Ethics of Cartominyx

This case highlights ethical issues in the management of drug and vaccine testing. The issues are real and all costs, percentages, and probabilities approximate real-life situations. However, the firm, the managers, the drug, the science, and the outcomes are fictitious and should not be interpreted as endorsing or challenging any decisions about existing drugs, vaccines, or pricing. The case does not advocate for any particular viewpoint. You may agree with some statements in this case and disagree with others. You should be prepared to defend the ethics of any decisions you advocate.

BreakTruDrugs (BTD)

Logan stared out the window of his office with a view of East Cambridge and the McGrath Highway. Logan, who believed strongly in the social responsibility of corporations, founded BreakTruDrugs (BTD) fifteen years ago after graduating from MIT’s Sloan School of Management. While at MIT Sloan, Logan met researchers in MIT’s Broad Institute who had developed a technology to target a newly identified dysregulated gene in cancer cells. This gene regulated cell growth, survival, and metabolism via a previously unknown pathway. Studies indicate that, without this gene, the cancer cells are less able to reproduce and spread and are more-easily targeted. BTD’s breakthrough was the ability to target this gene directly so that, slowly, the patient would return to a more-normal life. Together, BTD’s founders obtained funding for BTD’s first applications, drugs targeting a common cancer that killed slowly, but affected 500,000 patients each year in the US alone. A new drug, Cartominyx, could be a big win for BTD after fifteen years of investment and hard work.

Cartominyx does not cure this cancer, but if administered intravenously once per month, it promises to reduce the cancer’s spread and make dramatic improvements in patients’ quality of life. Cartominyx has side effects, but Logan believes the side effects were a small price to pay for the benefits of using Cartominyx. If Cartominyx succeeded, BTD had more drugs in the pipeline, some just as promising than Cartominyx. If Cartominyx failed or if BTD could not justify a profitable price, Logan may not be able to obtain sufficient funding to stay in business and the technology may never realize its potential to make life better for all cancer patients.
Clinical Trials of Cartominyx

Soon after its founding in 2007, BTD chose its initial target. Over 90% of the patients with this cancer are over 65 and, in addition to a mortality risk, this cancer has a huge impact on the quality of patients’ lives. As a result, patients (and their families) afflicted with this cancer represent a large target market and form a well-funded and vocal advocacy group. By 2007, the scientific theory for BTD’s method was well-established, but 50 drugs designed to target the gene had failed. When BTD’s pre-clinical tests on cells and on animal models succeeded where others had failed, BTD moved to Phase I safety testing in real patients. (Appendix A provides a review of the phases of clinical trials.)

Roughly one-third of the patients in Phase I experienced mild brain-swelling and microhemorrhages, but these neurotoxic side-effects abated on their own after about a month. Although somewhat painful, most side effects did not cause long-term damage. A Phase II trial of 57 patients suggested that 18 doses over one and a half years was about right. After the success of Phase II was announced, BTD was valued at $30 billion. Logan and team were able to obtain the necessary funds for a full-scale Phase III randomly controlled trial (RCT).

In the RCT protocol, approved by the FDA, BTD would treat 500 cancer patients with 18 doses of Cartominyx over 18 months (test group). An additional 500 cancer patients would receive the currently accepted treatment (control group). 1,000 total patients provided sufficient statistical power to observe an effect if there were indeed an effect. (Appendix B reviews statistical testing for RTCs.)

Patients were to be assigned randomly to test and control groups, selected from a representative set of demographic categories, and in various stages of cancer progression. Neither the patients, their doctors, nor BTD’s researchers would know who was assigned to which group (double-blind). The approved protocol defined the primary endpoint (outcome) as stopping the growth of the cancer after 18 months, although a number of secondary quality-of-life endpoints and biomarkers (successfully inhibiting the gene) were to be observed.

