MITx 6.419x Data Analysis
Statistical Modeling and Computation in Applications

# Analysis 1
## Stats Review

Jonathan Chang (JonathanChang6d41)

April 15, 2021

# Contents

# Problem 1.1: The Salk Vaccine Field Trial

| **NFIP Study** | | |
|---|---|---|
| | **Size** | **Polio rate per 100,000** |
| Grade 2 (vaccine) | 225000 | 25 |
| Grade 1 and 3 (no vaccine) | 725000 | 54 |
| Grade 2 (no consent) | 125000 | 44 |

Table 1.1: Polio vaccine offered to Grade 2 students in NFIP controlled experiment.

| **Randomized Controlled Double-Blind Experiment** | | |
|---|---|---|
| | **Size** | **Polio rate per 100,000** |
| Treatment (vaccine) | 200000 | 28 |
| Control (Salt Injection) | 200000 | 71 |
| No consent | 350000 | 46 |

Table 1.2: Subsequent double-blind RCT removes observer bias.

## 1 Double-blind procedures

**How would you run a randomized controlled double-blind experiment to determine the effectiveness of the vaccine?**

In a double blind study, neither the vaccine recipients nor anyone in contact with them should know who received the treatment or control. The following procedures should achieve that effect:

1. To limit interaction effects, schools in different districts with similar prior outcomes are identified.

2. The injection apparatus, vaccine and placebo, are produced with a randomized serial numbers that may be traced with a computer system, but otherwise visually indistinguishable. They are then distributed accordingly.

3. The medical staff that would be administering the study are instructed on its purpose and trained in the procedures of asking potential subjects for consent. Ethics regarding patient privacy and knowledge are of utmost concern.

4. Eligible patients are informed of the study purpose, expected risks from preliminary trials, and the possibility of receiving either treatment or control. Patients who consent are given an injection. Neither the staff nor the patient should know whether the batch or serial number corresponds to treatment or control.

5. Anonymized patient data are sent back for completion of the study.

6. After completion, patients are given the option to be informed if they have been given the placebo.

## 2 Vaccine effectiveness

**For each of the NFIP study, and the Randomized controlled double blind experiment above, which numbers (or estimates) show the effectiveness of the vaccine?**

The first NFIP study was problematic in a number of ways that may have obscured the apparent effectiveness of the vaccine. The students that did not give consent may be a different group than the consenting group in, for example, health attitudes, education or income levels. Doctors may be less likely to diagnose polio for low income families. Since polio is known to infect younger children at higher rates, infection rates between the grades are incomparable. The drop-off in rates is so drastic from Grade 1 to Grade 2[1][2] that it would be more prudent to compare with non-consenting students of the same grade, 25 against 44 per 100,000. However, in general, we should average the treatment group and those not offered vaccines to counteract bias of non-consenting students self-sorting. In that case, it would be 31.8 against 54 per 100,000.

With the double-blind RCT, randomization allows fair comparison between treatment and control groups, revealing how effective the vaccine really was, with 28 against 71 per 100,000.

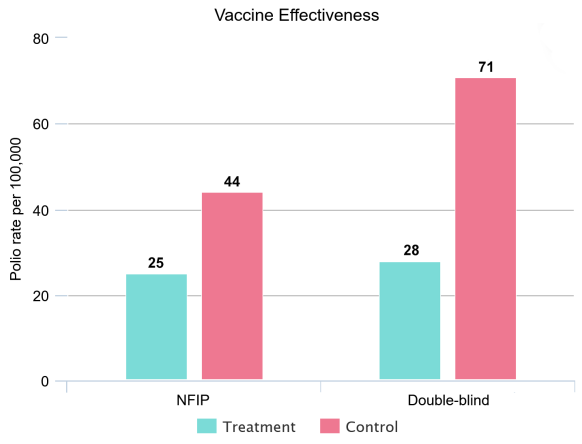


Figure 1.1: Comparison of polio rates between studies.

## 3 Problematic scenarios

### 3.1 Differential susceptibility

**What if Grade 1 and Grade 3 students are different from Grade 2 students in some ways? For example, what if children of different ages are susceptible to polio in different degrees? Can such a difference influence the result from the NFIP experiment?**

Yes. As previously described, younger students are more susceptible to polio[1][2]. Since Table 1.1 does not show the breakdown between the proportion of Grade 1 and Grade 3 students, it is hard to tell the bias from this effect. Greater proportion of Grade 1 students would bias for vaccine effectiveness, whereas more Grade 3 students would bias against it.

