What Petri dishes have to do with your research

This article, by **Douglas VanDerwerken**, was a runner-up in the Young Statisticians Group/Significance Writing competition. It nicely explains why throwing away data is not a good idea.

Pop quiz

Imagine yourself as an undergraduate sitting again in Microbiology 101 – as if the first time were not enough! (If you have never sat Microbiology 101, do not worry – this will become relevant to you later.) After another one of the professor's jokes about how he is such a "fun guy", he has you all take out a piece of paper for a pop quiz. Much of Microbiology 101 seems to be about counting colonies of bacteria on Petri dishes, and his question, borrowed from a popular US university's website[1], goes as follows:

> You plate several dilutions from a bacterial suspension onto nutrient agar plates, and count the numbers of colonies that have formed after 24 hours of incubation. Here are the results:
>
> A. 100 µl of the 10^{-4} dilution, TNTC (too numerous to count)
> B. 100 µl of the 10^{-5} dilution, 685
> C. 100 µl of the 10^{-6} dilution, 52
> D. 100 µl of the 10^{-7} dilution, 4.
>
> What is the concentration of bacteria in the original suspension?

Dutiful student that you are, you think back to the previous week's homework assignment. You remember something about a 30–300 rule. If the count on your Petri dish is below 30, it is plagued with "statistical inaccuracies". If it is above 300, there may be overcrowding. So answers D and B have to be discarded. That leaves C, 52, as the value to work from. You still need to divide by the dilution factor and convert the answer into colony-forming units (CFUs) per millilitre. You scribble down 5.2×10^8 CFUs/ml, and hand in the slip of paper just as the bell rings.

A similar scene can be found each semester at hundreds of universities across the globe. And hundreds of dutiful students each answer the same way.

Unfortunately, they are all wrong.

The 30–300 rule

The 30–300 rule is standard practice in industry and university research, wherever colonies of bacteria are cultured and counted. Its origins trace at least as far back as a 1916 paper by Breed and Dotterrer[2], in which they caution against using counts below 20 or greater than 400. When the count is less than 20, a few bacteria colonies from air contamination can drastically impact results; on the other hand, counts greater than 400 may be lower than expected due to overlapping colonies. Eventually these lower and upper bounds morphed to 30 and 300, probably because those are a little more user-friendly. (The change probably went something like this: "Was it 20-400? Or 40-200? Let's just split the difference!") The problem is not with the 30–300 (or 20–400) rule itself, but rather that it is always taught without explaining the important exceptions to when it should be applied. The result is that thousands of microbiology undergraduates are armed with the same broken tool. Or at least the same tool which is completely inadequate a fair percentage of the time.

If there are fewer than 30 colonies on a Petri dish the count is vulnerable to statistical errors. But why throw away that data?

So what are the exceptions? According to Breed and Dotterrer: "Plates having fewer than 20 or more than 400 colonies … should never be trusted unless checked by comparison with plates from different dilutions having more than 30 or less than 400 colonies." The implicit exception to the 30–300 rule, then, is that counts outside of this range *may* be used if they accord with counts inside the range from a different dilution.

What's the big deal?

If the dutiful student in the above scenario were to compute a typical 95% confidence interval having followed the 30–300 rule, he or she would obtain $5.2 \pm 1.96(\sqrt{52}/10) = (3.8, 6.6) \times 10^8$ CFUs/ml (assuming Poisson-distributed counts). The fear with using 685 (answer B) is that with counts that high, overcrowding can occur, which would cause CFUs to overlap and be indistinguishable from one another, resulting in too low a count. However, we see absolutely no evidence of this here since 685 is *greater* than $52 \times 10 = 520$, not less; our count is higher than the intermediate one, so clearly overcrowding in not an issue. By discarding the 685 CFUs and retaining only 52, you are effectively reducing your sample size by 90%. In fact, if the student had used 685 instead of 52, the resulting confidence interval would be $(6.34, 7.36) \times 10^8$, which is 36% as wide as the "right" answer (see Figure 1, intervals B and C).

