Qualitative assessment of the control arm in oncology randomized clinical trials

Olivier, Timothée

How to cite

OLIVIER, Timothée. Qualitative assessment of the control arm in oncology randomized clinical trials. Privat-docent Thesis, 2024. doi: 10.13097/archive-ouverte/unige:175385

This publication URL: https://archive-ouverte.unige.ch/unige:175385
Publication DOI: 10.13097/archive-ouverte/unige:175385

UNIVERSITÉ DE GENÈVE
FACULTÉ DE MÉDECINE

Department of Medicine

"QUALITATIVE ASSESSMENT OF THE CONTROL ARM IN ONCOLOGY RANDOMIZED CLINICAL TRIALS"

Thesis submitted to the Faculty of Medicine of the University of Geneva

for the degree of Privat-Docent
by

Timothée OLIVIER

Geneva

(2023)

TABLE OF CONTENTS

1 – SUMMARY

This thesis focuses on the pivotal, yet frequently underappreciated aspect of randomized controlled trials (RCTs) – the control arm. Our research highlights the critical importance of rigorously designing and implementing this component within RCTs.

The rise of the randomized controlled trial as one of the most robust methods for estimating the effect of a medical intervention is born out of a long history of ideas, particularly in statistical and medical sciences. Randomization has three main advantages: balancing potential confounding factors (known and unknown), avoiding or at least limiting the issue of analytic flexibility, and finally determining a time-zero - the time of randomization - which avoids numerous biases which can affect observational studies. The principle of equipoise, and uncertainty about the result of an experiment, is crucial as it justifies conducting a randomized trial. Estimates suggest that the frequency of suboptimal control arms, within modern clinical trials in oncology, varies from 11% to 17%.

However, based on our works, this proportion may be underestimated, as no systematic framework has been applied to assess for control arm's quality. Beyond factors classically assessed for, we described specific features which can unfavorably affect the control arm. For instance, when the control arm is left to the "physician's discretion", this choice may seem optimal, yet we show that it was restricted in 85% of cases, and such restriction can deprive the patients of essential treatment options. We also demonstrate that the treatment duration in the control arm, when compared to the duration in previous clinical trials in identical settings, can decrease. This raises the question of whether investigators are less incentivized to promote the control treatment, particularly in open-label trials. Lastly, we show that a majority (55%) of head-to-head registration clinical trials in oncology have dose modification rules or supportive care usage that penalize the control arm.

In the conclusion, we higlight that the qualitative analysis of the control arm is not taken into consideration by professional society evaluation scores (such as ASCO and ESMO) and other systematic evaluation methodologies from the evidence-based medicine (EBM) movement. We propose an evaluation framework that could prove beneficial to clinicians, patients, Health Technology Assessment (HTA) agencies, and regulators.

2 – INTRODUCTION

In this work, we will present four research studies, all connected to the qualitative evaluation of the control arm in randomized controlled trials in oncology. The central question underlying these studies can be summarized as follows: is the control arm optimal, adequate, and standard? We will discover that various biases can impact the quality of the control arm, making it suboptimal. When present, this leads to an intrinsic advantage for the experimental arm, which is often the innovative treatment, contradicting the very principle of conducting an investigation that adheres to the highest ethical standards.

The control arm should be approached within the context of what has been a significant advancement in science: the randomized controlled trial. In this introduction, we will be exploring five aspects.

Firstly, we will delve into the history of randomization and how it became the gold standard for evaluating the effectiveness of medical interventions during the 20th century, thanks to successive steps and synchronous evolutions in various scientific domains.

Secondly, we will emphasize on three strengths of randomization that can hardly be replicated by other research experiments.[1]

Thirdly, we will address the perception that random assignment in a trial can disadvantage patients by depriving them of innovative interventions. However, it is crucial to recognize that there is no certainty that the new intervention will be better, as numerous examples have shown. This is the concept of "equipoise," which signifies the state of uncertainty before initiating an investigation.

Fourthly, we will discuss the ethical principles associated with the control arm, focusing on the modern history of clinical trials and the Helsinki Declaration. These ethical principles are of utmost importance when evaluating the quality of the control arm.

Finally, in this introduction, we will describe previous studies that have assessed the prevalence of suboptimal control arms in randomized clinical trials in oncology.

2.1 – A History Of Control And Randomization

2.1.1 – The Concept Of A Control Arm Is Found Throughout History

Well designed, properly conducted randomized clinical controlled trials are nowadays widely considered the scientific gold standard in assessing the efficacy of novel medical intervention.[2] Surprisingly, it is not before our recent history, during the 20th century, that emerged this major advancement in experimental science: systematic randomization within randomized controlled clinical trials. However, a historical perspective allows us to approach its stages and better understand how a series of events enabled this major progress. We find traces in history where the physician refers to the use of a control arm to better delineate the efficacy of their treatment.

In the Old Testament, within the Book of Daniel, dated by historians around the 5th to the 2nd century BC, we find what strongly resembles a controlled trial. In the Book of Daniel, Chapter 1, King Nebuchadnezzar has conquered Jerusalem and has taken several young Jewish men, including Daniel, to Babylon to serve in his palace. The king ordered that these young men be fed the same food and wine that he himself ate, thinking this would keep them healthy and robust. However, Daniel and his friends, Hananiah, Mishael, and Azariah, refused to eat this food and drink this wine, as they believed it would make them impure according to Jewish dietary laws.

So, Daniel proposed a test to the chief of the eunuchs. For ten days, Daniel and his friends would eat only vegetables and drink water. At the end of these ten days, their appearance would be compared to that of the other young men who ate the king's food.

At the end of the ten days, Daniel and his friends appeared healthier and better nourished than the young men who had eaten the king's food. As a result, the chief of the eunuchs allowed Daniel and his friends to continue eating vegetables and drinking water. Below is the excerpt from the Bible recollecting this event.

“1:12 Put your servants to the test for ten days; let them give us grain for our food and water for our drink.
1:13 Then take a look at our faces and the faces of the young men who have food from the king's table; and, having seen them, do to your servants as it seems right to you.
1:14 So he gave ear to them in this thing and put them to the test for ten days.

1:15 And at the end of ten days their faces seemed fairer and they were fatter in flesh than all the young men who had their food from the king's table.
1:16 So the keeper regularly took away their meat and the wine which was to have been their drink, and gave them grain.
1:17 Now as for these four young men, God gave them knowledge and made them expert in all book-learning and wisdom: and Daniel was wise in all visions and dreams."

This story is often interpreted as a demonstration of Daniel and his friends' faith in God, and their determination to uphold Jewish laws, even while held captive in Babylon. Beyond the theological interpretation, the documentation of a controlled experiment holds significant importance.

Another fascinating example is found later, in the 10th century, when a Persian physician named Abu Bakr Muhammad ibn Zakariyaʾ Al-Razi (Rhazes) mentioned the importance of a control arm for evaluating the effectiveness of a treatment.

Here is an excerpt, translated by Selma Tibi, and quoted from the James Lind Library.[3]
We have added emphasis to the part referring to the control arm.

"When the dullness (thiqal) and the pain in the head and neck continue for three and four and five days or more, and the vision shuns light, and watering of the eyes is abundant, yawning and stretching are great, insomnia is severe, and extreme exhaustion occurs, then the patient after that will progress to meningitis (sirsâm) … If the dullness in the head is greater than the pain, and there is no insomnia, but rather sleep, then the fever will abate, but the throbbing will be immense but not frequent and he will progress into a stupor (lîthûrghas). So when you see these symptoms, then proceed with bloodletting. For I once saved one group [of patients] by it, while I intentionally neglected [to bleed] another group. By doing that, I wished to reach a conclusion (ra'y). And so all of these [latter] contracted meningitis."