One design to prevent this difference would be to stratify the participants by age to ensure that the same proportion of students are in each group. That way, susceptibility is similar with regards to age between the treatment and control. Another design would be to randomize chosen participants into the groups, so that a large sample would cause similar expected values in susceptibility.

### 3.2 Infectious disease

**Polio is an infectious disease. The NFIP study was not done blind; that is, the children know whether they get the vaccine or not. Could this bias the results?**

Yes. Since the students know whether they received treatment, it could alter their behavior. Students who received the vaccine might be more confident to socialize, whereas students who did not give consent might be more self-conscious and conservative. Another possibility is that vaccinated students would only interact with each other. On the other hand, since the vaccines were only offered to Grade 2 students, there could also be interaction effects, such as some herd protection of non-vaccinated Grade 2 students due to their classmates being vaccinated.

A blocking (or clustering) design could prevent this problem. For example, vaccines could be given to students of one school, to compare with another school. In that case, there would be less differences in socialization since the entire school community would be granted protection. A double-blind design with a placebo also prevents this, since there would be no expected difference in behavior between the students who received the treatment and control.

### 3.3 Causation

**Even if the act of "getting vaccine" does lead to reduced infection, it does not necessarily mean that it is the vaccine itself that leads to this result.**

Yes. There could be a confounding factor. For example, it may be that the parents of students who consent to vaccination are more highly educated or have higher incomes, and therefore live in cleaner neighborhoods. These socioeconomic factors could have a causal relationship to polio infection rates as well as vaccine consent. If cleaner neighborhoods reduced exposure to polio or prevented children from building immunity to pathogens, then this would lead to the treatment effect being overstated or understated respectively.

It may also be that wealthier students receive better medical care, and therefore a higher chance of polio diagnosis, since 72% of polio infections are without symptoms[3]. Or that students administered the vaccine are less likely to be diagnosed. The treatment effect would be understated or overstated respectively.

To prevent this, another district with similar socioeconomic makeup could be studied as the control, with parental income of students in both districts recorded, so that we could tell the causal effect of the confounding factor itself. Again, this may not be as practical as implementing a double-blind randomly controlled trial, in which the students who receive the placebo would be among the same group of students who had consented.

## 4 Lower rates without consent

**In both experiments, neither control groups nor the no-consent groups got the vaccine. Yet the no-consent groups had a lower rate of polio compared to the control group. Why could that be?**

In the NFIP study, the no consent group might have enjoyed some of the benefits of their classmates being vaccinated, reducing their own risk by proxy. The control group of the NFIP study may have also suffered due to higher rates of Grade 1 infection[1][2].

In general, however, it may be the case that students who did not consent had on average some property that caused them to be more protected from polio than the consenting students. For example, as previously speculated, non-consenting students could have come from poorer neighborhoods and have a more robust immune system, or are given less attention during diagnoses. Both possibilities would depress their recorded rates of polio infection.

## 5 Influence of prior studies

**In the randomized controlled trial, the children whose parents refused to participate in the trial got polio at the rate of 46 per 100000, while the children whose parents consented to participate got polio at a slighter higher rate of 49 per 100000 (treatment and control groups taken together). On the basis of these numbers, in the following year, some parents refused to allow their children to participate in the experiment and be exposed to this higher risk of polio. Were their conclusion correct? What would be the consequence if a large group of parents act this way in the next year's trial?**

Such a conclusion would be incorrect on two accounts. First, the consenting group may have a higher baseline of infection rate, as previously discussed. Second, if we assume the perspective of these concerned parents that the non-consenting group represents the ground truth, we could find that the difference in proportions is statistically insignificant.

$$
\begin{aligned}
Z &= \frac{\mu_1 - \mu_0}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0}}} \\
&= \frac{0.00049 - 0.00046}{\sqrt{\frac{0.00049(1-0.00049)}{200000} + \frac{0.00046(1-0.00046)}{350000}}} \\
&\approx \frac{0.00003}{\sqrt{\frac{0.00048976}{200000} + \frac{0.00048975}{350000}}} \\
&\approx \frac{0.00003}{\sqrt{\frac{0.00048976}{200000} + \frac{0.00048975}{350000}}} \\
&\approx \frac{0.00003}{0.000033691} \\
&\approx 0.89
\end{aligned}
\tag{1.1}
$$

The two-tailed p-value is roughly 0.37.

