expiry.

The process and methodology developed is highlighted using an example in pancreatic cancer. This was particularly impactful as, not only has it supported EOP2 decisions, it also changed clinical perception within Amgen on the importance of progression free survival hazard ratio as a phase 2 outcome measure. Moreover, the systematic review and modelling of the relationship between the phase 2 and 3 outcome measures has been utilised in discussions with regulatory agencies.

Broadly following the approaches developed enhances the strategic contribution and influence of statisticians in the EOP2 decision making process, ensuring key elements of proposed development plans are thoroughly investigated and decisions risk informed.

Table of Contents

List of Figures

List of Tables

List of Key Terms

Abbreviation	Definition
ACR	American College of Rheumatology
BLA	Biologic License Application
CI	Confidence Interval
CrI	Credible Interval
CRPC	Castrate resistant prostate cancer
CTA	Clinical Trial Authorisation
CTD	Common Technical Document
ECOG	Eastern Cooperative Oncology Group
EMA	European Medicines Agency
EMBASE	Excerpta Medica database
EOP2	End of Phase 2
EU	European Union
FIH	First in Human
HR	Hazard Ratio
HTA	Health Technology Appraisal
IND	Investigational New Drug
LRV	Lower reference value
MAA	Marketing Authorisation Application
mBC	Metastatic breast cancer
MCMC	Markov Chain Monte Carlo
mCRC	Metastatic colorectal cancer
MEDLINE	Medical Literature Analysis and Retrieval System Online
NCCN	National Comprehensive Cancer Network
NDA	New Drug Application
NHS	National Health Service
NICE	National Institute for Health and Clinical Experience
NSCLC	Non-small cell lung cancer
ORR	Objective response rate
OS	Overall Survival
P	Probability
Ph2	Phase 2
Ph3	Phase 3

Abbreviation	Definition
PFS	Progression Free Survival
PoS	Probability of Success
RADDM	Rheumatoid Arthritis Drug Development Model
SCLC	Small cell lung cancer
SE	Standard Error
STS	Soft tissue sarcoma
SUCRA	Surface Under the Cumulative Ranking
TV	Target Value
US	United States

1. INTRODUCTION

The development of new medicines is a long, complex and costly process. It typically takes at least 10 years, with an average research and development spend of $1.8 to $2.6 billion, for a new medicine to make it onto the market following its initial discovery. In Section 2.1 to 2.5, which follows, the clinical drug development process is defined. This implicitly highlights the many hurdles that need to be overcome, and the complex decisions that need to be taken before a drug reaches the market.

Section 2.6 summarises the multi-disciplinary structure of typical governance bodies involved in drug development decision making and the key decision points along the drug development pathway. Also provided is an appraisal of the publicly available literature on the phase to phase transition rates by therapeutic area over approximately the last two decades, and background data highlighting the financial impact of improving success rates. This literature clearly highlights that considerable attrition of investigational drug candidates occurs in late phase development (Kola and Landis 2004; Di Masi, Feldman, Seckler and Wilson 2010; Paul et al 2010; Arrowsmith 2011a,b; Thomas et al 2016; Wong 2018; Dowden 2019), a situation which results in companies incurring the vast majority of the drug development costs. This in turn translates to a higher cost of medicines and reduced overall profitability of the companies concerned.

Moreover, the literature review brings particular attention to the oncology therapeutic area where the big picture shows that over all oncology indications there is <10% chance that an oncology molecule entering clinical development will make it through to approval.

The financial impact of improving success rates highlights that one of the most critical decisions to be made in this process is whether or not to progress onto extensive ‘within human’ clinical trials (i.e. moving from phase 2 to phase 3 (late phase) in the research and development process). Incorrect decisions at this point result in a massive waste of resources and have a major impact on company finances and share-holder confidence. The criticality of this decision is such that it needs to be risk-informed by exploring, developing and implementing methodologies that enable a quantitative assessment of the available efficacy evidence thereby ensuring the highest possible probability of reaching the correct decision for both the business and patients.

The aims of this thesis are to:

1. Define and examine the existing processes.
2. Extend the decision making processes to include more sophisticated statistical and analytical techniques to enable the business to better understand the risks and thereby have greater confidence in the EOP2 decision.

3. To provide a framework to pharmaceutical statisticians that will enable them to enhance their contribution and strategic influence in the critical drug development decision making process.

To achieve these aims, this PhD thesis focuses on the development, evaluation and implementation of a practical quantitative process for enhancing drug development decisions which, when followed, will ensure that a more consistent and explicit evidence-based approach is used to make decisions for new drug candidates.

The research carried out for this PhD has resulted in the production of three academic articles, two of which have already been published, whose authorship and subject matter was led by Antony Sabin. These papers are listed at Annex E to this document and were released to ensure that the pharmaceutical industry profited from this research at the earliest possible opportunity. Some text in this thesis may also be found in these articles.

After understanding the key inputs into the entire clinical decision making process, target areas were selected which provided opportunities to enhance the quality of the decisions through either the development of new quantitative statistical approaches or the enhancement of existing methodologies in the selected areas. The key research areas identified were:

1. Enhancing the quality of the end of phase 2 decision using a practical model to quantify the probability of technical success (PoS) with respect to efficacy in phase 3 on completion of phase 2.

2. The incorporation of qualitative factors into the decision making process, and the implementation of a PoS framework in the situation where very limited or no prior clinical data is available.
3. The determination of the expected efficacy of the investigational treatment and how this compares and ranks to the current and future treatment options available for patients.

Key Research Area 1:

The research conducted starts by describing for the first time the key questions that require addressing when making an EOP2 decision and focuses on the systematic abstraction of the associated relevant data from the literature. This is then used in the development of a novel statistical modelling approach for predicting the probability of efficacy success in phase 3 at the end of phase 2 and informing the choice of PoS threshold (Section's 3.2 to 3.8).

The PoS model developed synthesizes the relationships between phase 2 and phase 3 study outcome measures on the relative treatment difference scale, the influence of prognostic factors on the relationship, the treatment difference observed for the phase 2 outcome measure in the phase 2 study, the prior opinion of key decision makers for the treatment difference in the phase 2 outcome measure, and knowledge of the proposed phase 3 study design. These results are used to predict the probability of success in a future phase 3 study to be analysed using frequentist statistical methods.

Also introduced within Section 3.3 is new research to develop reasonable bounds of belief for the PoS, generated by running the model prediction incorporating a range of subjective prior opinions representing different attitudes of key decision makers for the treatment difference in the phase 2 outcome measure.

Using the predicted PoS to inform a decision requires a threshold to be set, such that if the PoS is greater than this threshold a decision to move to phase 3 may be made. The selection of such a threshold will be specific to the funder/sponsor and their current portfolio. To ensure an informed choice of threshold can be made, the predicted PoS needs to be put into context with the operating characteristics of the model employed. This involves evaluating the probability of making a go decision and failing phase 3, and a go decision and being successful in phase 3 for a given development strategy based on different PoS commit to phase 3 decision thresholds. A simulation process to determine these probabilities and help decide among different development strategies is described in Section's 3.4 and 3.5.

Moreover, when interpreting the observed results from a phase 2 study, decision makers will carefully investigate the robustness of evidence provided. One key area of evaluation will focus on the similarity of the observed phase 2 control response with that which was expected from prior knowledge. An innovative approach to adjust for any potential discordance is described in Section 3.7 whereby the variance of the treatment difference in the phase 2

outcome measure in inflated as the observed phase 2 control group response departs from the expected.

To highlight the utility of a bayesian framework to predict PoS in practice, Section 3.10 discusses the factors influencing the formation of the framework and the practical advantages. Also evaluated is the PoS values for 63 completed phase 3 studies spanning the oncology, respiratory and cardiovascular portfolio that started after 2015 at AstraZeneca. This includes an assessment of the predicted PoS and observed phase 3 success rate for studies that used a bayesian framework, and those that used a more quantitative framework as discussed in key research area 2.

A wide ranging and interesting body of work that focuses on methodologies and models to predict the probability of success in phase 3 have been published over the last 15 years. A summary of this body of literature is described in Section 3.1.1.

Key Research Area 2:

Section 3.9 discusses the need to include additional qualitative considerations into the end of phase 2 decision, and also explores the development of a PoS framework in the situation where very limited or no prior clinical data is available. Such a situation may arise in the life cycle management (broadening the use of the drug into new lines of therapies or disease indications) of oncology products that target patients with specific genes or biomarker targets. An example from the 1st line ovarian cancer setting in presented.

Key Research Area 3:

For the third key area studied, new research is presented that builds on the methodology and framework developed in research area 1 to enable early predictions (to be used as part of the EOP2 decision) of the expected efficacy of the investigational treatment at the end of a successful phase 3 study and how this will compare and rank to the current and future successful treatment options available for patients (Section 3.11).

Following regulatory approval of a new treatment, applications for reimbursement need to be made in many geographical regions. If the application is approved this allows a drug company to be paid for all or part of a prescription. If the treatment is considered to have a high cost or a significant budget impact on the health care system, this application may require a form of pharmacoeconomic assessment called a Health Technology Appraisal (HTA) to take place. Part of this assessment usually requires making effectiveness comparisons between treatments not already compared in head to head trials (indirect treatment comparisons). The research therefore enables predictions of the outcomes of these indirect comparisons to be made at the end of phase 2. This includes enabling indirect treatment comparisons with all of the currently available marketed treatments and also with potential competitor treatments that are still in development. Moreover, the research enables the predicted treatment ranking to be established, for example, this could be the predicted probably of being ranked 1st, 2nd or 3rd amongst all of the treatment options that

will be available at the time the new drug would be ready to launch. This therefore greatly complements the modelling of the PoS and further supports the EOP2 decision. Such information could, for example, be used to stop the development of an investigational drug prior to investing in a costly phase 3 trial where it is unlikely to beat the competitors and become profitable. This could be despite having a high probability of success in a phase 3 trial designed to show superiority against the currently available standard of care. A summary of the prior published literature surrounding this research topic is discussed in Section 3.1.1.

The utility and benefits of the processes developed for this PhD are predominantly highlighted throughout by the use of an illustrative example from the pancreatic cancer indication (Section's 3.6 and 3.11). However, to further illustrate the utility of the methodology additional examples for gastric cancer, soft tissue sarcoma, non-small cell lung cancer and ovarian cancer are highlighted in Section's 3.8 and 3.9.

2. OVERVIEW OF DRUG DEVELOPMENT PROCESSES

2.1. Introduction

Developing new medicines is a long, complex and costly business. It typically takes at least 10 years for a new medicine to make it onto the market following its initial discovery, with the average cost of research and development (incorporating the cost of the many thousands of molecules that are rejected along the way) being in the region of $1.8 to $2.6 billion. The drug development process can be divided up into five steps, these being discovery research, pre-clinical research, clinical research, filing an application to market the drug, and post-approval research and safety monitoring (Figure 2.1). The process begins with the setting of the drug developer's vision for the company. For example, a company may wish to focus solely on discovering and developing innovative therapies to treat serious illnesses, such as cancer or inflammation diseases. Once this is agreed the process of discovery research can begin.

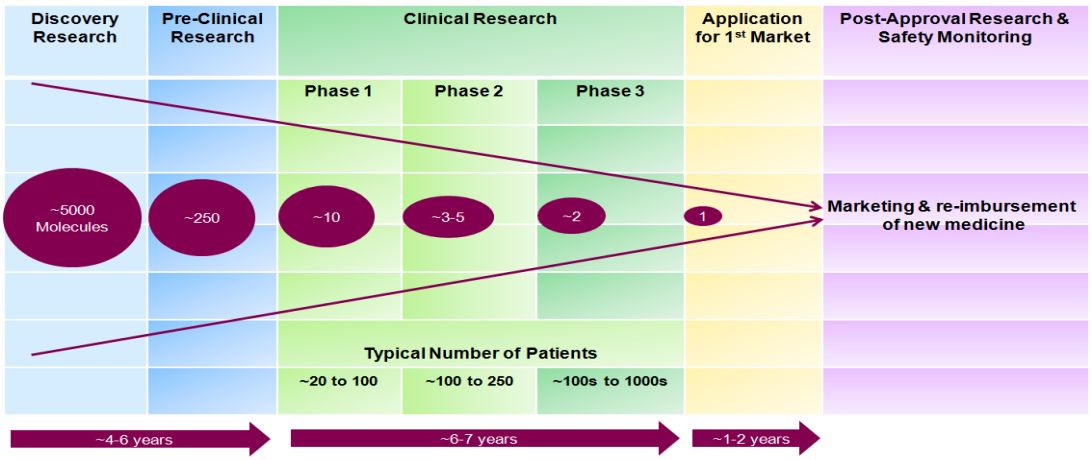


Figure 2.1: Overview of Drug Development

2.2. Discovery Research

The development of a new medicine begins in discovery research. Here researchers will first aim to understand the mechanism behind the disease and identify biological targets that hopefully will interact with a potential medicine to produce a clinically beneficial effect to patients. This will generally be done by conducting studies in cells, tissues and/or animal models. After learning about the mechanism and possible drug targets, depending upon the disease area, biotechnology or pharmaceutical chemistry may be used to identify potential proteins (large molecules) or small molecules respectively to treat disease. As the research progresses, the number of molecules is reduced from many thousands to a few hundred promising candidates that are further explored in pre-clinical testing. Such decisions are generally made using a combination of advanced computer modelling techniques and automated high-throughput screening facilities. Other considerations that also need to be taken into account in the decisions include:

- Can a quality assured manufacturing process and delivery technology be developed to produce the molecule in sufficient quantities?
- Does the molecule remain pure and stable under certain conditions and with time?
- What are the molecules components and impurities?

2.3. Pre-Clinical Research

Prior to testing a drug in humans, the safety and biological effectiveness of a molecule must be first established. Pre-clinical research is designed to:

- Validate animals as predictive of the human condition
- Establish a therapeutic window (the safety margin between therapeutically effective and toxic effects)
- Establish the toxicology in several animal species
- Understand the relationship between response and metabolism
- Optimise the formulation, dose range and frequency of administration.

During the pre-clinical research stage the promising molecules go through a series of tests to evaluate the pharmacodynamics (what the drug does to the body), pharmacokinetics (to determine how the molecule is absorbed in the body, distributed, metabolised and excreted) and toxicology of the drug (the safety in animal and cellular models).

Such investigations are normally performed in living cells (in-vitro), in animal models (in-vivo) or via computational modelling. At this time the researchers will again look to reduce the pool of candidate molecules, the successful ones selected based upon their pharmacokinetic properties, the balance between the maximum well-tolerated and minimum effective dose, the expected cost of goods and the commercial landscape. Molecules that pass these hurdles may then undergo further molecular optimisation and subsequent testing to improve

their safety and effectiveness. The end result will be a lead and a selection of backup molecules that will be put forward for clinical research in humans.

2.4. Clinical Research

Clinical research in human trials can only begin once the pre-clinical package has been reviewed and approved by the regulatory authorities. Companies wishing to perform studies in the US or EU must therefore successfully file for an Investigational New Drug (IND) application and/or a Clinical Trial Authorisation (CTA) respectively for any new candidate molecule. In addition to the regulatory approval, in order to ensure the patients are appropriately protected, all clinical trials must be approved by ethics committees at the establishments where the trials will take place. Clinical research is typically divided up into three phases which are described below.

2.4.1. Phase 1 Clinical Research

The aim of phase 1 clinical trials is to understand the biological characteristics of the molecule. Typically, this is where the molecule will be tested for the first time in humans. The studies are generally conducted in a small number (20 to 100) of normal healthy subjects, with the principal aim of evaluating the pharmacological response with increasing dose. Occasionally these may be done in patients when the potential toxicity excludes the use of healthy subjects. (e.g. refractory patients are often used in the oncology setting when studying cytotoxic drugs). Moreover, in addition to providing initial data on the safety, tolerability and activity, the human pharmacokinetics of the drug are evaluated.

On completion, the evidence gained from phase 1 studies is used by the drug company to decide if the molecule has a suitable therapeutic window (in terms of a dose range and schedule of dosing) with an acceptable safety, tolerability, activity and cost of goods profile for study in phase 2.

2.4.2. Phase 2 Clinical Research

Phase 2 studies look to investigate the effectiveness of a new molecule in a moderate number of patients with the targeted disease. Early phase 2 studies (often called proof of concept studies) look for preliminary evidence of efficacy (often in intermediary endpoints as opposed to those used in phase 3) and short-term safety. Multiple doses (and dosing schedules) may be explored in an attempt to identify the minimum and maximum effective safe dose range. Around this point in development the criteria for clinical and commercial success become better defined. Additional phase 2 trials are subsequently conducted to support the downstream regulatory filing, ideally mirroring the target population and control groups. Such studies provide evidence to determine the probability of meeting the clinical and statistical success criteria in phase 3. Those molecules meeting the end of phase 2 success criteria will move into the confirmatory phase 3 studies.

2.4.3. Phase 3 Clinical Research

Phase 3 studies are typically large-scale definitive studies enrolling patients from sites globally. The aim of phase 3 is to confirm the findings of phase 2 and

prove the drug is suitable for registration by demonstrating both short- and long-term safety and efficacy (benefit-risk).

The study designs must withstand scrutiny by regulatory authorities and are therefore often conducted as randomised, double-blind studies to minimize potential bias, and ideally go head to head against the current standard of care to prove their superiority.

The results of these studies form the pivotal information for marketing applications. The benefit-risk profile provides the basis for product labelling, which includes the diseases and patient population the drug is licensed for, the side effect profile, the doses schedule, and if there are any contra-indicated marketed drugs with which the new molecule may negatively interact.

2.5. Market Application

2.5.1. Introduction

Following a successful phase 3 programme, companies wishing to file for marketing applications follow the New Drug Application (NDA), or Biologic License Application (BLA) process in the US. The equivalent process in the EU is the Marketing Authorisation (MAA) or Common Technical Document (CTD). The filings contain the results from all of the research conducted on the new molecule. This is reviewed by the regulatory authorities who make a decision whether or not to give the drug a marketing licence based on its risk-benefit profile.

2.5.2. Post-Marketing Research and Safety Monitoring

After successfully obtaining marketing authorisation, the research on a new molecule continues to monitor safety and long-term side effects. This is done as long as the medicine is on the market, with periodic safety and tolerability reports provided to regulatory authorities. Often marketing authorisation will be granted under the proviso that the company fulfils certain conditions. These conditions are typically of the form that requires additional post marketing research. For example, it may be of interest to evaluate the risk benefit in specific subgroups of patients, or to obtain real world evidence on the clinical effectiveness and patterns of routine clinical use and toxicity management of the new molecule. Such studies typically are labelled phase 4 or Medical Affairs Studies.

2.5.3. Reimbursement

In many countries obtaining marketing approval does not automatically translate into gaining access to the market. In countries such as the UK where the healthcare system is primarily public funded (~80% of funding coming from taxation), the independent body NICE (National Institute for Health and Clinical Experience) selects certain new treatments for cost-effectiveness evaluation (a Health Technology Appraisal, HTA). Selection is generally based upon the following key criteria:

- Is the new treatment likely to result in a significant health benefit to patients of the National Health Service (NHS) as a whole?
- Is the new treatment likely to significantly impact on other government policies (e.g. reduction in health inequality)?
- Is the new treatment likely to have a significant impact on NHS resources?
- Can NICE add value by issuing national guidance over the interpretation of clinical evidence or cost effectiveness?

There are two types of HTA; (1) A multiple technology appraisal which examines the disease area or class of drugs incorporating the evidence from the new treatment, or (2) a single technology appraisal to provide early guidance on new drugs targeting a single indication.

As it is a mandatory requirement for National Health Authorities within England and Wales to fund treatments recommended by NICE, this appraisal process influences the uptake of a new treatment onto the market. This guidance is also known to influence the marketing decisions of other countries across the globe as it is often available ahead of approval in other countries. The outcome is that market access decisions are heavily influenced by cost-effectiveness and cost-containment arguments, as opposed to using purely safety and efficacy guided decisions. The likelihood of achieving reimbursement therefore plays an important role in the decisions made during the drug development process.

2.6. Decision Making During Drug Development

Implicitly highlighted in the drug development process described through Section 2, are the many hurdles a new molecule has to overcome before it reaches the market. Many, and often complex, decisions need to be taken along the way by multi-disciplinary flexible governance teams. For example, a typical governance body accountable for decisions across early and late development may include flexible representation and input from:

- Chief Executive Officer
- Head of Research of Development
- Early Stage Clinical
- Late Stage Clinical
- Drug Discovery
- Pharmaceutical Sciences
- Clinical Pharmacology
- Precision Medicine
- Safety Sciences / Pharmacovigilance
- Portfolio Management
- Biometrics / Statistics
- Finance
- Regulatory
- Legal
- Chief Medical Officer

- Commercial leaders across various regional areas
- Development Operations.

The principal aim of clinical drug development decision making is to stop the development of non-viable treatments as soon as possible. This avoids administering unsafe or ineffective medicines to patients, mitigates the drug development costs and makes both resources and patients available for the development of other potentially more promising treatments.

There are a number of decision points within the clinical drug development process, the key ones being:

- The decision to go into humans for the first time
- The transition to phase 2 after gaining evidence of biological activity
- The decision to initiate phase 3 (herein referred to as the end of phase 2 (EOP2) decision) after completion of dose ranging and finding
- The commitment to file with regulatory bodies
- Commitment to launch into the market.

These critical decision points are shown diagrammatically in Figure 2.2 below:

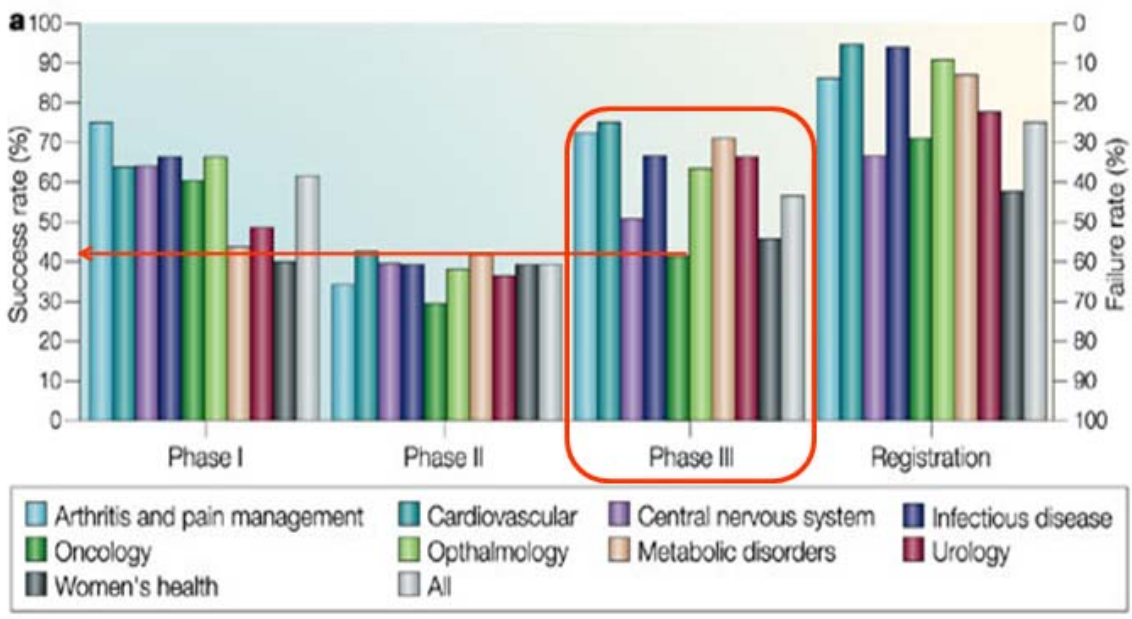
FIH: First in Human; P2: Phase 2; EOP2: End of Phase 2.

Figure 2.2: Critical Decision Points within the Clinical Drug Development Pathway

It is well documented that considerable attrition of potential drug candidates occurs in late phase development, with lack of efficacy being the main reason for failing phase 3 (Kola and Landis 2004; Di Masi, Feldman, Seckler and Wilson 2010; Paul et al 2010; Arrowsmith 2011a,b; Thomas et al 2016; Wong 2018; Dowden 2019). This is not a good scenario for patients or drug development companies. It results in companies incurring the majority of the drug development costs which, in turn, translates to a higher cost of medicines and reduced overall profitability of the companies concerned.

The oncology therapeutic area is a particularly noteworthy example, where success rates for transitioning between development phases have been lower than other therapeutic areas, with success rate in phase 3, arguably, being unacceptably low. The data shown in in Figure 2.3 (Kola and Landis 2004) was published in 2004 where we see that of the oncology products that successfully make it through phase 1 and 2 only 40% are successful in phase 3. Moreover, the overall success rate of oncology products entering phase 1 and achieving registration was approximately 5% (with phase transition rates of 60% success

in phase 1, 30% in phase 2, 40% in phase 3 and 70% of successful phase 3's achieving registration).



Kola and Landis. Nature Rev Drug Discov 2004;3:711.

Figure 2.3: Success by Phase and Therapeutic Area (Kola and Landis, 2004)

Data collected between 2006 and 2015 (Thomas et al 2016), which is the largest study of clinical success rates conducted to-date, showed the overall success rate in oncology to have remained very consistent at 5%. Table 2.1 highlights the phase to phase successful transition rate during this period. A comparison to the 2004 survey shows there to be a decrease in the percentage of studies successfully passing phase 2 (a change from 30% to 24.6%), no change in the percentage of successful phase 3 studies at 40%, and an

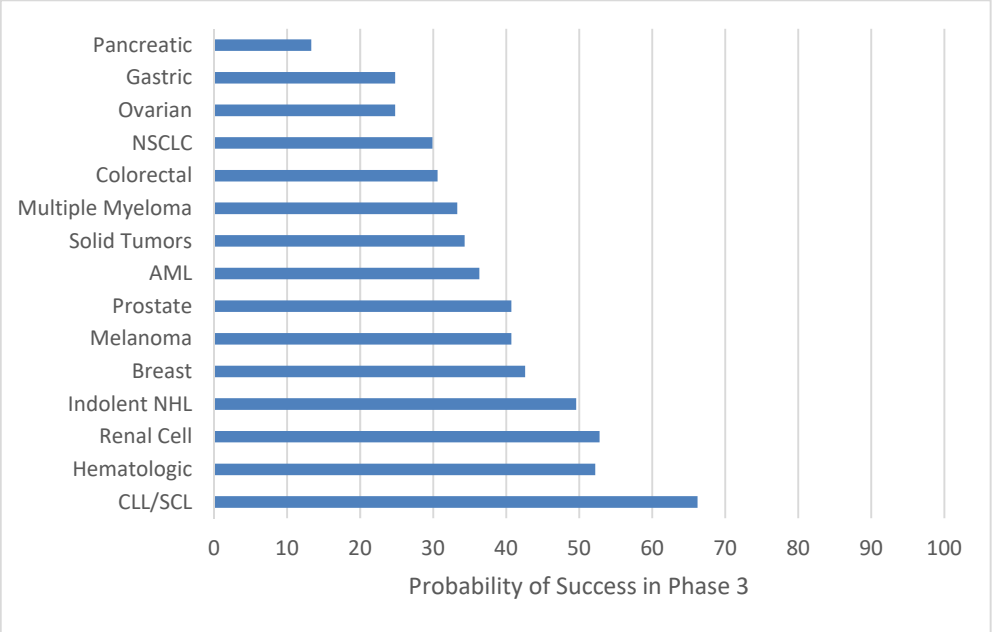
increase in the number of phase 3 successfully achieving registration (from 70 to 84.2%).

	Phase 1	Phase 2	Phase 3	Registration	Overall
Cardiovascular	58.9	24.1	55.5	84.2	6.6
Infectious dis.	69.5	42.7	72.7	88.7	19.1
Oncology	62.8	24.6	40.1	82.4	5.1
Opthalmology	84.8	44.6	58.3	77.5	17.1
Metabolic	61.1	45.2	71.4	77.8	15.3
Urology	57.1	32.7	71.4	85.7	11.4
All indication	63.2	30.7	58.1	85.3	9.6

Data obtained from Thomas et al (2016)

Table 2.1: Success rate by phase and therapeutic area 2006-2015

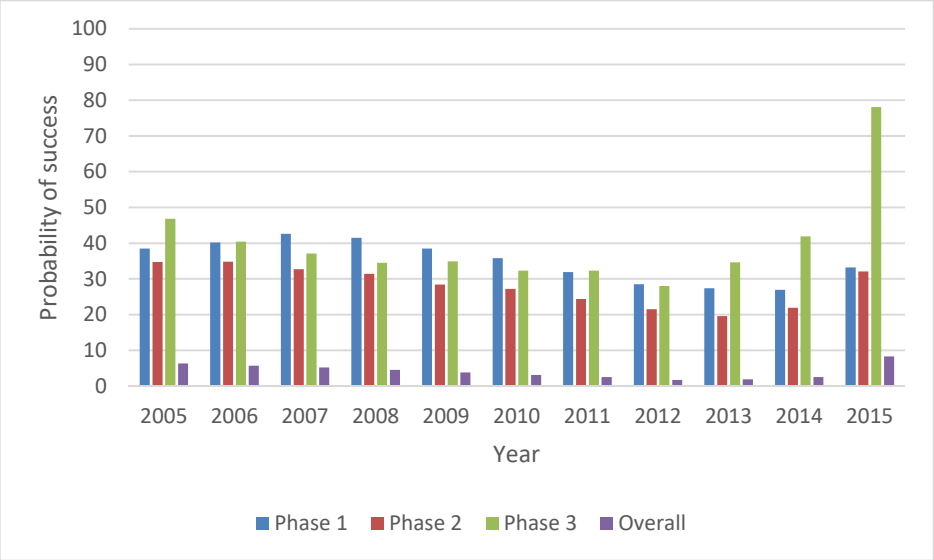
The Oncology therapeutic area is by nature very diverse, and phase 3 success rates have been shown to vary widely across indications. Thomas et al (2016), showed the indication with the lowest phase 3 success rate during 2006-2015 to be pancreatic cancer where only 13.3% of phase 3 studies were successful. In contrast to this 66.2% of chronic lymphocytic leukemia / small cell lymphoma phase 3 studies were successful which is more in line with non-oncology indications. Figure 2.4 shows the phase 3 success rates for a variety of oncology indications using data derived from Thomas et al (2016).



Data obtained from Thomas et al (2016)

Figure 2.4: Phase 3 success rates by Oncology Indication

The results produced by Thomas et al (2016) were also consistent with the findings of Wong et al (2018) with an average overall PoS of 4.1% over the time interval between 2005 and 2015. While these results did not incorporate the probability of registration, Wong also presented an estimate of the overall success rate in oncology each year (Figure 2.5). The data showed a success rate of 6.3% in 2005, with a slow and consistent decline to 1.7% in 2012, before increasing to 2.5% in 2014, followed by a jump to 8.3% in 2015. The increase in 2015 being largely driven by an increase in predicted success rate in phase 3 from 41.9% to 78.1% between 2014 and 2015.



Data obtained from Wong (2018)

Figure 2.5: Oncology success rates by phase 2005-2015

Dowden (2019) present data from CMR international who operate with a consortium of biopharmaceutical companies to compare R&D performance on a like for like basis. The data presented ranged over a period of 2010 to 2017 and indicated an average success rate of 59% in phase 3 oncology trials, with an average overall success rate for oncology molecules entering phase 1 and achieving approval to be 9%.

It is possible the increase in oncology success rate seen in Wong (2018) and Dowden (2019) are simply an artefact of the data collected or the oncology indications targeted over the last few years. However, it could also be the result of an increased proportion of biomarker targeted drug candidates working their way through the drug development lifecycle, and to growth in the number of approvals of drugs for orphan indications and rare diseases. While this latest

signal may be considered encouraging, the big picture remains that there is <10% chance that an oncology molecule entering clinical development makes it through to approval, and thereby there is room for improvement.

Arrowsmith (2011a) highlights the financial impact that improving success rates could bring. On the x-axis is the average capitalised cost per launch which highlights the price paid for the drug, development and launch costs. These costs are presented for various stages of the drug development process and the probability of technical success at each stage, as shown in the y-axis. The plot indicates that the average capitalised cost per launch is over $2B assuming the probability of success in phase 3 is 60%. The plot also shows the impact on cost associated with improving the phase 3 probability of success. If the probability of phase 3 success is improved by 20% (from 60% to 80%) costs reduce by approximately $500M. Making good EOP2 decisions is therefore critical to a company's profitability and the cost of medicines. This is shown diagrammatically in Figure 2.6 below.

The EOP2 decision is influenced by a number of factors, including the current condition of an organisation's development portfolio, the expected future cost of development / return on investment, the competitive landscape, and the expected safety and efficacy benefit of a new therapy. There is a pressing need to develop and implement methodologies and processes to enhance the EOP2 decision making capabilities within the industry.

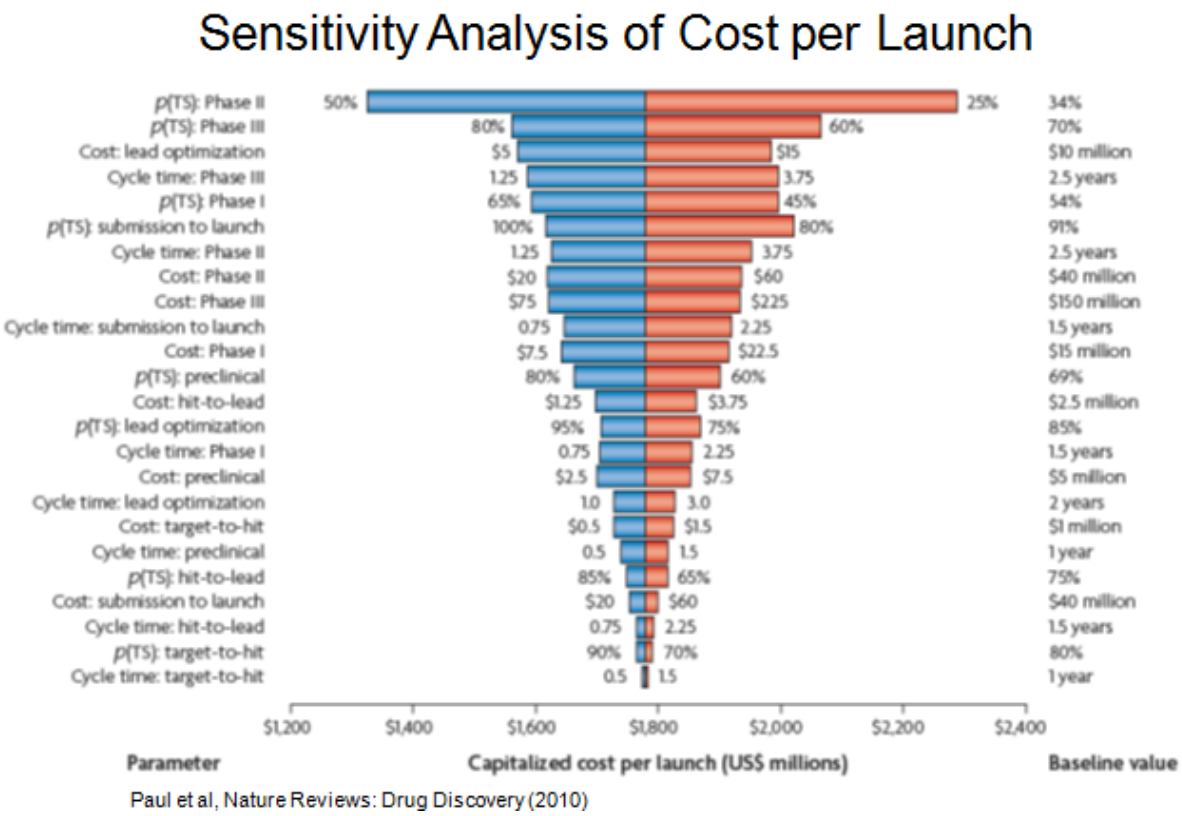


Figure 2.6: The Financial Impact of Improving Success Rates

The EOP2 decision should involve a systematic quantitative assessment of the available evidence. Statisticians with experience in systematic data reviews, quantitative modelling and experimental design have a fundamental role to play in the making of this vital decision which, if carried out successfully, will save considerable nugatory work and expense for their companies. Without a quantitative assessment decision making becomes more qualitative, and thereby more subjective, less data driven, more inconsistent and less decisive. This is not to say qualitative factors should not be taken into account. The results of such an approach may contain additional information that is not easily incorporated into quantitative assessment (e.g. company experience with

recruiting the desired population for a study). In practice combining both the qualitative and quantitative information will lead to the most appropriate decision.