2.1.2 – James Lind and Scurvy: Between Reality and Myth, a Trial That Shapes History

James Lind, a Scottish surgeon in the British Navy, is credited with conducting one of the earliest documented clinical trials in medical history in 1747. This experiment was carried out in the context of combating scurvy, a disease widespread among sailors of the time due to a deficiency in Vitamin C.

In his book "A Treatise of the Scurvy" published in 1753, Lind detailed the trial he conducted.[4] He chose 12 sailors suffering from scurvy and divided them into six pairs. Each pair received a different treatment, including:

- Cider.
- Twenty-five drops of sulfuric acid.
- Six spoonfuls of vinegar.
- Half a pint of sea water.
- Two oranges and one lemon.
- A spicy decoction of roots and sandalwood.

Lind observed the effects of these different treatments on the participants for 14 days. He noted that the group that received the oranges and lemons recovered much faster than the others. This group was even able to resume their duties aboard the ship, while the other groups remained severely ill.[5]

Stephen Brown, a biographical expert on Lind, suggests that Lind himself may have had doubts about the results of his trial.[6] This uncertainty could have stemmed from Lind's own experience when he attempted to concentrate citrus juice through cooking, aiming for easier transport and storage. Unfortunately, this cooking process destroyed the vitamin C, which was the unknown active ingredient at the time, rendering the boiled product ineffective.

The conclusion of Lind’s experiment was that citrus had a “peculiar advantage”. This lack of conviction was probably part of what explained that it took 42 years from the initial publication until the British admiralty ultimately mandated the inclusion of citrus in the sailors' diet in 1795.

The British Navy associated scurvy with poor organization, and some authors have reported an official culture aimed at downplaying the occurrence of scurvy itself.[7] For instance, the ship on which Lind conducted his clinical trial hardly reported any cases of scurvy. This has even led some authors to question the authenticity of the experimentation altogether.[8]

While it is difficult to definitively separate the myth from reality, the undeniable fact remains that Lind played a significant role in shaping and promoting the importance of a control arm in experimentation. His contributions have had a lasting impact on the evolution of ideas, as well as shaping the modern dissemination of the concept of randomization.

2.1.3 – The 20th Century, The Role Of Sir Ronald Aylmer Fisher

Sir Ronald Aylmer Fisher, a renowned statistician and geneticist, played a crucial role in integrating randomization into experiments. He emphasized the importance of randomization as a fundamental principle in conducting scientific experiments.

Fisher's seminal work, "The Design of Experiments," published in 1935, revolutionized the field of experimental design.[9] In this book, Fisher introduced the concept of randomization as a means to eliminate bias and increase the validity of experimental results. He advocated for the random allocation of treatments to participants or study units to ensure that the assignment is unbiased and free from any preconceived patterns or preferences.

By incorporating randomization, Fisher aimed to minimize confounding variables and enhance the ability to draw causal inferences from experimental data. He recognized that randomization helps create comparable groups, balances potential sources of bias, and allows for valid statistical analysis.

Fisher's influence was likely pivotal in the emergence of the randomized controlled trial in the 20th century. His expertise in statistics and genetics, combined with his unwavering determination, allowed him to effectively advocate for the widespread implementation of randomization in experimental research. This was evident in his pivotal book published in 1935, where he passionately conveyed the importance of randomization as a fundamental principle for obtaining reliable and unbiased results.

2.1.4 – The First Randomized Controlled Clinical Trial: Streptomycin In Tuberculosis

Often regarded as the first example of a randomized controlled clinical trial, the Medical Research Council (MRC) conducted a clinical study published in 1948, focusing on the treatment of pulmonary tuberculosis using streptomycin.[10]

The trial was carried out by a group of researchers affiliated with the MRC, including Dr. Charles Fletcher and Dr. Austin Bradford Hill. The objective of the study was to evaluate the efficacy of streptomycin in treating pulmonary tuberculosis by comparing it to a control group receiving the standard of care at that time: “bed rest”.[11]

The study employed a rigorous methodology that involved randomizing participants into different treatment groups. This random assignment ensured the reduction of potential biases and enabled more reliable comparisons. The results of this trial convincingly demonstrated the effectiveness of streptomycin in the treatment of pulmonary tuberculosis, marking a significant advancement in the use of randomized controlled trials to evaluate medical interventions. This pioneering study paved the way for numerous subsequent research endeavors utilizing similar methods and contributed to the evolution of scientific evidence in the medical field.

2.2 – Three advantages of randomization

2.2.1 – Balancing Known And Unknown Confounders

One of the most universally recognized benefits of randomization is its ability to balance, on average, known and unknown confounders. These are variables that can influence the relationship between the treatment and the outcome. Confounders can introduce bias into the results of a study, making it difficult to ascertain the true effect of the treatment. This is a particularly significant problem in observational or retrospective studies, where the confounders are not controlled for and can have an impact on the results.

Confounding by indication is a very telling example. If intervention A is supposedly more effective but more risky than medical intervention B, one could imagine that physicians will be more likely to propose A to patients who are fitter and more likely to tolerate it as compared to patients where physicians will consider intervention B instead. An observational study could conclude that the intervention A is superior to B, even after controlling for multiple factors such as age, comorbidities, etc. The problem is that it's impossible to control for all variables. It remains therefore possible that the effect seen in the observational study is related to the characteristics of the patients who received interventions, rather than the interventions themselves. These are the characteristics which guided the physician's indication to intervention A or B: thus the name “confounding by indication”.

There are numerous examples where an intervention was adopted or continued based on observational studies, and later, randomization demonstrated that the intervention was either ineffective or harmful. This phenomenon has been termed Medical Reversal by Adam Cifu and Vinay Prasad.[12] In a comprehensive study scrutinizing works published in the New England Journal

of Medicine between 2001 to 2010, the authors found that in the articles testing the standard of care, 40.2% resulted in a reversal of that practice.[13]

In conclusion, randomized controlled trials (RCTs) are designed to balance both known and unknown confounders, reducing the likelihood that these variables will influence the results of the study. This contrasts with observational or retrospective studies, which are often confounded by variables that are not controlled for or accounted for.

2.2.2 – Setting a Time-Zero

Setting a time zero is often an overlooked key feature of randomization. It allows for a clear and unambiguous division between the pre-intervention and post-intervention periods. In a randomized controlled trial, time zero marks the moment when participants are randomly assigned to either the treatment group or the control group.

In observational studies, various forms of bias related to time can emerge that may significantly impact the interpretation of results. Immortal time bias is a concept in cohort study designs where a certain segment of the observation period, known as “immortal time”, is such that the outcome under study could not have occurred. This bias tends to arise when cohort members are assigned to exposure groups (for instance, “treated” versus “untreated”) based on information that becomes apparent after the start of the study, or time-zero.

For instance, study participants may not be classified as 'treated' until they fill their first prescription at some point following their admission into the study. Given that these individuals must have stayed alive or remained free of the event between the time of their study enrollment and their first prescription, this time period is called “immortal time”. However, if this “immortal” period is not correctly accounted for or is left out in the analysis phase, it can give rise to immortal time bias. This tends to skew observed effects in favor of the treatment or exposure under examination.[14,15]

2.2.3 – Limit Analytic Flexibility (Multiple Hypothesis Testing)

Analytic flexibility, allowing for multiple hypothesis testing, has the risk of findings spurious results by chance alone. This is why, when conducting a research work, the hypothesis has to be predetermined before running the experiment.