Like many protocols, the Cartominyx trial would be examined by a Data and Safety Monitoring Board (DSMB) at six months and at twelve months. The DSMB had the ability to stop the test for futility or if the side effects were more frequent or more intense, continue the clinical trial, or seek immediate approval should the trial clearly indicate a significant benefit. BTD would seek FDA approval to market its drug if Cartominyx was significantly better than standard treatment options at the \( p = 0.05 \) level on the primary endpoint of stopping the cancer’s growth.
Phase III Crisis

At the six-month review, the DMSB voted to continue the Phase III trial. Congratulations all around. But at the twelve-month review, the reduction in cancer growth due to Cartominyx was not significantly better than standard treatment ($p = 0.09$) and side-effects were a concern. The side effects were not as frequent as in Phase II, but the brain swelling and microhemorrhages in the test group (21%) were significantly more than in the control group (9%). Three patients died in the test group and one patient died in the control group, but these deaths did not appear to be the result of Cartominyx. Despite progress on secondary endpoints and BTD’s protestations that patients were willing to risk the side-effects for the promise of reduced cancer growth, the DMSB voted to stop the test for futility. BTD could continue to provide treatment to those patients who opted in based on compassionate use. About two-thirds did so. BTD could continue to track all patients for 18 months.

The reaction was immediate. BTD lost half of its market value – down to $15 billion, and was forced to lay off the 1,000 employees who were helping to ramp up marketing and production. BTD was forced to abandon 150,000 square feet in Kendall Square. Logan would have to find a way to cut costs by about $1 billion per year to conserve funds so that BTD could move another promising drug to a Phase III trial. (The new drug, CartoMax, did slightly better than Cartominyx in Phase I and Phase II trials. But this is not a guarantee it will do better in Phase III.)

On the plus side, the patient advocacy group became a vocal supporter of Cartominyx. A survey of cancer patients indicated that two-thirds were eager to try Cartominyx, knowing fully the risk of side effects. The advocacy group argued further that without a viable revenue stream from Cartominyx, future drugs like CartoMax might not even make it to Phase III trials. The Wall Street Journal cited secondary endpoints as important. Although Cartominyx did not stop cancer growth it slowed cancer growth. Cartominyx enabled many patients to enjoy a more normal life. Testimonials were published of patients talking about spending more time with grandchildren or continuing to work in productive careers. The Wall Street Journal argued that the delay in the worst cancer symptoms meant a substantial savings in the cost of hospice care.

A Chance Meeting

Seven months after the Phase III trial, BTD’s Chief Statistician ran into an FDA director while picking up pizza at Za in Kendall Square. Such chance meetings are common given the large number of innovative drug companies in Kendall Square. The director revealed a scientific interest in the theory underlying Cartominyx and a belief that this theory should lead to many successful cancer-fighting drugs. The director encouraged BTD’s Chief Statistician to take another look at the Phase III data now
that eighteen months of observations were available. With Logan’s blessing, the Chief Statistician did just that.

The results were promising. The Chief Statistician discovered errors in implementation in the Phase III trial.

- 27% of the patients in the test group got fewer than 18 doses. Those that received all 18 doses showed much more positive results on all measures.
- the trial had been stopped too early. Cartominyx seemed to be much more effective at the end of 18 months that had been expected after 12 months.
- due to a quirk of randomization, more than twice as many “rapid progressors” were in the test group as the control group

Reanalyzing the data, focusing only on 30% of patients who were not rapid-cancer-progressors and those that received all 18 doses under compassionate use, the Phase III data suggested that

- the key gene was successfully targeted in 73% of the test patients, but none in the control
- after 18 months, cancer growth slowed in 38% of the test patients versus 10% in the control patients.
- after 18 months, the test patients experienced an 88% smaller decline in the ability to function normally, an 81% decline in burden on caregivers, and a 30% smaller reduction in the ability to stay at their job.
- after reanalysis, the differences in the primary endpoint (cancer growth stoppage) were close to significant, \( p = 0.051 \) level. If one more patient in the test had experienced cancer growth stoppage, the \( p \)-value would have been just below \( p = 0.049 \).

All deaths during the clinical trial were investigated carefully and attributed to unrelated underlying conditions. There was no evidence that they were caused by Cartominyx. Nonetheless, the investigation continues.