3. A QUANTITATIVE PROCESS FOR ENHANCING EOP2 DECISION MAKING

3.1. Introduction

In Section 3 the steps required to build the quantitative framework for any disease under investigation are described. Firstly, an overview of the published background literature is described in Section 3.1.1. Section 3.2 then introduces a set of core questions fundamental to the decision making process and the associated requirement to systematically review and abstract the relevant data from the literature to answer these questions. Central to any EOP2 decision is the need to predict the treatment difference in the phase 3 study from the treatment difference observed in the phase 2 trial. As the phase 2 outcome measures used for decision making are often different to the phase 3 regulatory outcome measures (particularly in the oncology setting) meta-regression techniques are used to model this relationship from the systematically abstracted data. Moreover, in situations where a decision is made to use a different phase 3 population to that used in the phase 2 study the meta-regression can be used to predict the phase 3 outcome in the planned phase 3 population.

A critical step in influencing the pharmaceutical decision makers to implement this framework in practice was incorporating the ability to take the prior belief of key decision makers into account. Section 3.3 describes how, at the end of phase 2, this prior belief can be synthesized with the observed phase 2 outcome

and the meta-regression to predict a range of posterior distributions for the phase 3 outcome measure and population.

Next in Section 3.3, the synthesis of the EOP2 posterior distributions in the phase 3 outcome measure with the proposed phase 3 design is described which enables the prediction of the probability of success in the phase 3 study. This includes the development of bounds of belief for the PoS through the inclusion of sceptical and optimistic priors. After showing how to determine the PoS, Section 3.4 then describes the selection of a PoS threshold (decision criteria) that if achieved would equate to a commit to phase 3 go decision. The method of calculating the operational characteristics of the PoS decision criteria through the use of clinical trial simulation is presented. Section 3.5 describes how the process can be used to further optimise the design of the phase 2 and 3 studies. Section 3.6 then goes through a worked example of the process described through Section 3.2 to 3.5, through application to the pancreatic cancer indication. Additional examples in gastric cancer, soft tissue sarcoma and non-small cell lung cancer are also provided in Section 3.8.

Another key part in making the EOP2 decision is the degree of comfort decision makers have with the reliability of the phase 2 data. If, for example, the control group response observed in the phase 2 was not as expected, this may cast doubt on the validity of using the treatment difference observed in the phase 2 study results for prediction purposes. Section 3.7 presents a methodology extension to incorporate this uncertainty into the EOP2 decision making by taking the perspective that the observed treatment difference seen in the

randomized phase 2 study is the best unbiased estimate available, whilst inflating the variance of the phase 2 treatment difference as the observed phase 2 control group response departs from the expected.

In areas such as oncology where drug development times are long, there may be limited patent life remaining post the approval in the first indication. In order to expand use into new additional indications prior to the patent expiry a decision may be required to initiate a phase 3 study with limited or no phase 2 data in the target indication or population. This situation, including an approach to determine a PoS, is discussed in Section 3.9. An example from 1st-line ovarian cancer is presented.

Section 3.10 then further explores the utility of a bayesian framework to predict the PoS in practice. The key factors influencing the formation of the framework, and the practical advantages of taking such a modelling approach are discussed. This is followed by a review of the PoS values and the associated methodology used to calculate them (a benchmarking or bayesian modelling approach), for 63 completed phase 3 studies that started during or after 2015 at AstraZeneca.

There are however additional considerations that pharmaceutical companies need to take into account above and beyond the predicted probability of the phase 3 trial being successful (PoS). One extremely important consideration is the competitive landscape. It is particularly important for the EOP2 investment decision to understand and predict how the new investigational treatment is

likely to compare with currently marketed and other potential treatment options being developed by other companies at the time that the new investigational drug is scheduled to reach the market. To complement the PoS modelling at the EOP2, Section 3.11 therefore expands the modelling framework further to incorporate network meta-analysis to enable early predictions of indirect treatment comparisons in the phase 3 outcome measure and the probability of being highly ranked amongst the treatment options that will be available at the time of market approval. The utility of this is shown through the expansion of the pancreatic cancer example introduced in Section 3.6.

3.1.1. Published Literature

In this section prior published literature around the 3 key research areas included in this PhD are discussed.

Key Research Area 1:

- Enhancing the quality of the end of phase 2 decision using a practical model to quantify the probability of technical success (PoS) with respect to efficacy in phase 3 on completion of phase 2.

A wide ranging and interesting body of work that focuses on methodologies and models to predict the probability of efficacy success in phase 3 has been published over the last 15 years.

In 2005, O'Hagan, Stevens and Campbell (2005) introduced the important concept of assurance; an unconditional probability that a phase 3 trial will

achieve a specific outcome (e.g., a statistically significant result according to a standard frequentist significance test) based on prior knowledge for the unknown true treatment effect. Here it is proposed that assurance should be a key measure of the practical utility of a proposed phase 3 study, suggesting that it may be more appropriate to choose the phase 3 sample size based upon a desired assurance rather than using a conventional power calculation. Conventional phase 3 sample size calculations are usually selected to achieve a desired power conditional on a pre-specified true underlying treatment effect. Given the uncertainty in this underlying assumption the power may therefore not accurately reflect the probability that the trial will be successful. In this PhD, I use a combination of these approaches where the PoS (probability of achieving a 'favourable' statistically significant result) is used to make a go/no-go decision to conduct a traditionally powered (e.g. 80% or 90%) and sample sized study.

Stallard, Whitehead and Cleall (2005) proposed a quantitative model-based approach in which the posterior probability that a future frequentist phase 3 study will be successful is calculated and used to inform the decision to initiate phase 3 at either the interim and final analyses of a phase 2 trial. In the paper the phase 2 outcome measure (PFS) was simply assumed to have the same hazard ratio as the phase 3 outcome measure (OS). This differs from the approach taken in this PhD whereby the uncertainty between the phase 2 and 3 outcome measures is incorporated using meta-regression techniques.

The paper by Stallard, Whitehead and Cleall (2005) was complemented by the publication by De Ridder (2005) who published a case study with a drug for

symptom relief in a chronic condition whereby a parametric dose-response model for the clinical response was generated using individual data from two phase 2 studies. Modelling and simulation were then used to predict the range of possible outcomes from three ongoing phase 3 trials. These outcomes were compared to the blinded data being concurrently generated in the phase 3 trials in order to assess the robustness of the ongoing studies with respect to the uncertainty of the true dose-response, patient variability in baseline severity and drug-response, and to assess the likelihood of achieving a clinically relevant response with a dose lower than those included in the trials. Although the decision to start the phase 3 studies was not impacted by this modelling this paper highlights the merits of using modelling and simulation to support the end of phase 2 decision making.

In 2006, Chuang-Stein (2006) built on the work O'Hagan (2005) by describing the distinction between statistical power and the probability of getting a successful trial. Chuang-Stein highlights that while a very high statistical power can be selected to detect a certain treatment effect, the high statistical power does not necessarily translate to a high PoS if the target treatment effect is based on the perceived ability of the drug candidate, rather than empirical clinical knowledge of the drug's ability to deliver the effect used to power the study. The paper then introduces a framework to calculate the 'average success probability' for a continuous endpoint and demonstrates how uncertainty about the treatment effect could affect the average success probability for a future confirmatory trial. In a later publication, Chuang-Stein

(2010), the concept of assurance was expanded to sample size re-estimation, the results highlighting again, that the traditionally designed phase 3 studies may have a much lower likelihood of success than originally anticipated.

In 2009, Nixon, Bansback, Stevens, Brennan and Madan (2009a) described a model to predict the six-month American College of Rheumatology (ACR) response rate based upon the ACR response rate collected at earlier time points, for trials in rheumatoid arthritis. Within this paper the data synthesis and modelling is performed within a treatment arm on the absolute scale which has the potential disadvantage of breaking the randomization within studies. This differs from the approach adopted within this PhD, which synthesizes the relationships between phase 2 and phase 3 study outcome measures on the relative treatment difference scale and therefore maintains the randomization within each study. Nixon et al (2009b) also expand on this research to develop a Rheumatoid Arthritis Drug Development Model (RADDM), which simulates proposed phase 2b and 3 trials based upon efficacy evidence on the ACR response rate at the end of phase 2a, evidence of efficacy from existing treatments and expert opinion on three key safety markers. Bayesian clinical trial simulation is then used to determine the assurances of licensing approval at the end of phase 3. Although the randomisation is broken in the Nixon example, this does provide a useful methodological example of predicting the PoS in the situation where single arm or non-randomised phase 2 trials are conducted with decision made from short term endpoints. It should also be noted that while the theoretical value of randomised phase 2's is presumed, primarily due to the

requirement for non-randomised designs to be interpreted relative to historical controls or through the use of indirect treatment comparisons which may be susceptible to selection bias and the impact of confounding factors (Cannistra et al (2009), Gan et al (2010), Sharma et al (2011)), a study by Monzon (2015) of 189 phase 3 studies found no evidence that randomised phase 2 studies were superior at predicting phase 3 success than single arm phase 2's.

In 2012 a method that uses predictive power to predict the probability of success in a phase 3 outcome measure (i.e. the Overall Survival (OS) log hazard ratio) based on a different phase 2 outcome measure (the Progression-Free Survival (PFS) log hazard ratio) was published by Hong (2012). This approach uses a bivariate normal model in phase 2 to estimate the joint distribution of the log hazard ratios on the two endpoints, and requires specification of a prior for the correlation between the treatment difference for the phase 2 outcome measure and the treatment difference for the phase 3 outcome measure. This approach aims to develops a relationship between endpoints estimated from a single trial which may not be sufficient to support predictions across trials. To overcome this issue, the modelling approach developed for the PhD differs from the Hong (2012) publication by its use of meta-regression techniques to estimate the relationship between the primary phase 2 outcome measure (PFS log hazard ratio) and the different primary phase 3 outcome measure (OS log hazard ratio).

Also in 2012, Claret (2012) synthesized observed phase 2 data with a public domain drug−disease modelling framework that had been built using the

individual patient level data relationship between changes in tumour size and overall survival (Wang, 2009), to simulate survival outcomes for phase 3 studies investigating motesanib or bevacizumab plus carboplatin/paclitaxel in first-line non−small-cell lung cancer. While an interesting approach, the model could not discriminate between failed and successful studies, suggesting that further enhancements were required to improve its utility for predicting phase 3 trial outcomes.

In 2013, while acknowledging the appealing nature of using the concept of assurance in phase 2 to 3 decision making, Carroll (2013) pointed out some of the nuances embedded within the assurance calculation to be aware of. One particularly interesting example being that the phase 3 assurance is capped at the 1 minus one-sided p-value observed in phase 2 even when phase 3 includes an infinite number of patients (assuming the phase 2 and 3 outcome measures are the same and additional prior opinion on the phase 2 outcome measure is not included). The literature on assurance was further developed in 2014 by Ren (2014), who enhanced the methodologies to accommodate survival outcome measures assuming both parametric and non-parametric models. Prior elicitation methods were also proposed for each survival model to enable the assurance calculations to be performed reliably and easily.

Also in 2013, Wang (2013) published on an approach from Eli-Lilly which involved the use of bayesian modelling to synthesize relevant data to create a posterior distribution for the treatment difference in the phase 3 endpoint which

was then used to determine the PoS in a phase 3 study. Three interesting examples from the oncology setting were presented:

1. An example where a randomized phase 3 study is in planning designed to compare an experimental regimen (E) with a control regimen (C) in a solid tumour. The primary endpoint is the overall survival (OS) time from randomization. Overall survival data from a preceding randomized phase 2 study was available to form the prior.
2. An example where two single arm studies of an experimental therapy (E) were available, in addition to historical data on the standard of care (C). The desire was to run a phase 3 trial to compare the combination of E+C against C. In this circumstance no prior data was available for the combination. The survival distribution for the historical data, and similarly the experimental treatment was determined using meta-analysis of the respective hazard functions (assuming an exponential distribution). These were then synthesised within a bayesian hierarchical model, and combined with various subjective assumptions about how the synergistic potential of the combination in order to generate a predicted posterior distribution of the treatment effect in a phase 3 study.
3. An example where a randomised phase 2 study had been conducted to compare the combination (E+C) versus control (C) in metastatic breast cancer. The primary endpoint of the phase 2 study was PFS. At the time of study completion only a few OS event were observed, however a changing regulatory landscape suggested the need to demonstrate OS

benefit in phase 3. In 2008, a meta-analysis (Sherrill, 2008) had been published that aimed to quantify the treatment effects with respect to the PFS HR and the OS HR in metastatic breast cancer. Using this data the relationship between the log PFS HR and the Log OS HR was modelled using meta-regression. This relationship was then use in conjunction with the PFS HR from the phase 2 to predict a posterior distribution for OS and thereby the PoS in phase 3. This later example closely resembles the methodology adopted to convert between phase 2 and phase 3 outcome measures within this PhD, and published within Sabin (2014 and 2015). Differences lie in the choice of functional form for the meta regression model, the choice of prior for the between study variance, and the additional incorporation of the prior opinion of key decision makers for the treatment difference in the phase 2 outcome measure.

Walley et al (2015) provide a further example on how the design and statistical analysis of a phase 2 study can be formulated into a bayesian framework using a worked example from chronic kidney disease. The bayesian framework incorporates methods to inform decisions on individual trials using informative priors on treatment effects and placebo response. The case study highlights the potential value of meta analysing and synthesising placebo data for use as an informative prior for the phase 2 study. Of note, this prior information is combined with the results of a randomised phase 2 study to predict the probability of success in phase 3 (where PoS is based on a bayesian analysis as opposed to a frequentist analysis of the phase 3 study) as long as there is

not a statistically significant difference between the expected result from the meta-analysis and the observed phase 2 placebo response (with a two-sided *p* - value of less than 0.05). If a difference was found this would imply discordance which would lead to the primary analysis being switched to one with a vague placebo mean prior. This would therefore lead to the observed phase 2 data being dominant in the decision making process and predicted probability of success in phase 3. Similarly, the use of borrowing historical data to enrich controls in clinical trials is further discussed in the publications by (Neuenschwander et al. 2010, and Viele et al. 2014). An alternative innovative approach explored within this PhD is to take the view that the observed treatment difference seen in the randomized phase 2 study is the best unbiased estimate available, whilst the variance of the randomised phase 2 treatment difference is inflated as the observed phase 2 control group response departs from the expected, thereby inserting additional uncertainty into the predicted probability of success in phase 3.

When implementing a bayesian decision framework one particular challenge is the construction of a prior for the treatment effect. Within this PhD sceptical, uninformative and optimistic priors for the phase 2 treatment difference are elicited from individual key decision makers. The sceptical and optimistic priors elicited are used to form bounds of belief for the probability of success in a future phase 3 study. If empirical evidence is available from previous studies then it may be natural for this to be used as the basis for a prior distribution. Spiegelhalter et al (2004) discuss several options for summarising external

evidence to form a prior distribution. These include various approaches for discounting the degree of influence of the prior data depending upon the degree of relevance or potential for bias. There is however the potential for there to be no or limited historical information on the treatment effect. Examples of this may occur when working with a treatment with a novel mechanism of action, or possibly a new outcome measure. In such circumstances subjective priors as used in this PhD (for the phase 2 treatment difference) may be required to be elicited from key decision makers. Walley (2015) follow a similar type of approach by eliciting subjective priors for the phase 3 treatment difference based upon the SHELF approach (Oakley 2010, 2017). This methodology has also been extensively adopted by GSK for any prior elicitation as described in Dallow (2018). The SHELF framework begins with the selection and training of experts to be involved and the generation of an evidence package to support the prior determination. Individual priors are then elicited from each individual expert in a masked fashion. These are then discussed and where possible a consensus prior created through agreement of what a rationale independent observer would determine after hearing the individual opinions.

When determining priors for the phase 3 study, consideration should be given to the potential need to adjust for over optimistic phase 2 results. The need to adjust the distribution of the observed phase 2 treatment effect was raised by Kirby (2012) and Wang (2006) because of two potential sources of bias. The first arises from selecting compounds with pre-specified favourable phase 2 results and using these results as the basis of treatment effect for phase 3

sample size planning. The second rises from projecting the phase 2 treatment effect onto the phase 3 population when this may be typically optimistic due to a more heterogeneous patient population used within phase 3. In an attempt to reduce the impact of these two sources of bias, Kirby (2012) proposes discounting the phase 2 estimate of treatment effect through making simple multiplicative or additive adjustments. Wang (2006) recommended an alternative approach discounting the observed treatment effect by one standard error of the treatment effect estimate. Additional approaches to downweight the phase 2 treatment effect were also considered by Saint Hiliary (2019) including the use of a mixture prior which combines a conjugate vague normal prior with the observed phase 2 distribution weighted by an assumed prior probability that the phase 2 outcome measure does not predict the phase 3 outcome measure. As briefly discussed earlier, within this PhD a different approach is taken whereby the observed treatment effect from a randomised phase 2 study is assumed to the best unbiased estimate of the treatment effect available, however additional uncertainty is incorporated by inflating the variance of the treatment effect based upon a function of difference between the expected and observed treatment effects on the control arm. Moreover, population changes are handled through the bayesian modelling framework. The PhD implicitly assumes that data is available to make such adjustments which may not always be the case. The concept of over-optimistic results within exploratory subgroups is discussed in Götte (2017) which is a result of the practice of performing exploratory biomarker driven subgroup analyses to identify patient

populations with a beneficial treatment effect. Selecting a subgroup based upon a large observed effect may lead to over-optimistic expectation on the probability of success of a phase 3 trial. Gȍtte highlights how the use of approximate bayesian computational techniques can be used to adjust for bias in this setting.

Frewer (2016) discuss an alternative decision making framework being applied within early development at AstraZeneca that builds on an approach originally proposed by Lalonde (2007). This methodology compares the observed upper and lower confidence intervals for the treatment effect with a lower reference value (LRV, the smallest clinically meaningful treatment effect leading to a commercially viable treatment) and a target value (TV, the commercially desired effect to establish the treatment as a lead treatment in the market). Typically a three-outcome decision framework is then constructed in line with pre-specified risks. For example,

- Go to the next phase : PCT_{20} >LRV and PCT_{90} >TV
- Consider / collect more data : $PCT_{20} \leq$ LRV and PCT_{90} >TV
- Stop development : $PCT_{90} \leq$ TV

where PCT_x denotes the xth percentile of the distribution of the observed treatment difference. This methodology is further expanded in Pulkstanis (2017) who applies a bayesian decision framework allowing for multiple levels of the targeted efficacy in line with the target product profile. Moreover, Dunyak (2018) expand this on this methodology presented in Frewer (2016) to include

dose response estimation into the decision framework. The approach incorporates a set of dose response models and uses model averaging. Whilst such approaches provide a useful standardised decision framework they are independent of the potential actions required to be taken in a subsequent phase of development, for example potential changes in endpoint or population. Further work expanding these concepts to account for such factors would be useful.

Similarly, Huang (2019) also presented a quantitative bayesian/frequentist decision framework for Go/No-Go criteria and sample size evaluation in phase 2 randomized studies with a time-to-event endpoint. This approach followed a similar integrated quantitative modelling approach to that used in the PhD with the exception that both the phase 2 and phase 3 trials shared a common endpoint while allowing a discount of the observed phase 2 data. This decision framework also incorporated criteria based upon the observed effect size and its precision against a target value and a minimally acceptable value in a similar way to Lalonde (2007) and Frewer (2016). Under the situation presented with no change of endpoint, the results showed that an increase in the sample size of a phase 2 trial will result in greater increase in the PoS of a phase 3 trial than increasing the phase 3 trial sample size by equal amount. Jiang (2011) also discussed the potential to optimise the phase 2 and 3 sample size using a combined bayesian and frequentist framework. This article proposed a PoS function to allow an integrated approach to the sample size and go/no-go decision criteria for a phase 2/3 program in which both the phase 2 and 3 trials

share a common, normally distributed response variable. Gȍtte (2015) also describe an approach for planning a phase 2 trial in a time-to-event setting that considers the whole phase 2/3 program. The article expands on the work on Jiang (2011) through the inclusion of stopping boundaries after phase 2 that minimise the number of events under select conditions for the conditional probability of a correct go/no-go decision after phase 2, as well as the conditional success probabilities for phase 3. Simulation is used show that unconditional probabilities of go/no-go decision as well as the unconditional success probabilities for phase 3 are influenced by the number of events observed in phase 2. Moreover, it highlights that a phase 2 sample size providing more than 150 events may not be necessary as the impact on these probabilities then becomes quite small.

Saint-Hilary (2019) extended a framework to determine the probability of success in a future phase 3 trial to the situation where prior information can be borrowed from multiple early phase endpoints, as opposed to a single phase 2 endpoint. In their approach the joint distribution of the treatment effects from the multiple early endpoints and the clinical endpoint are developed using a meta-analytic approach from external data. This is used in conjunction with the prior observations on the early phase endpoints to build an informative prior distribution for the phase 3 outcome measure. After incorporating the properties of the future phase 3 design, the predictive distribution of the phase 3 outcome measure is then used to calculate the PoS of the phase 3 trial. An example is shown from the multiple sclerosis.

In addition to becoming part of practice at Amgen and AstraZeneca, Crisp (2018) described how the use of assurance embedded within a quantitative bayesian modelling framework as a measure of the PoS has also now become routine practice at GlaxoSmithKline across all therapeutic areas. The key benefits highlighted being:

- The estimated PoS is more meaningful than focussing on the power of the phase 3 study, as it incorporates current knowledge about the treatment effect. An observed consequence of this is that it leads to focused discussions on study design and objectives.
- The probability of observing clinically meaningful effect sizes is explicitly characterised in advance.
- The assumptions that are driving the calculated PoS are transparent to all. The process of empirically determining or eliciting these assumptions can lead to important refinements in study design. For example, the relationship between PoS and sample size can be assessed such that an optimal sample size can be identified to balance risks and cost.

It is implicitly assumed through the drug development process that the selected phase 2 outcome measure will be predictive of the phase 3 outcome measure. In many circumstances the phase 2 outcome measure may not strictly fulfil the criteria (Prentice 1989) to be labelled as a formal surrogate endpoint. It is however expected that the choice of phase 2 outcome measure will have biological plausibility and that evidence of the prognostic value of the endpoint will be supported by epidemiological data or clinical trials. It is also important to

note that many of the proposed bayesian modelling frameworks (e.g., Nixon (2009b), Hong (2012), Claret (2012), Wang (2013), Sabin (2014, 2015), Saint-Hilary (2019)) incorporate the uncertainty in this relationship in their decision framework. While formal surrogacy is not deemed necessary for the modelling framework, a stronger relationship between the phase 2 and 3 outcome measures will provide greater confidence in the end of phase 2 decision. Buyse (2016) provide a statistical evaluation of surrogate endpoints for use in cancer trials that is a useful guide in selecting the phase 2 endpoints for a variety of cancer types. In addition, the relationship between the PFS HR and OS HR has been examined in several oncology cancer types (including pancreatic, gastric, STS, NSCLC, mBC, mCRC, esophageal, rectal, 1st-line ovarian, CRPC, SCLC, advanced NSCLC, advanced neuroendocrine neoplasms) in the literature. This work greatly facilitates both the use of a meta-analytic modelling approach to determine the PoS in future phase 3 studies, and further promotes understanding of the confidence drug developers have in the choice and interpretability of phase 2 endpoints.

Key Research Area 2:

- The incorporation of qualitative factors in the decision making process, and the implementation of a PoS framework in the situation where very limited or no prior clinical data is available.

Benda (2010) presents general considerations for comparing competing options for clinical programs, trial designs and analysis methods as a basis for decision making, highlighting that final decision making should incorporate qualitative factors as well as quantitative methods. This paper differs to the research presented in Section 3.9 and 3.10 as it does not broach the topic of how to synthesise the qualitative or quantitative factors into an overall PoS.

The situation where limited or no prior clinical data is available for an indication may arise in the life cycle management (broadening the use of the drug into new lines of therapies or disease indications) of oncology products that target patients with specific genes or biomarker targets. With the exception of the potential to use subjective formal priors (Walley (2015), Oakley (2010, 2017), Dallow (2018)) no further prior literature specifically focussing on predicting the PoS under this scenario was identified.

Key Research Area 3:

- The determination of the expected efficacy of the investigational treatment and how this compares and ranks to the current and future treatment options available for patients.

No prior literature focussing specifically on developing models to enable early predictions (to be used as part of the EOP2 decision) of the expected efficacy of the investigational treatment at the end of a successful phase 3 study and how this will compare and rank to the current and future successful treatment options available for patients was identified as part of the background review. Within

this PhD a bayesian re-sampling approach was used to determine the predicted probability that a treatment will rank 1st, 2nd or 3rd. Additional graphical and numerical approaches to summarise the results of these treatment ranking probabilities are described in Salanti (2011). Of note, in order to supplement a graphical display of cumulative ranking probabilities and as an alternative way to rank the treatments overall, Salanti uses the estimate of the surface under the cumulative ranking (SUCRA) line for each treatment as a simple numerical summary. Rücker (2015) built on the work by Salanti presenting a frequentist analogue to SUCRA called the P-Score. This approach for comparing treatments in a network meta-analysis does not use a re-sampling approach.

The main challenges of network meta-analysis and the use of indirect comparisons associated with evaluating the assumptions underlying the statistical synthesis of direct and indirect evidence follow through into the work conducted for this PhD. A comprehensive overview of these assumptions including the statistical and nonstatistical methodological considerations are discussed in Salanti (2012) and in Chapter 11 of the Cochrane handbook (Higgins et al (2011, 2019).

3.2. STAGE 1: Systematic Literature Review and Data Abstraction

For any disease area under investigation, the process begins (ideally prior to starting phase 2) with a literature review focused on identifying published trials relevant to addressing the ten core questions highlighted below. A thorough

systematic review of the literature is required to ensure that all appropriate historical trials and data are selected to support the decision making process. Guidance on conducting systematic reviews can be found in Moher et al. (2009), and Higgins et al (2011, 2019). The inappropriate inclusion or exclusion of trials will have a direct effect on the quality of decisions made.

The answers to the following questions form key inputs into the drug development decision making process. They provide information that supports the appropriate selection of study design, population and outcome measures to be used in the phase 2 and 3 studies. They enable the relationship between the phase 2 and 3 outcome measures to be determined, and indirect treatment comparisons against competitor drugs to be made.

The 10 core questions are:

1. Has the definition of disease changed over time? For example, in the hope of extending treatment benefits to early disease, over time common disease definitions may have been changed by lowering the threshold value or changing a component of the diagnosis.
2. What are the important prognostic markers for the targeted indication? Prognostic factors are important to help define the study population and objectives.
3. What is the expected absolute treatment effect of the current standard of care, and other drugs either marketed or in development, for both the phase 2 and primary phase 3 outcome measures? This information helps the

design and interpretation of the phase 2 and 3 study design and feeds into the target product profile (TPP) for the new investigational product.

4. What are the observed treatment effects of other drugs relative to the standard of care for both the phase 2 and primary phase 3 outcome measures? This information helps guide study design and interpretation. Understanding the relationship between phase 2 and 3 outcome measures is critical to the decisions to be made from the phase 2 study.
5. What is the impact of the important prognostic factors on the ability to detect a treatment effect relative to the standard of care in both the phase 2 and primary phase 3 outcome measures (i.e., is there a prognostic factor by relative treatment effect interaction)? If the phase 3 study enrols a different population to the phase 2 study this may lead to a different expectation of the treatment effect in the phase 3 study. This therefore needs to be considered when designing and determining the PoS of phase 3.
6. Have the absolute and relative treatment effects changed over time? Changes in clinical practice over time may also impact the expected treatment effects. This is critical to consider in the design and interpretation of results.
7. What are the most common side effects and their expected incidence rate? As well as efficacy it is also important to consider the safety and tolerability of the investigation drug and to be able to compare to competitor drugs.
8. What are the relationships between the phase 2 and primary phase 3 outcome measures for individual treatment groups?

9. Are these relationships between outcome measures likely to hold for drugs with different modes of action?
10. What is the impact of previous treatments on the relationships between the phase 2 and 3 outcome measures?

Questions 8, 9 and 10 are all important to design and interpretation. If the relationship between outcome measures differs with different treatments, drug classes or prior treatment exposure then this needs to be carefully considered in the design and interpretation of the phase 2 and 3 studies.

Moreover, it is very often desirable for phase 2 studies and the associated EOP2 decisions to utilise short term intermediary outcome measures for efficacy as opposed to the those requiring longer follow up used in phase 3. However, when selecting the outcome measures it is also important to consider the future context of use. For example, within oncology, outcome measures can serve different downstream purposes. They may be selected purely for internal sponsor decisions or to support accelerated or full approval. As discussed in the 2018 FDA guidance on the clinical trial endpoints for the approval of oncology drugs and biologics, the determination of outcome measure will be based on the specific disease and is highly dependent upon factors such as effect size and duration, the number of complete responses, other available therapy, the disease location and setting, the clinical consequences of delaying or preventing disease progression or delaying administration of more toxic therapies, and the risk-benefit relationship. Table 3.1 summarises some of the advantages and disadvantages associated with a selection of key endpoints (PFS, PFS at a

certain timepoint, objective response rate (ORR) and overall survival) often used in phase 2 oncology studies.

Endpoint	Advantages	Disadvantages
PFS (hazard ratio)	1. Likely leads to earlier decisions than overall survival studies. 2. Likely smaller sample size than overall survival studies. 3. Typically objectively and quantitatively assessed. 4. Includes stable disease in the assessment.	1. Potential for assessment and ascertainment bias. 2. Definitions vary between studies 3. May not always correlate with overall survival.
PFS Rate at certain timepoint	1. As PFS above. 2. Likely leads to earlier decisions than PFS studies.	1. As PFS above. 2. Doesn't factor in the entire PFS curve, so excludes information. 3. Likely not an approvable endpoint.
ORR	1. Leads to earlier decisions than OS studies. 2. Typically smaller sample size than OS studies. 3. Effect on tumor is attributable to drug and not natural history. 4. Typically objectively and quantitatively assessed.	1. Definitions vary between studies. 2. May not correlate to overall survival. 3. Tumor measurements may be imprecise in certain locations where there is a lack of demarcated margins.
OS	1. Easily, objectively and precisely measured.	1. Generally requires larger sample size than other oncology endpoints. 2. Requires long follow-up. 3. May be impacted by treatment switching post progression.

Table 3.1: Advantages and Disadvantages of Select Oncology Endpoints

As described in Section 3.1.1, while the uncertainty in the relationship is captured through the modelling process it is implicitly assumed that the selected phase 2 outcome measure will be a reasonable predictor of the phase 3 outcome measure. The choice of endpoint should have good biological plausibility, with its prognostic value supported by evidence from prior epidemiological data or clinical trials.

Once the outcome measures are defined, the relevant data are then systematically abstracted from the published literature and synthesized using standard meta-analytical techniques. If the data abstraction was first performed at the phase 2 design stage, the data should be updated in preparation for the EOP2 meeting to ensure that any new information is incorporated into the EOP2 decision. Given a large part of the EOP2 decision revolves around predicting the treatment difference in the phase 3 study from the treatment difference observed in the phase 2, the focus is placed on the development of a statistical (usually meta-regression) model to relate the treatment differences seen with the phase 2 outcome measure to treatment differences seen with the phase 3 outcome measure. This relationship forms part of the prior knowledge required for the statistical model used for predicting the probability of success in phase 3 described in stage 2.

In many situations where the phase 2 and phase 3 outcome measures are different, e.g., phase 2 studies may use short term intermediate outcome measures, rather than the longer-term outcome measures needed for regulatory approval, using meta-regression to investigate their relationship provides an

understanding of how good the outcome measure selected for phase 2 may be at predicting the phase 3 outcome measure. Determining this relationship, from completed studies where both the phase 2 and phase 3 outcome measures have been collected, on the relative treatment effect scale is particularly valuable. This enables the development of a model for estimating phase 3 outcome measure differences from any given fixed value for the phase 2 outcome measure difference. For example, assuming the relationship is linear and passes through the origin, the model to be fitted is:

$$\mu_i = \beta z_i \qquad (1)$$

where μ_i is the true phase 3 outcome measure treatment difference in the *ith* study, z_i is the phase 2 outcome measure treatment difference in the *ith* study, and β the slope of the regression line.

It should also be noted that the functional form of the model does not have to be linear or forced through the origin. In many scientific situations it may often be expected to choose a model that is forced through the origin. In the context of relating a treatment difference in phase 2 outcome measure to a phase 3 measure this would imply that no change in the phase 2 measure equates to no change in the phase 3 measure. However, as with any model it is simply a matter of choosing a functional form that makes sense. If the treatment differences are believed to be dependent upon certain important prognostic characteristics of the studies included in the meta-analysis, the above model may be further expanded to include the prognostic characteristics as trial level covariates. For example:

$$\mu_i = \alpha P_i + \beta z_i$$

where μ_i is the true treatment difference in the phase 3 outcome measure in the *ith* study, P_i is the prevalence of a key prognostic factor in the *ith* study, z_i is the phase 2 outcome measure treatment difference in the *ith* study.

The meta-data consist of *k* study estimates of phase 3 outcome measure treatment differences, $\hat{\vartheta}_i$, with variance ε_i^2 *(i=1 to k).* To utilize a random effect model, we assume $\hat{\vartheta}_i$ is normally distributed with mean μ_i, such that:

$$\hat{\vartheta}_i \sim N(\mu_i, \varepsilon_i^2), \qquad i = 1 \; to \; k.$$

and that μ_i is itself a realisation of a normally distributed random variable reflecting that the true effect might differ from study to study with variance τ^2, such that:

$$\mu_i \sim N(\varphi_i, \tau^2)$$

In my application of this method a bayesian approach is taken where β and τ are considered as hyperparameters with independent prior distributions. A non-informative prior N(0,10^4) is given to β. Lambert, Sutton, Burton, Abrams and Jones (2005) and Spiegelhalter et al (2004) highlight the importance of carefully selecting the prior for τ. The choice of prior for τ should be made following a review of the data, and sensitivity analyses conducted using a range of realistic vague prior distributions. For the pancreatic cancer example described later we selected a uniform (0,2) prior.

The posterior distributions for β and τ can be approximated using Markov Chain Monte Carlo (MCMC) methods.

While the use of meta regression to determine the relationship between the phase 2 and phase 3 outcome measures on the relative treatment effect scale is clearly valuable, there are a number of potential limitations to consider when applying the methodology.

