You want to be sure that the effective therapy, which produced no events at all, is matched by the new therapy. But it could also produce no events at all. That would be very reassuring, if you believe the population you've put into the trial are the people who would have had events, had they not been treated.

So you can't put people with a little viral pneumonia in the trial, and then learn anything, but if you put people with the bad pneumonias that led to these bad outcomes in the past, and now show that there are no bad outcomes, that's the whole point. So that's okay, if you're sure that you put sick people into the trial.

ACTING CHAIR TOWNSEND: Dr. Musher?

DR. MUSHER: I wanted to comment further on the point of the bacteriologic cure. Dr. Gitterman is, of course, correct, and I just was moving along too rapidly.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

It's a tiny bit more complicated than, at a ten day period, deciding that you want everybody who has been in your study to go ahead and cough, and provide a specimen. That's the kind of information that I was proposing was not relevant.

If a patient is not doing well clinically, and you repeat a sputum gram stain and culture, and you show lots of organisms, that is strongly supportive of the notion that you've got an inadequate drug. And that is the way that kind of a study is used. And I think I commented yesterday, it is in one of my backup slides, Dr. Finland, in the JB Amberson lecture, pointed out that, in the early days of penicillin, when they were using minuscule doses of penicillin, it would often take five or six days for the sputum to clear of pneumococci. And then he himself said, in the text of that lecture, he said that, as we increase the doses of penicillin, we eradicated the pneumococci much more rapidly.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

And see then what you are left with is the problem, at ten days, if you've got an asymptomatic patient, and he coughs up something, you might still grow your pneumococcus, because the person is colonized by it, or the same thing with a haemophilus.

So it happens to be kind of complicated. I was absolutely not meaning to dismiss the notion of failure to eradicate the organism from a patient who is remaining symptomatic.

ACTING CHAIR TOWNSEND: Thank you. Dr. Follmann?

DR. FOLLMANN: I guess I would like to amplify on Dr. Calhoun's question, and Bob Temple's reaction to it.

So, the FDA's position seems to be that, if we could do a placebo controlled trial, and show that the new drug beat placebo, we're happy. And I just wonder if that's the right way to think about it. And in some ways, it's the way we have to think

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

about it if we are going to be viewing these margins.

For example, some of the margins that Dr. Fleming proposed yesterday were on the order of ten percent margin, and so on. If we look at a relative risk, they might, like, allow a 50 percent increase in the mortality rate for some of them. Now, that doesn't seem reasonable really to me, if you're thinking, the new drug is up to a 50 percent increase in the death rate compared to standard, we are okay with that. The only way I think we can be okay with that view is to say, well, we would have beaten placebo.

And so it boils down to really, are we comfortable with the paradigm that we want to be assured that the new drug would have beaten placebo. And I have heard people talk today about placebo controlled trials are unethical in this situation. They can't be done. And so, why are we using the placebo controlled hypothetical study to justify a

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

non-inferiority margin?

Why don't we look at, you know, what it's doing relative to a comparator? To me, that seems to be the more relevant, because the comparator is, you know, within the armamentarium, it's the relevant question for 2010, isn't it? Not what would have happened in 1930.

DR. TEMPLE: You are looking at it relative to the comparator, but you have to know what it means.

Let's take the scary model. You put people into your trial who have no chance of having a bad outcome. You now study drug A and drug B, and nobody has a bad outcome. But it's not because either drug did anything. It's because it was a population that wasn't going to have a bad outcome. That means you've learned nothing about your new drug. There is no bulb in the colorimeter. There is nothing, there is no what we call assay sensitivity. You couldn't have learned

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

anything from that trial, because nobody would have had the outcome you're worried about. That is the worry.