If a large group of parents deny consent, then it may bias the results if the parents are biased toward any significant factor. In this case, we could assume such a group is either educated enough to pay attention to research findings, but perhaps not enough to correctly interpret them. It would also make the study more costly and increase time involved to produce a large consenting sample size.

# Problem 1.3: ASA Statement on P-values

## a-1

**Your colleague on education studies really cares about what can improve the education outcome in early childhood. He thinks the ideal planning should be to include as much variables as possible and regress children's educational outcome on the set. Then we select the variables that are shown to be statistically significant and inform the policy makers. Is this approach likely to produce the intended good policies?**

It is important to note that regression by itself depicts correlation and not causation, and cherry-picking statistically significant findings gives spurious impressions when the correlations only happen by chance. Policy in contrast need to be deliberate and based on causative relations. The full method and parameters of a formal study – the hypothesis – should be decided in advance, so that the data tests the hypothesis, rather than the converse. We should use evidence to decide whether to reject the null; otherwise, we have a version of confirmation bias. We must also adjust for the family-wise error rate (FWER), e.g. with Bonferroni correction, when we consider a large family of hypotheses, or we risk increasing the false positive rate.

The proposed design contradicts the second to fourth principles in ASA's position on p-values[4]. Most importantly, studies need to be fully transparent on all their analyses, or it renders the results uninterpretable.

## a-2

**Your friend hears your point, and think it makes sense. He also hears about that with more data, relations are less likely to be observed just by chance, and inference becomes more accurate. He asks, if he gets more and more data, will the procedure he proposes find the true effects?**

A large ratio of samples over features helps, especially if there are enough samples covering every combination of features. It would certainly decrease bias and variance in the inference.

But it would not resolve some fundamental flaws in the study design.

1. Failing to account for FWER increases Type I error, especially with a large number of hypotheses.

2. Regressing on superfluous variables increases the likelihood of confounding factors that camouflages the true effect. The reason for the correlation is unclear.

3. Some categories may not increase with increased sample size, such as the number of presidents or Nobel laureates.

4. Educational outcome may have caused a factor, rather than being the effect. There's no indication of whether the regressors are downstream or upstream variables.

5. As it suggests in the ASA statement, p-value cutoffs are arbitrary[4] (principle 3) and may not accurately reflect the significance needed for a particular domain. p-values are also statements about whether the data matches the hypothesized distribution, and not necessarily about either themselves[4] (principle 2).

## b-1

**An economist collects data on many nation-wise variables and surprisingly find that if they run a regression between chocolate consumption and number of Nobel prize laureates, the coefficient to be statistically significant. Should he conclude that there exists a relationship between Nobel prize and chocolate consumption?**

Broadly speaking, the regression establishes a correlation between nation-wide chocolate consumption and the number of Nobel prize laureates it produces. It doesn't explain the reason. There could be a confounding factor, perhaps that higher GDP causes both chocolate consumption per capita and better education and the opportunity to conduct research. Or that a higher population has both the opportunity to consume more chocolate in total and produce more Nobel prize candidates. But even a correlation by complete chance is still a relationship.

We could not, however, on this fact alone deduce a causative effect from chocolate to Nobel laureates, nor infer consumption by the laureates themselves. Therefore, even though there is a broad relation, it would be hard to interpret it to make meaningful policy or business decisions.

## b-2

**A neuroscience lab is interested in how consumption of sugar and coco may effect development of intelligence and brain growth. They collect data on chocolate consumption and number of Nobel prize laureates in each nation, and finds the correlation to be statistically significant. Should they conclude that there exists a relationship between chocolate consumption and intelligence?**

No. While it could be argued that Nobel prize laureates have high intelligence, such a correlation fails to account for the chocolate consumption in other segments within a nation. The number of laureates is both too small and a poor proxy for intelligence in general, since there should be many more highly intelligent individuals who are not Nobel laureates. It also doesn't account for the population who might be below average in intelligence that might also consume chocolate.