- The sample size needs to be sufficient. Where estimation methods are based upon asymptotical assumptions they can easily be biased when the sample size is small. It is recommended that meta-regression should generally not be considered when there are fewer than ten studies available (Higgins et al, Cochrane handbook 2019).
- Published papers may not always measure or appropriately report the information on covariates needed for the model, preventing the opportunity to appropriately adjust for confounding. Additionally, even if the number of studies is moderately large and the information on confounders is present the characteristics of the studies may be correlated leading to problems of collinearity.
- When applying meta-regression it should be noted that the while the association discovered may be real, it could also be driven by unmeasured confounders or other differences between selected studies.
- Literature reviews are suspectable to publication bias. Moreover, as discussed in the pancreatic cancer example in Section 3.6, care should be taken when including data from conference abstracts and presentations

as these can be susceptible to change. Whilst the choice to use these data for internal decision making is the sponsors risk it is recommended that sensitivity analyses are performed to explore the impact on the results and decisions.

- Meta regression uses the individual study as the unit of observation. There is however no logical imperative that the association seen at the study level reflects the association at the individual patient level, or vice versa (the ecological fallacy). It is therefore important to note that the association observed at the trial level should not be considered in the management of an individual patient (Korn 2005).
- Conventional random effects methods ignore the imprecision in the between-trial variance estimate. In the modelling described, a bayesian approach using a vague prior is adopted to account for the imprecision. While this is preferable in principle, especially when the number of trials is small or when the between-trial variance is estimated as zero, the resulting widening of the predicted confidence intervals may still be rather slight in practice.
- The model described assumes that the relationship between the treatment differences for the phase 2 and phase 3 outcome measures, determined from the systematic review, will apply to the new drug being tested. It is therefore recommended that sensitivity analyses are performed to assess how robust predictions are for departures from this relationship. The most practical way of doing this will be to split the data by subgroups. For

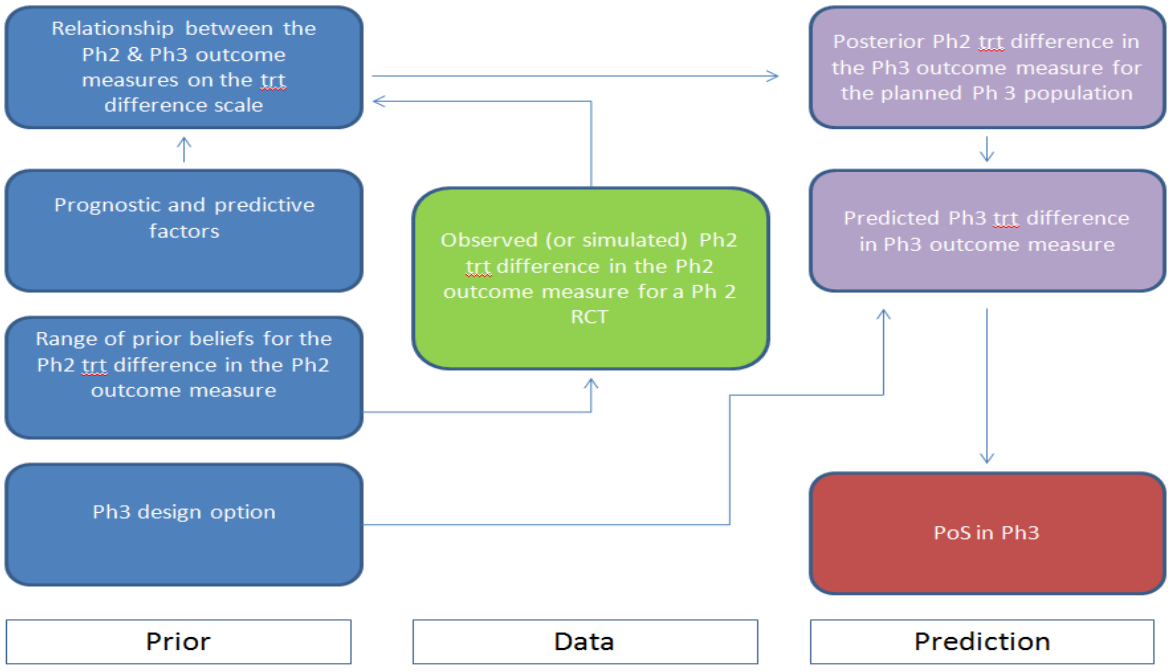
example, if data permit, relationship may be determined from selected studies on treatments with similar modes of action to the investigational drug.

3.3. STAGE 2: Statistical Model for Predicting the Probability of Success (PoS) in Phase 3

Prior to investing in a large costly phase 3 programme, it is important for pharmaceutical decision makers to understand the probability of success in the proposed phase 3 study. After completing the relevant data abstraction in accordance with the core questions described in Section 3.2, and conducting the meta-regression to determine the relationship between the treatment difference in the phase 2 and 3 outcome measures, the next step described herein is to synthesise the data collected into a model that enables the probability of success in a future frequentist phase 3 study to be calculated. The definition of success may of course differ depending upon the objectives of the phase 3 study. In general, this is likely to be the probability of achieving a favourable statistically significant efficacy result from an appropriate statistical test designed to reject, or not, the null-hypothesis of no difference between treatments, on the primary phase 3 outcome measure at the conventional 5% 2-sided level of significance.

A general overview of the statistical model used to estimate the PoS of the compound is provided in Figure 3.1. The model synthesises the following information:

- The relationships between phase 2 and phase 3 study outcome measures (on the relative treatment difference scale).
- The influence of prognostic factors on the relationship.
- The observed treatment difference for the phase 2 outcome measure in the phase 2 study.
- A range of prior opinions of key decision makers for the treatment difference in the phase 2 outcome measure.
- Knowledge of the proposed phase 3 study design.



Ph: Phase; trt: treatment; RCT: Randomized Controlled Trial; PoS: probability of success.

Figure 3.1 Statistical Model for Predicting the Probability of Success in Phase 3

A predominantly bayesian approach is used to the perform the prediction as described in the following 4 steps.

Step 1: The first step in the process is to obtain the phase 2 study results. The observed (or simulated) treatment difference in the phase 2 outcome measure may be expressed as:

$$d \sim N(\delta, \sigma_d^2) \qquad (2)$$

where d is the observed treatment difference, σ_d^2 is the observed variance of the treatment difference, and δ is the true treatment difference in the phase 2 outcome measure.

Step 2: A range of prior statistical distributions for the phase 2 outcome measure treatment difference reflecting differing opinions of key decision makers is elicited. Following Spiegelhalter, Freedman and Palmer (1993), the following three priors were chosen for δ:

- A non-informative prior distribution $\delta \sim \mathrm{N}(\delta_\mathrm{n}, \sigma_\mathrm{n}^2)$
- An optimistic prior opinion $\delta \sim \mathrm{N}(\delta_\mathrm{o}, \sigma_\mathrm{o}^2)$
- A sceptical opinion $\delta \sim \mathrm{N}(\delta_\mathrm{s}, \sigma_\mathrm{s}^2)$.

When combined with the observed phase 2 data, these priors lead to the development of reasonable bounds of evidence for the PoS in phase 3, as described in Step 3. Taking the prior belief of key decision makers into account was an important component in the successful implementation of this framework. In addition to using the observed evidence, the ability to state at the

EOP2 meeting that the PoS in phase 3 is still above a certain value taking the point of view of the most sceptical decision maker, or that the PoS is still below a certain value, taking into account the view of the optimist decision maker, brings valuable perspective to the go / no-go discussions. These priors are combined with the results from equation 2 to obtain a range of posterior statistical distributions for the treatment difference in terms of the phase 2 outcome measure which are given by:

$$\boldsymbol{p_j(\delta_j|d)} = \boldsymbol{N}\left[\frac{\frac{\boldsymbol{\delta_j}}{\boldsymbol{\sigma_j^2}} + \frac{\boldsymbol{d}}{\boldsymbol{\sigma_d^2}}}{\frac{\mathbf{1}}{\boldsymbol{\sigma_j^2}} + \frac{\mathbf{1}}{\boldsymbol{\sigma_d^2}}}, \quad \frac{\mathbf{1}}{\frac{\mathbf{1}}{\boldsymbol{\sigma_j^2}} + \frac{\mathbf{1}}{\boldsymbol{\sigma_d^2}}}\right]$$

where j = n, o, s (representing the non-informative, optimistic and sceptical distributions). It is assumed that each of these posterior distributions can be represented by the following notation, where the parameter d_{j2} represents the posterior distribution of the phase 2 outcome measure treatment difference at the end of phase 2, distributed with mean Υ_{j} and variance σ^2_{jd2}:

$$d_{\mathrm{j2}} \sim N\left(\Upsilon_{\mathrm{j}}, \sigma^2_{\mathrm{jd2}}\right) \qquad (3)$$

Step 3: The distribution of our expected phase 2 outcome measure treatment differences (equation 3) and the relationship between the treatment differences for phase 2 and phase 3 outcome measures determined using meta-regression from the systematic review (equation 1) are now used to form distributions for the phase 3 outcome measure treatment difference. This is constructed using MCMC simulations as follows:

- Sample a value $d_{j2}^{(n)}$ from the distribution d_{j2} (equation 3) and pass the result through the meta-regression model derived in Section 3.2 to determine the predicted phase 3 outcome measure treatment difference at the end of phase 2, represented by $d_{j23}^{(n)} = \beta . d_{j2}^{(n)}$. Here, β is the slope from the random effects meta-regression considered to be normally distributed with mean b and variance σ_b^2, that is $\beta \sim N(b, \sigma_b^2)$.
- The process is repeated (n) times, ensuring the uncertainty in the phase 2 outcome measure treatment difference is incorporated, to estimate the posterior predicted distribution of the phase 3 outcome measure treatment difference at the end of phase 2. This is represented by:

$$d_{j23} \sim N\left(\mu_j, \sigma_{j23}^2\right)$$

In this bayesian approach, three different prior distributions for the phase 2 treatment difference (representing sceptical, optimistic and non-informative opinions of key decision makers) are used to form three different predictive distributions for the phase 3 outcome measure treatment difference. If the relative treatment differences are believed to be dependent upon certain important prognostic characteristics of the studies included in the meta-analysis, then the model for the relationship between phase 2 and 3 outcome measures should also include the prognostic characteristics as trial level covariates. The predictive distributions of the phase 3 outcome measure treatment difference may then be estimated from the relationship by predicting the response for the planned characteristics of the phase 3 study.

Step 4: The range of predictive distributions for the phase 3 outcome measure treatment differences are used to simulate the results of the proposed phase 3 study design and estimate the PoS for the future frequentist phase 3 study. The earlier inclusion of sceptical and optimistic priors leads to the development of reasonable bounds of belief for the PoS.

The phase 3 study design parameters considered include the required level of statistical significance, the desired size of the treatment difference (which is based purely on the minimum clinically and commercially desirable effect, and not the posterior predicted distribution of the phase 3 outcome measure calculated in step 3), variance and the trial sample size.

The PoS may be determined by again using MCMC simulation. Assuming the predictive distributions for the phase 3 outcome measure treatment differences found in step 3 are normally distributed $d_{\mathrm{j}23} \sim N\left(\mu_{\mathrm{j}},\ \sigma_{\mathrm{j}23}^2\right)$, and the future phase 3 data are normally distributed $X_3 \sim N\left(d_{\mathrm{j}23}, \sigma_3^2\right)$ then:

1. Sample a value for the phase 3 treatment difference, $d_{\mathrm{j}23}^{(\mathrm{n})}$ from the posterior distribution in step 3.
2. Sample a value $X_3^{(\mathrm{n})}$ given $d_{\mathrm{j}23}^{(\mathrm{n})}$.
3. Calculate the confidence interval for $X_3^{(\mathrm{n})}$ and determine whether this represents a favourable statistically significant result.

4. Repeat (n) times and determine the proportion of statistically significant outcomes.

In addition to simulating the PoS, this approach can be extended to simulate the probability of observing a particular outcome in the phase 3 trial; that is, the probability $P(X_3 > y)$ that the phase 3 treatment difference is greater than a selected value (y) of interest.

3.4. STAGE 3: Informing the PoS Threshold for a Go or No-Go Decision

Section 3.3 focused on the calculation of the PoS in a future phase 3 study at the end of phase 2. In this section, methodology is presented that is designed to support the risk-informed selection of a PoS threshold. Determining a PoS threshold is a critical part of the decision making process because the following question will need to be answered by the sponsor/funder:

> "What magnitude of PoS (PoS threshold) should I be looking for to make a go decision?"

The PoS threshold therefore represents the minimum acceptable PoS that would lead to a 'commit to phase 3' decision. The selection of such a threshold will be specific to a funder/sponsor and their current portfolio. Factors including the unmet need of patients with a particular disease, the current financial state of the business, and the potential return on investment will determine if decisions makers will take more or less risk, and thereby select a higher or

lower PoS threshold. These considerations are further emphasised in Crisp (2018). Crisp also highlights that decision makers should not make the mistake of thinking they must seek to change the underlying study assumptions to force the assurance to be high. It may be useful instead to act on a low assurance estimate by considering options for futility interims, or in extreme cases, considering whether the study should actually go ahead.

Due to the important of this decision it should be risk informed. In this section, a simulation approach developed to support the risk informed selection of the PoS threshold is presented. The methodology developed focuses on simulating the following risks based on different 'PoS commit to phase 3' decision thresholds and hypothetical truths for the treatment difference in the phase 3 outcome measure:

- The probability of making a go decision and failing phase 3
- The probability of making a go decision and achieving success in phase 3.

It should be noted that while this methodology requires details of the phase 2 and 3 designs, it does not require the observed phase 2 results and, therefore importantly, can be done ahead of time in preparation for the end of phase 2 meeting. The range of hypothetical truths over which the simulations are conducted span the EOP2 predicted posterior distributions of the phase 3 outcome measures (sceptical, optimistic and uninformative). Moreover, they should also include the 'no treatment' effect and the 'minimum commercially viable' treatment effect.

The steps taken are as follows:

1. Select a hypothetical truth for the treatment difference, X, for the phase 3 outcome measure.
2. Calculate the corresponding treatment difference in the phase 2 outcome measure by utilizing the relationship between the treatment differences in phase 2 outcome measure and phase 3 outcome measure, $\delta = \mathrm{X}/\beta$.
3. In accordance with the phase 2 study design, sample a value for the phase 2 treatment difference, $d^{(n)}$, in the phase 2 outcome measure assuming $d \sim N(\delta, \sigma_d^2)$.
4. Follow Stage 2 and calculate the predicted PoS in the planned future frequentist phase 3 study selected to ensure the study has a fixed power (e.g. 80 or 90%) to detect the minimum clinical and commercially desirable treatment difference.
5. Repeat n times.
6. For any given PoS go / no-go threshold calculate the proportion of simulations resulting in go and no-go decision. This is the probability of making a go p(Go) or no-go P(No-go) decision.
7. For the selected frequentist phase 3 study calculate the probability that it results in a statistically significant phase 3 outcome, P(P3+ve), for the selected hypothetical truth. P(P3+ve) therefore represents the true PoS of the phase 3 study.
8. Calculate the probability of making a go decision at the EOP2 and being successful in phase 3 as P(Go+ve) = P(Go)*P(P3+ve).

9. Calculate the probability of making a go decision and failing in phase 3 as P(Go-ve) = P(Go)*(1-P(P3+ve)).
10. Repeat for a range of hypothetical truths for the treatment difference in the phase 3 outcome measure and PoS go / no-go criteria. Plot the P(Go+ve) and P(Go-ve) against the hypothetic truths for each PoS criteria. The sponsor may then select a criterion which meets the risks they are willing to take. It should also be noted that reasonable bounds of belief for P(Go+ve) and P(Go-ve) can be generated through the incorporation the sceptical and optimistic priors into step 4 above.

An example in the pancreatic cancer indication is shown in Section 3.6.3.

3.5. Further Optimising the Design of the Phase 2 and 3 Studies

The methodology presented to predict the PoS in Section 3.3 assumes that the phase 2 results have already been observed. While it may be more usual that the process begins after the phase 2 study has already started or been completed, ideally the first step would begin prior to starting phase 2. A range of plausible treatment differences for the phase 2 outcome measure can be simulated and in turn used to evaluate the PoS for the planned phase 3 study. Repeating this process using different design options for the phase 2 and 3 design can be used to optimize the development strategy around the PoS in phase 3 for a particular indication.

The approach can also help to decide among different development strategies for a specific indication. For example, strategy A could be to conduct a small phase 2 study and begin a phase 3 study but with an early futility analysis. Strategy B could be to conduct an interim analysis within a phase 2 study with a possible decision to begin the phase 3 study immediately, or wait until completion of the phase 2 study to decide on beginning a phase 3 study. Using the process to produce phase 3 predictions and EOP2 decisions in simulations for therapies that have a range of efficacy could be used to decide the best strategy. In some instances, it may be that only minimal phase 2 evidence will be enough to make an adequate decision, whereas other instances may require much larger phase 2 studies.

3.6. Worked Example: Predicting The PoS In Pancreatic Cancer

A worked example of predicting the probability of success is now presented in a pancreatic cancer study. Gemcitabine is indicated for use in the first-line treatment of advanced pancreatic cancer and is the most commonly used treatment in this setting. Here we assume it is being used as the control group for a randomized phase 2 study, and is planned to be used for a future phase 3 study. The primary outcome measure being used in the phase 2 study is progression free survival (PFS), a short-term outcome measure for the phase 3 regulatory outcome measure, overall survival (OS). The EOP2 decision is based upon treatment differences expressed in terms of hazard ratios. We

focus on predicting the distribution for the OS hazard ratio in a planned phase 3 study from the distribution of the PFS hazard ratio observed in phase 2. The modelling is performed on the log hazard ratio scale to allow the use of the normal distribution and is transposed back onto the hazard ratio scale for presentation purposes. The presence of metastases and Eastern Cooperative Oncology Group (ECOG) performance status are both considered to be prognostic factors for OS. The phase 2 study is being conducted in a 100% metastatic patient population with an ECOG performance status of 0 or 1 at randomization. In this example, the PoS in a future phase 3 study reflects the probability of rejecting the null hypothesis that the log OS hazard ratio (HR) =0 (in favour of the experimental treatment) at a 2-sided 5% level of significance.

3.6.1. STAGE 1: Systematic Literature Review and Data Abstraction

A thorough systematic literature review was conducted by Antony Sabin to identify all published randomized pancreatic cancer trials over the period from 2000 - 2012 in which gemcitabine was used alone or in combination with other therapies.

The specific objectives of the search were:

- To understand the progression-free survival and overall survival of subjects treated with gemcitabine alone
- To understand the relationship between the hazard ratio for progression free survival and the hazard ratio for overall survival

- To understand the distribution of treatment effects seen with other experimental drugs in combination with gemcitabine over gemcitabine alone.

Studies included in the literature review were identified using the following criteria:

- All randomized controlled comparative studies that were published in English in year 2000 or later, in which gemcitabine was used either alone or in combination with other therapies
- Adult patients (≥ 18 years of age) with locally advanced or metastatic pancreatic cancer
- Interventions of gemcitabine alone or gemcitabine-based combination chemotherapy
- First-line treatment for pancreatic cancer
- Outcome measures of overall survival (OS) data included in the report.

Studies to be excluded from the literature review were identified using the following criteria:

- Studies where patients were given concurrent radiotherapy or local regional modalities such as surgery, which might have influenced survival
- Cross over studies where the assessment of survival times was impaired
- Non randomized study
- Studies published in a language other than English
- Studies where information on patient survival times was not available.

Studies were identified by targeting MEDLINE, EMBASE, the American Society of Clinical Oncology web site, published meta-analyses and the internal knowledge of Amgen's clinical and regulatory groups. The following search terms were used to select the literature in MEDLINE and EMBASE:

- (Advanced or metastatic or unresectable) and (pancreatic and gemcitabine) and random*
- Year of publication 2000 to current date
- Remove duplicate information.

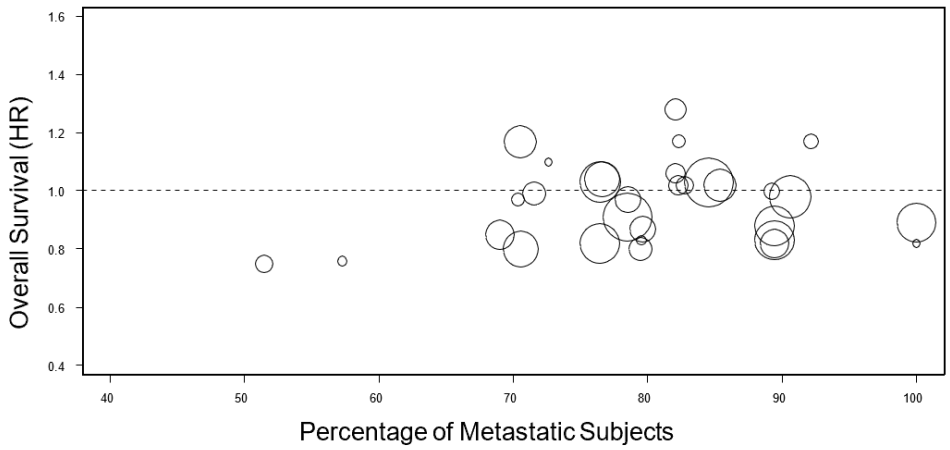
Once a final list had been determined the data were systematically abstracted from the literature and independently quality checked by statisticians Antony Sabin and Sarah Bray. The selected literature references can be found in Annex A and the associated data abstracted can be found at Annex C respectively.

In total forty-three studies were selected for detailed analysis. The methods in Tierney, Stewart, Ghersi, Burdett and Sydes (2007) were used to estimate the median survival, hazard ratio and their associated standard errors for the PFS and OS outcome measures. Data that were relevant to predicting the gemcitabine control group PFS and OS, the relationship between median PFS and median OS, the relationship between the PFS hazard ratio and the OS hazard ratio, and the pattern of treatment differences seen with important study level covariates were synthesized. It should be noted that data from 7 of the 43 studies (Cheverton 2004, Kindler 2007, Loehr 2009, Philip 2007, Riess 2005 & 2010, Viret 2004) were obtained from conference abstracts or publications. As

results obtained from conference abstracts and publications have the potential to change, a decision needs to be made regarding the inclusion of such data in the modelling. In the pancreatic example the data was included, but in order to alleviate concerns regarding the inclusion of such data in the meta regression sensitivity analyses were conducted excluding these studies from the analysis. The results of the sensitivity analysis are shown in Table's 3.3 and 3.4.

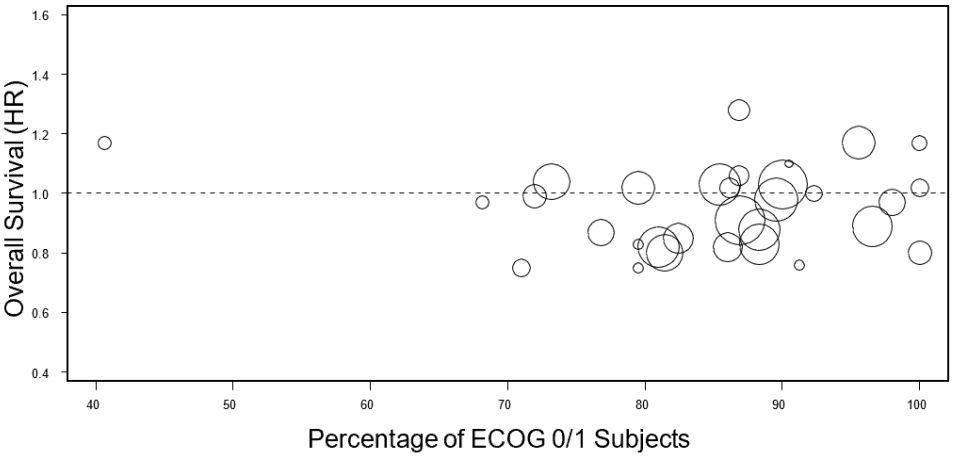
Key facts learned from this step pertinent to predicting the PoS included:

- Plots of the OS hazard ratio against the study level characteristics percentage of metastatic subjects (Figure 3.2), and percentage of ECOG 0/1 subjects (Figure 3.3) showed there to be a wide range of values for the percentage of metastatic subjects and percentage of subjects with ECOG=0/1 where positive treatment effects have been shown. Moreover, there is no range of values where a treatment effect has not been shown and no evidence of association between these study-level covariates and outcome.
- There is a strong association between the PFS hazard ratio and the OS hazard ratio (Figure 3.4).



The size of the circles is inversely proportion to the OS SE(LnHR)

Figure 3.2: Plot of the OS Hazard Ratio against the Percentage of Metastatic Patients



The size of the circles is inversely proportion to the OS SE(LnHR)

Figure 3.3: Plot of the OS Hazard Ratio against the Percentage of ECOG 0/1 Patients

3.6.2. STAGE 2: Statistical Model for Predicting the PoS in Phase 3

Step 1: Table 3.2 below presents an example assuming that the PFS HR observed in phase 2 is 0.8 (log PFS HR=-0.223, with variance 0.05). This represents a phase 2 study comparing a new treatment with gemcitabine analysed after 80 subjects have experienced a PFS event.

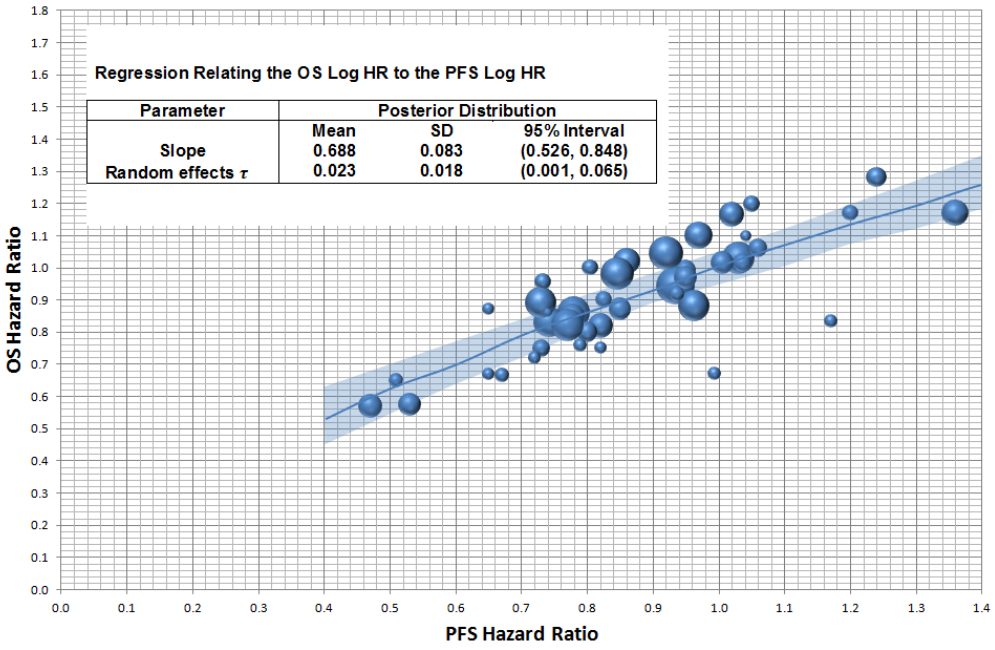
	Sceptical		Non-Informative		Optimistic	
	Mean	SE	Mean	SE	Mean	SE
Prior PFS log HR	0	0.217	0	10	-0.357	0.344
Observed Ph 2 PFS log HR	-0.223	0.224	-0.223	0.224	-0.223	0.224
Posterior PFS log HR	-0.108	0.156	-0.223	0.225	-0.263	0.188
Predicted Ph 2 OS log HR	-0.074	0.112	-0.152	0.158	-0.180	0.136
Probability of Ph 3 Success	0.198		0.395		0.447	

Table 3.2: Estimating the Probability of Success with a Phase 2 PFS Result (HR=0.8)

Step 2: Three prior distributions for the log PFS hazard ratio were elicited from key decision makers to represent non-informative $N(0\,,10^2)$, sceptical $N(0\,,0.2168^2)$, and optimistic $N(-0.357, 0.3441^2)$ prior opinions. The sceptical distribution reflects the opinion that the new treatment shows on average no benefit in PFS time relative to gemcitabine, but there is a 5% chance that the new treatment is better than gemcitabine with a hazard ratio ≤0.70. The optimistic distribution represents an average hazard ratio of 0.70, with a 15% chance that the new treatment is no better than control, i.e., a hazard ratio ≥ 1. Each of these distributions are then combined with the observed phase 2 results to obtain a range of posterior estimates for the PFS log HR (Table 3.2).

Step 3: The relationship between the PFS hazard ratio and OS hazard ratio using the systematically abstracted data were investigated using bayesian meta-regression (Figure 3.4 below). Each point on Figure 3.4 represents the results of one completed study, from which both the PFS hazard ratio and corresponding OS hazard ratio were abstracted. In this example we apply a no-intercept model forcing the regression through the origin. The potential for publication bias was minimized by including all randomized phase 2 and 3 studies in the model. Sensitivity analysis excluding the small earlier phase studies from the analysis was conducted and showed the small studies to have little impact on the parameter estimates. Also, given no evidence of association between the percentage of metastatic patients or percentage of

ECOG 0/1 patients, and the PFS log HR or OS log HR was observed, the selected model did not include these factors as study-level covariates.

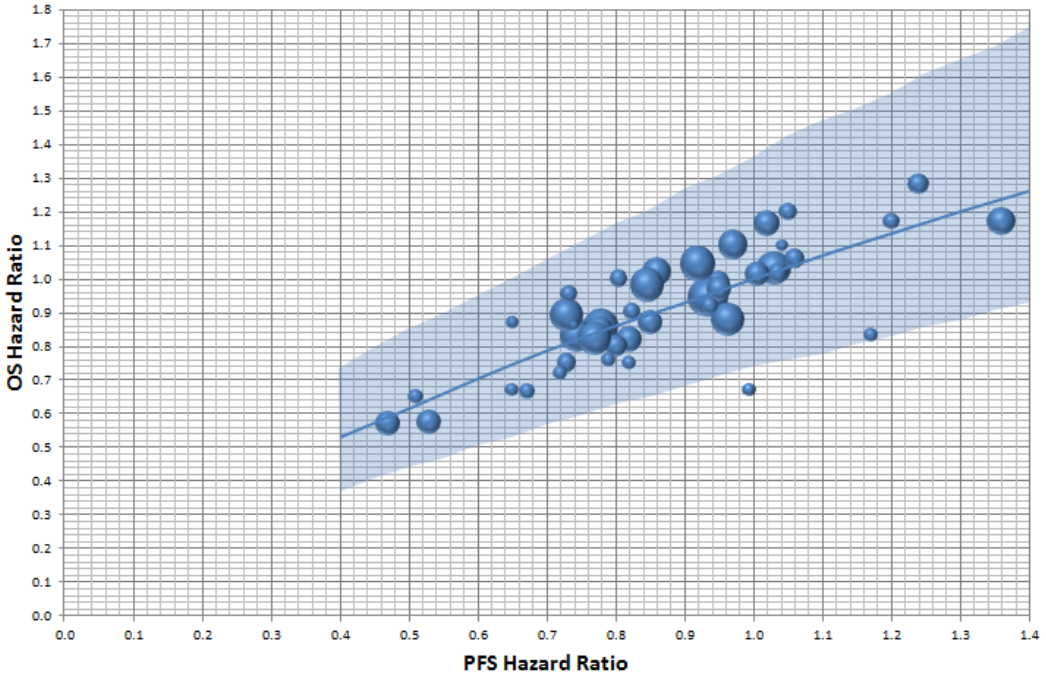


Axes are back transposed from a linear regression between OS log(HR) and PFS log(HR)
The diameter of the circles is inversely proportional to the SE of the OS log(HR) for each published study.

Plot shows the predicted mean and 95% CI of a new study for fixed PFS hazard ratios.

Figure 3.4: Random effects Meta-Regression for OS Hazard Ratio from PFS Hazard Ratio

Step 4: In this example it is assumed that the sample size for the frequentist phase 3 study requires 380 deaths to enable 80% power to detect an OS hazard of 0.75 or less with a statistical significance level of 0.05 (5%). In this step, the predictive distributions for the log OS HR are calculated and used to simulate the results of the proposed phase 3 study design, which in turn can be used to determine the probability of different patterns of study results, including the PoS.



Axes are back transposed from a linear regression between OS log(HR) and PFS log(HR)
The diameter of the circles is inversely proportional to the SE of the OS log(HR).
Plot shows the OS HR posterior predicted mean and 95% CrI assuming a phase 2 study with 80 observed PFS events and a non-informative prior for the PFS log HR.

Figure 3.5: Posterior Predicted OS Hazard Ratio with a Non-Informative Prior

Table 3.2 above completes the example determining the PoS for the planned phase 3 study assuming an observed phase 2 PFS hazard ratio of 0.8. Each of the posterior distributions for the phase 2 outcome measure (PFS log HR) are synthesized with the meta-regression to estimate the predictive distribution for the phase 3 outcome measure treatment difference, i.e., the OS log HR (Table 3.2 above). Within this step the meta-regression shown in Figure 3.4 is extended to ensure the uncertainty in the phase 2 outcome measure, PFS log HR, is incorporated. An example including the non-informative prior is shown in

Figure 3.5, which shows the posterior predicted OS HR, assuming that the PFS HR from a phase 2 study to be analysed after 80 PFS events are observed is unknown at this stage.

The sceptical and optimistic priors can be viewed as providing reasonable bounds of belief for the estimated PoS. Assuming we observe a PFS HR=0.8 in our phase 2 study we can conclude a PoS in phase 3 ranging from 19.8% to 44.7% depending on the prior belief. Figure 3.6 below expands on the example to show the estimated PoS across a range of potentially observed phase 2 PFS hazard ratios and prior distributions. It can be seen that a PoS of at least 60% in phase 3 would require the phase 2 PFS hazard ratio to be 0.7 or lower, with a non- informative or optimistic prior belief. Note that the lines for optimistic and uninformative prior will naturally cross at the point where the observed phase 2 results become more favourable than the optimistic prior.

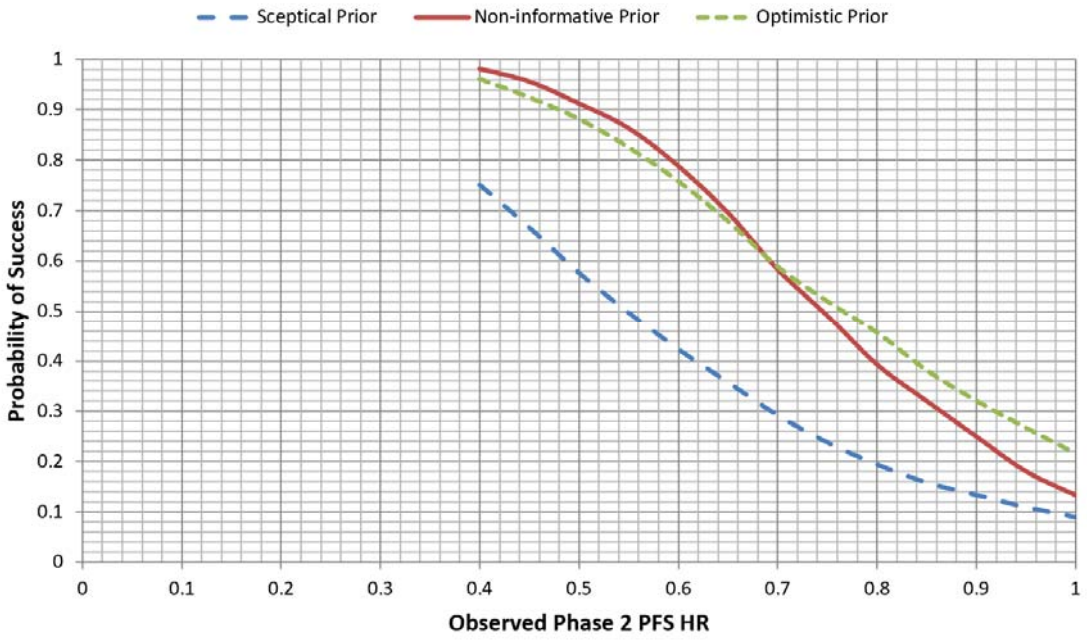


Figure 3.6 Probability of Success in a Phase 3 Study Analysed after 380 Deaths

3.6.2.1: Sensitivity Analysis

Seven of the 43 studies (Cheverton 2004, Kindler 2007, Loehr 2009, Philip 2007, Riess 2005 & 2010, Viret 2004) were obtained from conference abstracts or publications. In order to alleviate concerns regarding the inclusion of such data in the meta regression sensitivity analyses were conducted excluding these studies from the analysis.