So, the premise of the non-inferiority design is, look to the past to say, about a defined population, what would have happened in the absence of treatment. So the numbers Tom showed, and our people showed, suggest that maybe 30, for properly defined ill people, the bad outcomes might have been 30 percent, 40 percent, sometimes 80 percent. Now, you could, in some sense, set that as your non-inferiority margin and say, okay, you rule out a 50 percent mortality difference, then you have shown your drug is better than nothing. But nobody believes that would be acceptable. We call that margin, sorry about this, M1. The margin that shows that the drug is better than nothing, which, by the way, is the standard for most placebo controlled trials. If you tested a product at a P of 0.05, what you have shown is that it's better

than nothing. That's all you've shown. Now, there is a point estimate, which people actually believe, even though that hasn't been proved, but all you've really shown is better than nothing.

So in the present case, we define something called M-2. And the reason for that is, the whole reason you can't do a placebo controlled trial is that you don't want to leave people untreated, because you value the treatment so much. So the idea of losing all but the tiniest little bit of it is as offensive to the people who like these as it is to you.

So, you set M-2 as a small fraction, in this case, of what you believe the effect is. So if you believe there is a 40 percent reduction in mortality, you might set M-2 at ten percent. Then you're ruling out a difference of ten percent. Now, somebody could say, I don't want to be ten percent worse, that's no good. But always

remember, we are getting this level of assurance at a very high statistical level. We use, for example, the 0.05 when you try to show that a treatment is different from placebo, but there is also a point estimate that is larger than just better than zero, and the true value of the effect is somewhere along a 95 percent confidence interval. What you have done is make absolutely sure it's better than nothing.

So in the present case, you would make absolutely sure that you haven't lost more than the M-2 that you are willing to lose. But the likelihood that you've lost that much is low, and in fact, the point estimates of these effects are usually going to be on top of each other. So you're pretty reassured, and you are statistically very certain that you haven't lost that much.

But you absolutely are right. In these non-inferiority studies, we don't want to show that you've lost all but the tiniest

little bit, so we always ask that some fraction be preserved. The trouble in many areas, not this one, but in cardiovascular medicine, if you start to show that you have preserved almost all of the effect of something, you end up with a trial of a hundred thousand people, and you can't do it. Tom can elaborate.

DR. FLEMING: Dean, do you want to go first?

DR. FOLLMANN: Let me follow up. So, M-2 is really what's important, not necessarily what the new drug would have beaten placebo with.

Now, we've had a lot of evidence in terms of mortality for the margin M1, and we all agree that it's overwhelming. And I think, you know, to step back a little bit, part of the reason we are here today is because we had margins on clinical failure endpoints, which we felt were not justified by the data, and we wanted data to come up with

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

an M1 margin. And we've gone through that exercise. We found M1 really is too loose to be comfortable with. We want to have an M-2. And so that's sort of where we are at at this point.

Now, clinical failure wasn't examined in the '30s, and so we don't have any data on it. Yet, I would guess it's a fair assumption that, had we had data on clinical failure, we would have seen a whopping effect on clinical failure in the '30s, just as we saw on mortality in the '30s. And so, if we had that data, we would come around and say, M1 is really too big, let's talk about M-2. And so, we could be at the same place, really, on clinical failure talking about an M-2, had we had that data.

DR. TEMPLE: Yet, there is only one thing to remember. You have to know M1 for sure before you can talk about M-2. So for severe, severely ill people, or people over 50, or whatever it was, everyone seems

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

very comfortable that, whether it's mortality, and you could translate that into clinical failure, that's a judgment you guys have to make, that the effect is large. And then you are going to set M-2 as small. So you are very, very sure that M1 is smaller than M-2, and that you will be absolutely sure the drug has some effect, and you are going to put pressure on the study to make sure that you haven't lost more of that effect than you want to do.

But let me tell you, there has been a confusion of this. And sometimes people have said, for situations where they have no idea what M1 is, I'll be happy if I rule out a difference of ten percent. Well, if you don't know that the treatment difference of the active control was at least ten percent, that's completely meaningless. You haven't shown anything.