An example of this is illustrated by opposite results that can be concluded from the same dataset in nutritional science, the famous NHANES (National Health and Nutrition Examination Survey).[16] The NHANES dataset contains personal physical, health data and nutritional habits from thousands of people in the United States. The survey covers a broad range of topics, including demographic characteristics, health history, dietary behavior, and results of physical exams and laboratory tests. It is a unique dataset because of its ability to assess health factors at a population level using both self-reported and direct measures.

In a famous work led by Chirag Patel et al, they explored how the results of health-related studies can be influenced by the choice of factors that researchers take into account when they're analyzing their data.[17] When studying the relationship between diet and health, depending on which factors you consider - like exercise, age, gender, and many others - your results might look different. The researchers call this phenomenon the "Vibration of Effects" (VoE). When they looked at a wide range of health-related variables, they found that the results varied considerably based on which factors they adjusted for in their analysis. In fact, for almost a third of the variables, the effect could appear in the opposite direction depending on the adjustments made. For example, a certain type of vitamin E could be linked with either higher or lower risk of mortality depending on the analysis.

In another example, Brian Nosek's team conducted a fascinating experiment.[18] In their work, 29 teams, made up of 61 analysts, were all given the same task: to determine if soccer referees are more likely to give red cards to dark-skinned players compared to light-skinned players. Despite having the same data and the same question to answer, the teams used a wide range of approaches to analyze the data, resulting in a broad range of estimated effects. To put it into numbers, the estimated effect sizes ranged from 0.89 to 2.93 in odds-ratio units. About 69% (20 out of 29) of the teams found a significant positive effect - meaning dark-skinned players were more likely to get red cards, while 31% (9 out of 29) did not find a significant relationship.

Randomized clinical trials are not immune to the risk of analytical flexibility, or interpreting data in a preferentially favorable way, with the risk of p-hacking. We have detailed how this phenomenon could have occurred in modern randomized trials in oncology.[19,20] Nevertheless, having a pre-specified methodology outlined in a protocol, and ensuring that the statistical analysis is defined in advance, serves as a strong safeguard against potentially questionable research practices.

2.3 – The Concept Of Equipoise: Justification of Randomization

From the beginning of the RCTs era, the idea of conducting a randomized trial often originated from the situation where there was no preferred treatment option from the investigator's point of view, a concept referred to as "theoretical equipoise" by Benjamin Freedman in 1987.[21] However, Freedman posited that an investigator's true impartiality is exceedingly rare and could unnecessarily limit the number of trials if made a requirement. In response to this, he introduced a broader concept known as "clinical equipoise" which encapsulates the "genuine uncertainty within the medical expert community...about the preferred treatment".

While Freedman's "clinical equipoise" met some resistance, it has largely gained acceptance over time. The rationale for this acceptance is the idea that community-wide uncertainty is a more practical and realistic condition for initiating clinical trials, rather than requiring individual investigators to lack preference.

A famous quote from Dr. Thomas Chalmers may be helpful in reminding the uncertainty of the superiority of the novel intervention, which also render the principle of randomization more acceptable for clinicians and patients:

"One only has to review the graveyard of discarded therapies to discover how many patients have benefited from being randomly assigned to a control group.".

The concept of equipoise, particularly clinical equipoise, underscores the importance of addressing uncertainty in medical research. It ensures that clinical trials are initiated whenever there is uncertainty, in order to progressively build scientific evidence.

2.4 – Ethical Considerations: What Should Be The Control Arm?

Alongside the development of randomized clinical trials, ethical issues and their enforcement have emerged as central and closely linked with the methodology and designs of clinical trials.[22]

For example, it's clear that a control group that is less than optimal, meaning it does not represent the standard of care, can disadvantage patients participating in the clinical trial. This can be seen as a potential violation of ethical principles. Another crucial ethical rule for running a trial is that it

should ultimately address significant questions. If the trial's design doesn't allow this, it can be deemed unethical.[23]

The Helsinki Declaration is a set of ethical guidelines concerning medical research involving human subjects. It was adopted by the World Medical Association (WMA) in 1964, in Helsinki, Finland, hence its name.[24] Before the Helsinki Declaration, the Nuremberg Code (1947) was one of the first sets of guidelines established for human subject medical research, in response to the inhumane medical experiments conducted during World War II.

The Helsinki Declaration expanded on these principles, focusing on the necessity of informed consent, the balance of risks and benefits in research, and the protection of vulnerable populations. Since 1964, the Helsinki Declaration has been revised several times to respond to evolving research practices and ethical standards. Revisions have addressed issues such as the use of placebos, access to post-research interventions, and the obligation to account for research results. Today, nonetheless debates emanating from agencies like the FDA about adopting the most recent versions,[25] the Helsinki Declaration remains one of the most influential ethical standards for human subject medical research.

We identified two principles of the Helsinki Declaration, that could be related to the control arm. We have added emphasis to the part referring to the control arm:

“8. While the primary purpose of medical research is to generate new knowledge, this goal can never take precedence over the rights and interests of individual research subjects.”

“33. The benefits, risks, burdens and effectiveness of a new intervention must be tested against those of the best proven intervention(s), except in the following circumstances:
Where no proven intervention exists, the use of placebo, or no intervention, is acceptable; or
Where for compelling and scientifically sound methodological reasons the use of any intervention less effective than the best proven one, the use of placebo, or no intervention is necessary to determine the efficacy or safety of an intervention and the patients who receive any intervention less effective than the best proven one, placebo, or no intervention will not be subject to additional risks of serious or irreversible harm as a result of not receiving the best proven intervention.
Extreme care must be taken to avoid abuse of this option.”

In other words, the Helsinki Declaration serves as a comprehensive and simple guideline. According to its principles, enrolling a patient in a clinical trial and assigning them to a control arm with an inferior treatment compared to what they would have received outside the trial would be considered unethical.

Due to the inherent connection between ethical and methodological concerns, as highlighted with the example of the control arm, some proponents argue for the creation of a "methodological review board" that surpasses the responsibilities of an ethical review board.[26] This specialized board would evaluate multiple aspects, with a particular emphasis on the most demanding question: whether the study will yield significant and valuable findings. Moreover, it would offer methodological guidance to address this concern and enhance the overall quality of the study.

2.5 – Estimation Of Suboptimal Control Arms In Oncology

The qualitative assessment of the control arm has relatively recently appeared in the medical literature on oncology.

Industry sponsorship bias, also known as funding bias or sponsorship bias, refers to the tendency of a scientific study to be influenced by the financial interests of its sponsor. This bias can manifest through distorted research design, conduct, and publication methods. Examples include selecting non-representative study populations, or using suboptimal control arms. A Cochrane review found that results from industry sponsored trials were reporting more favorable results than trials sponsored by other entities. [27]

It is therefore possible that the issue of suboptimal control arm was less relevant at a time where industry was not dominating the sponsorship of randomized clinical trials. Del Paggio and colleagues found the percentage of trials funded by the pharmaceutical industry exhibited a significant rise. In the earlier cohort studied (1995-2004), 57% of RCTs were industry-funded. This percentage saw a substantial increase in the recent decade, with 89% of RCTs being funded by the pharmaceutical industry from 2010-2020.[28]

In a work published by Talal Hilal and colleagues in 2017, they empirically evaluated the quality of control arms within randomized clinical trials (RCTs) leading to marketing authorization of anticancer drug by the US Food and Drug Administration (FDA) between 2013 and 2018. The

researchers studies 143 drug approvals and found that 16 (17%) were using substandard control arms.