Additionally, a more recent meta-analysis of pancreatic cancer was published by Makris (2017). A sensitivity analysis was conducted using all of the data from this meta-analysis to check for consistency with the data published in Sabin

(2014). The results of these sensitivity analyses are shown in Table 3.3 and Table 3.4.

Parameter	Mean	SD	95% CI
All studies			
Slope	0.688	0.083	(0.526,0.848)
Random effects τ	0.023	0.018	(0.001, 0.065)
Sensitivity excluding conference abstracts and presentation			
Slope	0.701	0.084	(0.539,0.868)
Random effects τ	0.025	0.019	(<0.001, 0.071)
Sensitivity using the meta-analysis of Makris(2017)			
Slope	0.658	0.093	(0.474, 0.840)
Random effects τ	0.031	0.024	(0.001, 0.087)

Table 3.3: Sensitivity analysis of the meta regression relating the PFS Log HR to OS Log HR

The results showed the exclusion of the data from the conference abstracts and presentations led to a very small increase in the slope and between study variance. The slope increased from 0.688 when including all studies to 0.701. (Table 3.3) This translates into a 0.4%, 0.9% and 1.9% increase when

incorporating the sceptical, non-informative and optimistic priors into the predicted PoS at the end of phase 3 (Table 3.4).

The results using the meta-analysis by Makris (2017) showed a small reduction in slope and increase in the between study variance. The slope decreased from 0.688 to 0.658 (Table 3.3). This translates into a 1.0%, 1.3% and 1.4% decrease when incorporating the sceptical, non-informative and optimistic priors into the predicted PoS at the end of phase 3 (Table 3.4).

As discussed in Section 3.2, the choice of prior for the between study standard deviation, τ, should be subjected to sensitivity analyses using a range of realistic vague prior distributions. For the pancreatic cancer the range of vague prior distributions explored included a uniform (0,2) on τ, a uniform (0,0.6) on τ, a half normal, and a Gamma (0.001,0.001) on 1/ τ^2. The results are shown in Table 3.5.

The results of the sensitivity analyses for the choice of prior with the between study standard deviation, τ, were found to be highly consistent with the uniform(0,2) prior selected in the pancreatic model, providing reassurance on the robustness of the results.

	Sceptical		Non-Informative		Optimistic	
	Mean	SE	Mean	SE	Mean	SE
Original Analyses						
Prior PFS log HR	0	0.217	0	10	-0.357	0.344
Observed Ph 2 PFS log HR	-0.223	0.224	-0.223	0.224	-0.223	0.224
Posterior PFS log HR	-0.108	0.156	-0.223	0.225	-0.263	0.188
Predicted Ph 2 OS log HR	-0.074	0.112	-0.152	0.158	-0.180	0.136
Probability of Ph 3 Success	0.198		0.395		0.447	
Sensitivity excluding conference abstracts and presentation						
Prior PFS log HR	0	0.217	0	10	-0.357	0.344
Observed Ph 2 PFS log HR	-0.223	0.224	-0.223	0.224	-0.223	0.224
Posterior PFS log HR	-0.108	0.156	-0.223	0.225	-0.263	0.188
Predicted Ph 2 OS log HR	-0.075	0.113	-0.156	0.161	-0.188	0.138
Probability of Ph 3 Success	0.202		0.404		0.466	
Sensitivity using the meta-analysis of Makris(2017)						
Prior PFS log HR	0	0.217	0	10	-0.357	0.344
Observed Ph 2 PFS log HR	-0.223	0.224	-0.223	0.224	-0.223	0.224
Posterior PFS log HR	-0.108	0.156	-0.223	0.225	-0.263	0.188
Predicted Ph 2 OS log HR	-0.070	0.109	-0.146	0.155	-0.174	0.131
Probability of Ph 3 Success	0.188		0.382		0.433	

Table 3.4: Sensitivity Analyses estimating the Probability of Success with a Phase 2 PFS Result (HR=0.8)

Parameter	Mean	SD	95% CI
Uniform (0,2) on τ			
Slope	0.688	0.083	(0.526,0.848)
Random effects τ	0.023	0.018	(0.001, 0.065)
Sensitivity analysis using a Uniform (0,0.6) prior on τ			
Slope	0.683	0.070	(0.529,0.836)
Random effects τ	0.023	0.017	(<0.001, 0.065)
Sensitivity analysis using a Half Normal (0,1) prior on τ			
Slope	0.687	0.078	(0.530,0.836)
Random effects τ	0.022	0.017	(<0.001, 0.064)
Sensitivity analysis using a Gamma(0.001,0.001) prior on 1/ τ^2			
Slope	0.684	0.078	(0.527,0.834)
Random effects τ	0.020	0.014	(0.004, 0.055)

Table 3.5: Sensitivity analysis of the meta regression relating the PFS Log HR to OS Log HR to the choice of prior for τ

3.6.3. STAGE 3: Informing the PoS Threshold for a Go or No-Go Decision

In order to guide the sponsor organisation on the choice of PoS criteria to use to make a go/no-go decision for the pancreatic indication, the probability of making a go decision and failing phase 3, and a go decision and being successful in phase 3 was simulated using different PoS commit to phase 3 decision criteria (ranging between 0.1 and 0.8), and hypothetical truths for the treatment difference in the phase 3 outcome measure (a hazard ratio ranging between 0.5 and 1) for a selection of different development strategies. The different strategies explored in this example include the use of different phase 2 sample sizes ranging from a total of 40 to 320 subjects, followed by a phase 3 study with either 80% (Figure 3.7) or 90% (Figure 3.8) power to detect a minimum clinically and commercially desirable OS hazard of 0.75 with a statistical significance level of 0.05 (5%). Figures 3.7 and 3.8 (below) show the results with a non-informative prior for the phase 2 outcome measure treatment difference. As described in Section 3.3, reasonable bounds of belief for each scenario shown can also be generated by incorporating sceptical and optimistic priors. The plots down the left-hand side of each figure show the probability of making an EOP2 go decision but failing the subsequent phase 3 study. The plots down the right-hand side of each figure show the probability of making an EOP2 go decision and being successful in the subsequent phase 3 study. Each row of plots utilises different PoS thresholds for the go to phase 3 decision criteria.

Table 3.6 below summarizes the results shown in the figures for select PoS thresholds, true treatment differences in the phase 3 outcome measure, phase 2 sample sizes and choice of phase 3 power.

Table 3.6: Select PoS Operating Characteristics for Pancreatic Cancer

PoS Threshold	True OS HR	Pred. Ph.2 PFS HR	Ph. 2 Sample Size	Ph. 3 Study Power	Probability of Go (%)	Probability of Go+ve (%)	Probability Go-ve (%)
0.7	1.0	1.0	80	80	2.4	0.1	2.3
	1.0	1.0	80	90	3.4	0.1	3.3
	0.86	0.8	80	80	14.8	4.9	9.9
	0.86	0.8	80	90	21.8	9.2	12.6
	0.75	0.66	80	80	46.4	37.8	8.6
	0.75	0.66	80	90	55.8	50.8	5.0
	0.62	0.5	80	80	87.2	87.0	0.2
	0.62	0.5	80	90	90.4	90.4	0.0
	0.53	0.4	80	80	98.2	98.2	0.0
	0.53	0.4	80	90	99.8	98.8	0.0
0.6	1.0	1.0	80	80	4.4	0.1	4.3
	1.0	1.0	80	90	6.2	0.2	6.0
	0.86	0.8	80	80	26.4	8.7	17.7
	0.86	0.8	80	90	32.6	13.7	18.9
	0.75	0.66	80	80	61.8	50.3	11.5
	0.75	0.66	80	90	69.8	63.5	6.3
	0.62	0.5	80	80	93.2	92.9	0.3
	0.62	0.5	80	90	95.0	95.0	0.0
	0.53	0.4	80	80	99.2	99.2	0.0
	0.53	0.4	80	90	99.2	99.2	0.0
0.5	1.0	1.0	80	80	8.0	0.2	7.8
	1.0	1.0	80	90	10.6	0.3	10.3
	0.86	0.8	80	80	37.4	12.4	25.0
	0.86	0.8	80	90	44.6	18.8	25.8
	0.75	0.66	80	80	73.2	59.6	13.6
	0.75	0.66	80	90	78.2	71.1	7.1
	0.62	0.5	80	80	96.2	95.9	0.3
	0.62	0.5	80	90	97.4	97.4	0.0
	0.53	0.4	80	80	99.6	99.6	0.0
	0.53	0.4	80	90	99.6	99.6	0.0
0.7	1.0	1.0	160	80	0.2	0.0	0.2
	1.0	1.0	160	90	1.2	0.0	1.2
	0.86	0.8	160	80	10.0	3.3	6.7
	0.86	0.8	160	90	17.6	7.4	10.2
	0.75	0.66	160	80	53.0	43.2	9.8
	0.75	0.66	160	90	66.4	60.4	6.0
	0.62	0.5	160	80	96.2	95.9	0.3
	0.62	0.5	160	90	98.0	98.0	0.0
	0.53	0.4	160	80	99.8	99.8	0.0
	0.53	0.4	160	90	100	100	0.0

PoS Threshold	True OS HR	Pred. Ph.2 PFS HR	Ph. 2 Sample Size	Ph. 3 Study Power	Probability of Go (%)	Probability of Go+ve (%)	Probability Go-ve (%)
0.6	1.0	1.0	160	80	1.8	0.0	1.8
	1.0	1.0	160	90	2.5	0.0	2.5
	0.86	0.8	160	80	20.2	6.7	13.5
	0.86	0.8	160	90	29.2	12.3	16.9
	0.75	0.66	160	80	69.8	56.8	13.0
	0.75	0.66	160	90	77.8	70.8	7.0
	0.62	0.5	160	80	98.2	97.9	0.3
	0.62	0.5	160	90	99.2	99.2	0.0
	0.53	0.4	160	80	100	100	0.0
	0.53	0.4	160	90	100	100	0.0
0.5	1.0	1.0	160	80	2.8	0.0	2.7
	1.0	1.0	160	90	4.5	0.0	4.5
	0.86	0.8	160	80	32.6	10.8	21.8
	0.86	0.8	160	90	42.4	17.8	24.6
	0.75	0.66	160	80	79.6	64.8	14.8
	0.75	0.66	160	90	85.4	77.7	7.7
	0.62	0.5	160	80	99.2	98.9	0.3
	0.62	0.5	160	90	99.4	99.4	0.0
	0.53	0.4	160	80	100	100	0.0
	0.53	0.4	160	90	100	100	0.0
0.7	1.0	1.0	240	80	0.0	0.0	0.0
	1.0	1.0	240	90	0.2	0.0	0.2
	0.86	0.8	240	80	6.6	2.2	4.4
	0.86	0.8	240	90	13.2	5.6	7.6
	0.75	0.66	240	80	56.6	46.1	10.5
	0.75	0.66	240	90	72.6	66.0	6.6
	0.62	0.5	240	80	98.6	98.3	0.3
	0.62	0.5	240	90	99.2	99.2	0.0
	0.53	0.4	240	80	100	100	0.0
	0.53	0.4	240	90	100	100	0.0
0.6	1.0	1.0	240	80	0.2	0.0	0.2
	1.0	1.0	240	90	1.0	0.0	1.0
	0.86	0.8	240	80	16.2	5.4	10.8
	0.86	0.8	240	90	26.8	11.3	15.5
	0.75	0.66	240	80	73.4	59.8	13.6
	0.75	0.66	240	90	82.2	74.8	7.4
	0.62	0.5	240	80	99.2	98.9	0.3
	0.62	0.5	240	90	99.8	99.8	0.0
	0.53	0.4	240	80	100	100	0.0
	0.53	0.4	240	90	100	100	0.0
0.5	1.0	1.0	240	80	1.4	0.0	1.4
	1.0	1.0	240	90	2.4	0.0	2.4
	0.86	0.8	240	80	29.2	9.7	19.5
	0.86	0.8	240	90	41.0	17.2	23.8
	0.75	0.66	240	80	84.2	68.6	15.6
	0.75	0.66	240	90	89.2	81.1	8.1
	0.62	0.5	240	80	99.8	99.5	0.3
	0.62	0.5	240	90	100	100	0.0
	0.53	0.4	240	80	100	100	0.0
	0.53	0.4	240	90	100	100	0.0

The development strategy used throughout this example was to use 80 subjects in total (40/arm) in phase 2, followed by a phase 3 study with 80% power. Here we learn that if the drug has a minimal clinically and commercially desirable beneficial effect (OS HR=0.75), then when using a threshold for the PoS of 0.6, there is a 61.8% chance of making a correct EOP2 go decision, but with a 11.5% chance of making this decision and subsequently failing in the phase 3 study. If a PoS threshold of 0.5 is selected, there is a 73.2% chance of making a correct EOP2 go decision, with a 13.6% chance of making this decision and subsequently failing in the phase 3 study. Similarly, when selecting a PoS threshold of 0.7, there is a 46.4% chance of making a correct EOP2 go decision, with a 8.6% chance of making this decision and subsequently failing in the phase 3 study.

If the drug has no benefit (OS HR=1.0), when using threshold for the PoS of 0.7, 0.6, or 0.5 there is a 2.4%, 4.4% and 8% chance of making an incorrect EOP2 go decision. Similarly, if the drug has a marginal but not clinically meaningful effect (OS HR=0.86), when using threshold for the PoS of 0.7, 0.6, or 0.5 there is a 14.8%, 26.4% and 37.4% chance of making an incorrect EOP2 go decision. Moreover, there is a corresponding 4.9%, 8.7% and 12.4% chance of the phase 3 still being successful potentially resulting in the future population of patients receiving a drug with a poor benefit risk.

Other noteworthy points from Table 3.6 include:

- Assuming a true clinically meaningful OS HR of 0.75, as the size of the Phase 2 study increases from 80 (40/arm) to 160 the probability of making

a go decision increases by approximately 9-10% (depending upon choice of threshold). The magnitude of increase is not so marked when including a further 80 patients to a total of 240 with the probability of making a go decision increasing by another 4-6%.

- Over the range of scenario's presented in Table 3.6, assuming a true HR of 0.75 and utilizing a phase 3 study with 90% power as opposed to 80% increases the probability of a go and being successful between 12.1 to 20%, and decreases the probability of a go and failing by 3.6 to 7.5%
- The lower the PoS threshold the greater the probability of making a go decision. From the patient perspective keeping this probability low when there is no true benefit of the drug (HR=1) is vital. We see this probability is between 2.4% and 10.6%, 0.2% and 4.5%, and 0.0% and 2.4% for the various design options and thresholds with 80, 160 and 240 subjects respectively in phase 2.
- The lower the PoS threshold the greater the probability of making a go decision followed by subsequent phase 3 failure. From purely the sponsor perspective of avoiding failed phase 3 studies, keeping this probability low is important. When there is no true benefit of the drug (HR=1), this probability is between 2.3% to 10.3%, 0.2% to 4.5%, and 0% to 2.4% for the various design and threshold options with 80, 160, and 240 subjects respectively in phase 2.

- Importantly the probability of making a go decision followed by a successful phase 3 trial when in truth there is no benefit of the drug is negligible <0.3% for all scenarios on Table 3.6.

As discussed in Section 3.5, sponsor companies can therefore use this approach to trade off the respective probabilities to ensure an appropriate PoS threshold is selected to balance the risks they are willing to take with the costs and resource utilisation associated with various development strategies. A review of the PoS values for 63 completed phase 3 studies spanning the oncology, respiratory and Cardiovascular portfolio that started during or after 2015 at AstraZeneca in presented in Section 3.10. Here we learn that the average predicted PoS of the phase 3 trials undertaken was 61.4%, with 82% of trials subsequently successful if the PoS was ≥. 61.4%.

Figure 3.7: Informing the PoS Threshold for a Decision to conduct an 80% powered Phase 3 Study

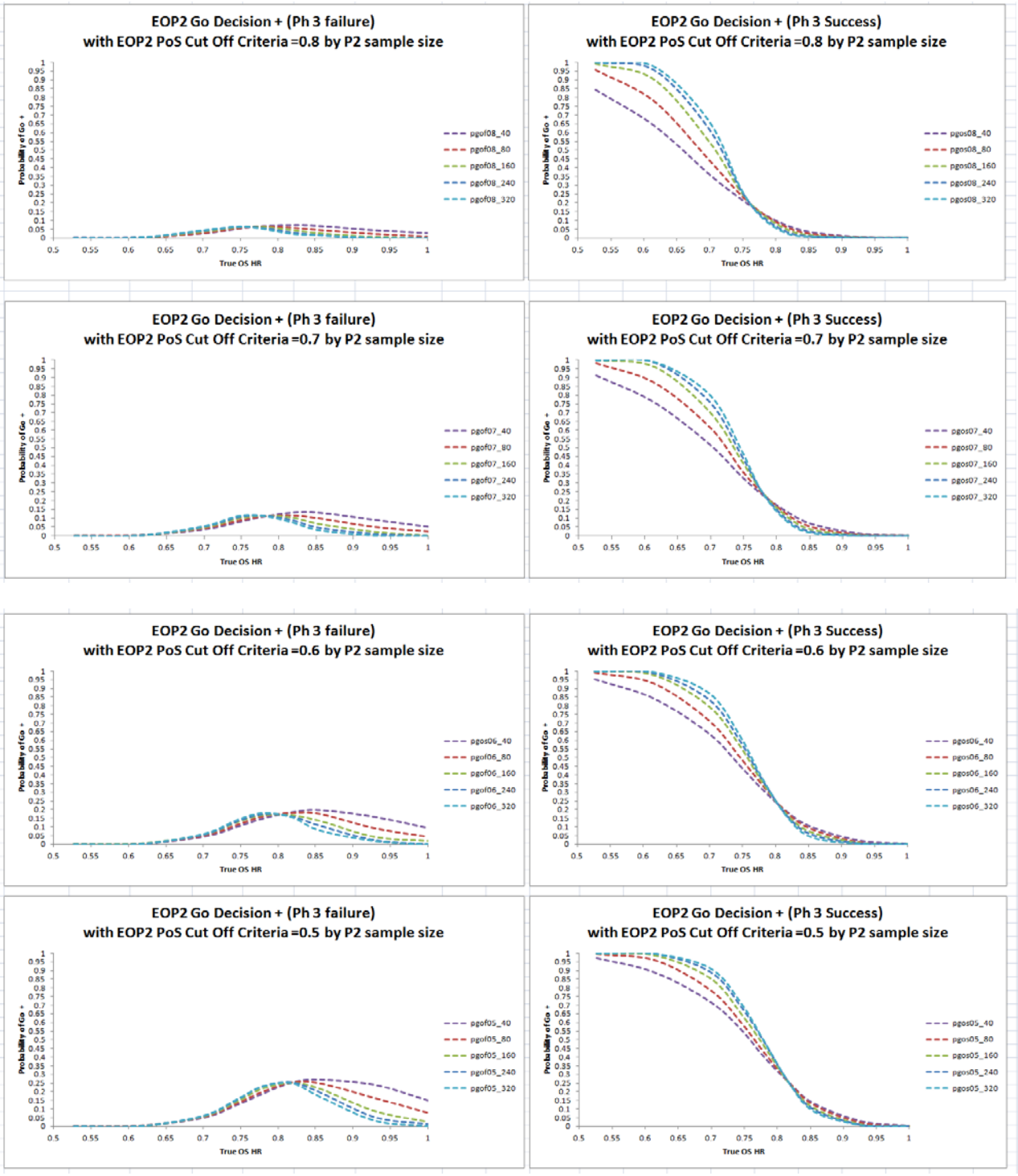


Figure 3.7 (cont.): Informing the PoS Threshold for a Decision to conduct an 80% powered Phase 3 Study

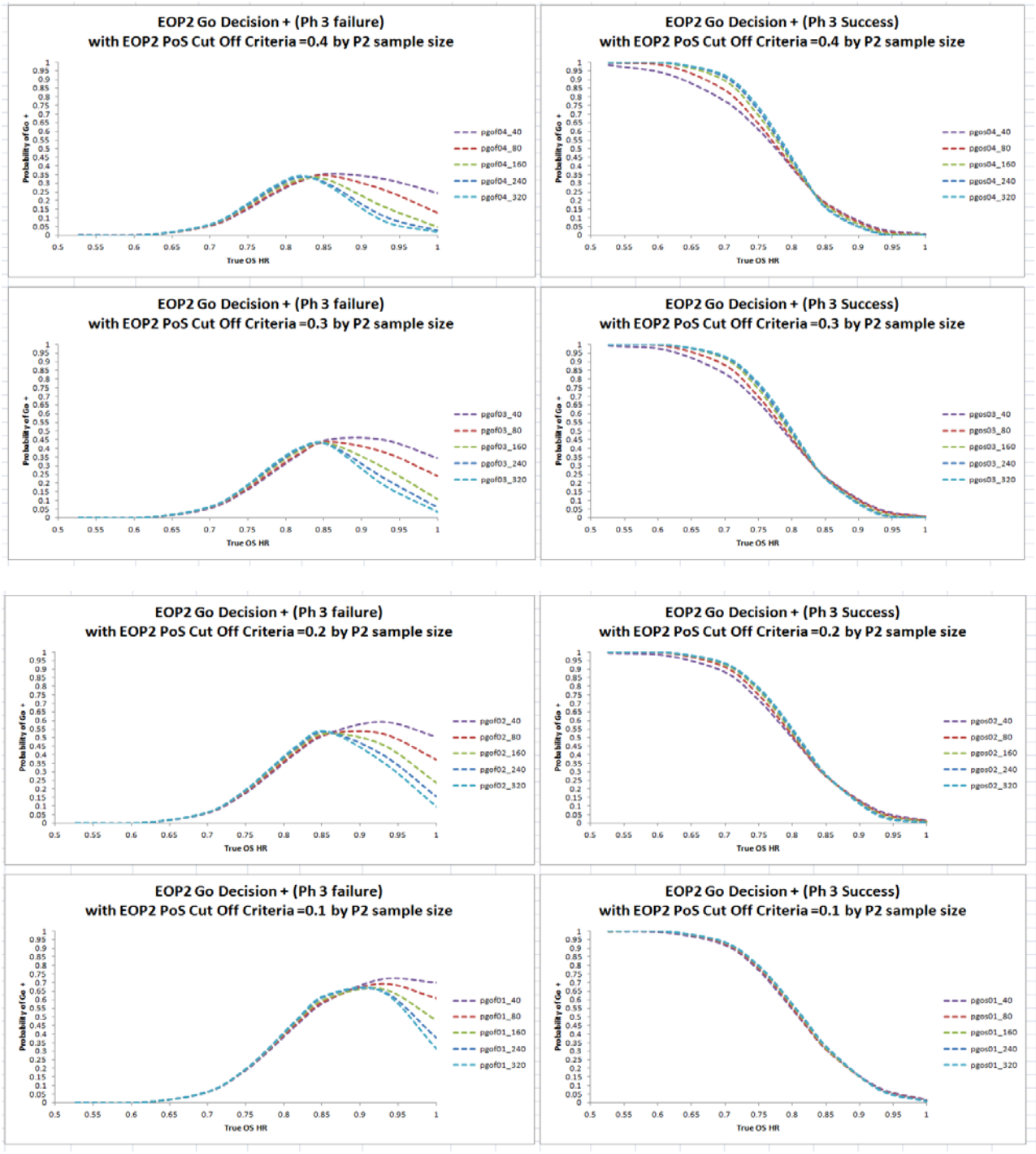


Figure 3.8: Informing the PoS Threshold for a Decision to conduct a 90% powered Phase 3 Study

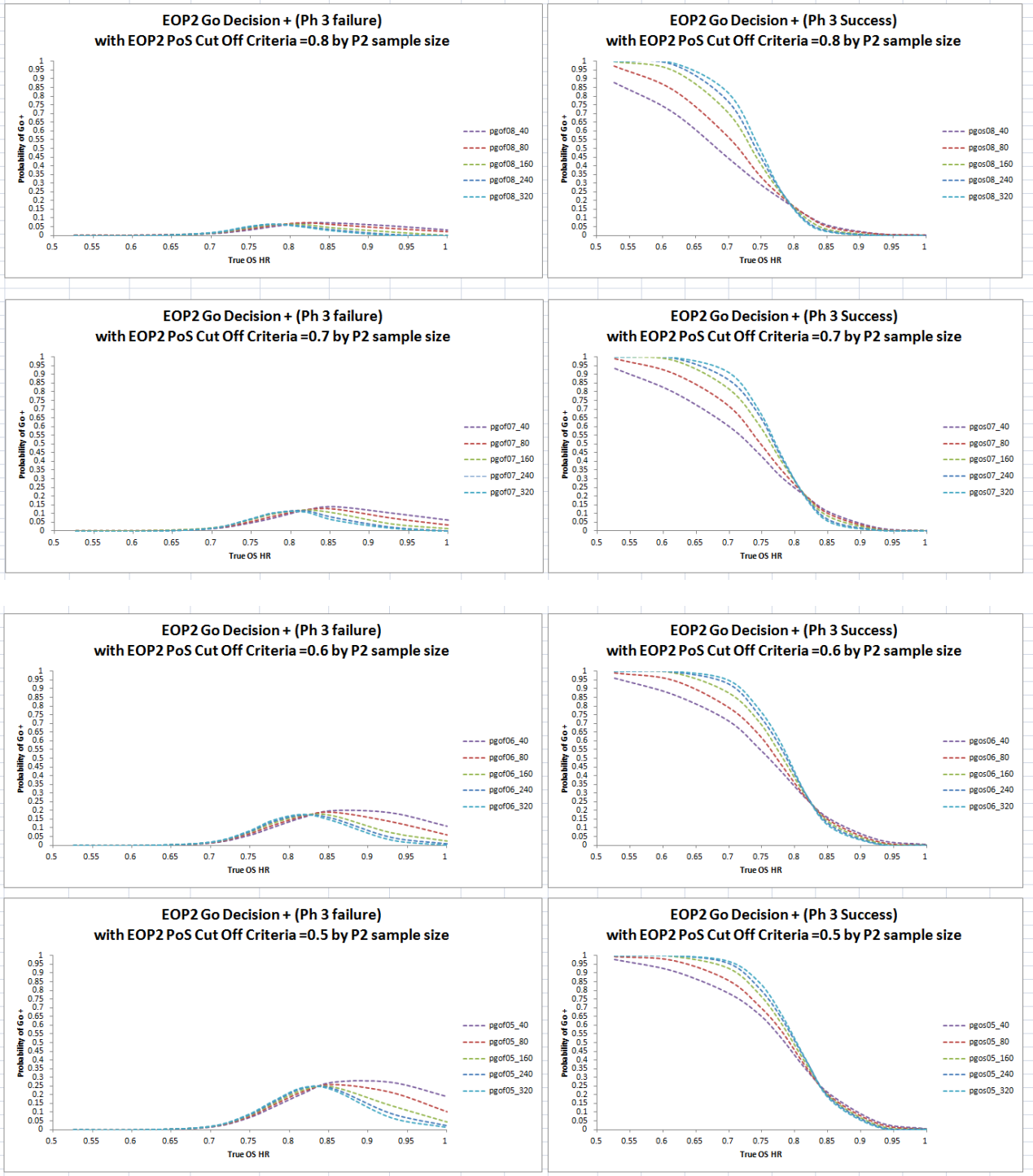
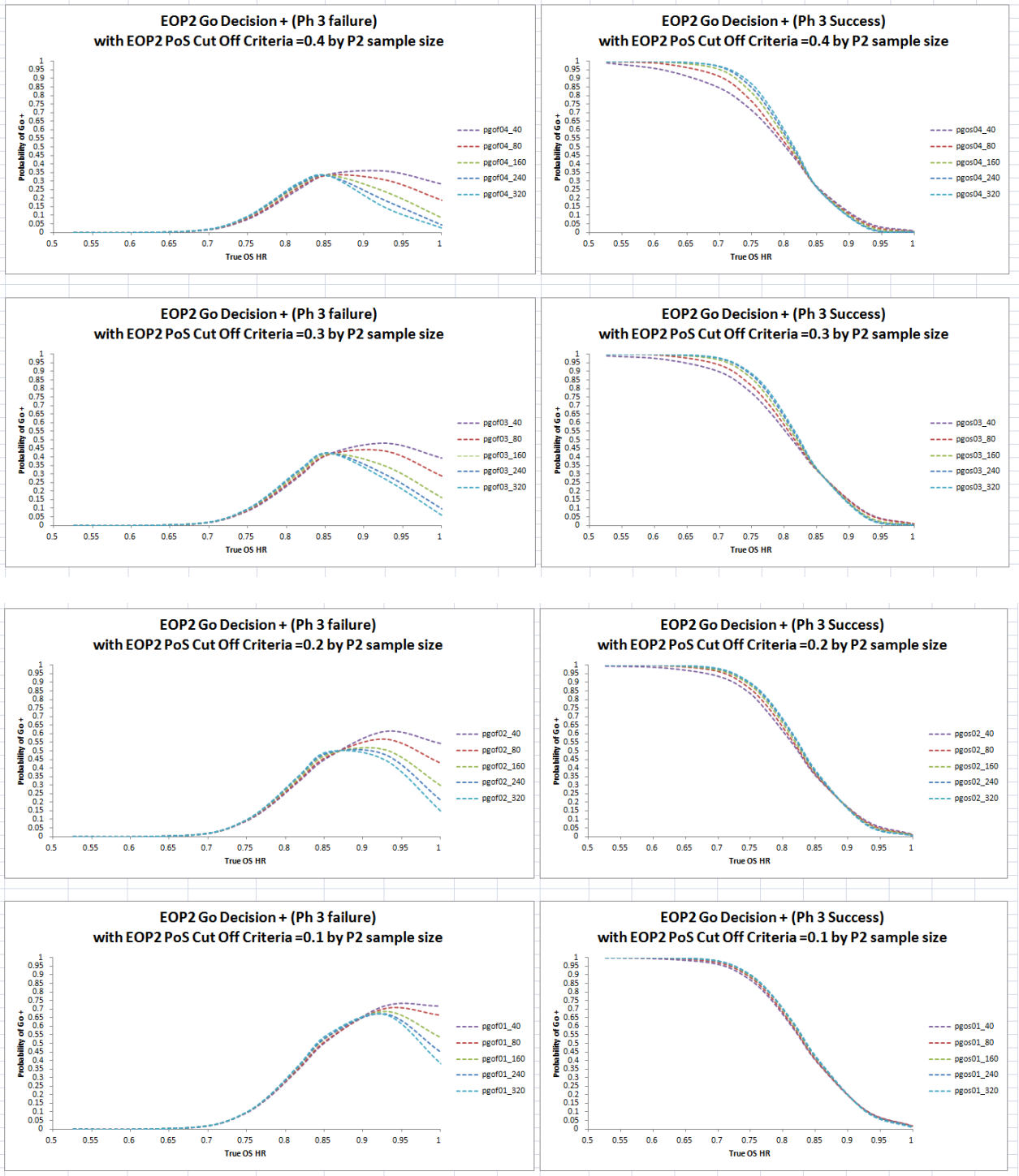


Figure 3.8 (cont): Informing the PoS Threshold for a Decision to conduct a 90% powered Phase 3 Study



3.7. Methodology Extension: Incorporation of a Variance Inflation Factor for the Treatment Difference in the Phase 2 Outcome Measure

When interpreting the observed results from a phase 2 study, decision makers will carefully investigate the robustness of the phase 2 evidence provided. One area of evaluation will focus on the similarity of the observed phase 2 control response with that which was expected from prior knowledge. The motivation behind the methodology presented in this section is to explore methods to build additional uncertainty into the phase 2 treatment difference distribution for the phase 2 outcome measure as the observed control departed from the prior expected result. This uncertainty would then pass through into the PoS modelling.

As part of the systematic review in step 1, a prior for the control group response in the absolute phase 2 outcome measure may be estimated using meta-regression by predicting the response for the observed characteristics of the phase 2 study. At the EOP2 meeting, after completing the phase 2 study, this prior will be compared to the actual observed control group response in the phase 2 study. If the observed response is similar to the prior then there will be support for using the observed treatment difference from the phase 2 study to predict the phase 3 outcome. If the control group response is not as expected, this may cast doubt on the validity of using the treatment difference observed in the phase 2 study for prediction purposes. Instead of using the control prior as

an external assessment of the trial's robustness, exploring alternative approaches to synthesize a prior for the control group response with the control arm in the phase 2 study may be useful.

One natural option to this is to assume that the control arms from the studies selected in the systematic review are compatible with the new phase 2 study control group data. Then, assuming the phase 2 study is comparative, a posterior expected treatment difference in the phase 2 outcome measure is calculated after initially combining the prior and phase 2 study control results together. If the phase 2 is randomized, such an approach would however break the randomization and potentially introduce bias.

An alternative innovative approach would be to take the view that the observed treatment difference seen in the randomized phase 2 study is the best unbiased estimate available, whilst inflating the variance of the phase 2 treatment difference as the observed phase 2 control group response departs from the expected. Such an approach acts as a quality index for the phase 2 study and may help to discount early optimistic phase 2 results (Kirby, Burke, Chuang-Stein and Sin, 2012).

Following discussions Antony Sabin had with the drug development decision makers whilst working at Amgen, a generalized approach (which constrains the variance inflation to a maximum of 2-fold) was to be considered as an optional extension and useful sensitivity analysis of the base methodology.

Assuming the observed (or simulated) treatment difference in the phase 2 outcome measure may be expressed as:

$$d \sim N(\delta, \sigma_d^2)$$

Then, the predicted distribution of the treatment difference in the phase 2 outcome measure including the uncertainty due to differences between the observed and expected result for the control arm is:

$$d_2 \sim N(\delta, \mathrm{V}\sigma_2^2)$$

with V, the variance inflation factor equal to

$$\mathrm{V} = 2 - e^{\left(\frac{-(d_h^2)}{2\sigma_h^2}\right)}$$

where d_h is the difference between the mean expected (or historical) and the observed control group response, and σ_h^2 is the variance of the difference.

