

Why Are Clinical Trials So Complicated?

By [Derek Lowe](#) 1 May, 2020

Since we've been dealing with a lot of clinical trial data around here, I thought I would field a question from my email that might get to some things that others are thinking about as well. Here (with permission of its sender) is the idea:

. . .it does seem, from the outside, like most of the time and expense of clinical trials really is just bureaucracy. I can easily think of several compounds (some FDA-approved but no longer being made, some approved outside the US, some naturally occurring and sold as supplements) that I could literally run a trial on from my apartment, if the only question were "does this actually work?". The synthesis is easy enough to do in your kitchen, one or two steps with basic equipment and common reagents. My friends have LC/MS and NMR machines available, or, if that didn't work, there are analytics labs that I've sent samples to for a few hundred bucks. I have an AdWords account that I could use for recruiting, or, heck, there's 80,000 people on /r/researchchemicals who take experimental compounds just for fun.

Making placebos is easy, Gwern even has a protocol for doing a placebo-controlled trial on yourself (<https://www.gwern.net/Nootropics#blinding-yourself>), although even that wouldn't be necessary if you mail people the compound (or placebo) and do the randomization at home. And there are already studies that test drug effects, collect information from subjects, etc. entirely online, with some pretty simple software (<https://selfblinding-microdose.org/>) – since psychedelics are still controlled, in this case the

*subjects are totally anonymous). Now, granted, doing it on this low of a budget cuts a bunch of corners. And I'm sure the data would be noisier than in a traditional Phase II trial, so you'd need a larger sample size. And lots of drugs are much harder to synthesize, drugs like Spinraza can't be self-administered, etc., which adds extra complexity. But even so, this really doesn't seem like it *ought* to require \$100 million, a team of lawyers, a lobbying firm and **special guanxi for every trial you do**. Am I totally off base here?*

OK, I can see that some of this query is coming out of the biohacking movement. And I can imagine that from that perspective, a drug-industry-style clinical trial does look like a needlessly complex, expensive tangle. Hear me out, though: there are reasons for doing it the way we do. Some of those reasons are better than others, I'm sure, but the good reasons are in fact really good. I'll take these suggestions in the order they're mentioned in the email, and let me say up front that I'm not trying to make fun of these ideas and I'm not trying to put down the idea of asking them – they're valid questions and deserve good answers, and my thinking is that doing so out here on the blog might be of interest to others.

First off, the part about “the synthesis is easy enough to do in your kitchen”. There are indeed quite a few drugs where the active pharmaceutical ingredient could, in fact, be produced that way, for some values of the verb “produce”. But remdesivir (to pick one that's in the spotlight) is not one of those, unfortunately. Here's [a look at its manufacturing](#), and while some of the difficulties come from having to make zillions of doses of the stuff, some of them scale right back down to that kitchen. How much drug are we looking at? The [NIH/NIAID trial](#) that is starting to report dosed 286 people with the drug, 200mg the

first day and 100mg/day after that for up to ten days. Let's call it an average of 7 days – over that number of patients you need just over 228 grams of remdesivir to meet the exact amount.

Even if this were a drug that you could make in your kitchen, though, you'd better not just make that amount and call it a day. There will be losses in formulation for starters and more losses right down the line. How much do you need? Well, there are [entire companies](#) that [exist](#) to help you calculate just how much “overage” you might need in your clinical trial – if you run out of drug, you've blown the whole trial, and if you make too much you're wasting money. Just the existence of such a business niche is food for thought.

But if you look like you're running out, can't you just make more? That brings up another big problem: [Good Laboratory Practices](#) and the [Current Good Manufacturing Practices](#). Those two are broadly similar, but have [different requirements](#) and are used for different purposes. GLP/CGMP does indeed add bureaucracy, because it adds [a lot](#) of standardized procedures and a ton of documentation. I've never worked under these standards (I do early stage research, not up next to the clinical batches), and I would not enjoy it. It would indeed make things move faster to ditch the stuff. But that's a bad idea. You cannot wing it when you're giving a drug to human beings. Outside of the sheer ethical problems of the “go on, take the drug, it's probably good enough for the likes of you” aspect, there's the need to make sure that everything is absolutely standardized so that your clinical data themselves have a better chance of meaning something. The amount of noise in human clinical data can defy belief if you haven't seen it in person. Every part of the process that can be ironed down flat should be, because the parts you can't iron down (like “does the darn stuff

work in humans”) will completely eat your drug development project if you don’t. The idea of GMP is to eliminate all the risk and variability that can be eliminated by testing and standardization – what’s left, ideally, is the risk of the investigational drug itself and God knows that’s still enough.