A recent study, led by Alessandro Rossi, focused on anti-cancer randomized controlled trials (RCTs) published between 2017 and 2021, and assessed the frequency of suboptimal control arms.[29] This analysis differed from the one conducted by Hilal et al. by examining trials published across 11 major oncology journals. The analysis encompassed 387 studies, out of which 43 (11.1%) control arms were judged as suboptimal. The study found that the rates of suboptimal control arms were higher in industry-sponsored trials compared to academic trials.

The lower frequency of suboptimal control arm in published trials (11.1%) as compared to registration trials (17%) is most probably explained by the fact that almost all registration trials are industry sponsored.[30] It is possible, however, that based on other specific features we will present in the following part, that these numbers are still underestimating the prevalence of suboptimal control arms in oncology.

3 – ORIGINAL ARTICLES

3.1 – The ASCENT trial analysis

Sacituzumab govitecan in metastatic triple negative breast cancer (TNBC): Four design features in the ASCENT trial potentially favored the experimental arm. Timothée Olivier, Vinay Prasad. Translational Oncology 2022 Jan;15(1):101248. [31]

In this article, we analyzed the ASCENT trial and found four biases that raise concerns about the applicability of the reported improvements in overall survival (OS) for patients with metastatic triple-negative breast cancer (TNBC) treated with sacituzumab govitecan compared to single-agent chemotherapy. The four biases identified were: 1) the open-label design, which may exaggerate the experimental arm's effect, 2) the choice of progression-free survival (PFS) as a primary endpoint, partly because this endpoint is subject to the risk of amplifying the benefit in cases of early-stopping rules, 3) the control arm was not a true "physician's choice" but a restricted one, preventing the use of important therapeutic options and leading to a substandard control arm, and 4) different dose reduction and G-CSF (supportive care) recommendations between the experimental arm and the control arm, potentially favoring the experimental arm. Those 4 features could have lead to an exaggeration of the reported survival benefit of sacituzumab govitecan over chemotherapy in the specific setting of the ASCENT trial (second and subsequent lines of therapy). Below is a figure from our work, illustrating how each of this feature could have play a role in amplifying the reported benefit.

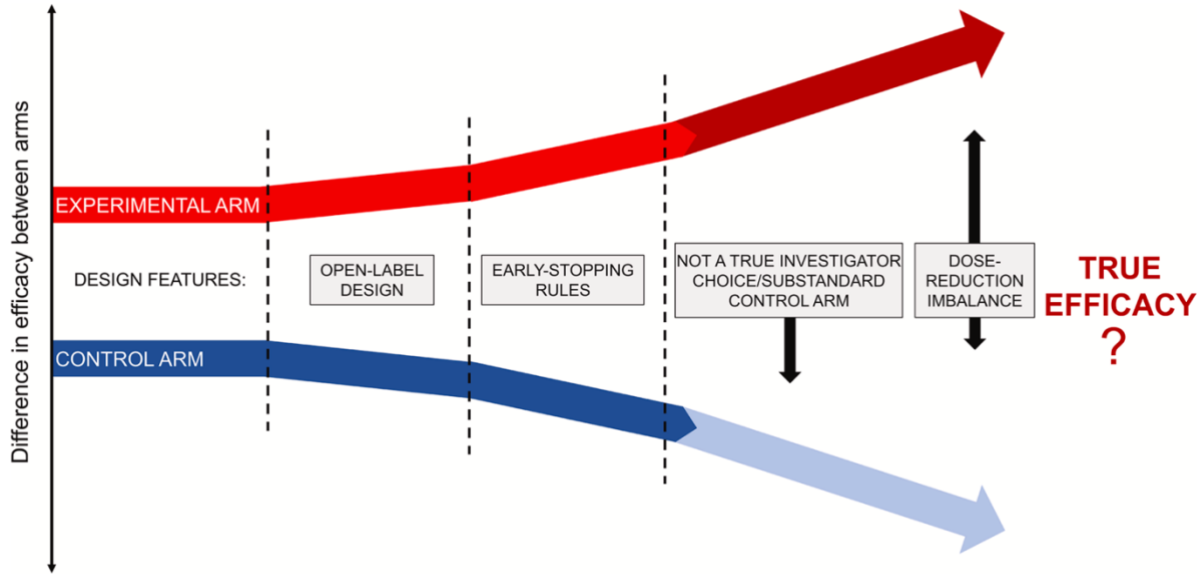


Fig. 2. Potential cumulative effect for each bias in design-features of the ASCENT trial.

Focusing on the two biases related to the control arm, we found issues with the concept of "physician's choice" and the problem of dose reduction and G-CSF rules.

Firstly, the term "single-agent chemotherapy of the physician's choice" is misleading as physicians could not choose platinum or anthracyclines, both agents that may have been preferred in this setting. In the control arm, 31% and 17% of patients had not been exposed to these therapies, respectively. This restriction led to a substandard control arm that may not accurately represent real-world treatment choices.

Secondly, we found imbalance in dose-reduction recommendations between arms. The trial report lacks transparency in documenting dose-modification recommendations: the authors refer to Fig. S8 for dose-modification recommendations for sacituzumab govitecan, which are the same as in the FDA label. However, within the trial, patients in the experimental arm were not treated according to these rules: the protocol did not advise dose reduction after the first episode of severe febrile neutropenia in the sacituzumab govitecan arm. In contrast, for the same toxicity, dose reductions were applied in the control arm, and G-CSF was not mandatory: these differences in dose-reduction and supportive care recommendations favored the experimental arm, allowing for higher dose-intensity in the experimental arm.

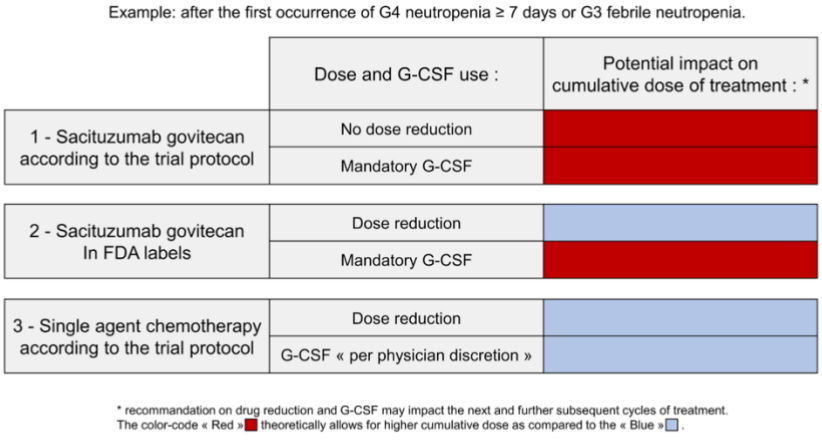


Fig. 1. Differences in recommendations in dose modification and G-CSF use for (1) sacituzumab govitecan within the protocol, (2) sacituzumab govitecan according to the FDA labels, (3) single agent chemotherapy in the ASCENT trial [1–15].
Example described here: after the first occurrence of G4 neutropenia ≥ 7 days or G3 febrile neutropenia.