**Ethical Dilemmas**

As CEO, Logan’s responsibility was to balance all interests, including those of the shareholders. There was every indication that the combined enthusiasm of the FDA director and the patient advocacy group would lead to an accelerated approval that would allow BTD to market Cartominyx while undertaking a new 18-month Phase III trial. As devastating as was the Phase III stoppage, an FDA approval based on the new data would return BTD’s value to at least $30 billion, perhaps more.
Logan faced two challenging decisions that balanced business decisions and ethical considerations. First, should BTD seek accelerated approval, wait another 18+ months for a new Phase III clinical trial, or abandon Cartominyx. Funding for the substantial cost of a new clinical trial was likely, but not guaranteed. Second, if the FDA approved Cartominyx, either through accelerated approval or a second clinical trial, what price per patient per year should BTD charge. Logan had built a good team and asked for input before making make-or-break decisions. Team members did not agree and each provided a unique perspective on the ethical dilemmas. Logan’s had to balance the attractive business opportunity with the ethical considerations.

Should Logan Seek Accelerated Approval?

Chief Financial Officer. We owe it to our shareholders. FDA approvals matter: when the FDA approved Biogen’s Alzheimer’s drug, Lecanemab, Biogen’s stock rose 37%. We are likely to see at least a 37% bump if we get an accelerated approval. An FDA approval validates our theories and technology. The market knows we have many more cancer drugs in the pipeline.

Chief Medical Officer. Standards are standards, don’t move the goalposts after the game has begun. The proper response to urgency is not to compromise the integrity of the process. The $p = 0.05$ standard was chosen to balance innovation and caution. If the FDA approves Cartominyx, then the next drug application will seek a lower standard until we have no standards at all. We could usher in a new era of costly treatments with no proven therapeutic benefit.

We need more time to investigate the deaths that occurred during the clinical trial to be absolutely sure they are not due to Cartominyx.

I am sure you heard of $p$-hacking (see Appendix C). There are many ways to achieve the illusion of a significant $p$-value by “fishing in the data.” How do we know whether or not the reanalysis was based on only one split of the data (slow- vs. rapid-cancer-progressors)? What if the successful split was cherry-picked from twenty potential splits?

Our protocol sought a defined primary endpoint; we should stick to it and not embrace secondary endpoints after the trial began. Were improvements in secondary endpoints the only ones that showed promise among a much larger set of potential secondary endpoints?

While our theory is strong and there are many peer-reviewed papers that support our medical theories, *Nature* just retracted a paper that supported our theory because its data could not be verified. There are new scientific papers in process that promote another theory about targeting cancer.
**Chief Human Relations Officer.** Think of the 1,000 people we had to let go and think of the morale among our top scientists. With FDA approval we can rebuild our marketing and production staff and retain our top scientists.

**Chief Legal Officer.** Targeting the gene is great news and an important development, but it is a surrogate measure. Our protocol called for stopping cancer growth as a primary endpoint. We need to worry about patient lawsuits if Cartominyx does not halt the growth of the patient’s cancer. We might get sued for marketing the drug, but we won’t get sued for not marketing the drug. Johnson & Johnson, AmerisourceBergen, Cardinal Health, and McKesson paid a combined $26 billion to settle opioid claims.²

**Head of Research and Development.** We’ve invested in our unique technology. Not only will we lose our investment if BTD folds, but much of our specific (not published) knowledge will be lost. We have more drugs coming down the pike that may be more effective and may have fewer side effects, but we have to get there from here. The first drug for multiple sclerosis was approved in 1993 and the US medical profession now has over two dozen options. Roche and Eli Lilly each have cancer drugs in the pipeline that may be sidetracked if Cartominyx is not profitable. And if that’s not enough, Cartominyx might be valuable for other diseases. For example, Onpattro was approved in 2018 to treat peripheral nerve disease – a modest market, but they are now seeking approval to treat cardiac issues – a huge market of about 30 million adults in the US.³ After all, the first treatments for AIDS were not very effective and had substantial side effect, but experience led to major improvements.

**Head of Public Relations.** It will come out that our Chief Statistician spoke to the FDA director through an informal back channel. It’s not a good look.