I won’t go into the details of CGMP, but it’s safe to say that the requirements cannot be met in a kitchen. Impurity profiles, batch documentation, tracking of all ingredients and starting materials, batch-to-batch variability. . .even just setting the right standards for those things before you start measuring them for real is work. So even if your friends have NMR and LC/MS machines available (I have some friends like that, too, and I wouldn’t have it any other way), who keeps the LC/MS running? Who works out the analytical method and shows that it’s optimized for detecting the most likely impurities as well as the less likely ones that could be most worrisome? What standards do you run, and how often? Can you be sure that the substance you’re measuring didn’t get contaminated along the way? If there’s a difference in one batch, can you trace everything back to find out where this occurred and why, so it doesn’t happen again? And so on.

Keep in mind that you’re watching both the manufacturing of the drug and its formulation, and how exactly do you formulate it? If it’s a solid, how do you make sure that you’ve made the same solid form every time? [Polymorphs](#) can and do appear seemingly out of nowhere and have ruined development timelines again and again. How fine should your particle size be for the substance, and how do you make sure that it’s the same in every pill? (That can make a substantial difference). What “excipients” should be in the pill as well (other compounds to help disperse the active ingredient in the pill, help it break up and

dissolve reproducibly in the gut, and so on? Those can make a big difference, too, and they need to be the same every time or your data could be trashed. An i.v. formulation like Remdesivir is, as you noted, a whole other order of difficulty: it's not just hard to self-dose, it's very hard indeed to manufacture reproducibly under sterile conditions, in a formulation that you're sure the drug is stable in over time, etc.

We can disagree at this point if you feel that CGMP (whose surface I have not penetrated very far) is just a way for the FDA to flex its power and turn itself into a bottleneck, and there are indeed people who think that – but to be honest, not many of those people have ever actually taken a drug into the clinic, which is also something to think about. I could imagine streamlining some of this stuff in an emergency, but I can't imagine just ignoring it, because the underlying issues that the whole system is trying to address are very real. The idea is to protect patients *and* to avoid having the data from the trial getting hashed up, and those two often overlap.

OK, now if you have your drug, it's time to recruit patients. Well, actually, it's time to [design a trial](#) before you start recruiting, and that's [a whole field](#) in itself. You'd figure that studying a single drug in an infectious disease setting might be one of the more simple trial designs, and that's actually true, but that's only simple on the relative scale. First off, what are your endpoints – what are you going to measure to see if the drug works? In the case of a coronavirus trial, do you want to measure viral load, lung function, overall severity of disease, days to discharge, days to death? Probably all of those, sure, but which one do you figure is the most important (primary) endpoint, the one the rest of your design decisions are aimed at answering?

Now, for those decisions. How many doses are you going to check?

How do you know you've picked the most useful ones? What sorts of patients are you going to recruit? In this case, you have several known variables to consider (or to set aside, if you can make a case for that too): severity of disease, time since onset of symptoms, age, gender, body weight, other medications being taken, known co-morbidities such as hypertension, and so on. While you're enrolling patients, you'll need to control for and match up/balance as many of these as you can in each arm of the study. How many arms are you going to have? How many patients do you figure you'll need for each one, in order for the results to be meaningful? Where will you find them? Who will examine them and collect all the initial data, and who will keep track of all the data collection along the way? Recruiting via AdWords or via the users of a subReddit would indeed be faster, but that unfortunately won't cut it. Even a Phase I study (normal individuals, no disease, mostly checking blood levels and looking for any wildly overt safety signals) needs to pick its members carefully to make sure that the curtain can come up smoothly in Phase II. And since you're talking about a "does this actually work" trial, that's Phase II by definition.

The NIAID study ended up recruiting (and dosing, and monitoring) at dozens of clinical centers on several continents just to enroll a properly matched set of 286 treatment patients and 286 control patients, or as properly matched as they could get them under the current conditions. The amount of noise in human clinical data can defy belief if you haven't seen it in person, and the last thing you want to do is go to the time, trouble, and expense of experimenting on human beings but not generating useful data at the end of it. Note that the data coming out of the trial are already being argued about, because the mortality data looked like they might be real, but didn't reach the usual statistical bar that we use. Perhaps the trial should have been larger, more

complex, more expensive – dosing hundreds of people around the world just wasn't enough, in the end?