And I can say this because I did it. We used to do this in oncology. We would

approve drugs that were within ten percent of the standard therapy, or 20 percent of the standard therapy, even if the standard therapy hadn't been shown to have any effect. And we finally figured that out, you know, so then we stopped doing that, with Tom's help, actually.

So, it's the same thing here. And my worry would be that, for the less ill people, it doesn't seem so obvious that we know how to say what the effect size is. So, it might be perfectly true that you would be willing to accept a drug that was within five percent of another drug, but you have to know for sure that the active control had a five percent effect to do that.

DR. FLEMING: So let me just kind of go back to some examples to reinforce what Bob is saying. We have the most evidence here about what the effects of antibiotics are on a mortality endpoint in a more severe patient population. And so to use a hypothetical, but not so far off of what the facts are, we may

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

well have had a population that had 50 percent mortality, and the existing antibiotics, the sulfonamides and penicillins at the time, could reduce that to 20 percent.

Because of the uncertainty about whether that 30 percent was, in fact, reliably estimated, i.e., taking into account the variability in any estimate, and this uncertainty as to whether that effect translates to the context of today, and because of what Dean is talking about as to how much excess mortality are you really going to allow before it's really clinically unacceptable, all of those together led to a margin of ten percent, okay, in the population that would have a 20 percent mortality on the existing therapies.

So now you come along with today's therapy, and you do a non-inferiority trial against that appropriate active comparator, we are using a ten percent margin. So, keeping the same example, if it is a high risk

population that has a 20 percent mortality, you are ruling out in excess of ten percent, going from 20 to 30. The point estimate that will be positive will be in excess of about three percent. You have to be in excess of three percent neutral or better to rule out ten percent.

So, to reassure Dean here, when we do this margin of ten percent, and you compare today's new antibiotic against an appropriate standard comparator, you are not going to declare victory unless you are in excess of mortality of about three percent. That's pretty reassuring when the historical data said it was 30 percent higher.

Now, here's the challenge. We would like to extend this to some lower risk patients. The data aren't inconsistent with the possibility that you have the same relative risk effect. But they are not conclusively establishing that, because it's much harder to do that in small numbers.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

But let's suppose, and the data seem potentially consistent with the scenario where a lower risk patient, untreated, using no specific therapy, had a nine percent mortality, and on the control antibiotics, the penicillin sulfonamides, they had a three percent mortality. Okay, that six percent, again, is estimated with some uncertainty, addressing the need to preserve half the effect, you would have a two percent margin. That is how the math would work out on that. That turns out to be a constant margin on a relative risk scale. That's what we talked about yesterday, using a relative risk margin of 1.67, you could then put patients into the trial that would have preferably the 15, 20, 25, 30 percent mortality, but at least 15 percent mortality using the ten percent margin, but it would allow you to put in some patients at lower risk, where you are preserving or ruling out the relative risk of 1.67.

And that was the approach that we were talking about yesterday that would allow the greatest flexibility to enter patients into trials, hoping, of course, that you would be able to put in patients at true high risk, because that's where we have the most confidence about the validity of the counts of the assumption. Of course, then, the payoff to the sponsor is that such trials would only take about 600 people. But if you ended up putting patients in that had half that death rate, half of a 15 to 17 percent death rate, about a six or seven percent, it's about a thousand person trial.

And so this is the concept. So coming back to defend or to amplify what Bob Temple is saying, if we do a non-inferiority trial with a ten percent margin, the concept behind that is, yes, we realize we have highly effective antibiotic therapies. We don't want to do a placebo control trial because we can't deprive patients of something that is really

effective. If we're going to approve something new, then the standard is, yes, we want it to be better than placebo, but we certainly also want to know we're not meaningfully losing or putting patients at risk of meaningfully losing the benefits of existing therapies. And that is the concept that has led to the non-inferiority margin.