**Chief Statistician.** I meet with the FDA all of the time. We all want the best for our patients and we all want the data to speak the truth. No RCT is perfect. It’s reasonable to use valid statistics to correct for the quirks of randomization. Furthermore, $p = 0.05$ is an arbitrary standard. All things considered, it is much more likely that Cartominyx is effective than it is not effective. For example, we can interpret $p = 0.10$ as meaning that, if there were no effect, there is a 90% chance that the primary endpoint would be less than the standard treatment. There is always the danger of false rejection (Appendix D).

**Head of Investor Relations.** If Cartominyx is not successful, how can we get investors to fund the Phase III trial for CartoMax and for the other drugs in our pipeline.

**Head of Public Relations.** If Cartominyx gets accelerated approval we have to invest in a new Phase III clinical trial. If Cartominyx fails that trial, we will lose goodwill with investors and the public, not to mention morale among our scientists.
Chief Marketing Officer. BTD was founded to cure diseases that matter to millions of people. We can save lives! If these “serious” drugs fail, then pharma will stop swinging for the fences and concentrate on mass-market drugs such as Viagra, Cialis, and cosmetic Botox. Why should we second guess the FDA? It it’s approved, leave it up to the doctors and their patients.

Patient Advocate. 73% of the increase in life expectancy in high-income countries from 2006 to 2016 is due to modern drugs. Accelerated approvals were effective for HIV, AIDS, sickle cell anemia, muscular dystrophy, and other cancers – 253 since 1992. Every day that Cartominyx is not available means thousands of patients will progress beyond the point where any drug can help them. I talk to these patients and their stories are wrenching. They beg for relief, for longer time with their loved ones, for a better quality of life, and for the hope of a cure. We cannot disappoint them even if it means taking a risk.

At What Price Should BTD Market Cartominyx?

Chief Financial Officer. Let’s look at the numbers. We need to charge a minimum $94,000 per year per patient. This cost comes from my rough calculations: our direct R&D costs and the cost of the clinical trial were about $3.9 billion. After we ramp up production, our marginal costs are about $10,000 per year per patient. Before Cartominyx was successful, we tried nine other drugs—a success rate of 10%. There are 500,000 potential patients and the patient will take Cartominyx for one and a half years. If we spend $500 million marketing Cartominyx, we’ll to achieve a penetration rate of 70%. And that doesn’t cover corporate overhead and profit.

Head of Public Relations. Whoa. Let’s be realistic. We only have an accelerated approval. And the accelerated approval is based on reanalysis and surrogate measures. Will doctors prescribe Cartominyx with only an accelerated approval. The Institute for Clinical and Economic Review (ICER) just published a study saying that the economic benefit to patients of Cartominyx is somewhere in the range of $10,000 to $20,000 per year. We’ll get pushback from the press, especially if we charge a higher price per year. Think of all of the patients who cannot afford the treatment. We’re talking lives here. Also, think of the goodwill that we will generate with a low price.

Head of Global Product Strategy and Communication. We are just building our salesforce and production capability. Both marketing and production costs will decrease overtime. What we learn from Cartominyx will translate to related drugs we have in the pipeline and will provide a common good for our competitors who are developing similar drugs.

Head of Global Safety & Regulatory Sciences. Even with a small profit, we’ll be charging $100,000 or more per year. Most patients are on Medicare. If we add 500,000 patients per year and
they are on treatment for one and a half years, then with a 70% penetration the total cost is $52.5 billion. That’s 170% of the yearly budget of NASA! Is it fair that we ask Medicare to pay back our development costs while Canada and other countries get the drug at a price that is closer to marginal costs?

**Chief Medical Officer.** Cartominyx has proven side effects – 21% of the patients had brain swelling. We should consider these side effects when pricing Cartominyx. The ICER makes a good point about the economic benefit to the quality of life. Without a randomized trial that follows a preset protocol, we cannot be sure of the benefits and this should affect the pricing.

**Patient Advocate.** A high price is unfair to patients, especially those who do not have insurance. If Medicare does not cover Cartominyx, only the rich will be able to afford treatment. Today, medical decisions are often decided by a cost-to-benefit ratio. If a new medication is too expensive, hospitals may not invest in the time, staff training, and equipment necessary for intravenous monthly treatment. BTD does not pay for these costs, but profits from Cartominyx. Do you want to force grandparents to choose between Cartominyx therapy and their grandchildren’s college education?