The variance inflation factor was determined as follows:

1. Calculate the probability density function for the PFS log HR comparing the observed control PFS hazard rate to the expected hazard rate.

$$\frac{1}{\sigma_h \sqrt{2\pi}} \; e^{\left(\frac{-(d_h^2)}{2\sigma_h^2}\right)}$$

2. Calculate the probability density function for the PFS log HR comparing the observed control PFS hazard rate to the expected hazard rate, assuming no difference.

$$\frac{1}{\sigma_h \sqrt{2\pi}} \; e^{\left(\frac{-(0)}{2\sigma_h^2}\right)}$$

3. Take the ratio of the two density functions and constrain to a maximum inflation of 2.

$$V = 2 - e^{\left(\frac{-(d_h^2)}{2\sigma_h^2}\right)}$$

Here it is assumed that the variance of the expected control group response is weighted equally to the observed control arm of the phase 2 study, implying the strength of information provided in the result of the external control data is equal to that determined from the phase 2 study.

Alternative choices of weighting for the expected control could be applied. The choice of weight will require judgement around the exchangeability assumption, i.e., the extent to which the historical control information can be considered contemporaneous to the data generated in the phase 2 study (Spiegelhalter, Abrams, Myles, 2004).

Incorporating this into the worked example from pancreatic cancer and assuming:

- The hazard function observed from R_0 PFS events in phase 2 for gemcitabine is λ_o, where R_0=40 and λ_o=0.173 (an observed median PFS of 4 months)
- An expected median PFS for gemcitabine of 3.7 months based upon historical evidence weighted equally to the R_0 PFS events. With the

added assumption of exponential survival times this equates to a hazard function of:

$$\lambda_e = 0.187 \quad (\text{Log2 / med PFS})$$

Then the log(hazard ratio), d_h = log(λ_o/λ_e)=log(0.925), with variance σ_h^2 = (R_0 + R_0)/(R_0^2) = 0.05, and the corresponding variance inflation factor = 1.06.

The following plot (Figure 3.9 below) expands this scenario to highlight the variance inflation factor for a range of potential PFS log hazard ratios comparing the expected and the observed result in the control arm of the phase 2 study for a selection of phase 2 sample sizes.

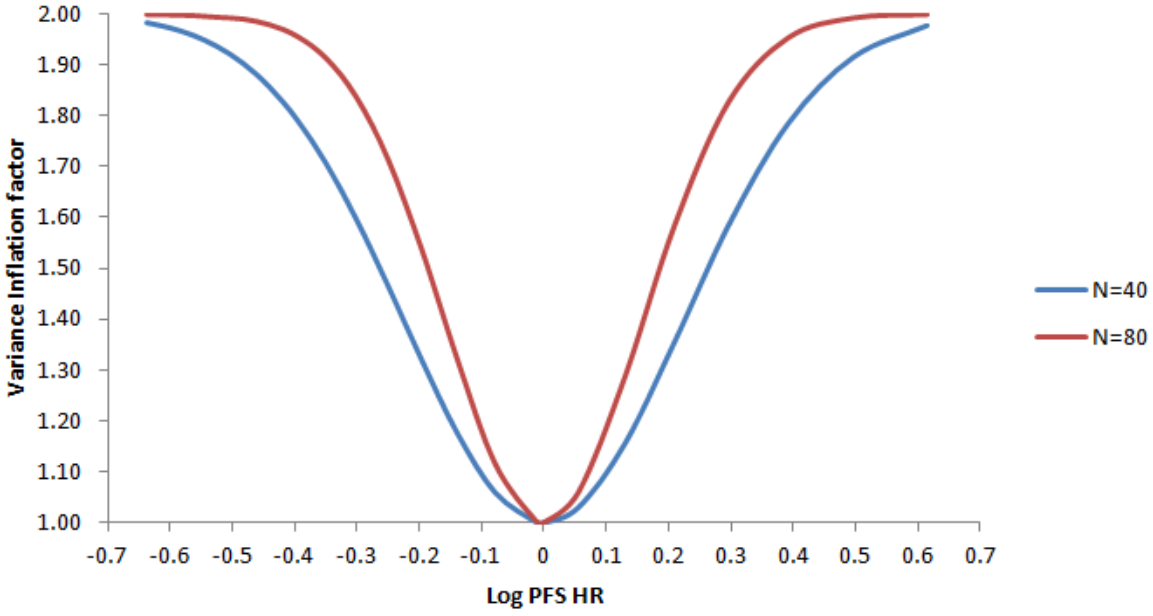


Figure 3.9: Variance Inflation factor by Log PFS HR

Figure 3.10 investigates the impact of the variance inflation factor on the predicted phase 3 results for the pancreatic example. It is assumed the median PFS is 4 months in the phase 2 control arm (N=40), while the historical expected estimate ranged from 2 to 8 months (log PFS HR ranging from -0.69 to 0.69). As used in Section 3.6.2 the phase 3 trial is designed to have 80% power to detect an OS hazard ratio of 0.75. In this example observed phase 2 PFS HR's (comparing test with control) of 0.5 and 0.8 are explored.

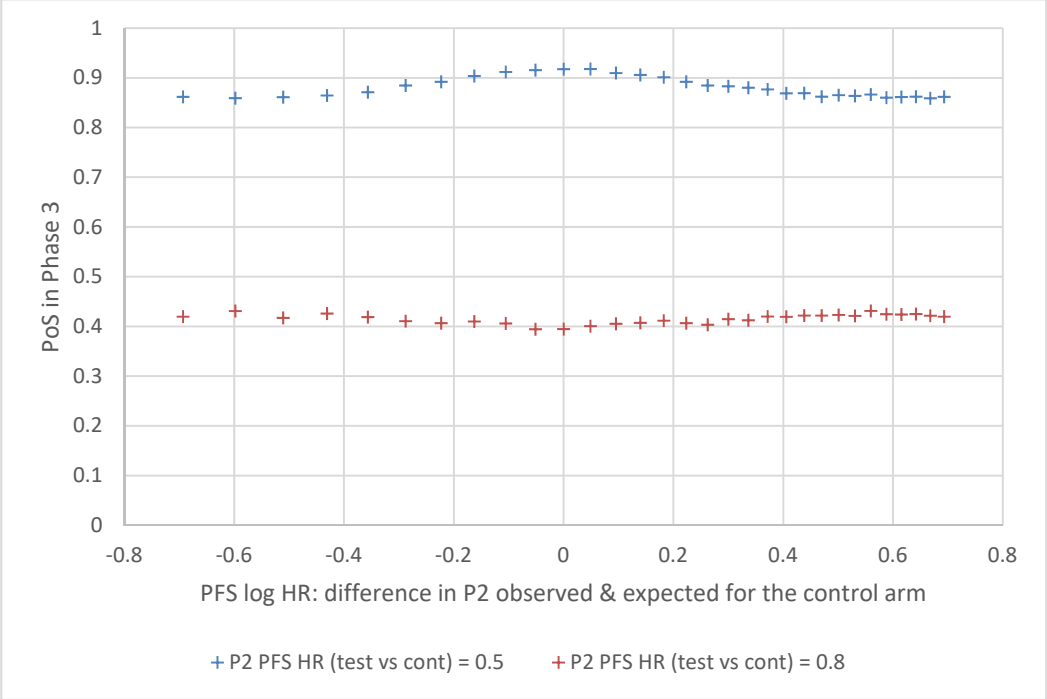


Figure 3.10: Assessing the impact of the Inflation Factor on PoS

As shown in Figure 3.10, in the scenario where the observed phase 2 PFS HR was 0.5 (leading to a predicted phase 3 OS HR of 0.61, stronger than the targeted phase 3 HR of 0.75), the variance inflation factor reduced the PoS from 91.7% down to 86%, as the difference between the observed and expected control log PFS HR moved away from 0 (no difference) towards -0.69 (HR=0.5)

or 0.69 (HR=2). In the scenario where the observed phase 2 PFS HR was 0.8 (leading to a predicted phase 3 OS HR of 0.86, worse than the targeted phase 3 HR of 0.75), the variance inflation factor moved the PoS up from 39.5% to 42% as the difference between the observed and expected control log PFS HR moved away from 0 (no difference) towards -0.69 (HR=0.5) or 0.69 (HR=2).

The implication of using this variance inflation factor approach is that additional uncertainty is incorporated into the posterior predicted phase 3 outcome measure. This has the consequence of reducing confidence in being successful or failing and thereby moving the PoS closer towards 50%. Assuming trials with a higher PoS (e.g. >60%) are selected to move forward in phase 3, this approach may help to discount early optimistic phase 2 results.

3.8. Additional Examples of Predicting the PoS through a Meta-Regression Modelling Approach

In this section additional examples following the meta analytical approach to determine the PoS outlined in Section 3.6.2 are presented. These complement the examples shown for pancreatic cancer in Section 3.6.2 and the breast cancer indication (Wang 2013).

An example using advanced gastric cancer:

In this example a randomised phase 3 study comparing olaparib in combination with paclitaxel versus paclitaxel alone in Asian patients with advanced gastric cancer was planned. The study was designed with two primary populations in

mind (all overall population, and a biomarker subgroup left unspecified at the study onset). The unspecified subgroup was to be defined and pre-specified during the course of the study prior to the interim analysis, and selected based upon efficacy data generated outside of the study and the estimated prevalence from patients recruited into the trial. Due to the uncertainty in the subgroup the test mass (alpha spending) was split with 4% assigned to the all-comers population and 1% to the subgroup. The primary endpoint required for regulatory approval in this setting was overall survival.

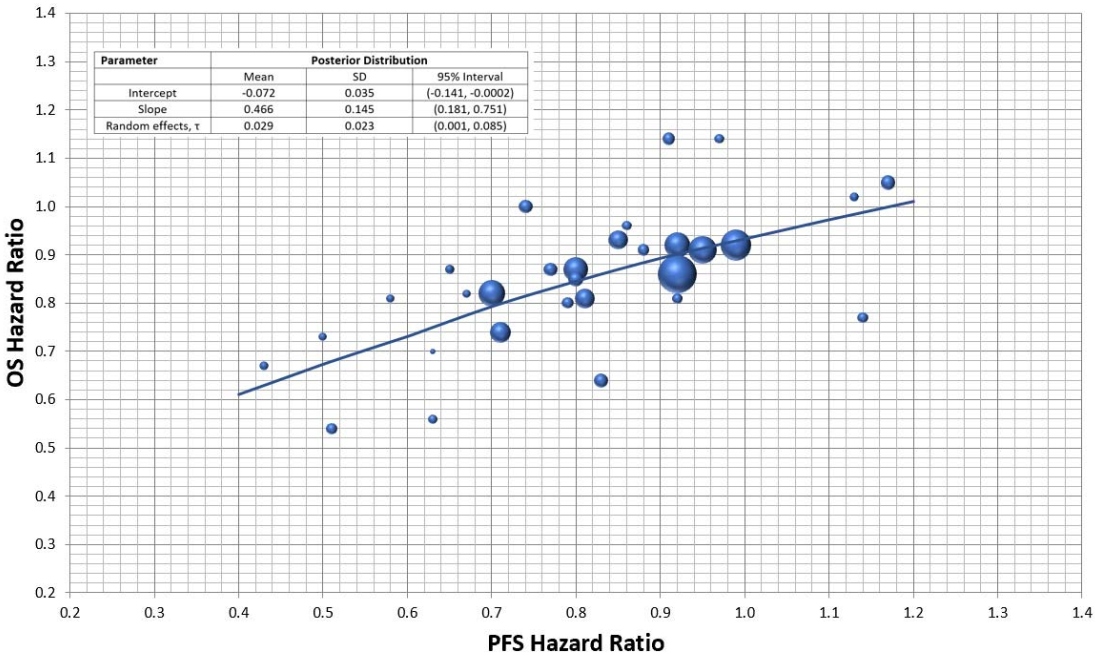


Figure 3.11 : Relationship between PFS and OS Hazard Ratio's in Advanced Gastric Cancer

The study was sized (500 patients with a 1:1 randomisation, with the primary analysis occurring after 353 deaths) to have had 90% power to detect a HR of 0.7 at a significance level of 4% in the intent to treat population. A single interim analysis for futility (both populations) and superiority (overall population) was

included at an information fraction of 50%. Prior data from a randomised phase 2 study (N=62/arm) conducted purely in Korea that compared the test drug in combination with paclitaxel with paclitaxel was available to support the decision to invest in phase 3. The phase 2 study had enriched for the biomarker effect. In order to adjust the result to the expected prevalence in the ITT population of the planned phase 3 study a weighted progression free survival hazard ratio was calculated. This was found to be 0.87 (95% CI 0.54 to 1.39, 110/122 PFS events). Some data on overall survival was also available, with the weighted OS hazard ratio found to be 0.69 (95% CI 0.41 to 1.17, 80/122 OS events).

In order to develop the PoS, the PFS HR and OS HR data from the meta-analysis and validation studies published in Paoletti (2013) were abstracted and a meta regression conducted to determine the relationship between PFS HR and OS HR (Figure 3.11).

Following the approach detailed in Section 3.6.2, passing the phase 2 PFS HR through the meta regression resulted in a phase 3 prior distribution for OS HR to be 0.95 (95% CI 0.638, 1.337), which in turn translates to a low PoS in the planned phase 3 study of 23%.

Due to the discrepancy between the OS and PFS findings in the phase 2 study the PoS was also calculated from the OS data. The discrepancy in results also led to a decision to include more scepticism in the predicted PoS for OS. A decision to add an additional 0.15 to the OS HR result was made by the project

team (i.e., an assumed HR of 0.84). The resulting PoS in the planned phase 3 study is then calculated to be 44%.

Although the PoS was low, a decision was made to still conduct the phase 3 study. The treatment of advanced gastric cancer, in second and later lines in particular, was considered to represent a high unmet medical need with poor outcomes and toxic treatments, and therefore a significant development and market opportunity with relatively low market entry and access hurdles compared to other areas. Moreover, the addition of olaparib to paclitaxel had been shown to have a manageable and predictable tolerability profile in phase 2 and a signal of efficacy in overall survival. However, the resulting phase 3 study was not successful. It resulted in an observed OS HR of 0.795 (95% CI 0.63, 1.00) and a median improvement of 1.9 months. Similarly, the secondary PFS endpoint resulted in a HR of 0.837 (95% CI 0.67, 1.04).

An example using soft tissue sarcoma:

In this example the PoS of a randomised phase 3 study comparing olaratumab in combination with doxorubicin versus doxorubicin alone in patients with advanced soft tissue sarcoma is determined. The phase 3 study was designed with two primary populations in mind (all overall population, and a Leiomyosarcoma subgroup). The test mass (alpha spending) was split with 4% assigned to the all-comers population and 1% to the subgroup. The primary endpoint required for regulatory approval in this setting is overall survival. The study was sized (460 patients with a 1:1 randomisation, with the primary

analysis occurring after 322 deaths) to have had 80% power to detect a HR of 0.723 at a significance level of 4% in the intent to treat population.

Prior data from a randomised phase 2 study (Tap, 2016) comparing olaratumab in combination with doxorubicin with doxorubicin was available to support the decision to invest in phase 3. The estimated progression free survival hazard ratio was 0.67 (95% CI 0.442 to 1.021). Some data on overall survival was also available. The OS hazard ratio found to be 0.46 (95% CI 0.301 to 0.710), which when comparing against the expected relative strength of the PFS and OS results seen in historical data was inconsistent to expectations.

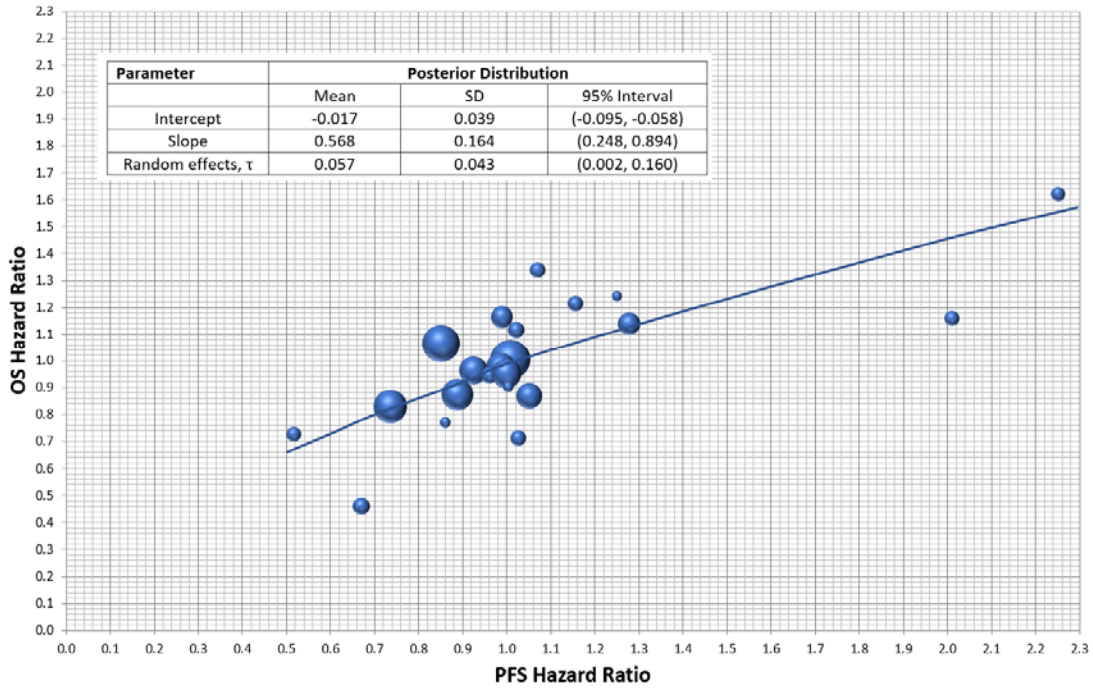


Figure 3.12 : Relationship between PFS and OS Hazard Ratio's in Advanced Soft Tissue Sarcoma

In order to develop the PoS, the PFS HR and OS HR data from the meta-analysis published in Tanaka (2019) were selected, and a meta regression conducted to determine the relationship between PFS HR and OS HR (Figure 3.12).

Following the approach detailed in Section 3.6.2, passing the phase 2 PFS HR through the meta regression resulted in a phase 3 prior distribution for OS HR to be 0.81 (95% CI 0.562, 1.06), which in turn translates to a PoS in the planned phase 3 study of 49.8%. The phase 3 study was not successful with an observed OS HR of 1.05 (95% CI 0.841, 1.303).

An example for Non-Small Cell Lung Cancer

In this example the PoS of a randomized phase 3 study in NSCLC comparing conatumumab in combination with carboplatin and paclitaxel against carboplatin and paclitaxel alone is established. The planned phase 3 study had 90% power (randomized 1:1 with the primary analysis after 844 events) to detect a clinically meaningful OS hazard ratio of 0.8 at the 2-sided 5% level of significance. Prior data from a randomised phase 2 study (Paz-Ares, 2013) comparing conatumumab in combination with carboplatin and paclitaxel against carboplatin and paclitaxel alone was available to support the decision to invest in phase 3. The estimated progression free survival hazard ratio from the phase 2 study was 0.84 (95% CI 0.57 to 1.24) in the 3mg/kg conatumumab dose and 0.93 (95% CI 0.64, 1.35) at the 15mg/kg dose.

In order to develop the PoS, a relationship between the PFS HR and the OS HR a systematic review was performed at the sponsor company. A meta regression was then conducted using the abstracted data. The parameters from the meta regression are shown in Table 3.7.

Parameter	**Posterior Distribution**		
	Mean	SD	95% Interval
Intercept	0.014	0.020	(-0.025, 0.053)
Slope	0.592	0.118	(0.362, 0.824)
Random effects, τ	0.031	0.022	(0.001, 0.083)

Table 3.7 Relationship between log HR (OS) and log HR (PFS) in NSCLC

The analysis indicates that the OS log(HR) = 0.014 + (0.592 x PFS log(HR)). Following the approach detailed in Section 3.6.2, passing the phase 2 PFS HR's through the meta regression resulted in a phase 3 prior distribution for OS HR to be 0.91 (95% CI 0.69, 1.15) for the 3mg/kg arm and 0.96 (95% CI 0.75, 1.22) for the 15mg/kg. These in turn translate to a PoS in the planned phase 3 study of 49.3% and 25.7% respectively. The planned phase 3 study did not go ahead.

In addition to the pancreatic, gastric, STS and NSCLC examples shown, the relationship between the PFS HR and OS HR has been examined in several other cancer types (including mBC, mCRC, esophageal, rectal, 1st-line ovarian, CRPC, SCLC, advanced neuroendocrine neoplasms) in the literature, which greatly facilitates both the use of such a modelling approach to determine the PoS in future phase 3 studies and promotes understanding of the confidence drug developers have in the choice and interpretability of phase 2 endpoints.

3.9. Qualitative Information

In this section the incorporation of qualitative factors into the decision making process is discussed.

A large portion of the research conducted for this PhD has focused on developing a quantitative framework to assess the PoS. The framework developed for the PoS modelling incorporated:

- The relationships between phase 2 and phase 3 study outcome measures (on the relative treatment difference scale)
- The influence of prognostic factors on the relationship
- The treatment difference observed for the phase 2 outcome measure in the phase 2 study
- A range of prior opinions of key decision makers for the treatment difference in the phase 2 outcome measure
- Knowledge of the proposed phase 3 study design.

While this is approach is undoubtably informative, it is predominantly driven through an assessment of efficacy. In practice, there will be additional situations and considerations that need to be taken account when making the EOP2 decision. While theoretically such uncertainties could be incorporated in the elicited priors, some of the more 'qualitative' considerations may not always be easily incorporated into the quantitative assessment. For example,

- Uncertainty in the feasibility of the target population being recruited
- The phase 2 formulation may not have used the final formulation
- Drug or class specific safety considerations
- Differences in regulatory expectations across regions.

In practice it is naturally expected that combining both qualitative and quantitative information will lead to the most appropriate decision. This may be achieved through the use of a bayesian model to predict the probability of efficacy success, followed by making subsequent modulations upwards or downwards depending upon the quantitative assessment of evidence to predict the probability of success on phase 3.

Moreover, in areas such as oncology where drug development times are long, there may be limited patent life remaining post the approval in the first indication. In order to expand use into new additional indications prior to the patent expiry a decision may be required to initiate a phase 3 study with limited or no prior phase 2 data in the target indication or population.

A proposed approach in the situation where there is very limited or no prior data may be to start with a company or industry benchmark for the probability of success based upon historical evidence, and then to modulate this upwards or downwards according to the strength of evidence actually available in the following key areas:

- Clinical efficacy (evidence from same/different drug, disease and patient population)
- Clinical safety (evidence from same/different drug, disease and patient population)
- Mode of action and dose choice (e.g. preclinical/clinical evidence linking the drug target and the disease, prior dose response data (including pharmacokinetic / pharmacodynamic relationships), changes in formulation and/or devices used)
- Risks associated with the phase 3 design (e.g. choice of comparator, primary endpoint, expected dropout rate, power and the multiplicity strategy adopted)
- Feasibility of study execution (company experience in recruiting the desired population).

However, as decision making becomes more qualitative, it becomes more subjective, less data driven, more inconsistent and less decisive. When following such an approach, care must be taken to ensure consistency in modulations across the portfolio so that all new drug candidates are evaluated on a level footing. It should also be noted that the use of a general benchmark as a starting point for the PoS has the disadvantage of not being directly related to the specific target product profile of a drug.

Example use of a Benchmarking approach in 1st-line Ovarian Cancer:

In this example the primary endpoint of the phase 3 study was PFS. The proposed phase 3 study was designed with a power of 90% to detect a hazard ratio of 0.62 (at the 2-sided 5% level, which translates into an 8 month benefit in median PFS over 13 months on placebo) in accordance with the target product profile. An overview of the study design in presented in Figure 3.13. Patients with gBRCAm 1st-line ovarian cancer who were in complete or partial response after platinum based therapy were randomised in a 2:1 ratio to either the Test Drug or placebo. Study treatment continued until progression, with the exception of patients with no evidence of disease at the 2 year point who's treatment was stopped.

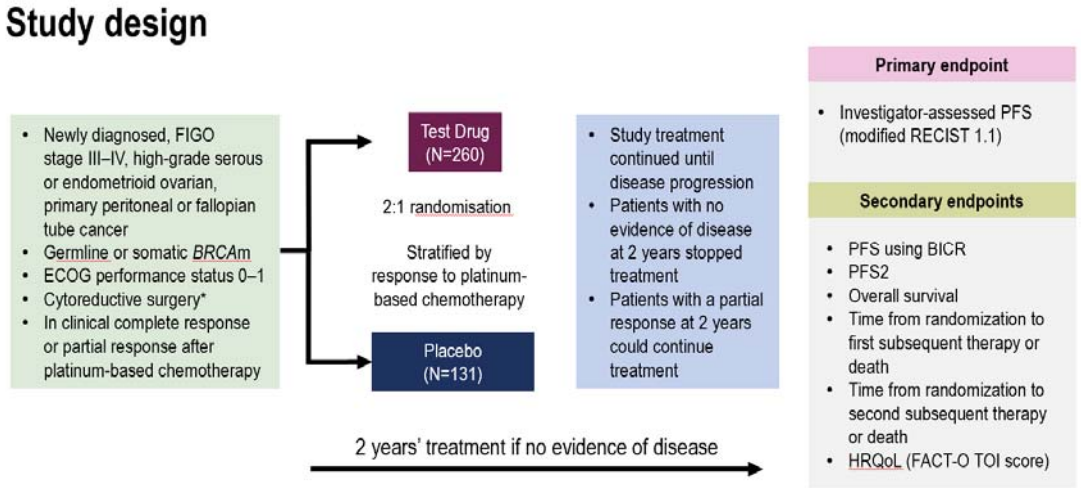


Figure 3.13: Design of a First Line Ovarian Cancer Study

At the time of initiating the study no prior phase 2 data was available in the population treated. The test drug was however approved in the USA as

monotherapy in patients with advanced ovarian cancer and a germline *BRCA*m who have received ≥3 lines of chemotherapy, and also had positive randomised

Benchmark PoS	65%
Acceptable safety data available from other studies in other tumor types and other lines of ovarian cancer at the selected dose level.	+5%
Strong evidence available to support the BRCAm scientific hypothesis from multiple tumor types and other lines of ovarian cancer. Positive results at the selected dose level in other studies.	+20%
No chemistry, manufacturing and quality control issues.	+0%
Uncertainty in the median PFS for the control group. Uncertainty in the cure rate after platinum-based therapy. Uncertainty on the impact of efficacy of stopping the test. treatment at 2 years for patients with non-evaluable disease. Targeted large magnitude of treatment benefit never previously observed in 1st line ovarian cancer setting.	-30%
Scientific advise sought from the FDA and EMA and partially followed. FDA had a preference for OS to be the primary endpoint, but such a study considered infeasible due to timelines.	-10%
Final PoS	50%

Table 3.8: Example benchmarking approach to determine PoS in 1st-line Ovarian Cancer

phase 2 data in patients with platinum-sensitive relapsed ovarian cancer. The internal AstraZeneca benchmark probability of success in phase 3 was set at

65% based upon historical experience. The modulations determined by decision makers for this scenario that lead to a final PoS estimate of 50% are described in Table 3.8. It is of interest to note the use of large modulations by decision makers, including +20 and -30%, within this example. While the final PoS estimate may of course still be appropriate, such large swings are contrary to some of the observations seen through the quantitative approach where wide variations in assumptions may only lead to small changes in the PoS. This emphasises the point that care must be taken to ensure consistency in qualitative modulations across the portfolio so that all new drug candidates are evaluated on a level footing.

3.10. The Utility of a Bayesian PoS Framework in Practice

In addition to becoming part of routine practice at Amgen and AstraZeneca, Crisp (2018) described how the use of assurance embedded within a quantitative bayesian modelling framework as a measure of the PoS has also now become routine practice at GlaxoSmithKline across all therapeutic areas. The publication by Wang (2013) also highlight that Eli-Lilly have also been utilising such an approach to support their development decisions. As highlighted in Section 3.1.1, the underlying choice of bayesian framework used varies in accordance with the available evidence. For example in practice not all phase 2 studies will be randomised. Data may need to be synthesised from randomised and/or non-randomised cohort studies. The modelling approaches however, all lead to the development of the posterior distribution for phase 2 treatment difference in the phase 3 outcome measure for the planned phase 3

population. Some of the key factors influencing the formulation of the framework include:

1. The need to elicit subjective priors versus the availability of empirical priors for the treatment effect in the phase 2 and phase 3 outcome measures.
2. The availability of prior data on the active and control treatment from a randomised, or non-randomised setting (e.g., a single arm study). In certain circumstances multiple prior studies may also be available to combine.
3. The need and availability of data to establish the relationship between the phase 2 and phase 3 study outcome measures. Where substantial historical data is available meta regression may be used to determine the relationship as proposed in the model within this PhD. If not, methods such as those proposed by Hong (2012) may be required to be used.
4. The need and availability of data to adjust for the influence of prognostic factors on the relationship between the phase 2 and 3 outcome measures.
5. The need and availability of data to adjust for changes in population studied between phase 2 and phase 3.
6. Knowledge of the proposed phase 3 study, including the population, objectives, choice of control group and design.

Although the building blocks of bayesian framework may vary depending upon the available evidence, the underlying use of such an approach to predict the

probability of success in a future phase 3 trial has a number of practical advantages:

1. The estimated PoS is more meaningful than focussing on the power of the phase 3 study, as it incorporates current knowledge about the treatment effect. Historically, despite phase 3 studies typically being powered at the 90% level, as described in section 2, the observed phase 3 success rates across all therapeutic areas have been much lower than this. This inconsistency highlights that power is not a good measure of PoS. To clarify further, this does not imply that a power calculation is inappropriate to use for the design of studies, however it should not be misinterpreted as a surrogate for the probability of success of the trial.
2. The assumptions that are driving the calculated PoS are transparent to all. The process of empirically determining or eliciting these assumptions can lead to important refinements in the phase 3 study design and objectives. For example, the relationship between PoS and sample size can be assessed such that an optimal sample size can be identified to balance risks and cost.
3. Each new treatment entering phase 3 will have an associated target product profile that define both the minimum and desirable clinical effect size required for commercialisation. Through the modelling, the probability of achieving these clinically meaningful effect sizes is explicitly characterised in advance.

Figure 3.14 highlights the PoS values for 63 completed phase 3 studies spanning the oncology, respiratory and Cardiovascular portfolio that started during or after 2015 at AstraZeneca. A bayesian modelling framework to predict the probability of success was used in 39/63 (62%) of cases. A benchmarking approach such as that described in Section 3.9 was taken in 24/63 (38%) of cases. Table 3.9 provides some basic summary statistics of the predicted PoS for the completed phase 3 studies by the methodology used to determine the PoS and the ultimate outcome of the phase 3 study.

It is evident from Figure 3.14 and Table 3.9. that those studies with a higher

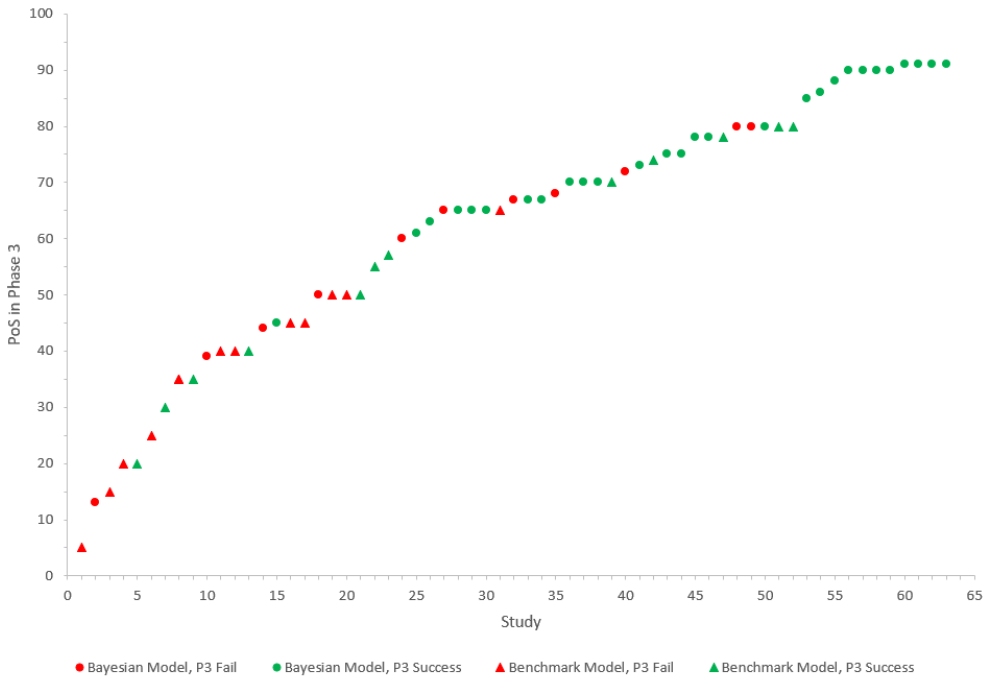


Figure 3.14: Predicted PoS by Methodology and Phase 3 Outcome

predicted PoS are more likely to lead to a positive outcome in the phase 3 study. Over all studies the average PoS that led to initiation of phase 3 studies was 61.4%. Of the studies initiated with a PoS ≥ 61.4%, 82% were successful, while only 36% of studies which initiated with a PoS < 61.4% were successful. The average PoS for studies that used the bayesian modelling approach was 71.5% (with 72% of studies being successful in phase 3), while the average PoS was 45.6% for the benchmarking (with 50% of studies successful in phase 3). The benchmarking approach has generally been used when data is limited to enable the formulation of a bayesian model. This reflects the lower PoS found within these studies. Such confounding in this data set limits the ability to draw comparisons between the utility of the two approaches.

Figure 3.11 also highlights that decisions to invest in phase 3 may still be taken when the estimated PoS is low. In addition to the PoS, additional factors including the competitor landscape, the cost, the unmet need in patients, and commercial opportunity will need to be considered in order to determine if the benefits of moving into phase 3 outweigh the risks. For example, a higher PoS may be required to move into a disease area that has well established treatment options with less unmet need. While trials may still go forward with a low PoS, through following such a modelling framework they do so with sponsors fully informed of the risk. This knowledge may lead to important risk mitigation strategies implemented in the phase 3 study such as the inclusion of interim analyses for futility.

All Studies	N	63
	Average PoS	61.4%
	%Successful Phase 3 ≥ 61.4%	(31/38) 82%
	%Successful Phase 3 < 61.4%	(9/25) 36%
PoS through bayesian modelling	N	39
	Average PoS	71.5%
	%Successful Phase 3	28/39 (72%)
	%Failed Phase 3	11/39 (28%)
	% Success with PoS ≥ 70%	20/23 (87%)
	% Success with PoS ≥ 65%	25/31 (81%)
	% Success with PoS ≥ 60%	27/34 (79%)
	% Success with PoS ≥ 50%	27/35 (77%)
	% Success with PoS < 50%	1/4 (25%)
PoS through benchmarks	N	24
	Average PoS	45.6%
	%Successful Phase 3	12/24 (50%)
	%Failed Phase 3	12/24 (50%)
	% Success with PoS ≥ 70%	5/5 (100%)
	% Success with PoS ≥ 65%	5/6 (83%)
	% Success with PoS ≥ 60%	5/6 (83%)
	% Success with PoS ≥ 50%	8/11 (73%)
	% Success with PoS < 50%	4/13 (31%)

Table 3.9 Summary of PoS in Completed Phase 3 Studies by Methodology and Outcome

It would also be useful to further break these data down by therapeutic area. However, such an evaluation is currently limited by the data provided being blinded across studies and therapeutic area.