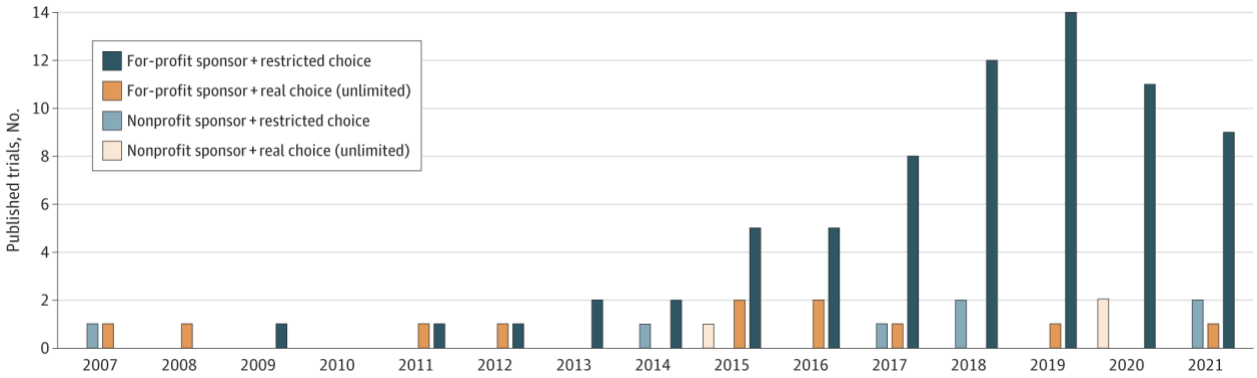
Identifying these issues within a landmark trial – the ASCENT trial – in breast cancer, served as the basis for future exploration of these issues in the works presented below.

3.2 – “Physician’s Choice”: Is It A Free Choice, Or An Illusory Choice?

Reporting of Physicians’ or Investigators’ Choice of Treatment in Oncology Randomized Clinical Trials. Timothée Olivier, Alyson Haslam, Vinay Prasad. JAMA Network Open. 2022;5(1):e2144770.[32]

In this article, we addressed more broadly the question we previously raised within the ASCENT trial, which is control arms defined as “physician’s choice”. Here, the question is the following: is the choice is free and unfettered, allowing for options including the best available care. Or in contrast, is the choice restricted, and preventing the use of important therapeutic options.

To investigate this, we carried out a cross-sectional study to analyze the use of "physician's choice" or "investigator's choice" in oncology randomized clinical trials (RCTs) and assessed whether the choice was a free choice (unrestricted) or instead restricted. We identified and included 92 oncology RCTs, they were published between 2007 and 2021, showing a clear increase in the use of such control arms, since the first report in 2007 (See the figure from our work).



Our study revealed that 89% of these trials were industry-sponsored, and 85% offered a restricted choice instead of a free and full choice. The use of the term "physician's choice" in these trials creates a false sense of free choice, which may lead to the perception of optimal treatment options. However, our findings indicate that the choice is often restricted. As a result, it is crucial for editors and regulators to demand clarification in the use of these terms within RCT protocols and reports to ensure better understanding and applicability of the trial outcomes. This will help prevent potential substandard control groups and enable physicians to make more accurate decisions when generalizing the reported results to their patients.

3.3 – The Duration Of Treatment May Vary Across Trials

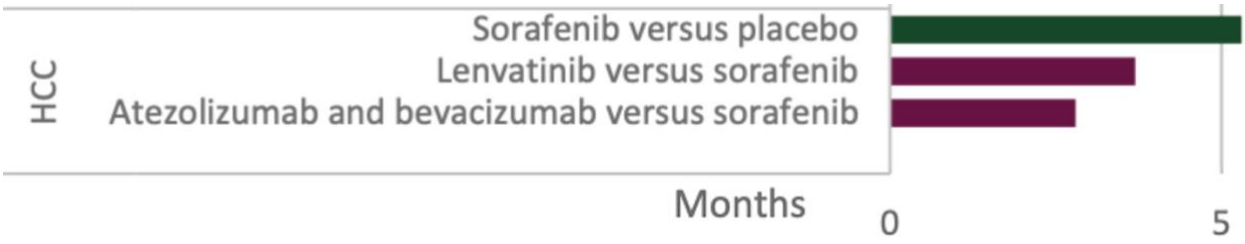
Duration of treatment in oncology clinical trials: does the duration change when the same drug moves from the experimental arm to the control arm? Alyson Haslam, Timothée Olivier, Rajat Thawani, Vinay Prasad. ESMO Open. 2022 Jun;7(3):100480.[33]

Between 2009 and 2020, we carried out a cross-sectional investigation into drug approval announcements, specifically focusing on those targeted at advanced, metastatic, or unresectable cancers. Our research scope encompassed studies that provided reports on a drug that received approval, as well as those where the same drug was employed as a comparative measure for other medications seeking marketing authorization from the FDA. We analyzed the treatment duration, in the context of both the drug's initial approval and its subsequent use as a control for other drugs seeking approval.

This research is an example of a "Bayesian approach" to provide insights by comparing the duration of a treatment when initially investigated as an experimental drug to the duration of the same treatment when later used as a control. Over time, one might expect that treatment durations would increase due to improvements in toxicity management. However, our findings demonstrated that the opposite could occur: drugs had a shorter duration when used later as a comparator in 48% of instances.

Also, we found that the median duration of treatment was 6.0 months (range: 2.2-12.7 months) in the trials first investigating the drug. When later used as a comparator, the median treatment duration was 4.9 months (range: 1.7-12.0 months), 1.1 months shorter.

An illustrative example is the first-line setting of hepatocellular carcinoma. In this setting, sorafenib, when tested as a novel therapy, was administered with a 5.3-month duration of treatment (SHARP trial).[34] When used later as a comparator in the non-inferiority study of regorafenib versus sorafenib (REFLECT trial)[35], the duration of sorafenib dropped to 3.7 months. More recently, in the IMbrave150 trial, testing atezolizumab and bevacizumab against sorafenib, the duration was even shorter (2.8 months)[36]. The main conclusion of this work is that trialists may not be incentivized to push the control arm drug as much as they are in the experimental arm. While this may be explained by several reasons, including justifiable ones, this provides an additional insight in appraising the “quality” of the control arm.



Abstracted from Figure 2: HCC stands for hepatocellular carcinoma.

The results of this study highlight the importance of considering treatment duration of control arm therapy across many trials in similar settings. In instances where the treatment duration is shorter than initially tested, the reasons for such phenomenon has to be questioned. Investigators may be incentivized not to push the control arm drug as much as in the experimental arm, particularly into open-label design. Of course, this may be explained by several reasons, some of which being absolutely justifiable. However, this piece of data provides a better understanding of the "quality" of the control arm when appraising a trial.

3.4 – Are Dose Modification And Supportive Care (G-CSF) Rules Fair?

Dose modification rules and availability of growth factor support: A cross-sectional study of head-to-head cancer trials used for US FDA approval from 2009 to 2021. Timothée Olivier, Alyson Haslam, Vinay Prasad. European Journal of Cancer. 2022 Sep;172:349-356.[37]

In a fourth research endeavor, we sought to determine the prevalence of imbalanced rules related to dose modification and the use of myeloid growth factors in comparative FDA registration trials. Both issues were identified and described in the ASCENT trial, and within other works,[38,39] but no comprehensive work has been conducted to assess for the frequency of such imbalances across trials in oncology.

We performed a cross-sectional analysis of all head-to-head registration randomized controlled trials that led to FDA approval between 2009 and 2021. These trials investigated anti-cancer drugs in advanced or metastatic settings where comparisons could be made between treatment arms in terms of dose modification rules or myeloid growth factor recommendations.

Of the 62 registration trials meeting our criteria, we discovered that 40 (65%) had imbalanced rules concerning dose adjustments, granulocyte colony-stimulating factor (G-CSF) utilization, or both. In 10% of the trials (6 out of 62), rules favored the control arm, while 55% (34/62) favored the experimental arm. ((below are illustrations from our work)).