The statistician's cardinal rule is: "Never throw away relevant data!" If there is no strong evidence of overcrowding on high plate counts or of contamination from the air on low plate counts, then feel free to use all of these in your estimate of the original concentration. Along these same lines, the budding microbiologist would do well to recall G. William Claus's[3] advice that "TNTC (Too numerous to count) is not the same as TLTC (Too lazy to count)"!

Once we have collected our data, the Poisson assumption implies that the maximum likelihood estimator is simply a weighted average of all valid observations, where the weights correspond to the inverse of the relative variances. For instance, in the pop quiz example, combining all the data gives us the weighted average $(100\times 6.85+10\times 5.2+1\times 4)/111 = 6.68\times 10^8$, which is approximately normally distributed about the true concentration. By incorporating all the data, we have effectively increased our sample size from 10 to 111, yielding a 95% confidence interval of (6.20, 7.16). Note that this interval is slightly smaller than the interval obtained using only 685, and 34% as wide as the "right" answer. These intervals are summarised in Figure 1.

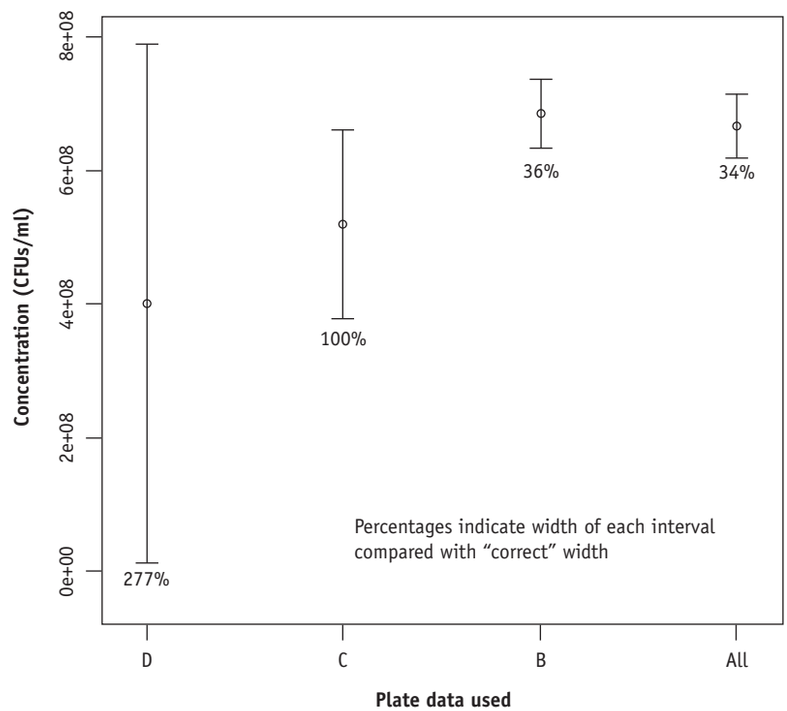

Figure 1: Comparison of the point estimates and confidence intervals obtained using different amounts of data.

Formally, for the cancer studies in the text, we say that $\hat{\theta}_i \sim N\left(\theta, \sigma_i^2\right)$. It can be shown that the weighted average

$$\tilde{\theta} = \frac{\sum_i w_i}{\sum_i w_i \theta_i},$$

with weights $w_i = \left(\hat{\sigma}_i^2\right)^{-1}$, is the best unbiased estimator for θ. A study's accuracy is determined by its standard error, which we can approximate from the reported confidence limits:

$$\hat{\sigma}_i = \frac{\log(\mathrm{UCL}_i) - \log(\mathrm{LCL}_i)}{2 \times 1.96}.$$

Importantly, we know that $\tilde{\theta}$ is approximately normal about the true θ, which allows for straightforward inference. From this we get the result that the one-sided *p*-value for testing H_0: $\theta = 0$ against H_a: $\theta > 0$ is 5.4×10^{-7}. In other words, the odds of cancer in group 1 are higher than in group 2. Using the data from all four studies gives a decisive answer that was not obtainable from any of the studies individually.