## b-3

**In order to study the relation between chocolate consumption and intelligence, what can they do?**

A better study design should utilize an intelligence proxy that influences a broader spectrum of society. Unfortunately, the tools we have to measure intelligence aren't perfect and are often subject to ethnic and socioeconomic factors. Regardless, IQ tests or standardized admission assessments like the SAT or GRE are often used.

One way is to have participants take annual IQ tests, but offer stratified treatment group a supply of chocolate. It could either be an observed study where the control is offered nothing, or double-blind where the control is offered some other snack, and participants are not informed that the study is about chocolate in particular.

Alternatively, the population could be randomly sampled to complete a survey of chocolate consumption, and stratified groups could be selected to take an IQ test.

## b-4

**The lab runs a randomized experiment on 100 mice, add chocolate in half of the mice's diet and add in another food of the equivalent calories in another half's diet. They find that the difference between the two groups time in solving a maze puzzle has p-value lower then 0.05. Should they conclude that chocolate consumption leads to improved cognitive power in mice?**

Yes, but it should be qualified to the extent that solving the maze puzzle involves cognitive power (as opposed to for example reaction time, high energy or speed), and also that the study only tests short-term effect. The study would be randomized and controlled, thereby establishing causal inference. The design also seems

to purposefully measure the effect of chocolate consumption without any superfluous and unrelated variables.While principles 2 and 3 of the ASA statement says that scientific conclusions shouldn't be based on a p-value threshold[4], it should be noted that science would be impossible if we disqualify all statistical statements from scientific knowledge. Instead, I think the point is that we keep in mind no scientific claim is ever absolute, but subject to our confidence in them, based on data and models from past studies.

## b-5

**The lab collects individual level data on 50000 humans on about 100 features including IQ and chocolate consumption. They find that the relation between chocolate consumption and IQ has a p-value higher than 0.05. However, they find that there are some other variables in the data set that has p-value lower than 0.05, namely, their father's income and number of siblings. So they decide to not write about chocolate consumption, but rather, report these statistically significant results in their paper, and provide possible explanations. Is this approach correct?**

This approach is not correct for the same reason as in a-1. The only difference here is that the statistically significant regressors sound more reasonable, and using individual data instead of nation-wise data might confer more power. However, the same problem occurs that the entire analysis would not be fully and transparently reported, which could make the study uninterpretable[4]. Instead, every variable should be reported along with the significant results, and the entire selection process so that they could be replicated.

## c

**A lab just finishes a randomized controlled trial on 10000 participants for a new drug, and find a treatment effect with p-value smaller than 0.05. After a journalist interviewed the lab, he wrote a news article titled "New trial shows strong effect of drug X on curing disease Y." Is this title appropriate? What about "New drug proves over 95% success rate of drug X on curing disease Y"?**

According to principle 5 of the ASA statement, the p-value does not measure the size of a treatment effect[4], and so neither of these titles would be formally appropriate. Instead, the p-value measures statistical significance, or the defined success (rejection of null hypothesis) of a treatment effect within a confidence interval. In other words, presuming a universe where non-deterministic events are possible, the treatment effect would succeed in at least 95% of possible worlds. However, this is not a statement about the parameters of what success means or its importance.

## d

**Your boss wants to decide on company's spending next year. He thinks letting each committee debates and propose the budget is too subjective a process and the company should learn from its past and let the fact talk. He gives you the data on expenditure in different sectors and the company's revenue for the past 25 years. You run a regression of the revenue on the spending on HR sector, and find a large effect, but the effect is not statistically significant. Your boss saw the result and says "Oh, then we shouldn't increase our spending on HR then". Is his reasoning right?**

The reasoning is incorrect, not the least because business decisions should not be made purely on statistical significance[4] (by principle 3), but also that the lack of significance does not necessarily imply a lack of effect. That's why we say that we could not reject the null hypothesis, instead of saying that we accept the null hypothesis. That we could not affirm the lack of effect, we are just not as sure whether there is an effect. In this case, more research might be necessary.