It is, in fact, preserving a substantial or an important fraction of the effect of the existing therapy, but it is also allowing for an approach that is an achievable, scientifically achievable and feasible design, in terms of the size of the trial. Of course, as Dr. Temple pointed out, there needs to be assay sensitivity. If you put patients in with viral pneumonia, then non-mortality isn't persuasive, or no difference in mortality, I should say, with the active comparator isn't persuasive.

So, coming back to what Dr. Musher is saying, it will be important, in my words

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

for what he's saying, is assay sensitivity. It will be important to have a substantial fraction of the population with microbiological confirmation to help on this assay sensitivity.

ACTING CHAIR TOWNSEND: Dr. Wiedermann.

DR. WIEDERMANN: This is, I guess, partly related to these discussions, but sort of away from mortality, and getting to the other endpoints, or potential endpoints or outcomes that might be more useful for the patients with milder disease.

And in the background information we were sent, there was a little bit of a discussion of hierarchical primary endpoints, but I haven't heard anything about that as a technique in the presentations here. We're talking about a lot of slippery slopes in study design, and I would be interested to hear what any of the experts think about potential slippery slopes when you invoke that

kind of measurement.

DR. COX: I can try and make some comments on that, and maybe others can fill in. The issue with the hierarchal endpoint is, I think, when you are looking at a composite to make sure that there isn't, you know, in the overall endpoints, if two, if the endpoints look the same across study arms, the question is, is there a difference in one of the components. And for instance, is there a difference in one of the important components, like mortality?

So, looking at things hierarchically, you know, to make sure there isn't a problem, or a difference in mortality before moving on to a composite endpoint, is to help to protect against that, because you wouldn't want to have a situation where the two drugs looked the same on a composite endpoint, but in fact, there is a mortality difference that is hidden within.

DR. MUSHER: So you still have to

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

go ahead and deal with the others. I mean, that is exactly the point. Just go beyond mortality. I can come back and ask you. I would still say to you, as a distinguished group of statisticians, guys, you pretend that you have no data from the 1930s. I think there is a totally different way to look at it. And you could design the thing differently, and you could show whether drug B is as good as, better than, or worse than drug A.

And I think there are ways to do it, and you don't have to refer to some baseline data from 60 years ago that you are thinking kind of looks like thus and so. I happen to think, my intuition tells me that, if you really put your mind to it, you could find something better. But leave that for now. We can come back to it. What about all the other parameters in the more mild cases of pneumonia? Rates of defervescence, rates to go on, time to going back to work. See, I am

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

looking at you. I am saying, give me some way to analyze it statistically. That is the problem, because that is what we are treating. I want to be able to compare how long it took them to go back to work, when they got out of bed, when they felt better, when they defervesced. You tell me how to do it.

DR. TEMPLE: It's not a statistical question. You need to know what the effect of the control agent is on those things.

So, I throw it back to you. If the past experience, or any place else you can get it, tells you what the control drug does on fever resolution, then you can set a margin, and then you can do all the comparisons.

DR. MUSHER: Okay. Control --

DR. TEMPLE: But if you don't know what the control drug did, then you don't know how to interpret similarity of the control and the test drug. You have to have --

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

DR. MUSHER: When you say control drug, you mean no drug. You mean placebo. What do you mean, control drug?

DR. TEMPLE: Well, no. You are talking now about a trial in which you are comparing one drug with another.

DR. MUSHER: That is correct. So you mean the B with A. Okay, I'm sorry.

DR. TEMPLE: Yes. And what you are always looking at is the difference, the difference you see between the control drug and the test drug.

DR. MUSHER: Correct.

DR. TEMPLE: What we call C minus T. And you are always trying to show that C minus T is less than some amount that would trouble you. And what would trouble you is if the difference between the two accounts for the entire effect of the control. So, if the test drug is worse by the whole effect of the control drug, well, you have lost the whole effect. So, that's bad. But you can't even

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

start that discussion until you know what the effect of the control drug is. And on mortality, or on mortality equivalent, maybe things going down the tube might be the same, we have pretty decent numbers. And Tom and others, and our people have shown those. If there is similar data for defervescence, then you could start doing that. But without that, you don't have a basis for setting a margin. That's the trouble.