**Chief Marketing Officer.** Patients are desperate. They’ll find a way to pay. At minimum, we should pay back our R&D costs, the costs of our clinical trial, and marginal costs. If we advertise Cartominyx right, it will pay back these costs and provide shareholders with a fair profit. A cancer drug, Keytruda, is priced at $150,000 per year; an anti-inflammatory drug, Humira, at $70,000 per year, and an ALS drug, Radicava at $171,000 per year. We should be in that ballpark.

**Chief Legal Officer.** Think of the goodwill if we tout the savings to Medicare because patients avoid other treatments and/or hospice care.

**Chief Human Relations Officer.** Our scientists are idealists. They want to help patients, not profit from their misery.

**Appendix A: Clinical Trials**

Before a new drug or vaccine can be approved for use in the general population, the drug is tested in one or more clinical trials. Clinical trials are necessary to establish whether the drug or vaccine is safe (acceptable side effects) and effective (better than either a placebo or the current best treatment). Clinical trials are also used for new uses of a drug, new medical devices, or changes in the methods of treatment. Because it may take months or years to observe an endpoint, say tumor shrinkage in cancer, clinical trials can take years to complete.
**Phases of a Clinical Trial**

Typically, clinical trials are run in three phases, plus a pre-clinical (or laboratory) study.

*Pre-clinical.* Before a drug, device, or treatment is tested on human subjects, it is tested in the laboratory with cells, animal tests, or other surrogates for human testing. If the drug, device, or treatment shows promise, researchers apply to the US Food and Drug Administration (FDA) for approval to move to clinical trials.¹¹

*Phase 0.* Some, but not all, clinical trials use small doses on a few people to examine if the laboratory studies apply to human beings.

*Phase I.* Safety is the main concern in Phase I. A few volunteers get low doses, then increasing doses, and are monitored closely for side-effects. Researchers look for promising signs that the drug, treatment, or device will be effective. The Moderna vaccine was tested on 105 patients in Phase I.¹²

*Phase II.* More patients receive the drug, treatment, or device using the doses found to be safest and most effective in Phase I, although sometimes doses are varied. Placebos are not typical in Phase II. Researchers use the data to decide if the drug, treatment, or device is promising for a more-expensive Phase III trial. The Moderna vaccine was tested on 600 patients in Phase II, split into two age groups.

*Phase III,* also known as an A/B test. If the drug, device, or treatment shows sufficient promise, researchers proceed to a much larger randomized controlled trial (RCT). Patients are assigned randomly to test and control “arms.” The test group is given the drug, device, or treatment and the control group is given either a placebo (no treatment) or the best available treatment.¹³ Patients, caregivers, and researchers do not know who is in the test group and who is in the control group (double-blind). Researchers seek to equitably enroll patients from a variety of regions, demographic groups, and with proportional variation on other indicators that might affect outcomes. The Moderna vaccine was tested on 30,000 patients split into two arms.

**Protocols**

Before a clinical trial begins, researchers submit a protocol detailing the methods, procedures, and analyses that are planned. The protocol includes plans for recruiting and randomizing patients (including inclusion and exclusion), a summary of the results of Phases I and II, a benefit/risk assessment, objectives (endpoints) and how endpoints are measured, a description of the study design, dosing plans, safety assessment, planned methods of analysis, monitoring committees, methods of blinding, plans for any interim analyses, and ethical considerations, among other things. For example, the third amendment of the Moderna vaccine protocol was 132 pages long. Protocols may be amended, but amendments need to be reported and, sometimes, cleared with an Internal Review Board.
Appendix B: Significance Testing\textsuperscript{14}

Suppose we run a pre-registered clinical trial for a vaccine or a new drug. (The same theory applies to A/B testing in eCommerce.) Assuming outcomes are normally distributed (bell-shaped curve), then we expect that if the null hypothesis were true, the outcome probabilities would be given by the red shaped curve below. An example outcome measure, pre-defined in the protocol, might be observed tumor shrinkage in the test arm minus observed tumor shrinkage in the control arm.