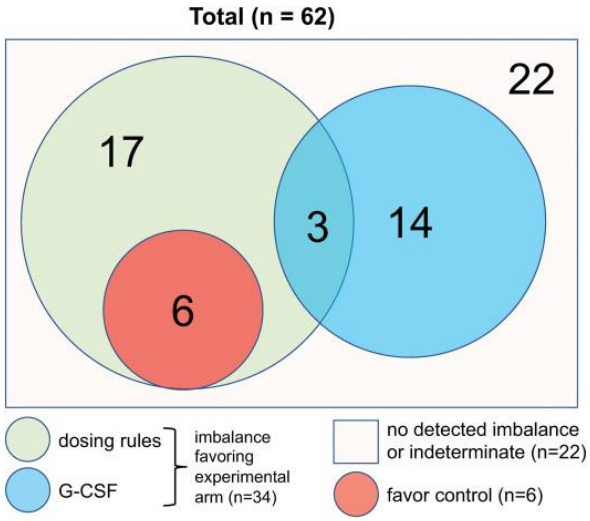
Table 2
All selected FDA head-to-head registration trials (n = 62), with number and proportion of (1) trials favouring the control arm, (2) with No detected imbalance and (3) favouring the experimental arm.

	Favour control	No detected imbalance[a]	Favour experimental	Total
Number of trials	6	22	34	62
Proportion	9.7%	35.5%	54.8%	100%

[a] Including eight trials coded as 'indeterminate'.

10% control

55% experimental



Our investigation found that 55% of the head-to-head trials featured rules for dose adjustments or G-CSF support that favored the experimental arm. Our conclusion was that, in those cases, it remains unclear whether the new drug is genuinely superior to the old one, or if better outcomes were obtained through unfair rules allowing for higher dose intensity.

Regulatory agencies must ensure that unnecessary imbalances in dose modification rules or growth factor support are avoided to prevent penalizing the control arm.

In conclusion, the findings of the "Dose modification rules and availability of growth factor support: A cross-sectional study of head-to-head cancer trials used for US FDA approval from 2009 to 2021" study show that the dose modification rules and the availability of growth factor support can have a significant impact on the outcomes of head-to-head cancer trials. The authors emphasize the importance of considering the impact of these factors when designing and conducting cancer trials, as they can have a significant impact on the trial results and the overall success of the trial. The authors also recommend further research to better understand the impact of these factors on the outcomes of cancer trials.

4 – CONCLUSION

4.1 – A Framework For Clinicians And Regulatory Bodies

The qualitative assessment of the control arm is paramount because a control arm that fails to meet standard and ethical norms, i.e. what we call the standard of care, introduces a bias that is impossible to correct. An analogy could be a 100-meter race: if one runner starts 20 meters ahead, the result is obviously unreliable. The objective here is not to cast a pejorative judgment on the occurrence of these biases, as we have demonstrated in our various studies, but quite the opposite: the identification of these elements and their systematic evaluation could enable a more fair and ethical overall research by avoiding, as much as possible, the introduction of such biases. With this goal, we will provide, at the end of this conclusion, a proposal for clinicians and regulators.

4.2 – Scores from Professional Organizations and Evidence-Based Medicine Ratings

4.2.1 – Scores Emanating From Medical Societies

To improve the accuracy and transparency of drug evaluations and provide guidance to healthcare providers, oncology professional societies have developed tools to assess the clinical benefits of new therapies. The two main scores are the European Society of Medical Oncology-Magnitude of Clinical Benefit Scale (ESMO-MCBS) and the American Society of Clinical Oncology Value Framework Net Health Benefit Score (ASCO Value Framework). Both scores greatly rely on the reported magnitude of benefit in trials (i.e., based on hazard ratio threshold) and the type of endpoints (a survival benefit being the most valuable). While these efforts are commendable, the ESMO and ASCO scores demonstrated discrepant results in their evaluation of the same drugs.[40] When applied to contemporary trials, less than one-third of RCTs with significant results met the ESMO thresholds for meaningful clinical benefit.[41] The ESMO score does not take into account some critical limitations in appraising the benefit of drugs, with shortcomings identified by the ESMO-MCBS working group.[42] Quality of the control arm is a key feature which is not evaluated in these scores. Also, they were developed by organizations which may not be free from financial conflict of interests with the industry, potentially undermining full independence in appraising new products.[43]

4.2.2 – GRADE and the Cochrane risk of bias tool (RoB)

GRADE was developed as a seamless extension of the EMB movement's philosophy, and it represents the acronym for "Grading of Recommendations, Assessment, Development, and Evaluation". [44,45] The GRADE methodology aims to assess the quality of evidence in order to make recommendations and is nowadays officially utilized by more than 100 organizations, including public bodies worldwide. Within the GRADE methodology, a first step is to define the clinical question under study, and then a detailed methodology is deployed to identify, rate, and conduct meta-analyses on selected studies to derive conclusions. In other words, GRADE's main goal is to provide an assessment regarding a clinical question. The resulting level of confidence has 4 categories: "very low", "low", "moderate" and "high".

Therefore, the GRADE scope is not primarily intended to assess individual trials, but rather to assess a body of evidence regarding a clinical question. Nonetheless, evaluating selected trial is an important step in GRADE, as well as in systematic reviews and meta-analyses done by the Cochrane Collaboration, another major international organization that promotes the use of EBM. The Cochrane Collaboration has developed a "risk of bias" (RoB) tool to systematically assess for the risk of specific biases.[46] The updated "RoB2" is the most commonly used nowadays.[47]

While being very useful in appraising certain aspects of randomized clinical trials, such tools may miss other important biases, of which not all researchers are aware. For instance, assessing the quality of the control arm is not built-in in those tools, yet being of major importance as we highlighted. Another limitation in applying GRADE for novel drugs, for instance in the oncology field, is that the vast majority are marketed based on a single trial, while it used to be based on 2 or more trials constituting a body of evidence.[48]

While GRADE provides valuable tools for assessing the evidence in medical literature, it may not be entirely appropriate in appraising distinct trials in the fields of oncology and hematology due to their unique complexities.

4.3 – Drug Regulation and Health Technology Assessment (HTA) Are Key

In a market where for-profit companies dominate, the incentives for drug development research are largely dictated by regulatory requirements. This is illustrated by simple examples, such as

the fact that if a composite or surrogate endpoint is not approved for drug approval, sponsors will not conduct trials based on such endpoints. Similarly, if a substandard control arm could prevent approval of the experimental drug, such control arms would never be included in the trial design.