DR. MUSHER: And I just don't accept that presupposition.

ACTING CHAIR TOWNSEND: Dr. Musher, if you can hold that thought.

DR. MUSHER: I'm sorry.

DR. TEMPLE: There is no presupposition. I'm just saying, if you can find what the margin is, those are perfectly fine endpoints. Nobody doubts that. But you have to have a basis for the endpoint. You have to know what the control drug does.

DR. MUSHER: You let him go, and

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

you wouldn't let me go, but that's okay.

(Laughter.)

ACTING CHAIR TOWNSEND: I was going to cut him off in a second, too.

Dr. Wiedermann has a follow-up question.

DR. WIEDERMANN: Or maybe an attempt to rephrase. So, say we design a non-inferiority trial that everyone is happy with, and the mortality rates are not meaningfully non-inferior, but we look at duration of fever, or something else, and the new Gorillacillin patients have longer duration of fevers. Are we now into a range of violating multiple comparison rules, or what do you do with that in a non-inferiority trial?

DR. TEMPLE: It's, as is pointed out in some of those documents, if you design a non-inferiority trial and win, you know, you beat the control, that's okay. That's interpretable. That's like setting a margin of 0.05, and winning at 0.001. We let you

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

claim it.

So, there's nothing that says, if the most important thing is shown to be non-inferior, for example, you rule out the margin you are worried about on survival, if you had pre-specified a secondary endpoint about defervescence, and you win, that's fine. Now, you won't get to that unless you show non-inferiority first, but probably you will be able to do that. But you can get to it. We have a lot of ways of dealing with additional endpoints. You do worry about having too many of them. They should be specified, but there is nothing that says a non-inferiority study can't lead to a claim of superiority for a secondary endpoint, such as time to defervescence.

Moreover, even if you didn't anticipate it, and it looked really worse on some of those endpoints, I think people would be very nervous, although statistically, it would be hard to say exactly how to figure

that. But you could also lose unexpectedly, too, I think.

ACTING CHAIR TOWNSEND: Dr. Fleming, did you have anything?

DR. FLEMING: Well, just to respond to Bud's question, the question you are asking, Bud, is relevant in any trial, whether it's a non-inferiority or superiority.

The primary endpoint of any trial should reflect substantially what is critically important, or very important to patients in a given setting. And obviously, there can be many different outcomes that are very important. I've talked about a hierarchy where mortality would be the most compelling endpoint. Major complications, breakthrough infections, would be next. Symptoms, such as cough dyspnea, chest pain, fatigue, other measures that, when Donna Lamping, in her PRO activities in these areas were validated by patients from a content validity perspective to be important to them, kind of a hierarchy.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

So ideally, you would be doing a trial on the measures that are the most persuasive, but also measures that you think are likely to be impacted by the intervention.

Now, if it's a non-inferiority trial, or a superiority trial, if your trial is assessing effects on those measures, you will always look at what the effects are on other measures, and on safety issues, and quality of study conduct, et cetera, you are going to look at totality of the data. You are going to factor those in. So, if you are doing a non-inferiority trial, and the results seem to be just the same thing in a superiority trial, the results seem to be consistent, but marginally with what you would consider a successful trial, you will be influenced in the totality of the data by what the rest of the measures show. So you do factor in all of those, but the ideal is to be able to, first and foremost, define what the principal interest is, what the principal

measure of importance to patients would be in this setting, reliably estimate and test that, either in non-inferiority or superiority, and then factor in these other measures that are collected in the totality of the data.

ACTING CHAIR TOWNSEND: Dr. Rex.

DR. REX: I recognize we are getting close to the end of this slot, so there are a lot of things to comment on, but there is one thing I particularly want to say right now. It has to do with this whole mortality debate we have been having.