Another factor: you also have to have people in all these locations whose job it is to dose the patients, unfortunately. Mailing participants the drug and asking them to take it will add a lot of unnecessary scatter to the data, because some people will forget to take it. And some of those will try taking double the amount the next day, and some of them won't be sure if they did the dose yesterday or not. Some of them will have eaten, and some won't, and some will suddenly remember as they're going to bed. The more people that are involved in the trial, the more of this you'll get, and it will likely run right alongside any advantage of that larger sample size and cancel it right out. I'm not making this up: ask any physician to speak candidly about patient compliance and get ready for an earful. Even things like taking a pill at breakfast with orange juice versus grapefruit juice can make a gigantic difference (which is something that for years wasn't even realized). But we're talking Phase II, though, so these patients might be pretty sick, or even full-time hospital patients (as is the case under coronavirus conditions). So you're automatically talking doctors and nurses, etc, and you have to figure out where you will find such people who are willing to add the burden of assisting your clinical trial to their daily duties. What is the distribution of available medical help compared to the distribution of enrollable patients? Who keeps track of all these folks and makes sure that their record keeping is what it should be?

You also need, for both ethical and practical reasons, another set of people who are watching over the whole thing. These folks, the monitoring committee, are the answer to the ancient question of *quis custodiet ipsos custodes*, and they will be off to the side of the whole affair,

empowered to look at the data being generated at set points during the trial to make sure that things aren't going off the rails. Most of the time they will say "carry on", but a reasonable number of trials end because they have said "You have no chance of meeting your endpoints – stop right now, it's futile and it's unethical to keep giving people your drug". Once in a while, they'll even say "Stop right now – your drug is so damn good that it's unethical not to give it to everyone in the trial", but years go by when that does not happen, anywhere in the industry. Either way, you need some outside experts who are not involved in the dosing and day-to-day parts of the trial, and who are not part of a group that has a vested interest in the results. These are the people who got a top Pfizer executive out of the shower on a Sunday morning one time to tell him that a review of the data showed that the company's biggest drug development project ever (a cardiovascular program to raise HDL) was definitely (and most unexpectedly) killing off the treatment group at a slightly higher rate than the controls in a multi-thousand-patient Phase III trial, which is the sort of thing you want to know about at the first opportunity.

The big overarching difficulty with running a human trial the way you're proposing to, though, is that you would open yourself up to criminal prosecution by doing so. The FDA wants to hear about your drug before you give it to human beings: what it is, what you expect it to do and why, the details (GMP!) of how you make it and how you're sure that you've made what you think you've made, what your trial design is and why you chose that, and what you're going to look for as it goes on. This is an IND, an Investigational New Drug application, and it is a pain in the rear. Under the pandemic conditions, the agency is indeed speeding that sort of thing up and cutting corners, but they're not just dropping it, either, for the same reasons that they're not just

dropping CGMP. Again, we can argue about whether what the appropriate background work is to start dosing human subjects with an investigational drug and what the penalties should be if you decide to ignore them – but those are details and degrees. I would very strongly stick to my position that you need *some* sort of written statement, plan, rationale, and documentation that you're going to do this right, though, and that it should be reviewed by an outside team that knows the business as well but isn't at the same time *in* the business. And that's the FDA.

You mention that under your proposal “*the data would be noisier than in a traditional Phase II trial, so you'd need a larger sample size*”, and I definitely agree about the first part. But as mentioned above, going to a larger sample size under these conditions runs a large risk of just generating every more and louder noise. The amount of noise in human clinical data can defy belief if you haven't seen it in person. Even under all the constraints and controls that I've been describing here, 90% of our drugs fail in the clinic: can you imagine the failure rate if we tried to go faster and noisier? Even controlling for everything we can imagine, we barely get anything through.

Well, I knew this was going to be a lengthy answer, and I hope it didn't come across too ranty. The procedures that have grown up around drug development are a sort of [Chesterton's Fence](#): we shouldn't pull them down without considering the reasons why they were put up. I do understand the argument that human lives are at stake when people are dying from a pandemic. But lives are at stake when you test human beings for their response to a new drug, too, and all of the biggest rules and controls mentioned above are, in fact, built on the hospital beds (and the graves) of patients who were tested in more sloppy ways and

for less well-worked-out reasons.