## e

**Even if a test is shown as significant by replication of the same experiment, we still cannot make a scientific claim. True or False?**

It should be noted that scientific claims are inductive and not absolute. Even the surest result (such as gravity) comes from replicated experiments increasing our confidence in a relation. That confidence could increase until it becomes colloquial fact, but it never becomes absolute.

In that light, we could make scientific claims from any study, but qualified by correct amounts of confidence and optimism. A replicated study improves the probability of the result from a Bayesian perspective, but not absolutely. The study methodology could be based on faulty assumptions, or have low power. It might even be the result of noise that future studies might contradict. Or, if teams are motivated by some external goal or racing to publish results, then either or both of the research could be biased.

## f

**Your lab mate is writing up his paper. He says if he reports all the tests and hypothesis he has done, the results will be too long, so he wants to report only the statistical significant ones. Is this OK? If not, why?**

This is *not* okay, for the same reason as a-1 and b-5. Reporting incompletely possibly causes the results to be uninterpretable[4]. If too many hypotheses have been considered, this may indicate a problem in the methodology. They should still all be reported, in a table if not the results section. No matter the motivation, ensuring that the report is interpretable is ethical and necessary to contribute positive value.

## g

**If I see significant p-values, it could be the case that the null hypothesis is consistent with truth, but my statistical model does not match reality. True or False?**

Principle 1 and 2 of the ASA statement suggests that the p-value summarizes incompatibility between data and a statistical model, that it is a statement about data in relation to the explanatory model[4]. Therefore, if the model considered in the study is inconsistent with the data, then it would be plausible that the p-value could be overstated or understated. In that case, the p-value would be uninterpretable, and it could be that the null hypothesis is true, or would have been found true if a consistent model was used.

# Problem 1.5: Why Most Published Research Findings Are False

## 8 Extent that repeated testing can reduce probability

**Show that the extent of repeated independent testing by different teams can reduce the probability of the research being true.**

The post-study probability of research being true could be stated as:

$$\begin{aligned}
\text{PPV} &= \frac{\mathbf{P}(\text{relation exists, at least one of the } n \text{ repititions finds significant})}{\mathbf{P}(\text{at least one of the } n \text{ repititions finds significant})} \\
&= \frac{\mathbf{P}(\text{relation, significant} \geq 1)}{\mathbf{P}(\text{relation, significant} \geq 1) + \mathbf{P}(\text{no relation, significant} \geq 1)} \\
&= \frac{\mathbf{P}(\text{relation})\mathbf{P}(\text{significant} \geq 1|\text{relation})}{\mathbf{P}(\text{relation})\mathbf{P}(\text{significant} \geq 1|\text{relation}) + \mathbf{P}(\text{no relation})\mathbf{P}(\text{significant} \geq 1|\text{no relation})} \\
&= \frac{\frac{R}{(1+R)}(1 - \prod_i^n \beta_i)}{\frac{R}{(1+R)}(1 - \prod_i^n \beta_i) + \frac{1}{1+R}(1 - \prod_i^n (1-\alpha_i))} \\
&= \frac{R(1 - \prod_i^n \beta_i)}{R(1 - \prod_i^n \beta_i) + (1 - \prod_i^n (1-\alpha_i))}
\end{aligned} \tag{3.1}$$

We consider a very large $n$ number of studies and find the PPV to be equivalent to the pre-study probability.

$$\begin{aligned}
\text{PPV} &\to \lim_{n\to\infty} \frac{R(1 - \prod_i^n \beta_i)}{R(1 - \prod_i^n \beta_i) + (1 - \prod_i^n (1-\alpha_i))},\ 0 < \alpha,\ \beta < 1 \\
&\to \frac{R}{R+1}
\end{aligned} \tag{3.2}$$

## 9 Condition that PPV does not decrease

**What would make bias or increasing teams testing the same hypothesis not decrease PPV? (Assuming $\alpha = 0.5$)**

It is easy to see from the previous result that if $c_1 = \prod_i \beta_i$ and $c_2 = \prod_i (1-\alpha_i)$, then

$$\begin{aligned}
\text{PPV} &\stackrel{?}{=} \frac{R(1 - \prod_i^n \beta_i)}{R(1 - \prod_i^n \beta_i) + (1 - \prod_i^n (1-\alpha_i))} \\
&\stackrel{?}{=} \frac{R(1-c_1)}{R(1-c_1) + (1-c_2)},\ 0 < c_1,\ c_2 < 1 \\
&\stackrel{?}{=} \frac{R}{R + \frac{1-c_2}{1-c_1}}
\end{aligned} \tag{3.3}$$