3.11. Early Predictions of Indirect Treatment Comparisons and Treatment Ranking

The predicted PoS of the future phase trial is no doubt a very important component of the EOP2 decision. However, it is also important for sponsors/funders to predict and understand how their new investigational treatment is likely to compare with currently marketed and other potential treatment options being developed by other companies at the future time that the sponsors/funders new investigational drug is scheduled to reach the market.

To complement the PoS modelling at the EOP2, Section 3.11 therefore expands the modelling framework further to incorporate network meta-analysis to enable early predictions of indirect treatment comparisons in the phase 3 outcome measure and the probability of being highly ranked amongst the treatment options that will be available at the time of market approval to be evaluated.

Moreover, following regulatory approval of a new treatment, applications for drug reimbursement need to be made in many different geographical regions. If the treatment is considered to have a significant budget impact on health care systems often this will require a form of pharmacoeconomic assessment called a Health Technology Appraisal to take place. Part of this assessment requires

making effectiveness or indirect treatment comparisons between treatments not already compared in head to head trials. Therefore, in addition to predicting how the new investigational treatment will compare to its competitors, the methodology presented in this section also provides results that enable the sponsor to perform early pharmacoeconomic assessments of the investigational drug and thereby consider the likelihood of reimbursement and the future drug costs.

This methodology therefore greatly complements the modelling of the PoS and further supports the EOP2 decision. Such information could, for example, be used to stop the development of an investigational drug prior to investing in a costly phase 3 trial where it is unlikely to beat the competitors and become profitable. This could be despite having a high probability of success in a phase 3 trial designed to show superiority against the currently available standard of care.

No prior literature focussing specifically on developing models to enable early predictions (to be used as part of the EoP2 decision) of the expected efficacy of the investigational treatment at the end of a successful phase 3 study and how this will compare and rank to the current and future successful treatment options available for patients was identified as part of the background review. Within this PhD a bayesian re-sampling approach has been used to determine the predicted probability that a treatment will rank 1st, 2nd or 3rd. Additional graphical and numerical approaches, including the SUCRA score, that may be used to

summarise the results of these treatment ranking probabilities are described in Salanti (2011).

The main challenges of network meta-analysis and the use of indirect comparisons associated with evaluating the assumptions underlying the statistical synthesis of direct and indirect evidence follow through into the work conducted for this PhD. A comprehensive overview of these assumptions including the statistical and nonstatistical methodological considerations are discussed in Salanti (2012) and in Chapter 11 of the Cochrane handbook (Higgins et al 2011, 2019).

The methodology is presented in Section 3.11.1 and illustrated in Section 3.11.2 through extending the pancreatic cancer example introduced in Section 3.6.

3.11.1. Methodology and Framework

The framework of the bayesian model designed to calculate the indirect treatment comparisons and treatment ranking is shown in Figure 3.15. This shows there are four key steps in the process which are fully described below:

Step 1: Change of Endpoint between Phase 2 and 3

The relationship between phase 2 and phase 3 study outcome measures (on the relative difference scale), the influence of prognostic factors on this relationship and the observed treatment difference in the phase 2 outcome measure between the test product and the phase 2 control in the phase 2 study

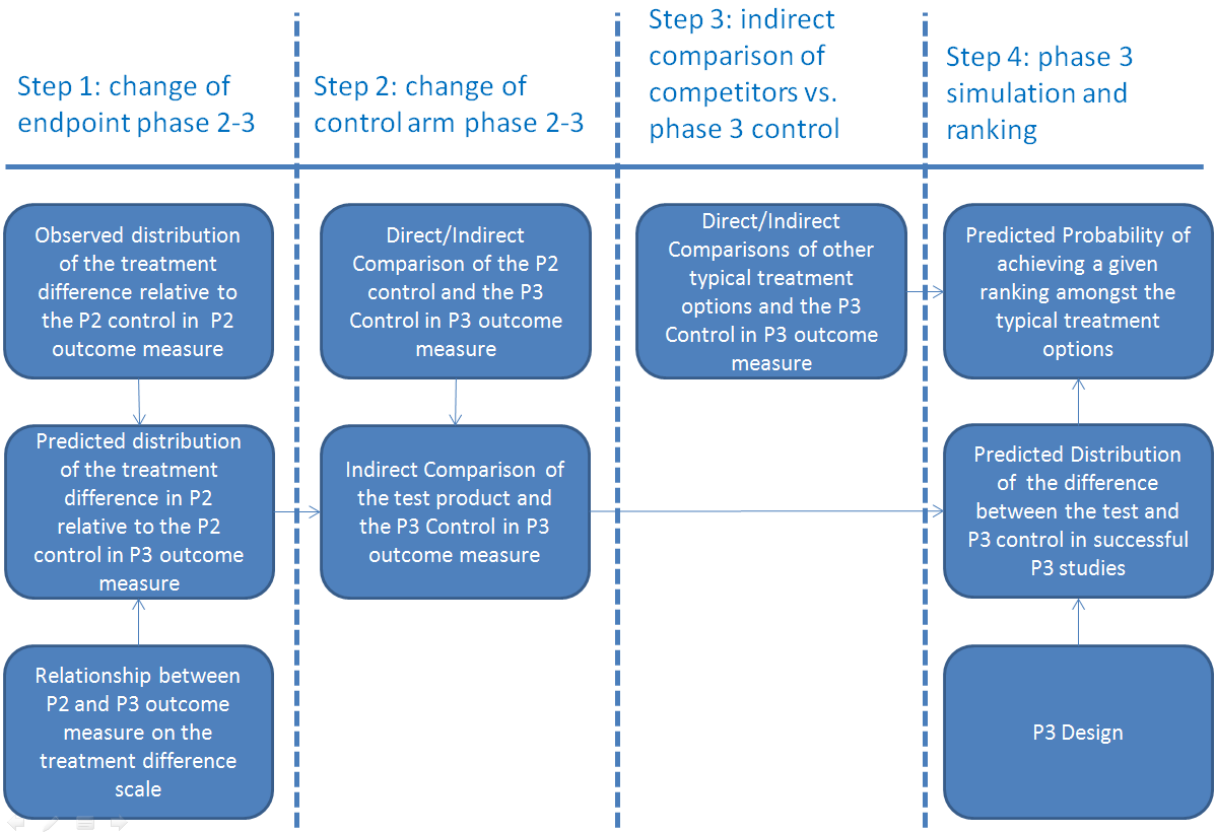


Figure 3.15 Statistical Model for Predicting the Indirect Comparisons and Treatment Ranking

are initially synthesized to predict the difference in the phase 3 outcome measure between the test product and the phase 2 control at the end of phase 2. It should be noted that sceptical and optimistic priors for the treatment difference in the phase 2 outcome measure could also be included at this stage to enable reasonable bounds of belief for the treatment ranking probability to be generated.

Step 2: Change of Control Arm between Phase 2 and 3

Due to a changing clinical landscape (e.g., where an alternative new treatment option becomes standard of care) a different control group may be required for the phase 3 study than used in the phase 2 study. To allow for this the bayesian model synthesizes direct and indirect comparisons of the phase 2 and phase 3 control groups in the phase 3 outcome measure with the predicted distribution of the difference in the phase 3 outcome measure between the test product and the phase 2 control calculated in step 1. This enables the determination of the distribution of the treatment difference in the phase 3 outcome measure between the investigational product and the chosen phase 3 control group.

Step 3: Comparisons of Competitor Treatment Options versus Phase 3 control

The predicted distribution of the treatment difference in the phase 3 outcome measure between other typical treatment options and the phase 3 control are calculated using direct and/or indirect comparisons.

Step 4: Phase 3 Simulation and Treatment Ranking

The properties of the planned phase 3 study design are synthesized with the phase 2 predicted distribution of the treatment difference in the phase 3 outcome measure between the test product and the phase 3 control (as determined in step 2). This enables the development of the predictive

distribution of the treatment difference between the test product and phase 3 control in phase 3 conditional on achieving a successful phase 3 trial. We require this conditional distribution because in practice a health technology appraisal of the test product will not take place unless the phase 3 trial is successful.

Results are concurrently sampled from each of the predicted distributions of the treatment difference comparing the various treatment options against the phase 3 control (step 3) in the phase 3 outcome measure, and the corresponding predictive distribution of the test product against the phase 3 control conditional on achieving a successful phase 3 trial. The sampled results are then ranked and through repeated sampling the probability of achieving a certain efficacy ranking amongst all available treatment options is determined. Although not presented within this PhD, the calculated probabilities can be directly used to calculate the SUCRA score for each treatment and used as an alternative way to rank the treatments overall (Salanti 2011).

It should be noted that prior to constructing such a model a thorough systematic review of the literature is required to ensure that all appropriate historical trials and data are selected to support the decision making process (Moher et al. (2009); Higgins (2011, 2019)). The inappropriate inclusion or exclusion of trials will have a direct effect on the quality of decisions made.

3.11.2. Worked Example in Pancreatic Cancer

To illustrate the methodology, the pancreatic cancer example presented in Section 3.6 is extended to enable early predictions of the outcomes of indirect comparisons and treatment ranking assuming a successful frequentist phase 3 study is achieved, using the predictions of phase 3 outcomes at the end of phase 2 (Path A in the Figure 3.16 below). The competitor situation is a key component of EOP2 decision making and this approach can be used to quantify not only the magnitude of the indirect treatment comparisons, but also their levels of uncertainty at the phase 2 stage of the development (path B in Figure 3.16 below).

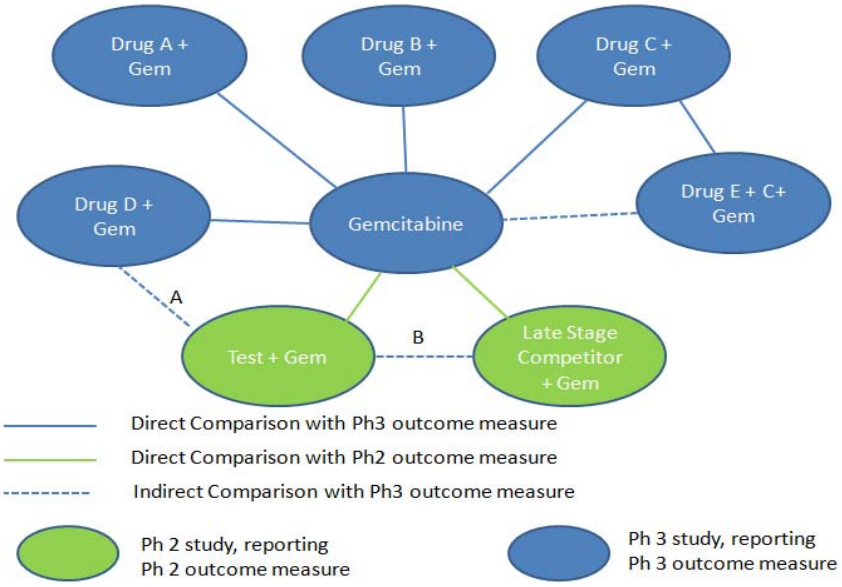


Figure 3.16: Planned Network Diagram for Indirect Comparisons of Pancreatic Cancer Treatments

Recall that Gemcitabine is indicated for use in the first-line treatment of advanced pancreatic cancer. At the time of designing the phase 2 study,

gemcitabine was considered the most commonly used treatment in this setting and hence was used as the control group in a randomized phase 2 study. The phase 2 study was conducted in a 100% metastatic patient population, with this also being the target population for phase 3.

In Section 3.6 we assumed Gemcitabine was also to be the control in the phase 3 study. However, whilst conducting the phase 2 study, new treatment options became available (abraxane in combination with gemcitabine, and FOLFIRINOX).

It is assumed in this case study that gemcitabine + abraxane combination will be selected as the control group for the phase 3 study. In addition to the aforementioned treatment options and in light of advances in clinical practice, the treatment options tarceva in combination with gemcitabine, and oxaliplatin in combination with gemcitabine (GEMOX) were also incorporated into the modelling framework. In some cases, the published studies for these other treatment options were conducted in a mixture of metastatic and locally advanced patients. In these instances, the results in the 100% metastatic subgroup were selected to ensure the indirect comparisons were made in our target population.

The primary outcome measure being used in the phase 2 study is again assumed to be progression free survival (PFS), commonly used as a short-term outcome measure for the phase 3 regulatory outcome measure, overall survival (OS). The EOP2 decision is based upon treatment differences expressed in

terms of hazard ratios. We therefore focus on predicting the distribution for the OS hazard ratio of the test treatment versus the gemcitabine plus abraxane combination in the 100% metastatic population.

The ranking of the test treatment against all other treatment options will be predicted conditional on achieving a successful phase 3 study, starting from the distribution of the PFS hazard ratio (test versus gemcitabine alone) observed in phase 2. The modelling is performed on the log hazard ratio scale to allow the use of the normal distribution and is transposed back onto the hazard ratio scale for presentation purposes.

Within the example presented the impact of the choice of statistical framework for the phase 3 design (superiority or non-inferiority) on the predicted indirect comparisons and treatment rankings is explored. While in practice non-inferiority phase 3 designs have rarely been done in oncology, this is included to highlight how this modelling approach can help to decide between different development strategies.

In these two frameworks, success in phase 3 is defined as achieving a statistically significant result in favour of the test drug (superiority), or that the upper limit of the two-sided 95% confidence interval comparing the test drug to the control is below the largest acceptable non-inferiority margin (non-inferiority).

3.11.3. Step 1: Change of Endpoint between Phase 2 and 3

The first step in the process is to predict the distribution of the treatment difference relative to the phase 2 control (gemcitabine alone) for the phase 3 outcome measure (OS HR) in the phase 3 population (100% metastatic) from the short-term outcome measure (PFS HR) being used for the end of phase 2 decision. In the published data from the systematic literature review described in Section 3.6.1 of all published advanced pancreatic cancer randomized trials over the period 2000 to 2012 in which gemcitabine was used alone or in combination with other therapies (containing a mixture of both metastatic and locally advanced patients), where both the phase 2 and phase 3 outcome measures had been collected, were used to develop a model for predicting the OS HR from the PFS HR in pancreatic cancer. Here it is assumed our phase 2 study was conducted in a 100% metastatic patient population with this also being the target population for the phase 3 study.

The regression presented in Section 3.6.2 was therefore extended to additionally include a study level covariate for the proportion of metastatic patients included in each study, thus enabling the estimation of the posterior predicted OS hazard ratio in the 100% metastatic population.

The model selected was: $\mu_i = \alpha M_i + \beta Z_i$

where μ_i is the true phase 3 outcome measure treatment difference (OS log HR) in the *ith* study, M_i is the proportion of metastatic patients in the *ith* study, Z_i is the phase 2 outcome measure treatment difference (PFS log HR) in the *ith*

study, and β is the slope of the regression line. The meta-data consist of study estimates of phase 3 outcome measure (OS log HR) treatment differences, $\hat{\vartheta}_i$, with variance ε_i^2.

To utilize a random effect model, it is assumed $\hat{\vartheta}_i$ is normally distributed with mean μ_i, such that:

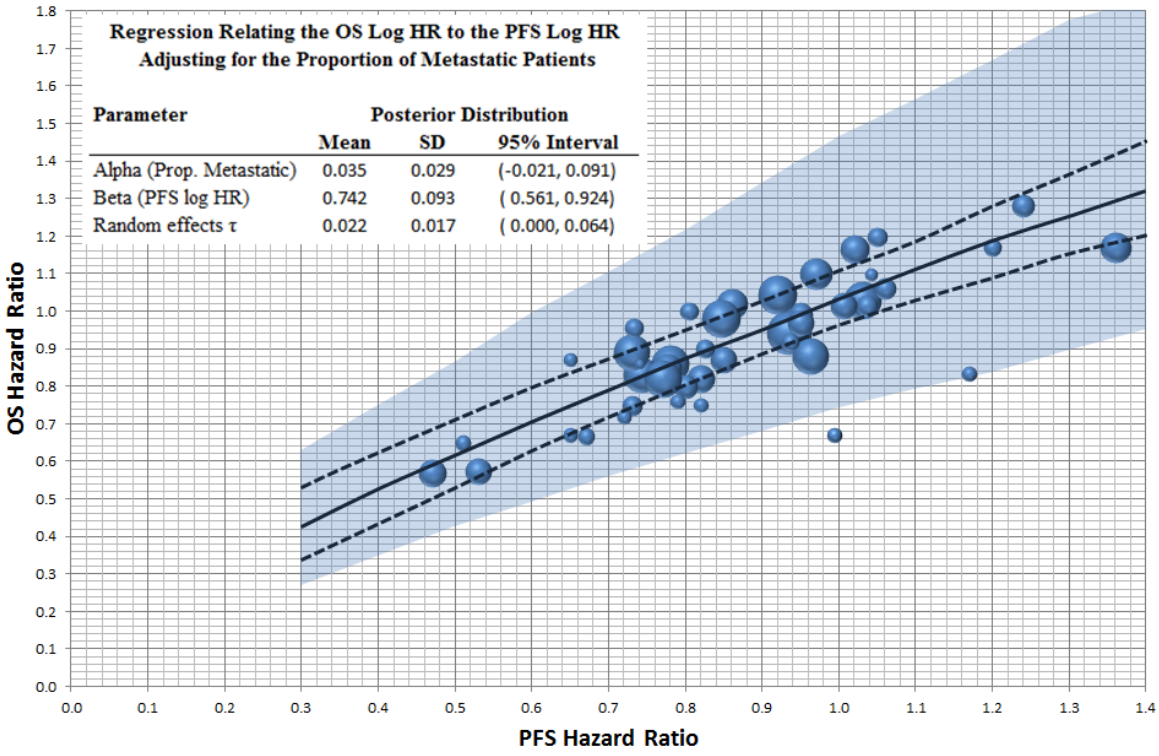
$$\hat{\vartheta}_i \sim N(\mu_i, \varepsilon_i^2)$$

and that μ_i is itself a realisation of a normally distributed random variable such that:

$$\mu_i \sim N(\varphi_i, \tau^2)$$

In the application of this method, a bayesian approach is adopted where α, β and τ are considered as hyperparameters with independent prior distributions. A non-informative prior N(0,10^4) is given to α and β, and a uniform (0,2) prior for τ (Lambert, Sutton, Burton, Abrams and Jones 2005). It should be noted that the potential for publication bias was minimized by including all randomized phase 2 and 3 studies in the model. Additionally, sensitivity analysis excluding the smaller earlier phase studies from the analysis were conducted and showed the small studies to have little impact on the parameter estimates. Figure 3.17 below presents the results of this modelling, with the predicted mean and the 95% CI for a new study in 100% metastatic patients represented by the solid and dotted black lines respectively. Also shown (and represented by the shaded area) is the posterior predicted 95% CrI for the OS HR in the metastatic

population, assuming the phase 2 study had been analysed after observing 80 PFS events. Here, the uncertainty in the phase 2 outcome measure is incorporated within the MCMC simulation process by assuming the phase 2 outcome measure (PFS log HR) is a normally distributed random variable.



Axes are back transformed from a linear regression between OS log(HR) and PFS log(HR) adjusting for the proportion of metastatic patients. The diameter of the circles is inversely proportional to the SE of the OS log(HR) for each published study. The solid and dotted lines represent the predicted mean and 95% CI of a new phase 2 study for fixed PFS HR's. The shaded area shows the posterior predicted 95% CrI for the OS HR in the metastatic population, assuming a phase 2 study with 80 observed PFS events.

Figure 3.17: Random effects meta-regression for OS HR from PFS HR Adjusting for the Proportion of Metastatic Patients

The first two rows of Table 3.10 below show the results of this modelling assuming that a PFS HR of 0.45 was observed in a randomized phase 2 study analysed after observing 80 PFS events (40 per treatment arm). The predicted

phase 2 OS HR (95% CrI) comparing test+gemcitabine versus gemcitabine alone for this example was 0.57 (0.39 to 0.81).

3.11.4. Step 2: Change of Control Arm between Phases 2 and 3

In step 1, the predicted distribution of the OS log HR comparing the test product+gemcitabine versus gemcitabine alone in metastatic patients was calculated. In this next step it is assumed the control group for the future phase 3 trial will be abraxane+gemcitabine. It is wished, therefore, to determine the predicted distribution of the treatment difference between test+gemcitabine and abraxane+gemcitabine. A phase 3 trial containing a direct comparison of abraxane+gemcitabine versus gemcitabine alone within the metastatic population (row 3 of Table 3.10) is available for the phase 3 outcome measure (OS log HR) from the published literature (Von Hoff, 2013).

Assuming the distributions of the treatment difference for OS log HR from the phase 2 study and the abraxane study are normally distributed, the predicted distribution of the indirect treatment difference between test+gemcitabine and abraxane+gemcitabine is computed using MCMC as follows:

1. Sample a value $t_2^{(n)}$, from the distribution for the treatment difference between test+gemcitabine and gemcitabine alone
2. Sample a value $a_3^{(n)}$, from the distribution for the treatment difference between abraxane+gemcitabine and gemcitabine alone
3. Calculate $t_2^{(n)} - a_3^{(n)}$

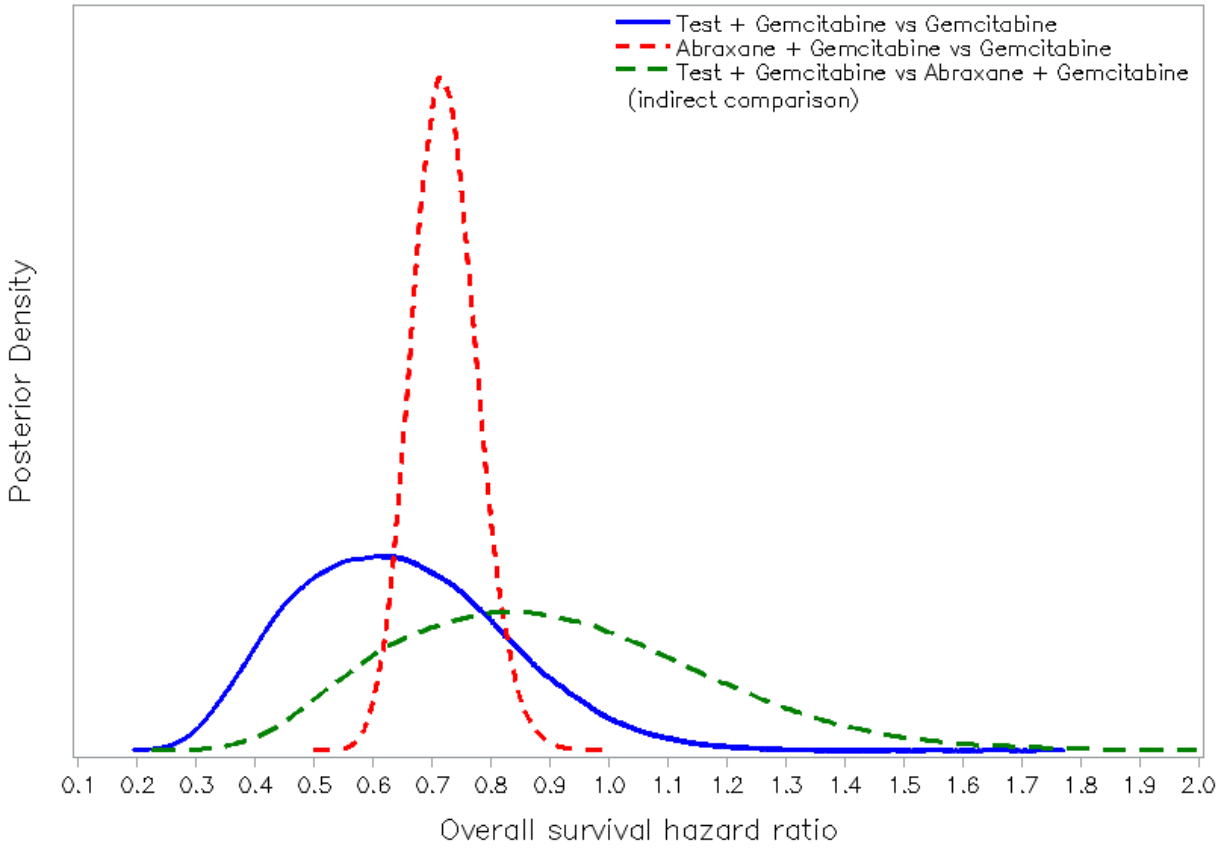
4. Repeat (n) times, to determine the required distribution.

Continuing with the example in Table 3.10, the predicted OS HR (95% CrI) comparing test+gemcitabine versus abraxane+gemcitabine is 0.82 (0.52 to 1.14). This is shown in Figure 3.18 and on the log scale in row 4 of Table 3.10.

	Comparison	**Mean (HR)**	**SD**
Observed Ph2 PFS log HR	Test+Gem vs Gem	-0.799 (0.45)	0.224
Predicted Ph2 OS log HR	Test+Gem vs Gem	-0.557 (0.57)	0.184
Observed Ph3 OS log HR	Abraxane+Gem vs Gem	-0.329 (0.72)	0.076
OS log HR Ind. Comp	Test+Gem vs Abraxane+Gem	-0.228 (0.82)	0.220
OS log HR Ind. Comp	GEMOX vs Abraxane+Gem	0.217 (1.24)	0.129
OS log HR Ind. Comp	Folfirinox vs Abraxane+Gem	-0.232 (0.79)	0.143
OS log HR Ind. Comp	Tarceva+Gem vs Abraxane+Gem	0.094 (1.10)	0.128
Superiority PoS in Ph 3 Prob. Ph3 success and Ranked 1st Prob. Ph3 success and Ranked 2nd	Test+Gem vs Abraxane+Gem	0.583 0.431 0.151	
Non-Inferiority PoS in Ph 3 Prob. Ph3 success and Ranked 1st Prob. Ph3 success and Ranked 2nd	Test+Gem vs Abraxane+Gem	0.810 0.487 0.312	

Ind. Comp = indirect comparison; Gem=Gemcitabine; Ph=Phase.

Table 3.10: Estimating the Ranking Probability for a given phase 2 PFS result (HR=0.45 with 80 PFS events)



This assumes a PFS HR of 0.45 was observed in a randomized phase 2 study analysed after 80 observing PFS events. Plots are back transformed to the HR scale from the log HR scale.

Figure 3.18: Indirect Comparison of Test+Gemcitabine vs abraxane+Gemcitabine for the phase 3 outcome measure (OS HR) at the end of phase 2.

3.11.5. Step 3: Comparisons of Competitor Treatment Options versus Phase 3 control

In addition to the abraxane + gemcitabine combination (the phase 3 control group), the other key treatment options identified were FOLFIRINOX, tarceva in

combination with gemcitabine, and oxaliplatin in combination with gemcitabine (GEMOX). In this step, (with no direct head to head trials available) we use indirect comparisons to determine the distribution of the treatment difference in the phase 3 outcome measure between each of these treatment options and the abraxane + gemcitabine phase 3 control group.

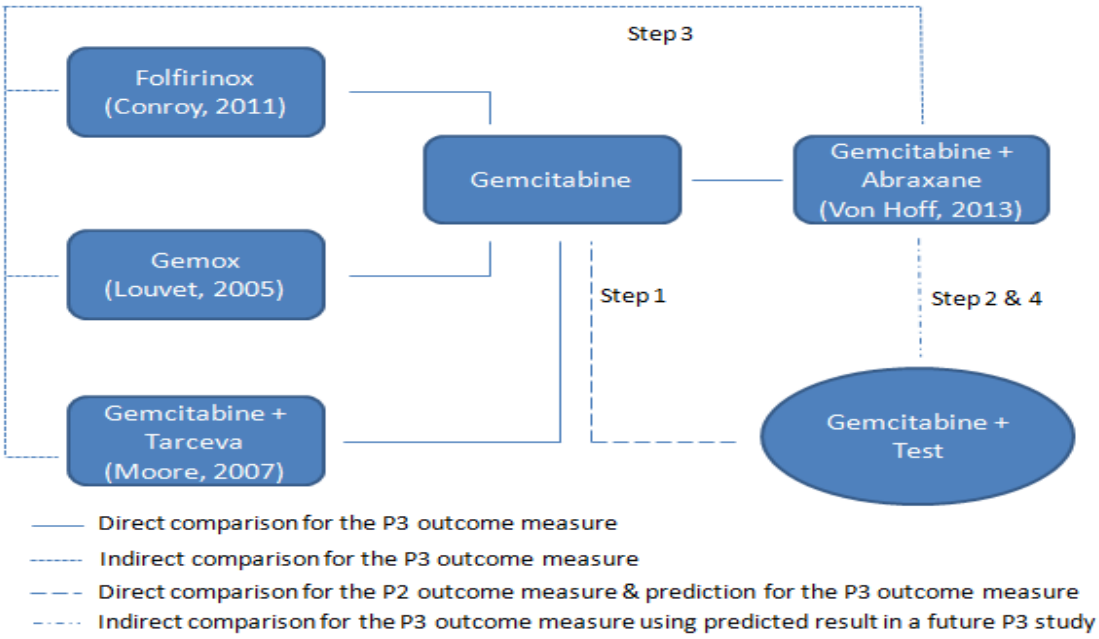


Figure 3.19: Network diagram enabling the indirect comparisons of key treatment options and the phase 3 control group

These distributions are then used in step 4 to predict the probability of achieving a certain efficacy ranking of our test product amongst all of these available treatment options. As no direct head to head trials were available for the comparison with the phase 3 control, the indirect comparisons were calculated using the published results from randomized phase 3 trials that directly compare the various treatment options with the common comparator gemcitabine

monotherapy in the metastatic population. The systematic reviews of pancreatic cancer published by Sabin (2014), Hu (2011), Sultana (2008), and key references identified for each treatment option in the National Comprehensive Cancer Network (NCCN) Version 2.2014 guidelines were used to select the studies and develop the required network (Figure 3.19 above).

The results of the direct comparisons from the selected studies comparing each treatment option with gemcitabine in metastatic patients are presented in Table 3.11. Assuming each of the OS log HR distributions for the treatment differences presented in Table 3.11 (on the OS HR scale) are normally distributed, the predicted distribution of the indirect treatment difference between each treatment option and the phase 3 control (abraxane+gemcitabine) is computed using MCMC as follows:

1. Sample a value $to_i^{(n)}$, from the OS log HR distribution for the treatment difference between the *ith* treatment option *(i= Folfirinox, Tarceva and GEMOX*) and gemcitabine alone.
2. Sample a value $a_3^{(n)}$, from the distribution for the treatment difference between abraxane+gemcitabine and gemcitabine alone.
3. Calculate $ic_i^{(n)} = to_i^{(n)} - a_3^{(n)}$
4. Repeat (n) times, to determine the required distributions, ic_i.

The results of these indirect comparisons are presented on the OS log HR scale in Table 3.10, rows 5 to 7.

Study	Comparison	OS HR (95% CI)
Conroy (2011)	Folfirinox vs gemcitabine	0.57 (0.45, 0.73)
Moore (2007)	Tarceva + gemcitabine vs gemcitabine	0.79 (0.65, 0.97)
Louvet (2005)	GEMOX vs gemcitabine	0.89 (0.61, 1.10)
Von Hoff (2013)	Abraxane + gemcitabine vs gemcitabine	0.72 (0.62, 0.83)

Table 3.11: Direct comparisons of key treatment options with a common comparator (gemcitabine) in the metastatic population

3.11.6. Step 4: Phase 3 Simulation and Treatment Ranking

In this step we select the predictive distribution of the treatment difference between test+gemcitabine and abraxane+gemcitabine in the phase 3 outcome determined in step 2. This is used in conjunction with the properties of the proposed phase 3 study to simulate the results of phase 3 and the distributions from the indirect comparison of the various treatment options and abraxane+gemcitabine (step 3) to predict the probability that the test treatment will be successful in phase 3 and ranked either first or second in terms of efficacy amongst the various treatment options.

Within the example we explore two different frameworks for the phase 3 study, these being superiority and non-inferiority. The superiority framework selected (test+gemcitabine vs abraxane+gemcitabine) requires 459 deaths to enable 90% power to detect an OS hazard ratio of 0.74 or less (a 3 month increase in median OS from 8.5 to 11.5 months) with a 2-sided statistical significance level of 0.05. The non-inferiority framework was set up following the Rothman (2003) 95% two-sided confidence interval procedure, assuming that the new phase 3 study is required to demonstrate that 50% of the active control effect is retained. Following this approach and using the phase 3 OS HR (log HR= -0.329, se =0.076) results from Von Hoff (2013) comparing abraxane+gemcitabine versus gemcitabine, the non-inferiority margin (for test+gemcitabine vs abraxane+gemcitabine) was selected to be 1.094 (OS HR scale). The resulting non-inferiority framework selected required 824 deaths to enable 80% power to detect the alternative hypothesis that the hazard ratio is 0.9 with a one sided 2.5% non-inferiority test. For each design the following steps are then taken:

- Sample a value, $\mathrm{id}_2^{(n)}$, from the predicted distribution of the OS log HR for test+gemcitabine vs abraxane+gemcitabine determined in Step 2
- Assuming the future phase 3 study OS log HR is normally distributed, $X_3 \sim N(id_2, \sigma_3^2)$, sample a value, $X_3^{(n)}$, where σ_3^2 is the variance in the respective phase 3 superiority or non-inferiority study

- Sample a value, $ic_i^{(n)}$ from each of the OS log HR indirect comparison distributions between the *ith* treatment option and abraxane+gemcitabine determined in step 3.

For the superiority framework, calculate the 2-sided 95% confidence interval for $X_3^{(n)}$ and determine whether this indicates a favourable statistically significant result. Note, through repeating this for (n) simulations, the probability of success can also be determined by summarizing the proportion of favourable statistically significant results. If the upper 2-sided 95% CI for exp($X_3^{(n)}$) <1, then order the values of $X_3^{(n)}$ and $ic_i^{(n)}$ from smallest (rank 1) to largest.

For the non-inferiority framework, calculate the 2-sided upper 95% confidence limit for $X_3^{(n)}$ and determine whether exp($X_3^{(n)}$) is less than the non-inferiority margin. Note, through repeating this for (n) simulations the probability of success can be determined by summarizing the proportion of times non-inferiority was achieved. If the upper 2-sided 95% CI for exp($X_3^{(n)}$) < 1.094, order the values of $X_3^{(n)}$ and $ic_i^{(n)}$ from smallest to largest. Note, if $X_3^{(n)}$< 1, and $X_3^{(n)} < ic_i^{(n)}$ then the ranking of the test treatment will be 1.

Repeat this procedure for (n) simulations and summarize the proportion of times the test product is ranked 1st or 2nd.

In summary, the treatment ranking is therefore determined within the MCMC simulations by sampling and comparing estimates of the treatment effect from the predictive distributions of the OS log HR for test+gemcitabine vs

abraxane+gemcitabine in the phase 3 study (assuming the phase 3 trial is successful), and the various treatment options vs abraxane+gemcitabine (the phase 3 control). Following the example shown in Table 3.10 through to completion we learn that, assuming a PFS HR of 0.45 (comparing a test product+gemcitabine to gemcitabine alone) was observed in a randomized phase 2 study analysed after observing 80 PFS events (40 per treatment arm) and we plan to conduct a superiority phase 3 study as described in Step 4, the probability that the trial is successful and the test drug ranked 1st or 2nd for efficacy amongst the competitor treatment options for metastatic pancreatic cancer is 43.1% and 15.1% respectively (Table 3.10 row 8).