Regulatory authorities often evaluate the value, potential benefit and danger of new medicines through health technology assessments (HTA) bodies. Based on HTA, regulatory agencies may approve or reject marketing authorization for novel therapies, and thus play a significant role in shaping medical practices. The network, organization, and relationships between HTA bodies are complex, as demonstrated by the new regulation of Health Technology Assessment, which came into effect in 2022 in the European Union. This regulation includes the evaluation of clinical data by multiple countries in a collaborative effort (Joint Clinical Assessment), along with several other proposals.[49,50]

Another illustration is Project Orbis, which was launched by the US FDA Oncology Center of Excellence. It is a global initiative that aims to expedite patient access to cutting-edge cancer therapies in multiple countries. Switzerland, as well as countries such as Singapore, the United Kingdom, Canada, Israel, Australia, and Brazil, are participating in the project.[51]

However, HTAs have limitations, and proposals for regulating these agencies have been discussed.[52] Among these limitations, they do not systematically assess for limitations within trials that have the potential to bias their results.[53] The main risk is that novel therapies may reach the market when their purported benefits may have limited or no benefits in real-life settings, or may even be harmful. The bar for approving cancer drugs has been repeatedly criticized because it has allowed a growing number of low-value and costly drugs to enter the market.[54] The accelerated approval pathway, initially designed to allow innovative and promising drugs to be prescribed while awaiting for more robust data, has also derailed from its initial goals.[55] The US Food and Drug Administration (FDA) regulatory agency has immense influence in drug regulation globally, and similar issues have been described within the European Medical Agency (EMA).[56] Beyond efficacy assessments, similar limitations were described for quality of life (QoL) evaluation, both within the US FDA and EMA assessments.[41]

4.4 – The THEOREMM Project

The THEOREMM Project aim to integrate the assessment of control arm's quality among other key features in appraising registration trials. (https://www.theoremm.com/)

The goal of the project is to develop and implement a structured, transparent, academic, and independent framework for evaluating clinical trials in hematology and oncology. This framework, using meta-research methods, will use a novel score to assess trials and provide detailed reviews explaining the score evaluation. This novel score could allow to better differentiate, as compared to existing methods, between unequivocal benefit from situations where patient's quality of life may be altered while providing limited benefit.

Within the THEOREMM score, assessing the control arm's quality will be a key component.

The project kick-starts by forming a team of expert professionals from across the globe. Following this, the focus shifts to developing a new score utilizing a pre-determined, pre-registered, and structured methodology, a strategy that creates a foundation for the later validation of the score. Concurrently, the project begins to share research findings through a process that includes peer-reviewed publications, a freely accessible online platform, and crucially, the active involvement of patients and the public, a factor that is expected to be present throughout the project. This process guides us to the crucial stage of scoring registration trials and offering in-depth reviews, an essential phase within the project timeline. The project finally concludes with the vital step of integrating the findings with drug regulation agencies.

The proposal may be of public interest and has the potential to benefit patients, which are expected to be involved at every stage of the project. The improved evaluation of trials will help identify treatments with the greatest potential to benefit patients, contributing to better-informed clinical decision-making. Through connection with drug regulation agencies, the score implementation could modify current incentives in drug development toward higher standards, thus ultimately benefiting patients and the society at large. Positive and constructive interactions with the industry are expected while maintaining financial and scientific independence. Beyond the hematology and oncology fields, methods developed during the project could be later utilized in other disciplines: the potential for academic collaboration is large.

4.5 – Overall Conclusion

We have demonstrated the steps and key concepts that led to randomized controlled trials becoming the gold standard method in generating medical evidence during the 20th century.

By definition, the control arm is central to this method. Alongside the increasing numbers of clinical trials, the ethical rules and philosophical concepts underpinning their conduct, like equipoise, have been defined and refined.

Despite this, and clear ethical guidelines, numerous studies have shown that the presence of suboptimal control arms is a common issue in oncology, estimated to be between 11% and 17% of clinical trials.

In our research works, we have identified complex factors which can lead to a suboptimal control arm, like a limited yet seemingly “free choice”, a duration of treatment that varies across trials, or unfair rules of dose modifications or use of growth factors. The description of these biases will be integrated into a systematic approach to the qualitative analysis of the control arm in oncology, which is currently not taken into account in professional scores (ESMO, ASCO), nor in classic tools for evaluating EBM. The integration of this systematic approach within an academic, independent score, as part of the THEOREMM project, could allow the use of this framework by regulatory bodies, practitioners, and the public, to refine the evaluation of drugs used in registration trials.

5 – REFERENCES

1. Gill J, Prasad V. Improving observational studies in the era of big data. The Lancet. 2018;392(10149):716-717. doi:10.1016/S0140-6736(18)31619-2
2. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. JAMA. 2000;284(10):1290-1296. doi:10.1001/jama.284.10.1290
3. minervation. al-Razi (10th century CE; 4th century AH). The James Lind Library. Published May 26, 2010. Accessed June 4, 2023. https://www.jameslindlibrary.org/al-razi-10th-century-ce-4th-century-ah/
4. Lind J. A treatise of the scurvy : in three parts, containing an inquiry into the nature, causes, and cure, of that disease, together with a critical and chronological view of what has been published on the subject. Bull World Health Organ Int J Public Health 2004 8210 793-796. Published online 2004. Accessed June 3, 2023. https://apps.who.int/iris/handle/10665/72991
5. Headland RK. Vitamin C in Exploration - The history of scurvy and vitamin C. K. J. Carpenter 1986. Cambridge, Cambridge University Press. 288p, illustrated, hard cover. ISBN 0 521 32029 1. £27.50. Polar Rec. 1987;23(146):606-607. doi:10.1017/S0032247400008184
6. Bown SR. Scurvy: How a Surgeon, a Mariner, and a Gentlemen Solved the Greatest Medical Mystery of the Age of Sail. First Edition. St. Martin's Griffin; 2005.
7. Sutton G. Putrid gums and "dead men's cloaths": James Lind aboard the Salisbury. J R Soc Med. 2003;96(12):605-608. doi:10.1177/014107680309601213
8. Baron JH. Sailors' scurvy before and after James Lind--a reassessment. Nutr Rev. 2009;67(6):315-332. doi:10.1111/j.1753-4887.2009.00205.x
9. Fisher RA. The Design of Experiments. Oliver & Boyd; 1935:xi, 251.
10. Group BMJP. Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation. Br Med J. 1948;2(4582):769-782. doi:10.1136/bmj.2.4582.769
11. Crofton J. The MRC randomized trial of streptomycin and its legacy: a view from the clinical front line. J R Soc Med. 2006;99(10):531-534.
12. Prasad VK, Cifu AS. Ending Medical Reversal: Improving Outcomes, Saving Lives. 1st edition. Johns Hopkins University Press; 2015.
13. Prasad V, Vandross A, Toomey C, et al. A decade of reversal: an analysis of 146 contradicted medical practices. Mayo Clin Proc. 2013;88(8):790-798. doi:10.1016/j.mayocp.2013.05.012

14. Suissa S. Immortal Time Bias in Pharmacoepidemiology. Am J Epidemiol. 2008;167(4):492-499. doi:10.1093/aje/kwm324
15. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. BMJ. 2010;340:b5087. doi:10.1136/bmj.b5087
16. NHANES Questionnaires, Datasets, and Related Documentation. Accessed June 5, 2023. https://wwwn.cdc.gov/nchs/nhanes/
17. Patel CJ, Burford B, Ioannidis JPA. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J Clin Epidemiol. 2015;68(9):1046-1058. doi:10.1016/j.jclinepi.2015.05.029
18. Silberzahn R, Uhlmann EL, Martin DP, et al. Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. Adv Methods Pract Psychol Sci. 2018;1(3):337-356. doi:10.1177/2515245917747646
19. Olivier T, Migliorini D. Autologous tumor lysate-loaded dendritic cell vaccination in glioblastoma: What happened to the evidence? Rev Neurol (Paris). 2023;179(5):502-505. doi:10.1016/j.neurol.2023.03.014
20. Olivier T, Haslam A, Prasad V. Sotorasib in KRASG12C mutated lung cancer: Can we rule out cracking KRAS led to worse overall survival? Transl Oncol. 2023;28:101591. doi:10.1016/j.tranon.2022.101591
21. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. 1987;317(3):141-145. doi:10.1056/NEJM198707163170304
22. de Melo-Martín I, Sondhi D, Crystal RG. When Ethics Constrains Clinical Research: Trial Design of Control Arms in “Greater Than Minimal Risk” Pediatric Trials. Hum Gene Ther. 2011;22(9):1121-1127. doi:10.1089/hum.2010.230
23. Prasad V, Grady C. The Misguided Ethics of Crossover Trials. Contemp Clin Trials. 2014;37(2):167-169. doi:10.1016/j.cct.2013.12.003
24. WMA - The World Medical Association-WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. Accessed June 5, 2023. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/
25. Wolinsky H. The battle of Helsinki: Two troublesome paragraphs in the Declaration of Helsinki are causing a furore over medical research ethics. EMBO Rep. 2006;7(7):670-672. doi:10.1038/sj.embor.7400743