Therefore,

$$\text{PPV} > \frac{R}{R+1}, \text{ if } c_1 < c_2 \tag{3.4}$$

$$\text{PPV} = \frac{R}{R+1}, \text{ if } c_1 = c_2 \tag{3.5}$$

$$\text{PPV} < \frac{R}{R+1}, \text{ if } c_1 > c_2 \tag{3.6}$$

Finally, for PPV *not* to decrease as $n \to \infty$, it must be true that PPV $\leq R/(R+1)$ for some finite $n$, so $c_1 \geq c_2$. Equivalently, $\prod_i \beta_i \geq \prod_i (1-\alpha_i)$. Assuming $\beta = \beta_i, \ \forall i$ and $\alpha_i = 0.05, \ \forall i$, then PPV does not decrease when $\beta \geq 0.95$, or stated another way, $\mathbf{1 - \beta < 0.05}$.

To find the result under bias,

$$\begin{aligned}
\text{PPV} &= \frac{uR + (1-u)R(1-\beta)}{u(1+R) + (1-u)R(1-\beta) + (1-u)\alpha} \\
&= \frac{R - R\beta + uR\beta}{u + R - R\beta + uR\beta + \alpha - u\alpha} \\
&= \frac{R(1-\beta) + uR\beta}{R(1-\beta) + \alpha + u(1 + R\beta - \alpha)}
\end{aligned} \tag{3.7}$$

In order for PPV to increase when $u$ increases, it must be true that

$$\begin{aligned}
\frac{R(1-\beta)}{R(1-\beta)+\alpha} &\leq \frac{R\beta}{1+R\beta-\alpha} \\
R(1-\beta)(1+R\beta-\alpha) &\leq (R(1-\beta)+\alpha)R\beta \\
R(1-\beta)R\beta + R(1-\beta)(1-\alpha) &\leq R(1-\beta)R\beta + \alpha R\beta \\
R(1-\beta)(1-\alpha) &\leq \alpha R\beta \\
\frac{1-\beta}{\beta} &\leq \frac{\alpha}{1-\alpha} \\
\frac{1}{\beta} - 1 &\leq \frac{1}{1-\alpha} - 1 \\
1 - \alpha &\leq \beta \\
1 - \beta &\leq \alpha
\end{aligned} \tag{3.8}$$

Given that $\alpha = 0.05$, PPV does not decrease when bias increases if $\mathbf{1 - \beta \leq 0.05}$.

## 10 Unanimous replication

**Read critically and critique! Remember the gold rule of science, replication? For the third table in the paper, if researchers work on the same hypothesis but only one team finds significance, the other teams are likely to think the results is not robust, since it is not replicable. In light of this, how would you model the situation when multiple teams work on the same hypothesis and the scientific community requires unanimous replication? What would be the PPV?**

Under unanimous replication,

$$\begin{aligned}
\text{PPV} &= \frac{\mathbf{P}(\text{relation exists, } n \text{ repititions unanimously finds significant})}{\mathbf{P}(n \text{ repititions unanimously finds significant})} \\
&= \frac{\mathbf{P}(\text{relation, significant} = n)}{\mathbf{P}(\text{relation, significant} = n) + \mathbf{P}(\text{no relation, significant} = n)} \\
&= \frac{\mathbf{P}(\text{relation})\mathbf{P}(\text{significant} = n|\text{relation})}{\mathbf{P}(\text{relation})\mathbf{P}(\text{significant} = n|\text{relation}) + \mathbf{P}(\text{no relation})\mathbf{P}(\text{significant} = n|\text{no relation})} \\
&= \frac{\frac{R}{(1+R)}\prod_i^n(1-\beta_i)}{\frac{R}{(1+R)}\prod_i^n(1-\beta_i) + \frac{1}{1+R}\prod_i^n \alpha_i} \\
&= \frac{R\prod_i^n(1-\beta_i)}{R\prod_i^n(1-\beta_i) + \prod_i^n \alpha_i}
\end{aligned} \tag{3.9}$$