Similarly, within the non-inferiority framework selected in step 4, these probabilities are 48.7% and 31.2% respectively (Table 3.10, row 9). These probabilities implicitly assume the chosen framework is acceptable from a regulatory perspective. For the non-inferiority setting this would likely mean there is perceived toxicity or cost-effectiveness benefits and that the efficacy of the test drug is not inferior to the reference by 'too large' an amount.

Figure 3.20 further expands on the superiority example highlighting the impact of the phase 2 sample size and a range of potential phase 2 PFS HR's comparing the test drug in combination with gemcitabine to gemcitabine alone on the probability that the phase 3 trial is successful and the test drug ranked 1st amongst the competitor treatment options.

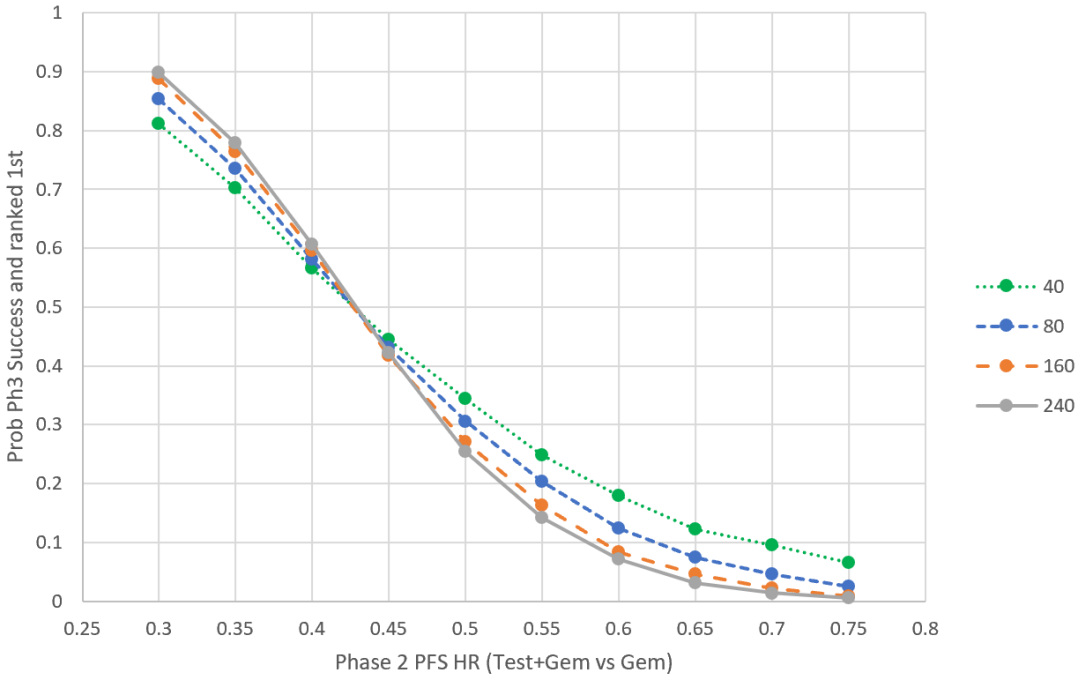


Figure 3.20: Probability of Phase 3 Success and Being Ranked First by Phase 2 sample size and observed PFS HR

Some key takeaways from these additional simulations are:

- The shape of the curve indicates that higher phase 2 sample size leads to greater certainty in being ranked first or not. However, there appears to be little extra certainty to be gained from increasing the phase 2 sample size from 160 to 240 patients.
- If a small phase 2 sample size is used (N=40, 20/arm), over a selected range of potential PFS HR's the predicted probability of success and being ranked first may be up to 10% higher or lower than if a larger phase 2 sample size has been used (e.g. 160 or 240).

- If the observed phase 2 PFS HR is 0.35 or better the probability of success and being ranked 1st will be >70% for all of the sample sizes explored.
- If the observed phase 2 PFS HR is 0.6 or worse the probability of success and being ranked 1st ranges from 7 to 18% over the sample sizes explored. It is expected that knowledge of this these probabilities is likely to be informative to the EOP2 go/no-go decision.

4. DISCUSSION

This PhD has focussed upon enhancing EOP2 decisions through the development of a quantitative process designed to ensure a consistent and explicit evidence-based approach is used to inform drug development decisions for new drug candidates. The process has broadly involved:

- The systematic abstraction of data to support the choice of phase 2 and 3 study design and population, the determination of the relationship between phase 2 and 3 outcome measures, and the understanding of the efficacy in competitor drugs.
- Elicitation of prior belief of the treatment effect in the phase 2 outcome measures.
- The inclusion of a variance inflation factor to incorporate additional uncertainty into the phase 2 result as the observed control effect departs from the prior expected result. This approach may also help to discount early optimistic phase 2 results.
- Evaluation of the probability of achieving the required statistical criteria for efficacy success (PoS) in a future phase 3 study.
- Determination of a threshold for probability of success that is indicative of a go to phase 3 decision.
- The incorporation of qualitative factors into the decision making process, and the implementation of a PoS framework in the situation where very limited or no prior clinical data is available.

- Evaluation of how the expected efficacy of the new investigational treatment compares and ranks to the current and future treatment options available for patients.

Also highlighted is how the process, if applied prior to phase 2 starting, can be used to optimise the drug development program (i.e. the design of the phase 2 and 3 studies) around the PoS in phase 3 for a particular indication.

Following the process therefore leads to the creation of a data package that, in totality, provides direct evidence on the likelihood of success of the development programme, the expected clinical value of the new treatment, and enables the optimisation of the overall development strategy. Moreover, the information developed can also be used as inputs to evaluate the pharmacoeconomic value of a new treatment. It therefore contributes widely to the value assessment undertaken at the EOP2.

Evaluating the probability of efficacy success in phase 3 using the quantitative methodology outlined in this PhD requires the relationship between treatment differences seen using the phase 2 study outcome measure and treatment differences seen using the phase 3 outcome measure to be developed from prior studies through meta-regression. It also requires knowledge of how prognostic factors could influence the treatment difference. Utilising the meta regression approach requires a sufficient sample size. Where estimation methods are based upon asymmetrical assumptions they can easily be biased when the sample size is small. It is recommended that meta-regression

approaches should not be used when there are fewer than 10 studies available. Where insufficient data is available to determine the relationship methods such as those proposed by Hong (2012) should be considered as a useful alternative approach. The bayesian PoS framework presented also assumes that the observed phase 2 treatment difference in the phase 2 outcome measure comes from a randomised phase 2 trial. The phase 2 result is then passed through the meta regression to determine the posterior distribution for phase 2 treatment difference in the phase 3 outcome measure for the planned phase 3 population. In practice not all phase 2 studies will be randomised. Data may come from non-randomised cohort studies, or potentially even a mixture of randomised and non-randomised studies. The underlying choice of bayesian framework that synthesizes this information will therefore need to be appropriately adjusted according to the evidence available. However, all approaches should lead to the development of the posterior distribution for phase 2 treatment difference in the phase 3 outcome measure for the planned phase 3 population, which in conjunction with knowledge of the planned phase 3 trial is used to determine the PoS.

Although the building blocks of the quantitative framework may vary depending upon the available evidence, as detailed in Section 3.10, the underlying use of such an approach to determine the PoS has a number of practical advantages including:

1. The estimated PoS is more meaningful than focussing on the power of the phase 3 study, as it incorporates current knowledge about the treatment effect.

2. The assumptions that are driving the calculated PoS are transparent to all decision makers. The process of determining these assumptions can lead to important refinements to the phase 3 design (for example, inclusion of early futility analyses).

3. Each new treatment entering phase 3 will have an associated target product profile. The probability of achieving this profile is explicitly characterised. Moreover, the process emphasizes the need to target a treatment difference in phase 3 that is clinically worthwhile, realistic and cost effective. The PoS implicitly assumes that this has been done.

It should also be noted that the PoS modelling is predominately driven through an assessment of efficacy. As discussed in Section 3.9, in practice there may be additional quantitative considerations that need to be taken into account when making the EOP2 decision. It would be natural to expect that combining both quantitative and qualitative information will lead to the most appropriate decision. This may be achieved this by incorporating the considerations into the formally elicited sceptical or optimistic priors, or using the bayesian model to predict the probability of efficacy success, followed by manual modification to the PoS to account for the quantitative considerations. Moreover, there will be situations when a decision to start a phase 3 study is required with limited or no

prior data in the target indication or population. In such situations one option discussed in Section 3.9 is to start the PoS evaluation with a company or industry benchmark and then modulate upwards or downwards according to the strength of available evidence. However, as decision making becomes more qualitative it becomes more subjective, less data driven, more inconsistent and less decisive. Section 3.10 presents the PoS, including the methodology used (modelling vs benchmarking) and the ultimate success or failure rate for 63 completed phase 3 studies. While it is evident that trials still may go forward into phase 3 with a low PoS, they are doing so with sponsors fully informed of the risk. It is clear that those studies with a higher predicted PoS are more likely to lead to a positive phase 3 study outcome. Of the phase 3 studies initiated with a predicted PoS ≥61.4%, 82% of them were successful, while only 36% of phase 3 studies initiated with a PoS <61.4% were successful.

By broadly following the concepts highlighted, statisticians can contribute greatly to project strategy and the decision making processes. In many ways the structure and process presented herein are just making explicit many of the implicit assumptions and decisions that are made when deciding whether to move on from a phase 2 result to a phase 3 trial. Additionally, as highlighted in the publication by Sargent et al (2005), if convincing evidence of a strong relationship between a short-term outcome measure and the currently used phase 3 outcome measure is found, there is potential to validate the use of the phase 2 outcome measure as a surrogate for the phase 3 outcome measure

and therefore influence current practice. This may subsequently translate into reduced drug development times.

Operationally, following such an approach is getting easier over time. With study results now registered on clinicaltrials.gov, more complete data is available for meta-analyses which should translate to more robust analyses.

5. FUTURE WORK

The methodology presented considers the size of the phase 3 study to be fixed at that required to show a frequentist success (e.g. for a superiority study with α=0.05). When designing the phase 2 study the PoS for this phase 3 study is evaluated based upon a range of potential results from a phase 2 study of fixed size. An alternative approach could be to fix the PoS for the phase 3 study (at say 65%) and determine the size of the phase 2 study required to achieve this for a range of plausible phase 2 results. This could indicate how reliant a good EOP2 decision could be on the size of the phase 2 study.

Additional work exploring alternative approaches to synthesize a prior for the control group response with the control arm in the phase 2 study may be useful. The primary approach used in the pancreatic cancer example takes the view that the observed treatment difference is the best unbiased estimate available and uses the control prior simply as an external assessment of the trial's robustness. An alternative innovative approach explored takes the view that the observed treatment difference seen in the randomized phase 2 study is the best unbiased estimate available, whilst the variance of the phase 2 treatment difference is inflated as the observed phase 2 control group response departs from the expected. If the phase 2 is randomized, such an approach would however break the randomization and potentially introduce bias. Moreover, the degree of variance inflation chosen is fairly arbitrary. However, this approach

may help to discount early optimistic phase 2 results (Kirby, Burke, Chuang-Stein and Sin, 2012).

Additionally, our pancreatic model assumes that the baseline hazard survivor function is consistent across studies. Whilst this may be a reasonable assumption for a model that uses the relative treatment effects from randomized trials, incorporating methods to adjust for differences in the baseline hazard would be an important attribute to develop for an approach that combines the absolute treatment effects of trial arms across different studies.

The pancreatic model assumes proportional hazards within each study. Whilst there was no reason to doubt this assumption in this indication, a potentially beneficial alternative but more resource intensive approach would be to use methods of data abstraction that reproduce the individual patient data (Guyot, Ades, Ouwens and Welton 2012). This would facilitate selection of an appropriate model from a wide set of parametric survival distributions. Moreover, in indications where the proportional hazards assumption is not so robust and the hazard ratio changes over time, additional work to adjust for the duration of follow up in the abstracted studies would be valuable.

The modelling presented has focused on the methods for enhancing decision making at the EOP2 with respect to the likely efficacy of a new treatment. With some adjustment, the general approach used could be applied to the evaluation of comparative safety data. This would require investigating the relationships between phase 2 and phase 3 safety outcome measures to predict phase 3

safety outcomes given the results of phase 2 safety assessment. If prediction of the likely efficacy and safety outcomes for a phase 3 study can be achieved, then it may also be possible to investigate the benefit-risk of a new drug by employing one of the benefit-risk methods that are currently being developed (EMA Benefit-Risk Methodology Project).

One of the key factors in making a go/no-go decision, and determining the phase 3 development strategy, is the expected efficacy of the investigational treatment and how this compares and ranks to the current and future treatment options available for patients. This is a vitally important consideration at the end of phase 2 because downstream after regulatory approval of a new treatment, gaining agreement for reimbursement in many different geographical regions may require a health technology appraisal to take place, which usually necessitates the use of indirect treatment comparisons between treatments not compared in head to head trials.

In Section 3.11, through using a worked example in pancreatic cancer, the statistical model has been expanded to predict the outcome of such indirect comparisons and evaluate how the efficacy of an investigational product will rank against the typical treatment options for a given indication and population of interest following a successful phase 3 program. The simulation process for treatment ranking can again naturally be extended to any potential development strategy and therefore be used to evaluate the value of different phase 2 and 3 study designs. Similarly, the process does not need to be restricted to efficacy endpoints. Evaluating the treatment ranking associated with key safety or other

outcome measures that are important to assessing the comparative value of treatments would also be a value addition to the end of phase 2 decision.

Again in a similar vein to the approach described in Section 3.4, while not specifically discussed, a threshold for the treatment ranking probability could also be used as the ‘go-to-phase 3’ decision criteria, and the inclusion of sceptical and optimistic priors into the simulations would enable reasonable bounds of belief for the treatment ranking probability to be generated and incorporated in the end of phase 2 decisions.

In summary, assessing the probability of achieving a particular ranking supports end of phase 2 decisions by providing evidence on the expected clinical value of the investigational treatment and may be used as an input to assess the potential risk-benefit and economic value of a new treatment. This is arguably a more explicit way of approaching the end of phase 2 decision problem and is more aligned to the thought process being used in general practice than one which uses the probability of success in phase 3 alone. Additional simulations and worked examples would be useful to further explore the operating characteristics of this approach.

Section 3.9 also acknowledges that, in practice, there are additional qualitative considerations that factor into the EOP2 decision. Further work synthesizing this information with the quantitative data to further refine the PoS may be useful. Moreover, the publication of additional examples on developing a PoS framework where there is limited or no prior information would be informative.

One potential option could be to evaluate the quantitative PoS modulations made in prior studies, and to build a bayesian network to drive consistency and learning in the modulations made for future decisions.

6. CONCLUSION

In conclusion, this PhD outlines a process for drugs entering phase 2 that is designed to enable the application of a consistent and explicit evidence-based approach to EOP2 decision making for new drug candidates. It also provides a structured approach for collecting and synthesizing prior data with the phase 2 data for a new treatment to predict outcomes in future phase 3 studies.

The modelling approach developed complements other published approaches through working on the relative treatment difference scale to maintain the randomization within each study, and its use of meta-regression techniques to estimate the PoS. Reasonable bounds of belief for the PoS are then determined using prior opinions representing different attitudes of key decision makers. The methodology and framework have then been expanded to quantify the expected efficacy of the investigational product and how this ranks against the current and future treatment options in development, while incorporating the levels of uncertainty at the phase 2 design stage. Finally, the development of a PoS framework in the situation where very limited or no prior clinical data is available was also explored.

Broadly following the approaches highlighted will enhance the way statisticians contribute to the EOP2 decision making process, ensure key elements of proposed development plans are investigated through modelling and simulation, enable statistical properties of phase 3 clinical trial designs to be considered relative to downstream evidence synthesis evaluations and enhance the

strategic influence provided by statisticians in drug development. Clinical development leaders will undoubtedly highly value statisticians who are able to make these types of contributions.

To help statisticians implement the approaches developed herein sample code is included in Annex B.

7. ACKNOWLEDGEMENTS

I would like to thank all of my family (in particular Debbie, James, Joseph, John and Vida Sabin), James Matcham, Sarah Bray, Andrew Copas and Max Parmar for all of their support, encouragement and the opportunity to spend time on this research.

ANNEX A - BIBLIOGRAPHY

1. Abou-Alfa 2006, Journal of Clinical Oncology: Randomized Phase III study of Exatecan and gemcitabine compared with gemcitabine alone in untreated advanced pancreatic cancer.

2. Arrowsmith, J. (2011a), “Trial watch: Phase III and submission failures: 2007–2010.” Nature Reviews Drug Discovery, 10: 87.

3. Arrowsmith, J. (2011b), “Trial watch: Phase II and submission failures: 2008–2010.” Nature Reviews Drug Discovery, 10: 328-329.

4. Benda, N., Branson, M., Maurer, W, Friede, T., Aspects of modernizing drug development using clinical scenario planning and evaluation. Drug Information Journal, Vol 44, pp 2099-315, 2010.

5. Berlin 2002, Journal of Clinical Oncology: Phase III study of gemcitabine in combination with Flourouracil versus gemcitabine alone in patients with advanced pancreatic carcinoma: Eastern Cooperative Oncology Group Trial E2297.

6. Boeck 2008, Annals of Oncology: Capecitabine plus oxaliplatin versus capecitabine plus gemcitabine versus gemcitabine plus oxaliplatin: final results of a multicenter randomized phase II trial in advanced pancreatic cancer.

7. Bramhall 2002, British Journal of Clinical Oncology: A double-blind placebo-controlled, randomised study comparing gemcitabine and marimastat with gemcitabine and placebo as first line therapy in patients with advanced pancreatic cancer.

8. Buyse, M., Molenberghs, G., Paoletti, X., Oba, K., Alonso, A., Var der Elst, W., Burzykowski, T., Statistical evaluation of surrogate endpoints with examples from cancer trials. Biometrical Journal 58 (2016) 1, 104–132.

9. Cannistra SA. Phase II trials in journal of clinical oncology. J Clin Oncol 2009;27(19):3073–6.

10. Cantore 2004, Journal of Chemotherapy: Gemcitabine versus FLEC regimen given intra-arterially to patients with unresectable pancreatic cancer: A prospective, randomised Phase III trial of the Italian Society for Integrated Locoregional Therapy in Oncology.

11. Carroll, K.J. 2013, Decision making from phase II to phase III and the probability of success: reassured by assurance. Journal of Biopharmaceutical Statistics, 23:5, 1188-1200.

12. Cascinu 2008, The Lancet: Cetuximab plus gemcitabine and cisplatin compared with gemcitabine and cisplatin alone in patients with advanced pancreatic cancer: a randomised, multicenter, phase II trial.

13. Chen, Cong and Beckman, Robert A. (2009), “Optimal Cost-effective designs of phase II proof of concept trials and associated go-no go decisions”, Journal of Biopharmaceutical Statistics, 19: 424–436, 2009.

14. Cheverton 2004, 2004 ASCO Annual Meeting: Phase III results of exatecan versus gemcitabine in chemotherapy-naïve patients with advanced pancreatic cancer.

15. Chuang-Stein, Christy (2006), sample size and the probability of a successful trial. Pharmaceut. Statist. 2006; 5: 305–309.

16. Chuang-Stein, Christy., Yang, Ruoyong (2010), “A revisit of sample size decisions in confirmatory trials” Statistics in Biopharmaceutical research 2010,2,2.

17. Claret L, Lu J-F, Bruno R, Hsu C-P, Heiz Y-J, Sun Y-N (2012), Simulations using a drug-disease modelling framework and phase II data predict phase III survival outcome in Non-Small Cell Lung Cancer. Clinical pharmacology & Therapeutics | VOLUME 92 NUMBER 5.

18. Colluchi 2002, Gemcitabine alone or with cisplatin for the treatment of patients with locally advanced and/or metastatic pancreatic carcinoma: A prospective, randomized phase III study of the Gruppo Oncologico dell'Italia Meridionale. Cancer. 94 (4) (pp 902-910), 2002.

19. Colluchi 2010, Randomized phase III trial of gemcitabine plus cisplatin compared with single-agent gemcitabine as first-line treatment of patients with advanced pancreatic cancer: The GIP-1 study. Journal of Clinical Oncology. 28 (10) (pp 1645-1651), 2010.

20. Conroy 2011, FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer, N Engl J Med 2011;364:1817-25.

21. Crisp, A., Miller, S., Thompson, D., Best, N., “Practical experience of adopting assurance as a quantitative framework to support decision making in drug development”, Pharmaceutical Statistics. 2018;17:317–328.

22. Cunningham 2005. European Journal of Cancer Supplements Vol 3: Phase III randomised comparison of gemcitabine versus gemcitabine plus capecitabine in patients with advanced pancreatic cancer.

23. Dallow, N., Best, N., Montague, T.H., (2018), “Better decision making in drug development through adoption of formal prior elicitation.” Pharmaceutical Statistics. 2018;17:301–316.

24. De Ridder, F., 2005, Predicting the outcome of phase III trials using phase II data: A case study of clinical trial simulation in late stage drug development. Basic Clin Pharmacol 96:235-41.

25. Di Costanzo 2005, British Journal of Cancer: Gemcitabine with or without continuous infusion 5-FU in advanced pancreatic cancer: a randomized phase II trial of the Italian oncology group for clinical research.

26. DiMasi, J. A., Feldman, L., Seckler, A., Wilson, A. (2012), “Trends in risks associated with new drug development: Success rates for investigational drugs.” Clinical Pharmacology & Therapeutics, 87 3, 272–277.

27. Dowden and Munro (2019), “Trends in clinical success rates and therapeutic focus.” Nature Reviews | Drug Discovery, July 2019 volume 18 https://doi.org/10.1038/d41573-019-00074-z.

28. Dunyak, J., Mitchel, P., Hamrén, B., Helmlinger, G., Matcham, J.,Stanski, D., Al-Huniti, N., Integrating dose estimation into a decision making framework for drug development decision making, Pharmaceutical Statistics. 2018;17:155–168.

29. Frewer, P.,Mitchell, P., Watkins, C., Matcham, J., “Decision making in early clinical development”. Pharmaceut. Statist. 2016, 15 255–263.

30. Friess 2006, BMC Cancer: A randomized multi-center phase II trial of the angiogenesis inhibitor Cilengitide (EMD 121974) and gemcitabine compared with gemcitabine alone in advanced unresectable pancreatic cancer.

31. Gan HK, Grothey A, Pond GR, et al. Randomized phase II trials: inevitable or inadvisable? J Clin Oncol 2010;28(15):2641–7.

32. Götte, H., Kirchner, M., Sailer, M.O., Kieser, M., Sample size planning for phase II trials based on success probabilities for phase III. Pharmaceut. Statist. 2015, 14 515–524.

33. Götte, H., Kirchner, M., Sailer, M.O., Kieser, M., Simulation-based adjustment after exploratory biomarker subgroup selection in phase II, Statist. Med. 2017, 36 2378–2390.

34. Guyot, P., Ades, A., Ouwens, M. J. N. M., Welton, M. J. (2012), “Enhanced secondary analysis of survival data. Reconstructing the data from published survival curves” BMC Medical Research Methodology 2012, 12:9.

35. Heinemann 2006, Journal of Clinical Oncology: A Phase III trial of gemcitabine plus cisplatin compared with gemcitabine alone in advanced pancreatic cancer.

36. Herrmann 2007, Journal of Clinical Oncology: Gemcitabine plus capecitabine compared with gemcitabine alone in advanced pancreatic cancer: A randomized, multicenter, phase III trial of the Swiss Group for Clinical Cancer Research and the Central European Cooperative Oncology Group.

37. Higgins, J. P. T., Green, S. (2011), "Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]." The Cochrane Collaboration, www.cochrane-handbook.org.

38. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition. Chichester (UK): John Wiley & Sons, 2019.

39. Hong, S., Shi, L. (2012) "Predictive power to assist phase 3 go/no go decision based upon phase 2 data on a different endpoint" Statistics in Medicine, DOI: 10.1002/sim.4476.

40. Huang, B., Talukder, E., Han, L., Kuan, P-F (2019) Quantitative decision-making in randomized Phase II studies with a time-to-event endpoint, Journal of Biopharmaceutical Statistics, 29:1, 189-202, DOI: 10.1080/10543406.2018.1489400.

41. Jiang, Kailhong. (2011), "Optimal Sample Sizes and Go/No-Go decisions for Phase II/III development programmes based upon probability of success" Statistics in Biopharmaceutical Research 2011, 3:3.

42. Kindler 2007, 2007 ASCO Annual Meeting: A double-blind, placebo-controlled, randomized phase III trial of gemcitabine (G) plus bevacizumab (B) versus gemcitabine plus placebo (P) in patients (pts) with advanced pancreatic cancer (PC): A preliminary analysis of Cancer and Leukemia Group B (CALGB).

43. Kindler 2010, A placebo-controlled, randomized phase II study of conatumumab or ganitumab (AMG 479) or placebo plus gemcitabine in patients with metastatic pancreatic cancer. J Clin Oncol 28:15s, 2010 (suppl; abstr 4035).

44. Kindler 2011, Axitinib plus gemcitabine versus placebo plus gemcitabine in patients with advanced pancreatic adenocarcinoma: A double-blind randomised phase 3 study. The Lancet Oncology. 12 (3) (pp 256-262), 2011.

45. Kirby, S., Burke, J., Chuang-Stein, C., Sin, C., Discounting phase 2 results when planning phase 3 clinical trials. Pharmaceut. Statist. 2012, 11 373–385.

46. Kola, I., Landis, J. (2004), "Can the pharmaceutical industry reduce attrition rates?" Nature Reviews Drug Discovery, 3, 711-716.

47. Korn, Edward L., Albert, Paul S., McShane, Lisa M. (2005), “Assessing surrogates as trial endpoints using mixed models”. Statist. Med. 2005; 24:163–182.

48. Kulke 2009, Randomized phase II study of gemcitabine administered at a fixed dose rate or in combination with cisplatin, docetaxel, or irinotecan in patients with metastatic pancreatic cancer: CALGB 89904. Journal of Clinical Oncology. 27 (33) (pp 5506-5512), 2009.

49. Lalonde RL, Kowalski KG, Hutmacher MM, Ewy W, Nichols DJ, Milligan PA, Corrigan BW, Lockwood PA, Marshall SA, Benincosa LJ, Tensfeldt TG, Parivar K, Amantea M, Glue P, Koide H. Miller R. Model-based Drug Development. Clinical Pharmacology & Therapeutics 2007; 82:21–32.

50. Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., Jones, D.R. (2005), “How vague is vague? Simulations study of the impact of the use of vague prior distributions in MCMC using WinBUGS.” Statistics in Medicine, 24:2401-2428.

51. Li 2004, 2004 ASCO Annual Meeting: A prospective randomized trial of gemcitabine alone or gemcitabine + cisplatin in the treatment of metastatic pancreatic cancer.

52. Loehr 2009, Cationic liposomal paclitaxel in combination with gemcitabine in patients with advanced pancreatic cancer: A phase II trial. Conference: 2009 Annual Meeting of the American Society of Clinical Oncology, ASCO.

53. Louvet 2005, Journal of Clinical Oncology: Gemcitabine in combination with oxaliplatin compared with gemcitabine alone in locally advanced or metastatic pancreatic cancer: Results of a GERCOR and GISCAD phase III trial.

54. Lutz 2005, Journal of Clinical Oncology: Docetaxel Plus Gemcitabine or Docetaxel Plus Cisplatin in Advanced Pancreatic Carcinoma: Randomized Phase II Study 40984 of the European Organisation for Research and Treatment of Cancer Gastrointestinal Group.

55. Makris E.A, MD, MacBarb R., Harvey D.J., Poultsides G.A., Surrogate End Points for Overall Survival in Metastatic, Locally Advanced, or Unresectable Pancreatic Cancer: A Systematic Review and Meta-Analysis of 24 Randomized Controlled Trials. Ann Surg Oncol (2017) 24:2371–2378 DOI 10.1245/s10434-017-5826-2.

56. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. (2009) “The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.” British Medical Journal, 339.

57. Monzon JG et al (2015) Correlation of single arm versus randomised phase 2 oncology trial characteristics with phase 3 outcome. Eur J Cancer 51:2501-7.

58. Moore 2003, Comparison of gemcitabine versus the matrix metalloproteinase inhibitor BAY 12-9566 in patients with advanced or metastatic adenocarcinoma of the pancreas: A phase III trial of the National Cancer Institute of Canada Clinical Trials Group. Journal of Clinical Oncology. 21 (17) (pp 3296-3302), 2003.

59. Moore 2007, Journal of Clinical Oncology: Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: A phase III trial of the National Cancer Institute of Canada Clinical Trials Group.

60. Nakai 2010, Multicenter randomized controlled trial of gemcitabine (G) alone versus gemcitabine and S-1 combination therapy (GS) in patients with unresectable advanced pancreatic cancer (PC): GEMSAP study. Conference: 2010 Annual Meeting of the American Society of Clinical Oncology, ASCO.

61. Neuenschwander B, Capkun-Niggli G, Branson M, Speiegelhalter DJ. Summarizing historical information on controls in clinical trials. Clinical Trials 2010; 7: 5– 18.

62. Nixon, R. M., Bansback, N., Stevens, J. W., Brennan, A., Madan, J. (2009), "Using short term evidence to predict six-month outcomes in clinical trials of signs and symptoms in rheumatoid arthritis." Pharmaceutical Statistics, 8: 150-162.

63. Nixon, R. M., O'Hagan, A., Oakley, J., Madan, J., Stevens, J. W., Bansback, N., Brennan, A. (2009), "The rheumatoid arthritis drug development model: A case study in bayesian clinical trial simulation." Pharmaceutical Statistics, 8: 371-389.

64. Oettle 2005, Annals of Oncology: A phase III trial of pemetrexed plus gemcitabine versus gemcitabine in patients with unresectable or metastatic pancreatic cancer.

65. O'Hagan, A., Stevens, J. W., Campbell, M. J. (2005), "Assurance in clinical trial design." Pharmaceutical Statistics, 4: 187-201.

66. Oakley JE, O'Hagan A. SHELF: the Sheffield elicitation framework (version 2.0), School of Mathematics and Statistics, University of Sheffield, 2010 http://tonyohagan.co.uk/shelf.

67. Oakley J. SHELF. R package version 1.3.0. 2017. https://cran.r-project.org/web/packages/SHELF/SHELF.pdf.

68. Paoletti, X., Oba, K., Bang, Y. J., Bleiberg, H., Boku, N., Bouch´e, O., Catalano, P., Fuse, N., Michiels, S., Moehler, M.,Morita, S., Ohashi, Y., Ohtsu, A., Roth, A., Rougier, P., Sakamoto, J., Sargent,D., Sasako,M., Shitara, K., Thuss-Patience, E., Van Cutsem, E., Burzykowski, T., Buyse, M. on behalf of the GASTRIC group. (2013). Progression-free survival as a surrogate for overall survival in patients with advanced/recurrent gastric cancer: a meta-analysis. Journal of the National Cancer Institute 5, 1608–1612.

69. Paul, S. M., Mytelka, D.S., Dunwiddie, C.T., Persinger, C. C., Munos, B. H., Lindborg, S. R., Schacht, A. L. (2010), "How to improve R&D productivity: the pharmaceutical industry's grand challenge." Nature Reviews Drug Discovery, 9:203-14.

70. Patel, Nitin R., Ankolekar, Suresh (2007) "A bayesian approach for incorporating economic factors in sample size design for clinical trials of individual drugs and portfolios of drugs" Statist. Med. 2007; 26:4976–498.

71. Paz-Ares, L., Bálint, B., de Boer, R.H., van Meerbeeck, J.P., Wierzbicki, R., De Souza, P., Galimi, F., Haddad, V., Sabin, T., Hei, Y., Pan, Y., Cottrell, S., Hsu. C., RamLau, R., A Randomized Phase 2 Study of Paclitaxel and Carboplatin with or without Conatumumab for First-Line Treatment of Advanced Non–Small-Cell Lung Cancer. Journal of Thoracic Oncology • Volume 8, Number 3, March 2013.

72. Philip 2007, 2007 ASCO Annual Meeting: Phase III study of gemcitabine [G] plus cetuximab [C] versus gemcitabine in patients [pts] with locally advanced or metastatic pancreatic adenocarcinoma [PC]: SWOG S0205 study.

73. Poplin 2009, Journal of Clinical Oncology: Phase III, randomized study of gemcitabine and oxaliplatin versus gemcitabine (fixed dose-rate infusion) compared with gemcitabine (30-minute infusion) in patients with pancreatic carcinoma E6201: A trial of the Eastern Cooperative Oncology Group.

74. Prentice, R. L. (1989). Surrogate endpoints in clinical trials: definitions and operational criteria. Statistics in Medicine 8, 431–440.

75. Pulkstenis, E., Patra, K., Zhang, J. (2017) A bayesian paradigm for decision-making in proof-of-concept trials, Journal of Biopharmaceutical Statistics, 27:3,442-456, DOI: 10.1080 / 10543406.2017.1289947.

76. Reni 2005, Gemcitabine versus cisplatin, epirubicin, fluorouracil, and gemcitabine in advanced pancreatic cancer: A randomised controlled multicentre phase III trial. Lancet Oncology. 6 (6) (pp 369-376), 2005.

77. Renn, Shijie and Oakley, Jeremy E., Assurance calculations for planning clinical trials with time to event outcomes. Statist. Med. 2014, 33 31–45.

78. Richards 2006, Annals of Oncology: Gemcitabine plus CI-994 offers no advantage over gemcitabine alone in the treatment of patients with advanced pancreatic cancer: results of a phase II randomized, double-blind, placebo-controlled multicenter study.

79. Richards 2009, Invest New Drugs: Gemcitabine plus enzastaurin or single agent gemcitabine in locally advanced or metastatic pancreatic cancer: Results of a phase II, randomized, non-comparative study.

80. Riess 2005, 2005 ASCO Annual Meeting: A Randomised, Prospective, Multicenter, Phase III Trial of Gemcitabine, 5-Fluorouracil (5-FU), Folinic Acid vs. Gemcitabine alone in Patients with Advanced Pancreatic Cancer.

81. Riess 2010, Double-blind, placebo-controlled randomized phase III trial of aflibercept (A) plus gemcitabine (G) versus placebo (P). Conference: 12th World Congress on Gastrointestinal Cancer, ESMO Barcelona Spain.

82. Rocha Lima 2004, Journal of Clinical Oncology: Irinotecan plus gemcitabine results in no survival advantage compared with gemcitabine monotherapy in patients with locally advanced or metastatic pancreatic cancer despite increased tumor response rate.

83. Rücker G., Schwarzer G., Ranking treatments in frequentist network meta-analysis works without resampling methods. BMC Med Res Methodol 15, 58 (2015).

84. Sabin, Tony., Matcham, James., Bray, Sarah., Copas, Andrew., Mahesh K. B. Parmar; A Quantitative Process for Enhancing End of Phase 2 Decisions, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2013.852617, 2014.

85. Sabin, Tony., Matcham, James., Copas, Andrew., Mahesh K. B. Parmar; Assessing End of Phase 2 Decision Criteria, Statistics in Biopharmaceutical Research, 2015.

86. Saif, Randomized double-blind phase II trial comparing gemcitabine plus LY293111 versus gemcitabine plus placebo in advanced adenocarcinoma of the pancreas. Cancer Journal. 15 (4) (pp 339-343), 2009.

87. Saint-Hilary G., Barboux V., Pannaux M., Gasparini M., Robert V., Mastrantonio G., Predictive probability of success using surrogate endpoints. Statistics in Medicine. 2019;38:1753–1774.