How it relates to you

The 30–300 rule, as currently employed, throws away useful data. This is especially true when higher counts are discarded which show no signs of overcrowding. Even in the best-case scenario, where the counts retained were the highest observed, incorporating lower plate counts still improves precision. But what does this have to do with your research? Well, if you are a microbiologist and you have been faithfully obeying the 30–300 rule, I hope I have convinced you to be at least a little more rebellious. For the rest of us, the connection is that statistics can and should be used to combine data from apparently incompatible (at least at first glance) sources. Just as microbiologist undergraduates have been using only a small portion of the available data, we as researchers do not always make the best use of prior literature. Meta-analysis seeks to combine data from disparate sources in an attempt to discern effects with more power or state hypotheses with more confidence than can any of the individual studies. But the exact same principles used in the Petri dishes can be applied in your discipline, whatever it is.

In the rest of this article I would like to convey the small epiphany I had while trying to combine data from different sources in a setting very different from Microbiology 101.

I was confronted with a table of summary statistics as part of a homework assignment for a graduate class in statistical modelling. Although these studies were performed at various times by different teams, they were trying to measure the same thing – the odds ratio of cancer between two groups. As is common in this type of research, logistic regression had been used to obtain estimates and confidence intervals on the log scale (where things are normally distributed), and these were then exponentiated. The question of interest was what evidence there was that the log odds ratio exceeds 0. When the log odds ratio θ is greater than zero then the odds of cancer are higher in the first group.

Suppose your summary data looked like the table below, in which "log(OR)" stands for the natural log of the odds ratio, and "log(LCL)" and "log(UCL)" represent the 95% lower and upper confidence limits.

Study	*log(LCL)*	*log(OR)*	*log(UCL)*
1	−0.20	0.05	0.29
2	0.14	0.23	0.32
3	−0.73	−0.04	0.66
4	−0.06	0.13	0.33

First of all, notice that in one of the four studies, the estimated log odds ratio was below 0; and in three studies, negative numbers were included in the confidence interval. On the surface it may appear that there is some evidence that θ is greater than zero, but it does not appear overwhelming.

My epiphany came when I realised that these studies are just like the Petri dishes plated with bacteria from different dilutions. That is, the studies are all trying to measure the same thing (the log odds ratio), but some are more accurate than others. The optimal contribution of each study is determined by the accuracy of that study – the more accurate the study, the more say it has in the final answer, so it is weighted accordingly. In this example, when we have taken all four studies into account and weighted them, the one-sided *p*-value for testing H_0: $\theta = 0$ against H_a: $\theta > 0$ is 5.4×10^{-7} (see box) That is, there is extremely decisive evidence that the log odds ratio of cancer is greater than 0, or equivalently, that the odds of cancer in group 1 are higher than the odds in group 2. The corresponding 95% confidence interval for the log odds ratio is (0.12, 0.27), which is smaller than that obtained by any of the individual studies. Using all four studies has clearly paid off.

Going forward

The moral here is not that a weighted average is the answer to every statistical question. Instead, I have shown that two seemingly unrelated problems deal with the same fundamental issue of combining information from disparate sources. As statisticians, we seek to answer questions about the world around us using all the relevant information we can get our hands on. Statistical methods are the tools we use to answer these questions. It is critical that we think about when the same tools should (and should not) be used to answer different questions. You may never find yourself in Microbiology 101 again, but you never know when tools picked up for one problem will lend themselves very nicely to solving another. That said, it is important that we not follow the status quo so blindly that we forget the limitations and assumptions of the status quo's methods. Going forward, we should strive to use existing methods when they are applicable and to develop new ones when they are not.

A weighted average is not the answer to every question, but it can help us to use all the data we can get our hands on

References

1. http://www.cbs.umn.edu/sites/default/files/public/downloads/dilutiontheory.pdf
2. Breed, R. and Dotterrer, W. D. (1916) The number of colonies allowable on satisfactory agar plates. *Journal of Bacteriology*, **1**, 321–331, 1916.
3. Claus, G. W. (1989) *Understanding Microbes: A Laboratory Textbook for Microbiology*. New York: W.H. Freeman.

Douglas VanDerwerken is at the Department of Statistical Science, Duke University, North Carolina. Some of this research was conducted while he was at Brigham Young University.