26. Lakens D. Is my study useless? Why researchers need methodological review boards. Nature. 2023;613(7942):9-9. doi:10.1038/d41586-022-04504-8
27. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. Cochrane Database Syst Rev. 2017;2017(2):MR000033. doi:10.1002/14651858.MR000033.pub3
28. Del Paggio JC, Berry JS, Hopman WM, et al. Evolution of the Randomized Clinical Trial in the Era of Precision Oncology. JAMA Oncol. 2021;7(5):728-734. doi:10.1001/jamaoncol.2021.0379
29. Rossi A, Aimar G, Audisio M, et al. Analysis of the adequacy of control arms in oncology randomised clinical trials published between 2017 and 2021: a meta-research study. Eur J Cancer. 2023;0(0). doi:10.1016/j.ejca.2023.05.008
30. Olivier T, Haslam A, Prasad V. Anticancer Drugs Approved by the US Food and Drug Administration From 2009 to 2020 According to Their Mechanism of Action. JAMA Netw Open. 2021;4(12):e2138793. doi:10.1001/jamanetworkopen.2021.38793
31. Olivier T, Prasad V. Sacituzumab govitecan in metastatic triple negative breast cancer (TNBC): Four design features in the ASCENT trial potentially favored the experimental arm. Transl Oncol. 2022;15(1):101248. doi:10.1016/j.tranon.2021.101248
32. Olivier T, Haslam A, Prasad V. Reporting of Physicians' or Investigators' Choice of Treatment in Oncology Randomized Clinical Trials. JAMA Netw Open. 2022;5(1):e2144770. doi:10.1001/jamanetworkopen.2021.44770
33. Haslam A, Olivier T, Thawani R, Prasad V. Duration of treatment in oncology clinical trials: does the duration change when the same drug moves from the experimental arm to the control arm? ESMO Open. 2022;7(3). doi:10.1016/j.esmoop.2022.100480
34. Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in Advanced Hepatocellular Carcinoma. N Engl J Med. 2008;359(4):378-390. doi:10.1056/NEJMoa0708857
35. Kudo M, Finn RS, Qin S, et al. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. The Lancet. 2018;391(10126):1163-1173. doi:10.1016/S0140-6736(18)30207-1
36. Finn RS, Qin S, Ikeda M, et al. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. N Engl J Med. 2020;382(20):1894-1905. doi:10.1056/NEJMoa1915745
37. Olivier T, Haslam A, Prasad V. Dose modification rules and availability of growth factor support: A cross-sectional study of head-to-head cancer trials used for US FDA approval from 2009 to 2021. Eur J Cancer Oxf Engl 1990. 2022;172:349-356. doi:10.1016/j.ejca.2022.06.023

38. Prasad V, Massey PR, Fojo T. Oral anticancer drugs: how limited dosing options and dose reductions may affect outcomes in comparative trials and efficacy in patients. J Clin Oncol Off J Am Soc Clin Oncol. 2014;32(15):1620-1629. doi:10.1200/JCO.2013.53.0204
39. Caquelin L, Gewily M, Mottais W, et al. Tivozanib in renal cell carcinoma: a systematic review of the evidence and its dissemination in the scientific literature. BMC Cancer. 2022;22(1):381. doi:10.1186/s12885-022-09475-7
40. Cherny NI, de Vries EGE, Dafni U, et al. Comparative Assessment of Clinical Benefit Using the ESMO-Magnitude of Clinical Benefit Scale Version 1.1 and the ASCO Value Framework Net Health Benefit Score. J Clin Oncol. 2019;37(4):336-349. doi:10.1200/JCO.18.00729
41. Arciero V, Delos Santos S, Koshy L, et al. Assessment of Food and Drug Administration- and European Medicines Agency-Approved Systemic Oncology Therapies and Clinically Meaningful Improvements in Quality of Life: A Systematic Review. JAMA Netw Open. 2021;4(2):e2033004. doi:10.1001/jamanetworkopen.2020.33004
42. Gyawali B, Vries EGE de, Dafni U, et al. Biases in study design, implementation, and data analysis that distort the appraisal of clinical benefit and ESMO-Magnitude of Clinical Benefit Scale (ESMO-MCBS) scoring. ESMO Open. 2021;6(3). doi:10.1016/j.esmoop.2021.100117
43. Saleh RR, Majeed H, Tibau A, Booth CM, Amir E. Undisclosed financial conflicts of interest among authors of American Society of Clinical Oncology clinical practice guidelines. Cancer. 2019;125(22):4069-4075. doi:10.1002/cncr.32408
44. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-926. doi:10.1136/bmj.39489.470347.AD
45. What is GRADE? | BMJ Best Practice. Accessed January 28, 2023. https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/
46. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928. doi:10.1136/bmj.d5928
47. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898. doi:10.1136/bmj.l4898
48. Haslam A, Prasad V. Confirmatory Trials for Drugs Approved on a Single Trial. Circ Cardiovasc Qual Outcomes. 2019;12(6):e005494. doi:10.1161/CIRCOUTCOMES.119.005494
49. Radley-Gardner O, Beale H, Zimmermann R, eds. Fundamental Texts On European Private Law. Hart Publishing; 2016. doi:10.5040/9781782258674

50. Drummond M, Tarricone R, Torbica A. European union regulation of health technology assessment: what is required for it to succeed? Eur J Health Econ. 2022;23(6):913-915. doi:10.1007/s10198-022-01458-6
51. de Claro RA, Spillman D, Hotaki LT, et al. Project Orbis: Global Collaborative Review Program. Clin Cancer Res Off J Am Assoc Cancer Res. 2020;26(24):6412-6416. doi:10.1158/1078-0432.CCR-20-3292
52. Vella Bonanno P, Bucsics A, Simoens S, et al. Proposal for a regulation on health technology assessment in Europe - opinions of policy makers, payers and academics from the field of HTA. Expert Rev Pharmacoecon Outcomes Res. 2019;19(3):251-261. doi:10.1080/14737167.2019.1575730
53. Prasad V, Berger VW. Hard-Wired Bias: How Even Double-Blind, Randomized Controlled Trials Can Be Skewed From the Start. Mayo Clin Proc. 2015;90(9):1171-1175. doi:10.1016/j.mayocp.2015.05.006
54. Schnog JJB, Samson MJ, Gans ROB, Duits AJ. An urgent call to raise the bar in oncology. Br J Cancer. 2021;125(11):1477-1485. doi:10.1038/s41416-021-01495-7
55. Powell K, Lythgoe MP, Prasad V. The Oncologic Drugs Advisory Committee Votes of April 2021—Implications for the Fate of Accelerated Approval. JAMA Oncol. 2021;7(11):1607-1609. doi:10.1001/jamaoncol.2021.3046
56. Naci H, Davis C, Savović J, et al. Design characteristics, risk of bias, and reporting of randomised controlled trials supporting approvals of cancer drugs by European Medicines Agency, 2014-16: cross sectional analysis. BMJ. 2019;366:l5221. doi:10.1136/bmj.l5221