We want to be reasonably confident that the observed data are based on outcomes that are different from the outcomes we would observe if patients received the control-arm treatment. The null hypothesis is no effect. The accepted scientific standard is that the results be significant at the $p$-value of 0.05 as shown by the red-shaded area under the curve. (Some scientific journals allow researchers to report “marginally significant” results if they are significant at the $p = 0.10$ level.) In pictures, if the observed endpoint difference is any point in red area, then the observed drug effectiveness is unlikely (5% or less) to have occurred by chance. If the clinical trial is completed according to appropriate protocols and side-effects are acceptable, the drug is likely to be approved.

![null hypothesis](null-hypothesis.png)

Null hypothesis of no effect

$p = 0.05$, significance level (probability that we falsely observe an effect if the null hypothesis is true)

We also want to avoid false rejection. If the proposed hypothesis is true, then we want a sample size large enough to give a reasonable chance that the clinical trial will find a (true) positive effect. For example, the protocol might call for tumor shrinkage in 60% of the patients. The potential outcomes for a drug that is 60% more effective than the control are shown by the grey curve. The probability that we observe a positive outcome if the drug is truly 60% effective more effective than the control is shown by the area under the grey curve, which is 75% of the total area. This area is known as the \textbf{power} of the test. Clinical trials are often designed to achieve a target power. In this example, the sample size would be selected so that there is a 75% chance of rejecting the null hypothesis.

Statistical significance focuses on avoiding false positives while power focuses on avoiding false negatives.
Appendix C: \( p \)-Hacking

Payton recently graduated from a top technology school with a PhD and was hired by NewCoPharm, a promising startup working on a new food supplement, Neenohurt. Neenohurt was designed to eliminate knee pain. Payton was asked to oversee a clinic trial in which 5,000 patients were given the new food supplement and 5,000 patients were given a placebo. The food supplement did not require FDA approval for planned claims, although any advertising was subject to laws on false advertising.

When the trial was complete, Payton undertook standard statistical analysis and reported a positive result, but at the 0.08 level. In the same analysis, Payton also examined side effects. Joint deterioration, headaches, stomach aches, blurred vision, lowered cognitive ability, ringing in the ears, weight gain, and swelling were all negligible and likely due to reasons other than Neenohurt. The only side effect of concern was bone loss.

After reading the report, the CEO of NewCoPharm was livid. NewCoPharm had staked its entire IPO on the success of Neenohurt. The observed \( p \)-value of 0.08 did not meet the accepted standard that the results be statistically significant at the \( p = 0.05 \) level or lower. The CEO asked Payton to take another look at the data. Upon looking at the details of the clinical trial, Payton discovered that Neenohurt had been mistakenly manufactured in 25 different colors. Payton’s training suggested that the sample-size (200 test, 200 placebo) for each color met accepted standards for statistical power. Payton reanalyzed the results separately for each color used in the trial. The \( p \)-values ranged from \( p = 0.04 \) for red to \( p = 0.12 \) for grey. Twenty-four of the 25 colors had \( p \)-values greater than 0.05.

The CEO was thrilled and immediately funded a marketing campaign – “Red Neenohurt cures runner’s knee!” The marketing campaign emphasized few side effects, listing no joint deterioration, headaches, stomach aches, blurred vision, lowered cognitive ability, ringing in the ears, weight gain, and swelling. There was no mention of bone loss. The CEO justified the statement because the results were
statistically significant at accepted levels – there was but a 4% chance the positive result was due to chance.

**Temptations**

A test with a *p*-value of 0.05 of less means that if the null hypothesis is true, then there is less than a 5% chance (1 in 20) that the results are due to random chance alone. If we test 25 colors, it is likely that the effectiveness of at least one color on Neenohurt is significant by chance alone. If only one of 25 color-tests is significant, then our confidence is much lower. The same issue applies to testing side effects, but we tend to be more conservative for side effects. Typically, the Neenohurt results would call for more testing to determine if bone loss due to Neenohurt was a major concern.