88. Salanti G., Ades A.E., Ioannidis J.P.A, Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. Journal of Clinical Epidemiology, Volume 64, Issue 2, February 2011, Pages 163-171.

89. Salanti G., Indirect and mixed-treatment comparison, network, or multiple treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res. Syn. Meth. 2012, 3 80–97.

90. Sargent, D. J., Wieand, H. S., Haller, D. G., Gray, R., Benedetti, J. K., Buyse, M., Labianca, R., Seitz, J. F., O'Callaghan, C. J., Francini, G., Grothey, A., O'Connell, M., Catalano, P. J., Blanke, C. D., Kerr, D., Green, E., Wolmark, N., Andre, T., Goldberg, R. M., De Gramont, A. (2005), "Disease-Free Survival Versus Overall Survival As a Primary End Point for Adjuvant Colon Cancer Studies: Individual Patient Data From 20,898 Patients on 18 Randomized Trials." Journal of Clinical Oncology, Vol 23, No. 34.

91. Scheithauer 2003, Annals of Oncology: Biweekly high-dose gemcitabine alone or in combination with capecitabine in patients with metastatic pancreatic adenocarcinoma: a randomized phase II trial.

92. Sharma MR, Stadler WM, Ratain MJ. Randomized phase II trials: a long-term investment with promising returns. J Natl Cancer Inst 2011;103(14):1093–100.

93. Sherrill B, Amonkar M, Wu Y, et al. Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. Br J Cancer 2008; 99: 1572–78.

94. Smith 2003, A phase II/III study comparing intravenous ZD9331 with gemcitabine in patients with pancreatic cancer. European Journal of Cancer. 39 (10) (pp 1377-1383), 2003.

95. Spano 2008, The Lancet: Efficacy of gemcitabine plus axitinib compared with gemcitabine alone in patients with advanced pancreatic cancer: an open-label randomised phase II study.

96. Spiegelhalter, D. J., Abrams, K. R., Myles, J. P. (2004), "Bayesian approaches to clinical trials and health care evaluation." Wiley, ISBN 0-471-49975-7.

97. Spiegelhalter, D.J., Freedman, L.S., Parmar, M. K. B. (1993), "Applying bayesian ideas in drug development and clinical trials." Statistics in Medicine, Vol 12, 1501-1511.

98. Stallard, N., Whitehead, J., Cleall, S. (2005), "Decision making in a phase II trial. A new approach combining bayesian and frequentist concepts." Pharmaceutical statistics, 4: 119–128.

99. Stathopoulos 2006, British Journal of Cancer: A multicenter Phase III trial comparing irinotecan-gemcitabine with gemcitabine monotherapy as first-line treatment in patients with locally advanced or metastatic pancreatic cancer.

100. Tanaka K. , Kawano M., Iwasaki T., Itonaga I., Tsumura H., Surrogacy of intermediate endpoints for overall survival in randomized controlled trials of first-line treatment for advanced soft tissue sarcoma in the pre- and post-pazopanib era: a meta-analytic evaluation. BMC Cancer (2019) 19:56 https://doi.org/10.1186/s12885-019-5268-2.

101. Tap W.D, Jones R.L, Van Tine B.A, Chmielowski B. , Anthony D Elias A.D, Adkins D., Agulnik M., Cooney M.M., Livingston M.B., Pennock G., Hameed M.R., Shah G.D., Qin A., Shahir A., Cronier D.M., Ilaria Jr R. , Conti I., Cosaert J., Schwartz G.K., Olaratumab and Doxorubicin Versus Doxorubicin Alone for Treatment of Soft-Tissue Sarcoma: An Open-Label Phase 1b and Randomised Phase 2 Trial, Lancet 2016 Jul 30;388(10043):488-97. doi: 10.1016/S0140-6736(16)30587-6.

102. Tierney, J. F., Stewart, L. A., Ghersi, D., Burdett, S., Sydes, M. R. (2007), "Practical methods for incorporating summary time-to-event data into meta-analysis." Trials, 8:16.

103. Thomas, David., Burns, Justin., Audette, John., Carroll, Adam., Dow-Hygelund, Corey., Hay, Michael, Bio Industry Analysis (2016): "Clinical development success rate 2006-1015 by BIO, BioMedTracker and Amplion."

104. Van Cutsem 2004, Journal of Clinical Oncology: Phase III trial of gemcitabine plus tipifarnib compared with gemcitabine plus placebo in advanced pancreatic cancer.

105. Van Cutsem 2009, Journal of Clinical Oncology: Phase III trial of Bevacizumab in combination with Gemcitabine and Erlotinib in Patients with Metastatic Pancreatic Cancer.

106. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, Hobbs B, Ibrahim JG, Kinnersley N, Lindborg S, Micallef S, Roychoudhury S, Thompson L (2014) Use of historical control data for assessing treatment effects in clinical trials. Pharm Stat 13:41–54.

107. Viret 2004, 2004 ASCO Annual Meeting: Gemcitabine in combination with cisplatin (GP) versus gemcitabine (G) alone in the treatment of locally advanced or metastatic pancreatic cancer: Final results of a multicenter randomized phase II study.

108. Wang SJ, Hung HMJ, O'Neill RT. Adapting the sample size planning of a phase III trial based on phase II data. Pharmaceutical Statistics 2006; 5:85–97.

109. Wang, Y. et al. (2009) Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. Clin. Pharmacol. Ther. 86, 167–174.

110. Wang, Y., Fu, H., Kulkarni, P., Kaiser, C., Evaluating and utilizing probability of study success in clinical development Clinical Trials 2013; 10: 407–413.

111. Whitehead, A. (2002), “Meta-analysis of controlled clinical trials.” Wiley, ISBN 0-471-98370-5.

112. Wong, Chi Heem., Siah, Kien Wei., Lo, Andrew W. (2018), “Estimation of clinical trial success rates and related parameters”, Biostatistics, Vol 20, Issue 2, Pages 273-286.

ANNEX B - KEY CODE FRAGMENTS

SAS Code to calculate the predicted probability of success in a future Frequentist Phase 3 study

```
**************************************************************************************************;

8.
9.   * Example SAS code to simulate datasets of possible phase 2 results (log PFS HR's);
10. %macro datasim(nsim=);
11.  data p2result;
12.        do trp2_hr_pfs = 0.4 ;  *range of true PFS HR to examine;
13.        do nsub = 40 ;       * range of P2 sample sizes;
14.        do sim= 1 to &nsim;
15.        ind+1;
16.        flag=1;
17.        trp2_lhr_pfs = log(trp2_hr_pfs);
18.        p2_selhr_pfs=sqrt(4/nsub);
19.        p2_lhr_pfs = trp2_lhr_pfs + (p2_selhr_pfs*rannor(ind));
20.      output;
21.    end;
22.    end;
23.    end;
24.   run;
25.
26. proc sort data=p2result out=indx;
27.   by flag ind;
28. run;
29.
30. data ind;
31.   set indx;
32.   by flag ind;
33.   if last.flag ;
34. run;
35.
36. data ind2;
37.  set ind;
38.  %global indmax;
39.  call symput('indmax',ind);
40. run;
41.
42. %mend datasim;
43.
44. %datasim(nsim=500);
45.
46. ********************************************************************************;
47.
48. /* Example SAS code to combine the simulated P2 results with uninformative, sceptical and
    optimistic priors
49.    to create posterior distributions for the phase 2 result (log PFS HRs);
```

```
50. */
51.
52.
53. %macro p2_post_ind(dset,ind,p_mean,p_se,postfix) ;
54.   data _p2result ;
55.     set &dset ;
56.     if ind eq &ind ;
57.   run ;
58.
59.  proc mcmc data=_p2result nbi=1000 thin=20 nmc=1000000 seed=21339
60.       monitor=(p2d) /* plots=(trace density autocorr)*/
61.       statistics(percentage=(2.5 5 25 50 75 95 97.5))=all ;
62.       ods output PostSummaries=mcmc_sum ;
63.       prior p2d ~ normal(&p_mean,sd=&p_se) ;
64.       parms p2d 0 ;
65.       model p2_lhr_pfs ~ normal(p2d,sd=p2_selhr_pfs) ;
66.  run ;
67.
68.  data _p2out (keep=ind p2_diff_&postfix p2_diffse_&postfix) ;
69.    set mcmc_sum ;
70.    ind=&ind ;
71.    p2_diff_&postfix   =mean ;
72.    p2_diffse_&postfix =stddev ;
73.  run ;
74.
75.  data &dset ;
76.   merge &dset _p2out ; by ind ;
77.  run ;
78. %mend ;
79.
80. %macro p2_post(dset,npred,p_mean,p_se,postfix) ;
81.   %do i=1 %to &npred ;
82.      %p2_post_ind(&dset,&i,&p_mean,&p_se,&postfix) ;
83.   %end ;
84. %mend ;
85.
86. %p2_post(p2result,&indmax.,0.00,10,uninf) ;  *uninformative prior;
87. %p2_post(p2result,&indmax.,0.00,0.216843,scept) ; *sceptical prior;
88. %p2_post(p2result,&indmax.,-0.35667,0.344137,optim) ;  *optimistic prior;
89.
90. ********************************************************************************;
91.
92. /* Example SAS code to pass the posterior P2 results through a meta-regression of the
93.    Phase 2 and 3 endpoints to create predictive distributions for the treatment difference
94.    in the phase 3 outcome measure, at the end of phase 2.
95. */
96.
97.
98. data p3result ;
99.   set p2result ;
100.     run ;
101.
102.     %macro p3_post_ind(dset,ind,beta_m,beta_se,tau,postfix) ;
103.       data _p3result ;
104.         set &dset ;
105.         if ind eq &ind ;
```

```
106.      run ;
107.
108.      proc mcmc data=_p3result nbi=1000 thin=20 nmc=1000000 seed=21339
109.          monitor=(p3_diff) /*plots=(trace density autocorr)*/
110.          statistics(percentage=(2.5 5 25 50 75 95 97.5))=all ;
111.          ods output PostSummaries=mcmc_sum ;
112.          prior beta  ~ normal(&beta_m,sd=&beta_se) ;
113.          prior theta ~ normal(0,sd=10);
114.          p3_diff_m=beta*theta ;
115.          prior p3_diff ~ normal(p3_diff_m,sd=&tau) ;
116.             parms beta 0 ;
117.             parms theta 0 ;
118.             parms p3_diff 0 ;
119.          model p2_diff_&postfix ~ normal(theta,sd=p2_diffse_&postfix ) ;
120.      run ;
121.
122.
123.     data _p3out (keep=ind p3_diff_&postfix p3_diffse_&postfix p3_diffll_&postfix
    p3_difful_&postfix) ;
124.        set mcmc_sum ;
125.        ind=&ind ;
126.        p3_diff_&postfix   =mean ;
127.        p3_diffll_&postfix =P2_5 ;
128.        p3_difful_&postfix =P97_5;
129.        p3_diffse_&postfix =stddev ;
130.     run ;
131.
132.     data &dset ;
133.       merge &dset _p3out ; by ind ;
134.     run ;
135.     %mend ;
136.
137.
138.     %macro p3_post(dset,npred,beta_m,beta_se,tau,postfix) ;
139.       %do i=1 %to &npred ;
140.          %p3_post_ind(&dset,&i,&beta_m,&beta_se,&tau,&postfix) ;
141.       %end ;
142.     %mend ;
143.
144.     %p3_post(p3result,&indmax.,0.6883,0.08253,0.02341,uninf) ;
145.     %p3_post(p3result,&indmax.,0.6883,0.08253,0.02341,scept) ;
146.     %p3_post(p3result,&indmax.,0.6883,0.08253,0.02341,optim) ;
147.
148.     data p3posterior (keep=ind nsub trp2_hr_pfs trp2_lhr_pfs p2_lhr_pfs p2_selhr_pfs
149.                       p3_diff_uninf p3_diffse_uninf p3_diffll_uninf p3_difful_uninf
150.                       p3_diff_scept p3_diffse_scept p3_diffll_scept p3_difful_scept
151.                       p3_diff_optim p3_diffse_optim p3_diffll_optim p3_difful_optim) ;
152.       set p3result ;
153.     run;
154.
155.     ********************************************************************************;
156.
157.     /* Example SAS macro combining the phase 3 predictive distribution with a planned
    phase 3 study design
158.     and working out the predictive probability of success.  This is used as part of the end of
    phase 2 decision.
```

```
159.    */
160.    %macro p3_predict_ind(dset,ind,p3_diff,p3_diffse,p3_se,postfix) ;
161.     data _p3predict ;
162.       set &dset ;
163.       if ind eq &ind ;
164.     run ;
165.
166.     proc mcmc data=_p3predict nbi=1000 thin=20 nmc=1000000 seed=88715
167.           monitor=(p3_result p3_result_lcl p3_result_ucl p3_result_sig p3_result_assure)
168.          /* plots=(trace density autocorr)*/
169.           statistics(percentage=(2.5 5 25 50 75 95 97.5))=all ;
170.        ods output PostSummaries=mcmc_sum ;
171.        prior theta  ~ normal( &p3_diff ,sd=&p3_diffse ) ;
172.        parms theta 0 ;
173.        prior p3_result ~ normal(theta,sd=&p3_se ) ;
174.        parms p3_result 0 ;
175.        model general(0) ;
176.           p3_result_lcl = p3_result-1.96*&p3_se ;
177.           p3_result_ucl = p3_result+1.96*&p3_se ;
178.           if     p3_result_lcl gt 0 then p3_result_sig=1 ;
179.           else if p3_result_ucl lt 0 then p3_result_sig=1 ;
180.           else                       p3_result_sig=0 ;
181.           if     p3_result_ucl lt 0 then p3_result_assure=1 ;
182.           else                       p3_result_assure=0 ;
183.     run ;
184.
185.     data _p3out (keep=ind p3_&postfix p3_lcl_&postfix p3_ucl_&postfix p3_sig_&postfix
    p3_assure_&postfix) ;
186.       set mcmc_sum ;
187.       ind=&ind ;
188.          if parameter eq 'p3_result'        then p3_&postfix=mean ;
189.          if parameter eq 'p3_result_lcl'    then p3_lcl_&postfix=mean ;
190.          if parameter eq 'p3_result_ucl'    then p3_ucl_&postfix=mean ;
191.          if parameter eq 'p3_result_sig'    then p3_sig_&postfix=mean ;
192.          if parameter eq 'p3_result_assure' then p3_assure_&postfix=mean ;
193.       retain p3_&postfix p3_lcl_&postfix p3_ucl_&postfix p3_sig_&postfix
    p3_assure_&postfix ;
194.     run ;
195.
196.     data _p3out ;
197.     set _p3out end=lastobs ;
198.     if lastobs ;
199.     run ;
200.
201.     data &dset ;
202.     merge &dset _p3out ; by ind ;
203.     run ;
204.    %mend ;
205.
206.    %macro p3_predict(dset,npred,p3_se,postfix) ;
207.     %do i=1 %to &npred ;
208.     %p3_predict_ind(&dset,&i,p3_diff_&postfix,p3_diffse_&postfix,&p3_se,&postfix) ;
209.     %end ;
210.    %mend ;
211.
212.    data p3prediction ;
```

```
213.      set p3posterior ;
214.    run ;
215.
216.    *N=380 over survival events required;
217.
218.    %p3_predict(p3prediction,&indmax.,0.10260,uninf) ;
219.    %p3_predict(p3prediction,&indmax.,0.10260,scept) ;
220.    %p3_predict(p3prediction,&indmax.,0.10260,optim) ;
221.
222.    proc sort data=p3prediction out= p3preds40hr04;
223.      by ind;
224.    run;
225.    ********************************************************************************;
226.
227.    *Example SAS code to work out the probability a P2 go decision is made for    different
    EOP2" criteria,
228.    *The probability of go+P3 failure and the probability of go+P3 success;
229.
230.    data Pos1;
231.      set p3preds40hr04;
232.      p01=0; p02=0; p03=0; p04=0; p05=0; p06=0; p07=0; p08=0; p09=0; *decision criteria
    for PoS;
233.      if p3_assure_uninf ge 0.1 then p01=1;
234.      if p3_assure_uninf ge 0.2 then p02=1;
235.      if p3_assure_uninf ge 0.3 then p03=1;
236.      if p3_assure_uninf ge 0.4 then p04=1;
237.      if p3_assure_uninf ge 0.5 then p05=1;
238.      if p3_assure_uninf ge 0.6 then p06=1;
239.      if p3_assure_uninf ge 0.7 then p07=1;
240.      if p3_assure_uninf ge 0.8 then p08=1;
241.      if p3_assure_uninf ge 0.9 then p08=1;
242.      run;
243.
244.      proc sort data=pos1 out=pos2;  by trp3_hr_os nsub; run;
245.
246.      proc univariate data=pos2 noprint;
247.        by nsub;
248.        var p01 p02 p03 p04 p05 p06 p07 p08;
249.        output out=pos1out  sum=s01 sum=s02 sum=s03 sum=s04 sum=s05
250.                        sum=s06 sum=s07 sum=s08;
251.        run;
252.
253.    *probability of a go decision;
254.    data po2out;
255.    set pos1out;
256.    ps01=(s01/500); ps02=(s02/500); ps03=(s03/500); ps04=(s04/500); ps05=(s05/500);
257.    ps06=(s06/500); ps07=(s07/500); ps08=(s08/500);  *500 simulations were run in this
    code;
258.    run;
259.
260.    data p3out;
261.      merge p3power po2out;  *merge with a dataset containing the power of the frequentist
    P3 trial (var p3pos);
262.      **Probability of go and success in P3;
263.      pgos01 = ps01*p3pos; pgos02 = ps02*p3pos; pgos03 = ps03*p3pos; pgos04 =
    ps04*p3pos; pgos05 = ps05*p3pos;
```

```
264.       pgos06 = ps06*p3pos; pgos07 = ps07*p3pos;pgos08 = ps08*p3pos;
265.       **Probability of go and fail in P3;
266.       pgof01 = ps01*(1-p3pos); pgof02 = ps02*(1-p3pos); pgof03 = ps03*(1-p3pos); pgof04
     = ps04*(1-p3pos);
267.       pgof05 = ps05*(1-p3pos); pgof06 = ps06*(1-p3pos); pgof07 = ps07*(1-p3pos); pgof08
     = ps08*(1-p3pos);
268.       run;
269.
270.      ************************************************************************************************
     ***;
```

Winbugs Code Fragments for a meta-regression between log PFS and log OS adjusting for the proportion of metastatic patients:

```
model{
 for(i in 1:nstudy){
      lhr_os_w[i]      <- 1/(lhr_os_se[i]*lhr_os_se[i])
      lhr_os[i]        ~ dnorm(lhros[i],lhr_os_w[i])
      lhros[i] ~ dnorm(mu[i], tau)
      mu[i]   <-  (beta*lhr_pfs[i]) + (mpar*meta_p[i])
           }
   beta  ~ dnorm(0,0.001)
   mpar  ~ dnorm(0,0.001)
   sd ~ dunif(0,2)
   tau  <- 1/(sd*sd)
          }
  list(nstudy=44,

lhr_os=c(-0.200,-0.010,-0.288,0.157,-0.058,0.02,-0.182,0.03,-0.151,-
0.139,0.043,-0.163, -0.02,0.016,-0.223,-0.186,-0.128,-0.03,-0.198,0.157,-
```

0.274,0.081,-0.288,-0.117,0.247,0.058,0.094,-0.405,-0.151,0.555,-0.431,0.180,-0.046,0.00, -0.104,-0.084,-0.329,-0.4,0.153,-0.139,-0.4,0.095,0.014,-0.562),

lhr_os_se=c(0.12,0.14,0.17,0.11,0.077,0.11,0.23,0.091,0.089,0.13,0.087,0.12,0.09,0.16,0.14,0.095,0.093,
0.13,0.092,0.19,0.23,0.208,0.24,0.094,0.15,0.16,0.27,0.205,0.35,0.13,0.213,0.177,0.185,0.185,0.186,0.19,0.229,0.233,0.12,0.253,0.24,0.106,0.13,0.123),

lhr_pfs = c(-0.198,-0.051,-0.315,0.307,-0.068,-0.151,0.157,0.03,-0.248,-0.163, -0.083,-0.252, -0.167,0.035, -0.223, -0.297, -0.038, -0.051, -0.261,0.182, -0.236, -0.036, -0.198,-0.315,0.215,0.058,0.041,-0.399, -0.301,0.635, -0.673,0.049, -0.31,-0.217,-0.192,-0.066,-0.329,-0.006,0.02,-0.431, -0.431,-0.03,0.006, -0.755),

meta_p =
c(0.9,0.72,0.51,0.7,0.79,0.85,0.8,0.76,0.71,0.79,0.85,0.69,0.91,0.82,0.8,0.89,0.89,0.79,

0.77,0.82,0.57,0.89,1,1,0.82,0.82,0.73,0.54,0.77,0.64,
0.71,0.88,1,1,1,0.79,0.79,0.79,1,1,1,0.84,0.72,1)

#Output

node	mean	sd	MC error	2.5%	median	97.5%
beta	0.7419	0.09284	1.733E-4	0.5606	0.7419	0.9237
mpar	0.03483	0.02852	5.418E-5	-0.02115	0.03494	0.0905
sd	0.02244	0.01719	3.217E-5	8.759E-4	0.01883	0.064

SAS code for predicting the treatment ranking probability at the end of phase 2

**********;

*dataset of potential phase 2 results in the phase 2 outcome measure PFS log HR;

data p2result;

input ind pfshr events;

```
cards;
1 0.3 80
2 0.35 80
3 0.4 80
4 0.45 80
5 0.5 80
6 0.55 80
7 0.6 80
8 0.65 80
9 0.7 80
10 0.75 80
;
run;
data p2resulta;
  set p2result;
  p2_diff=log(pfshr);
  p2_diffse = sqrt(4/events);
```

```
run;

**********************************************************************;

*Predicting the distribution of the phase 3 outcome measure OS log HR
at the end of phase 2;

*using the regression from the winbugs code;

ods graphics on ;

  proc mcmc data= p2resulta nbi=1000 thin=50 nmc=2000000
seed=21339

          monitor=(p3_diff) plots=(trace density autocorr)

          statistics(percentage=(2.5 5 25 50 75 95 97.5))=all ;

              by ind;

          ods output PostSummaries=mcmc_sum ;

          prior beta  ~ normal(0.7419,sd=0.09284) ;

              prior metap ~ normal(0.03483, sd=0.02852);

              prior theta ~ normal(0,sd=10);

       *predicting at 100% metastatic;

          p3_diff_m=(beta*theta) + (metap*1) ;

          prior p3_diff ~ normal(p3_diff_m,sd=0.02244) ;
```

```
        parms beta 0 ;

            parms metap 0;

        parms theta 0 ;

        parms p3_diff 0 ;

        model p2_diff ~ normal(theta,sd=p2_diffse) ;

run ;

ods graphics off ;




   *collecting the data;

data _p3outmeta

(keep= ind p2_diff_os p2_diffse_os p2_diffll_os p2_difful_os p2_hr_os) ;

      set mcmc_sum ;

      p2_diff_os   =mean;

      p2_diffll_os =p2_5;

      p2_difful_os =p97_5;

      p2_diffse_os =stddev;
```

```
            p2_hr_os= exp(mean);
  run ;
  data post_p2;
     merge _p3outmeta p2result;
     by ind;
  run;
**********************************************************************;
*SAS code for steps 3 and 4 (assuming superiority framework in phase
3);
ods graphics on ;
proc mcmc data=post_p2 outpost=simsup
          nbi=1000 thin=20 nmc=300000 seed=887152
monitor=(p2_diff p2_diff_hr abr_res_hrg oxi_res_hrg fol_res_hrg
tar_res_hrg
abr_res oxi_res fol_res tar_res  lhr_a_p3  hr_a_p3 h2h_res h2h_all
h2h_succ hr_o hr_f hr_t lhr_o lhr_f lhr_t rnk1 rnk2 /*rnk3 rnk4 rnk5*/
h2h_res_lcl h2h_res_ucl h2h_res_assure)
          plots=(trace density autocorr)
```

```
        statistics(percentage=(2.5 5 25 50 75 95 97.5))=all ;
        by ind;
ods output PostSummaries= pos_mcmc_p3post_p80_sup;
*entering phase 2 predictive result in phase 3 endpoint;
        prior p2_diff  ~ normal( p2_diff_os ,sd=p2_diffse_os ) ;
        parms p2_diff 0 ;
        p2_diff_hr = exp(p2_diff);
*entering the abraxane analysis result versus gem control;
        prior abr_res ~ normal(-0.329 ,sd=0.076) ;
        parms abr_res 0 ;
        abr_res_hrg = exp(abr_res);
*indirect comparison of test versus abraxane (based upon P2 predictive
dist);
        lhr_a_p3 = (p2_diff - abr_res);
            hr_a_p3 = exp(lhr_a_p3);
*predicted indirect comparison of test versus abraxane at end of phase 3;
*Assume a superiority phase 3 head to head trial powered for OS HR of;
```

```
*0.739 = 8.5 to 11.5 median increase. 90% power = 459 events,
se=0.093;
        prior h2h_res ~ normal(lhr_a_p3 ,sd=0.093) ;
        parms h2h_res 0 ;
*entering the gemox vs gem result;
        prior oxi_res ~ normal(-0.11328 ,sd=0.105913) ;
        parms oxi_res 0 ;
            oxi_res_hrg = exp(oxi_res);
*entering the folfirinox vs gem analysis result;
            prior fol_res ~ normal(-0.562 ,sd=0.123) ;
        parms fol_res 0 ;
            fol_res_hrg = exp(fol_res);
*entering the tarceva+gem vs gem analysis result in distant metastases;
            prior tar_res ~ normal(-0.23572 ,sd=0.104726) ;
        parms tar_res 0 ;
            tar_res_hrg = exp(tar_res);
        model general(0) ;
*determining the success of the phase 3 trial;
```

```
h2h_res_lcl = h2h_res - (1.96*0.093) ;

h2h_res_ucl = h2h_res + (1.96*0.093) ;

if h2h_res_ucl lt 0 then h2h_res_assure=1 ;

else h2h_res_assure=0 ;

*indirect comparison between test and abraxane at end of phase 3
assuming; *that the phase 3 trial was successful;

if h2h_res_assure=1 then do;

h2h_succ = exp(h2h_res);

end;

*indirect comparison irrespective of success;

h2h_all = exp(h2h_res);

*indirect comp between abraxane and oxaliplatin;

hr_o = exp( oxi_res - abr_res); *select percentiles for CI's;

lhr_o = ( oxi_res - abr_res);   *select percentiles for CI's;

*indirect comp between abraxane and folirinox;

hr_f = exp(fol_res - abr_res); *select percentiles for CI's;

lhr_f = (fol_res - abr_res); *select percentiles for CI's;

*indirect comp between abraxane and tarceva;
```

```
            hr_t = exp(tar_res - abr_res); *select percentiles for CI's;

            lhr_t = (tar_res - abr_res); *select percentiles for CI's;

*working out the ranking following a successful phase 3 trial;

    rnk1=0;

    if (h2h_succ < hr_o) and (h2h_succ < hr_f) and (h2h_succ < hr_t)then

    rnk1=1;

    if h2h_res_assure=0 then rnk1=0;

        rnk2=0;

        if (h2h_succ < hr_o) and (h2h_succ < hr_f) and (h2h_succ ge
hr_t)then

    rnk2=1;

    if (h2h_succ < hr_o) and (h2h_succ ge hr_f) and (h2h_succ <
hr_t)then

    rnk2=1;

        if (h2h_succ ge hr_o) and (h2h_succ < hr_f) and (h2h_succ <
hr_t)then

    rnk2=1;

    if h2h_res_assure=0 then rnk2=0;
```

```
    run ;
  ods graphics off ;
*combining the results;
data sup;
  merge pos_mcmc_p3post_p80_sup p2result;
  by ind;
run;
```

ANNEX C - PANCREATIC CANCER LITERATURE SEARCH RESULTS:

Estimates of treatment effects from the pancreatic cancer literature search

				Overall Survival			Progression Free Survival		
Reference	Author	Control	Experimental	ln HR	SE (lnHR)	HR	ln HR	SE (lnHR)	HR
5	Berlin	Gemcitabine	gemcitabine+fluorouracil	-0.200	0.120	0.820	-0.198	0.088	0.820
7	Bramhall	Gemcitabine	gemcitabine+marimastat	-0.010	0.140	0.990	-0.051	0.130	0.950
10	Cantore	Gemcitabine	FLEC	-0.288	0.170	0.750	-0.315	0.180	0.730
14	Cheverton	Gemcitabine	Exatecan	0.157	0.110	1.170	0.307	0.100	1.360
51	Li	Gemcitabine	gemcitabine+cisplastin	*-0.197*	*0.270*	*0.821*			
72	Philip	Gemcitabine	gemcitabine+cituximab	-0.058	0.077	0.943	-0.068	0.073	0.935
82	RochaLima	Gemcitabine	gemcitabine+irinotecan	0.020	0.110	1.020	-0.151	0.100	0.860
107	Viret	gemcitabine	gemcitabine+cisplastin	*-0.182*	*0.230*	*0.834*	*0.157*	*0.170*	*1.170*
104	Van Cutsem	gemcitabine	gemcitabine+tipifarnib	0.030	0.091	1.030	0.030	0.086	1.030
22	Cunningham	gemcitabine	gemcitabine+capecitabine	-0.151	0.089	0.860	-0.248	0.087	0.780
26	Di Costanzo	gemcitabine	gemcitabine+ CI fluorouracil	*-0.031*	*0.200*	*0.969*			
36	Herrmann	gemcitabine	gemcitabine+capecitabine	-0.139	0.130	0.870	-0.163	0.091	0.850
42	Kindler	gemcitabine	gemcitabine+bevacizumab	0.043	0.087	1.044	-0.083	0.069	0.920
53	Louvet	gemcitabine	gemcitabine + oxaliplatin	-0.163	0.120	0.850	-0.252	0.130	0.777
64	Oettle	gemcitabine	gemcitabine+pemetrexed	-0.020	0.090	0.980	*-0.167*	*0.090*	*0.846*
80	Riess	gemcitabine	gemcitabine+fluorouracil+folinic	0.040	0.100	1.040			
99	Stathopoulos	gemcitabine	gemcitabine + irinotecan	*0.016*	*0.160*	*1.016*	*0.035*	*0.160*	1.036
35	Heinemann	gemcitabine	gemcitabine + cisplatin	-0.223	0.140	0.800	-0.223	0.120	0.800
73	Poplin	gemcitabine	gemcitabine ([FDR])	-0.186	0.095	0.830	*-0.297*	*0.095*	*0.743*
73	Poplin	Gemcitabine	gemcitabine + oxaliplatin	-0.128	0.093	0.880	*-0.038*	*0.093*	*0.963*
1	Abou-Alfa	Gemcitabine	gemcitabine+exatecan	-0.030	0.130	0.970	-0.051	0.100	0.950
59	Moore	Gemcitabine	gemcitabine+tarceva	-0.198	0.092	0.820	-0.261	0.093	0.770
30	Friess	Gemcitabine	gemcitabine+cilengitide	0.157	0.190	1.170	0.182	0.200	1.200
78	Richards	Gemcitabine	gemcitabine+CI-994	0.020	0.170	1.020			
91	Spano	Gemcitabine	gemcitabine+axitinib	-0.274	0.230	0.760	-0.236	0.310	0.790
78	Richards	Gemcitabine	gemcitabine+enzastaurin	0.081	0.208	1.084	-0.036	0.207	0.965
91	Scheithauer	Gemcitabine	gemcitabine+capecitabine	-0.288	0.240	0.750	-0.198	0.170	0.820
105	Van Cutsem	gemcitabine+erlotinib	gemcitabine+erlotinib+bevacizumab	-0.117	0.094	0.890	-0.315	0.088	0.730
54	Lutz	gemcitabine+docetaxel	cisplatin+docetaxel	0.157	0.210	1.170			
6	Boeck	gemcitabine+capecitabine	gemcitabine+oxaliplatin	0.247	0.150	1.280	0.215	0.150	1.240
6	Boeck	gemcitabine+capecitabine	capecitabin+oxaliplatin	0.058	0.160	1.060	0.058	0.160	1.060
12	Cascinu	gemcitabine+cisplatin	gemcitabine+cisplatin+cetuximab	0.094	0.270	1.099	0.041	0.240	1.042
18	Colucci	Gemcitabine	gemcitabine+cisplatin (CDDP)	*-0.405*	*0.205*	*0.667*	-0.399	0.207	0.671
94	Smith	Gemcitabine	ZD9331	-0.151	0.350	0.860	-0.301	0.268	0.740

				Overall Survival			Progression Free Survival		
Reference	Author	Control	Experimental	ln HR	SE (lnHR)	HR	ln HR	SE (lnHR)	HR
58	Moore	Gemcitabine	BAY 12-9566	0.555	0.130	1.742	0.635	0.128	1.887
76	Reni	Gemcitabine	PEFG	-0.431	0.213	0.650	-0.673	0.219	0.510
86	Saif	Gemcitabine	gemcitabein+LY293111	0.180	0.177	1.197	0.049	0.149	1.050
48	Kulke	gemcitabine(FDR)	gemcitabine+cisplatin	*-0.046*	*0.185*	*0.955*	*-0.310*	*0.185*	*0.733*
48	Kulke	gemcitabine(FDR)	gemcitabine+docetaxel	*0.000*	*0.185*	*1.000*	*-0.217*	*0.185*	*0.805*
48	Kulke	gemcitabine(FDR)	gemcitabine+irinotecan	*-0.104*	*0.186*	*0.901*	*-0.192*	*0.186*	*0.825*
52	Loehr	Gemcitabine	gemcitabine+endoTAG-1 low dose	-0.084	0.190	0.919	-0.066	0.165	0.936
52	Loehr	Gemcitabine	gemcitabine+endoTAG-1 mid dose	-0.329	0.229	0.720	-0.329	0.184	0.720
52	Loehr	Gemcitabine	gemcitabine+endoTAG-1 high dose	-0.400	0.233	0.670	-0.006	0.175	0.994
81	Riess	Gemcitabine	gemcitabine+aflibercept	0.153	0.120	1.165	0.020	0.104	1.020
60	Nakai	Gemcitabine	gemcitabine+S-1	-0.329	0.205	0.720			
43	Kindler	Gemcitabine	gemcitabine+conatumumab	-0.139	0.253	0.870	-0.431	0.240	0.650
43	Kindler	Gemcitabine	gemcitabine+ganitumab (AMG 479)	-0.400	0.240	0.670	-0.431	0.237	0.650
19	Colucci	Gemcitabine	gemcitabine+cisplatin	0.095	0.106	1.100	-0.030	0.101	0.970
44	Kindler	Gemcitabine	gemcitabine+axitinib	0.014	0.130	1.014	0.006	0.130	1.006
20	Conroy	Gemcitabine	Folfirinox	-0.562	0.123	0.570	-0.755	0.119	0.470

Values in italics were estimated using the exponential distribution

ANNEX D - PUBLICATIONS

1. Tony Sabin, James Matcham, Sarah Bray, Andrew Copas, Mahesh K. B. Parmar; A Quantitative Process for Enhancing End of Phase 2 Decisions, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2013.852617, 2014.
2. Tony Sabin, James Matcham, Andrew Copas, Mahesh K. B. Parmar; Assessing End of Phase 2 Decision Criteria, Statistics in Biopharmaceutical Research, 2015.

Currently Drafted to be submitted

3. Incorporating Indirect Treatment comparisons in End of Phase 2 Decisions.