Therefore for a large $n$,

$$\text{PPV} \to 1, \qquad \text{if } \prod_i^n(1-\beta_i) > \prod_i^n \alpha_i \tag{3.10}$$

$$\text{PPV} = \frac{R}{R+1}, \text{ if } \prod_i^n(1-\beta_i) = \prod_i^n \alpha_i \tag{3.11}$$

$$\text{PPV} \to 0 \qquad \text{otherwise} \tag{3.12}$$

Therefore, if $\beta_i = \beta,\ \forall i$ and $\alpha_i = \alpha,\ \forall i$, then PPV continues to increase with each replication if $1-\beta > \alpha$, decreases if $1-\beta < \alpha$, and remains at the pre-study probability if both sides are equal.

## 11 No poor practices

**Suppose there is no bias and no teams are racing for the same test, so there is no misconduct and poor practices. Will publications still likely to be false than true?**

In the absence of bias, racing teams, and misconduct, research is more likely to be false if $(1-\beta)R \leq \alpha$[5], derived as such.

Assume that PPV must be $\leq 0.5$.

$$\begin{aligned} \text{PPV} &= \frac{R(1-\beta)}{R(1-\beta)+\alpha} \\ 0.5 &\geq \frac{R(1-\beta)}{R(1-\beta)+\alpha} \\ 0.5(R(1-\beta)+\alpha) &\geq R(1-\beta) \\ R(1-\beta)+\alpha &\geq 2R(1-\beta) \\ \alpha &\geq R(1-\beta) \end{aligned} \tag{3.13}$$

The answer therefore depends on the strictness of how we judge poor practice. The probability depends on power, pre-study relationships, and significance. $\alpha$ is usually fixed at 0.05, although the ASA has criticized this practice as arbitrary[4]. Then, there are fields in biological, social sciences, and data science where large amounts of hypotheses are generated, most of them false. There are fields where the sample sizes must be small, or the effects small, leading to small power. These limitations increase likelihood of research being false.

## 12 Application to Problem 1.3

**In light of this paper, let's theoretically model the problem of concern in Problem 1.3! Suppose people base the decision to making scientific claim on p-values, which parameter does this influence? $R$, $\alpha$, or $\beta$? Describe the effect on the PPV if scientists probe random relations and just look at p-value as a certificate for making scientific conclusion.**

A significant p-value depends on $\alpha$, but it doesn't influence the significance, $\alpha$. Cherry-picking and reporting only significant values, however, might produce a family-wise error rate (FWER) $\geq \alpha$, or an increase in Type I errors. Typically, such research fails to apply a Bonferroni correction. Such methodological flaws effectively increases $\alpha$.

Loannidis considers this bias in his Corollary 3-4 and says that it decreases $R$ since it increases the number of non-relationships[5].

Both an increase in $\alpha$ and decrease in $R$ increases the likelihood of false publications by $(1-\beta)R \leq \alpha$, decreasing PPV.

# List of Figures



# List of Tables



# References

[1] J. Singh, R.S. Sharma, MNAMS, and T. Verghese, "Epidemiological Considerations on Age Distribution of Paralytic Poliomyelitis," *Journal of Tropical Pediatrics*, vol. pp. 237-241, August 1996. doi:10.1093/tropej/42.4.237

[2] Paul Horstmann, "Age distribution of poliomyelitis during recent epidemics in Copenhagen," *Journal of Internal Medicine*, vol. 124, iss. 5, pp. 382-493, January 1946. doi:/10.1111/j.0954-6820.1946.tb09220.x

[3] "What is Polio?", *Centers for Disease Control and Prevention*, October 2019. [Online]. Available: https://www.cdc.gov/polio/what-is-polio/index.htm. [Accessed March 7, 2021].

[4] Ronald L. Wasserstein and Nicole A. Lazar, "The ASA's Statement on p-Values: Context, Process, and Purpose," *The American Statistician*, vol. 70, iss. 2, pp. 129-133, June 2016. doi: 10.1080/00031305.2016.1154108

[5] John P.A. Loannidis, "Why Most Published Research Findings Are False," *PLoS Medicine*, vol. 2, iss. 8, pp. 696-701, August 2005. doi: 10.1371/journal.pmed.0020124