The fundamental temptation underlying *p*-hacking is that there are strong incentives to “look the other way” when *p*-values are manipulated. The CEO of NewCoPharm stands to make millions of dollars if the test is significant at the *p* = 0.05 level, but incentives can also be subtle. Academic researchers advance their careers with journal publications and journals tend to publish significant results, especially for interesting non-intuitive findings. This effect is exacerbated because the popular press prefers to highlight surprising findings.

Many studies in the social sciences have been retracted when they do not replicate or when *p*-hacking or fraud is suspected. For example, a paper was retracted from one of the top scientific journals, *Nature*, even though that paper supported the prevailing theory of the cause of a disease that affects millions of patients.16

**How *p*-values are manipulated?**

- **cutting the data too many ways** (Neenohurt example). If we cut the data by gender, age category, region, income category, and education category we could easily have more than 1,000 subcategories.

- **re-analyzing the data.** Empirical data are never perfect and researchers often “clean” the data to remove outliers or data points that are suspect. Researchers may have a subconscious bias toward deleting data that do not support their hypothesis.

- **file-drawer problem.** Researchers have incentives to publish surprising findings and firms have the incentives to accept good financial outcomes. The results of failed tests do not get reported.

- **many outcome measures.** In a cancer drug the gold standard is to test whether the cancer is cured, but people might die while we wait for the gold standard. Instead, a test might examine surrogate measures such as tumor shrinkage, slower growth of cancer cells, slower growth of a
precursor to the cancer, or reduction in blood flow to the cancerous cells. If a researcher examines enough surrogates, the clinical trial might find one is significant by chance.

- **many independent experiments.** If a researcher runs many experiments simultaneously, one experiment may be significant by chance.
- **manipulate sample size.** As data come in, the outcome measure fluctuates for both the test arm and the control arm. If the fluctuation is sufficient, then at some point the test will fluctuate up and the control will fluctuate down. If a researcher stops the test before its planned endpoint, the decision to stop may exploit random error. A researcher might also manipulate sample size by adding sample to a non-significant test.
- **multiple statistical methods.** There are a variety of statistics that can be computed, each with different assumptions. They do not always agree. A famous data set, the “Lydia Pinkham” data on advertising and sales, was said to be the basis of more articles than the number of data points in the time-series data.
- **demand artifacts.** If the data are collected face-to-face and the data collectors know the hypothesis, the data collectors might communicate the hypothesis by body language or by the questions they ask. The experiment “demands the result.” Demand artifacts can be buried anywhere, such as the wording of an online survey that communicates the researcher’s intent.
- **self-selection.** Suppose we offer treatment to a random group of patients and allow patients to opt-in for treatment. We then compare those who opt-in to those who do not and find a significant positive effect of the treatment. The test does not necessarily show causality, because those patients who opted into the treatment may be different from those who do not. For example, opt-in patients may comply better with the treatment (favors a positive effect) or are more in need of treatment (favors a null effect).

**Guarding against p-hacking**

- **pre-registration.** Many journals require that experimental procedures and analyses be preregistered. Drug and vaccine trials require lengthy protocols on procedure, analyses, and controls.
- **reporting requirements.** Researchers are asked to fully report all analyses and experiments that were completed.
- **statistical corrections.** There are a variety of statistical methods to correct for multiple tests on the same data. The Bonferroni test is popular, but conservative.
data “pre-looks.” Sometimes it is ethical to stop a clinical trial for futility or severe side-effects. Sometimes it is ethical to stop a trial and approve a drug/vaccine because it is clearly superior. Pre-looks often require both pre-registration and anti-p-hacking analysis that corrects for the probability that multiple looks at the data can lead to false positives.

- **double-blind design.** If neither the data collectors nor the subjects know the hypotheses, demand artifacts can be reduced.
- **pretesting.** In behavioral and survey-research studies, respondents complete the protocol and are interviewed to see if they can guess the purpose of the research. Pretests continue until all demand artifacts are identified and removed.
- **commitments to publish/reward researchers.** Some journals pre-commit to publish a paper whether the findings are significant or not. Firms might send strong signals to researchers that their compensation and promotion are not dependent on significant results.

