Problem 1.1

1.
   a) Inform the parents that if they want their children to be part of this double-blind experiment, they will receive two injections, separated by a year (or whatever period is optimal for this type of vaccine). One of these trials would be a real vaccine, while the other – the salt injection.
   b) Ask all the participants to sign the agreement for participating in this double-blind experiment. Those who agreed to participate but do not want their children to be vaccinated, form the «No consent» group.
   c) Assuming that the effect of the vaccine is permanent, the results of the study should be analyzed just before the second trial.
2. In the NFIP study, the effectiveness of the vaccine should be compared within Grade 2 students, as they differ at least in age from Grade 1 and Grade 3 students. Thus, numbers 25 and 44 should be compared to judge the effectiveness of the vaccine.
   In Randomized Controlled Double-Blind Experiment, it is best to compare treatment and control group, as the participants were randomly assigned to one of this group. Moreover, the placebo effect is also handled due to double-blindness of the experiment. The “No consent” group may be having different baseline features, such as good immune system and low illness rate history, so that their parents preferred to refuse the vaccination.
3.
   - Yes, there is no evidence that polio is independent from individual’s age. Therefore, to exclude possible bias, the vaccine should be offered to Grade 1 and Grade 3 students also. One should then compare the polio rates within each Grade.
   - Indeed, knowing the fact of vaccination may change child’s behaviour. For example, he can start to take less care of himself, being subject to the higher contamination risk.
     SOLUTION: say to the patients, that the effects appears only one year after vaccination, so that they need to be as careful as before 😊
   - The act of getting vaccine, linked to the lower infection rate, may be caused by the fact that those who agreed taking the vaccine are more attentive to their health, than those who refused. It is then not the vaccine itself that results in fewer polio rate, but individual’s behaviour type. Double-blinded experiment with two-stage vaccination (with one fake-placebo injection) would be a possible remedy for this issue.
4. Either due to the different behaviour of individual’s in “Control” and “No consent” group, where the ones from the first were more careless about their health than the ones from the latter. It may also be that “No consent” group has some prior knowledge of the robustness of their immune system, permitting them to be less in danger from this new infection.
5. Parent’s conclusion is correct in a sense, that even if the vaccine shows possible effectiveness, due to the fact that half of the children receive placebo, the average contamination rate is higher compared to “No consent” group. If large group of children would consent the vaccination, it may affect the general distribution of the infection, affecting, in the first way, the “Control” group, increasing its polio rate. The “Treatment” group polio rate may be affected differently, as it can be principally determined by the vaccination effectiveness.

Problem 1.3

a-1. No, this is an erroneous approach. Selecting only significant results to be reported severely affects the true outcome of the study, leading to the false conclusions. Having a lot of hypothesis along with some significance level α (let say 0.05) will inevitably make 5% of these variables appear to be statistically significant, just by chance. There is a need to correct significance level (or p-value) while using multiple hypothesis testing.
a-2. It is true, that bigger sample sizes tend to give stronger significant results, and p-values will decrease, but the approach of reporting only significant results is still to be avoided, as it is not a valid remedy for "p-hacking" during multiple hypothesis testing.
b-1. No, he should not. Correlation does not mean causation. Testing multiple hypothesis inevitably makes some percentage of them to appear statistically significant just by chance.
b-2. Here the setting of the experiment is better, as the hypothesis is formed before observing the outcome, but still, one should not make a conclusion about the relationship between Nobel Prize laureates and chocolate consumption, as these two variables are very distinct from each other and there may be a lot more factors in play. Moreover, nothing guarantees that it is just a coincidence, and actually none of the laureates of a particular country was actually consuming chocolate, even if on average its consumption among the population is high. To test this hypothesis, the setting of the experiment should be completely changed.
b-3. To study the real relationship between chocolate consumption and intelligence, ideally there should be two groups (treatment and control) of long surveyed individuals, initially having the same IQ tested rate. Treatment group regularly receive chocolate, while the control group will be deprived from it, substituting it with some other placebo-like thing. The two groups should then pass an IQ test once again after some period of time, and the results should be compared. All the other differences between two groups should be controlled or excluded.
b-4. Yes, they should conclude that chocolate consumption leads to improved cognitive power in mice.
b-5. No, as stated in a-1, selective result reporting, a.k.a. data dredging, should be avoid by all means, as it leads to the wrong conclusions. Concretely, the significance level of 0.05 is no longer valid in multiple hypothesis testing problem, and appropriate correction (e.g. Bonferroni) should be applied.
c. The word "strong" in the first title should be avoided, as being relative. P-value of 0.05 is just a common value, which should not be used as strong reference to make claims on significance.
d. His reasoning is right.
e. False.
f. It is not OK, as some of the statistically significant hypothesis may be such by chance, only due to the specific significance level chosen in the beginning. There should be correction to the p-value, which depend on number of hypothesis tested.
g. Yes, it still could be the case, as the probability of type I error is never zero.

Problem 1.5

8. $PPV = \frac{P(relation\ exists, at\ least\ one\ of\ the\ n\ repetitions\ finds\ significant)}{P(at\ least\ one\ of\ the\ n\ repetitions\ finds\ significant)} = \frac{R(1-\beta^n)}{R+1-(1-\alpha)^n-R\beta^n} = \frac{R(1-\beta^n)}{R(1-\beta^n)+1-(1-\alpha)^n} = 1 - \frac{(1-(1-\alpha)^n)}{R(1-\beta^n)+(1-(1-\alpha)^n)}$

The term $(1-(1-\alpha)^n)$ tends to increase with n increasing, so the whole second relation is increasing, making the overall PPV decreasing. In other words, the extent of repeated independent testing by different teams can reduce the probability of the research being true.

9. Reporting all the results by all the teams, and accounting for all the results while analysing PPV. And taking into account the proportion of true relationships existing in the field.

10. The finding should be transparent, and for this, there should be a registration process, taking place before studies begin, so that we can estimate pre- and post-study odds of the tested hypothesis being true, decreasing significance due to a bias or a chance.
The PPV could be defined as:

$$PPV = \frac{R(1-\beta^n)}{R(1-\beta^n)+1-(1-\alpha)^n}$$

11. In this setting,

$$PPV = \frac{R(1-\beta)}{R-\beta R+\alpha}$$

The research finding is thus more likely to be true than false if $(1-\beta)R > \alpha$, so the answer to this question depends on the power β, significance level α and pre-study odds R.

12. It actually influences parameter R (which is the number of true relations divided by not-true relations), because disregarding high-p-valued relations from results decreases the number of not-true relations, making R greater. Recalling that

$$PPV = \frac{R(1-\beta)}{R-\beta R+\alpha} = 1 - \frac{\alpha}{R(1-\beta)+\alpha}$$

We can see that such an increase in R makes PPV falsely increase.