**Appendix D: The Danger and Social Cost of False Negatives**

Diseases such as ALS, Alzheimer’s, and many forms of cancer can be devastating. Patients (and their families and friends) learn that the patient has little hope as the disease progresses. The symptoms are often painful. The patients and their families want hope.

Suppose a clinical trial is completed in which a new cancer drug is found to have not a non-significant impact on mortality at the $p = 0.10$ level—there is a 10% chance that the observed effectiveness (against a placebo) is observed by chance. Such a trial would lead to disapproval. But another interpretation is that there is a 90% chance that the observed effectiveness is not due to chance.

Consider the following scenarios, each of which might affect the ethical choice of whether to allow patients to opt into the drug or vaccine.

- There are no side effects and the cost of treatment is small.
- There are no side effects and the cost of treatment is so high that a lottery is required. Only a select few get the treatment paid for by insurance or the government.
- Side effects are substantial, but a well-informed patient would choose to take the risk given the severity of the disease and the lack of other cures.
- Opting into the new drug or vaccine requires opting out of a known treatment regiment.
- The clinical trial is one of many clinical trials, all of which are significant at the $p = 0.10$ level.
• The drug is the first of many drugs in development, but the industry will not fund further development if the at-issue drug is not approved.

• Medical theory suggests strongly that a surrogate measure is an indicator of whether the drug is effective and the surrogate measure is impacted significantly by the drug. For example, the drug might lower blood pressure and the drug is approved for lowering the risk of a heart attack. Or medical theories suggest that lowering cholesterol prevents heart problems—a drug might be approved if the clinical trial demonstrates that it lowers cholesterol.

**Discussion Questions**

1. *Should Logan seek accelerated approval for Cartominyx?*

2. *Should the FDA grant accelerated approval?*

3. *What price per patient per year should BTD charge for Cartominyx?*

4. *Should Medicare pay for Cartominyx treatment?*
2 Brian Mann, “4 U.S. companies will pay $26 billion to settle claims they fueled the opioid crisis,” *NPR*, February 25, 2022.
6 Hooper and Henderson, op. cit. According to the FDA inflation adjusted fixed costs are about $7B in 2022. The number in the case is chosen to encourage ethical discussion.
7 Calculated as (R&D cost + Clinical Trial Costs + Marketing Costs)/(Patients*CLV *MarketPenetration*SuccessRate) + Marginal Costs, where CLV = customer lifetime value = 1.5 years.
8 Two measures are helpful in determining the economic value to a patient of a drug, the “equal value of life years gained (eVLYG)” and the “quality-adjusted life year (QALY).” ICER uses QALY, which attempts to measure the amount by which a treatment improves the quality of a patient’s live while subtracting any decrease for side effects. ICER’s governing boards include both health experts and insurance executives. https://icer.org/. ICER evaluation can influence decisions on whether and how much insurance companies will pay for treatment.
9 Calculated as (Patients*MarketPenetration*YearlyPrice*CLV. NASA’s 2022 budget is $30.62 billion. According to the Congressional Budget Office, Medicare, Medicaid, and related expenses are $1.463 trillion in 2022.
11 Because the case is set in the US, this appendix covers US requirements. Each country or region has its own requirements, but they are usually similar.
12 The Moderna COVID vaccine was based on a technology known as mRNA. In October 2022, Moderna formed a partnership with Merck to develop an mRNA cancer vaccine. Its stock jumped 11% on the announcement. Hannah Miao, “Moderna stock jumps after cancer vaccine announcement,” *Wall Street Journal*, October 12, 2022.
13 If a known treatment has been shown to be effective in previous trials, it would be unethical to give patients a placebo rather than the treatment accepted as most effective.
14 This appendix describes the concepts behind a simple significance test. These concepts are sufficient for discussing the ethical considerations for Cartominyx approval and pricing. Protocols may allow for more advanced statistical tests and/or tests based on Bayesian statistics.
15 If the tests are independent and we test each at the $p = 0.05$ level, then there is a 72.3% chance of finding at least one test significant by chance along. Calculated as $0.723 = 1 - (1 - 0.05)^2$.