Yesterday, Dr. Fleming and Dr. Powers gave us a very nice demonstration of how there were very strong mortality benefits in people with strong evidence of a bacterial pneumonia. A really very attractive summary. And we heard from Dr. Musher the importance of knowing that you actually have the disease in your study.

And so, if you think about what those people had, they had a very strong case

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

for a bacterial pneumonia. And let me just read to you from one paper how they diagnose. Gaisford. The diagnosis of lobar pneumonia has been based throughout on the same essentials: sudden onset of rigor or vomiting, fever, pain in the side of the chest, cough, often rusty sputum, physical signs of consolidation. So, a very strong syndrome that every physician recognizes.

Now, the mortalities that were seen in the '30s and '40s were quite striking. And even with treatments, you still had a 20 percent mortality. But let me read you one case of someone who died, and then point out why mortality now is different than mortality then.

A 31 year old man comes in, and is proven to have pneumococcal pneumonia. He starts on a drug. He actually defervesces over a period of three or four days, and then he starts to have a little nausea. So they stop the drug, and then the fever comes back.

And then they put him back on the drug, but they say the patient was very toxic, and the condition became worse. He died 21 days later.

At autopsy, it turned out that this young man, 31-year-old, had a chest full of pus. He had an empyema. Now, in the modern era, he either wouldn't have had it, or we would have found it and drained it. Okay, so this guy would not have died in 2008, because we would have said, your fever has come back, do another chest x-ray. Dr. Musher would have noticed the infiltrate. That would have led to a tap of the chest. We would have found the pus. We would have drained it. He wouldn't have died.

So the mortalities that we saw in the '30s and '40s, some of them are going to go away. And so this is why it is actually very important that we not focus just on mortality. And Dr. Temple said it well. It could well be that, in a very high risk group,

I drive the mortality, not down to 20 percent, maybe even down to ten percent, maybe even down to five percent, because I am working really hard. I am putting people in the ICU. I am not letting them die. They are going on pressers. As a matter of fact, anybody who winds up in a modern ICU on a vent was dead in 1939, because there weren't ICUs. There weren't vents in 1939. They all died.

So, it is an important contrast to the modern era. So if we focus just on mortality, we are going to very frustrated, because we are going to end up with smaller mortalities.

And then the statisticians can see what's coming. If now the mortality rate is not ten percent, but five percent, and now I'm forced to exclude a margin around that, and the numbers become very large, if I believe that the only thing I know is this low mortality.

So, it's actually very important to

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

have a clinical combination of the clinical event of response without complications, for which you also have to not die. So that is the theme I want to point out. It can't be just mortality. But that mortality thing, that huge mortality benefit that Dr. Fleming showed us yesterday, is there in everybody who doesn't die. Okay? I mean, it sounds like a silly thing to say, but we always have kind of death the same way. And I like that phrase from yesterday. But these days, we don't let people die for the same easy reasons. We would not let a 31-year-old die of empyema in 2008. At least, not without trying real hard.

So, that's the comment I want to make about not being too hung up on proving, at a high level of statistics, that the mortalities of three percent exclude plus or minus one and a half percent. If that was the only data we had, if we had no biological prior probability, if we knew nothing about the drug, if we had never seen this disease

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

before, then you would be absolutely right. But it's actually not the only thing we know. We know that the dog that didn't bark, the patients that didn't die, the empyemas that didn't happen, is actually a very real thing. So, I'll shut up.

ACTING CHAIR TOWNSEND: Thank you, Dr. Rex. Dr. Venitz.

DR. VENITZ: I want to follow up on the discussion that we had about non-inferiority margins, and I think what we haven't discussed yet is the experience that Dr. Nambiar shared with us that apparently, currently since 2000, you had multiple registration trials designed as non-inferiority trials with non-inferiority margins of 10 to 15 percent.

To me, that implies that the Agency, number one, believes there's assay sensitivity, meaning, if you had used clinical cure 50, 60 years ago, you would have seen a significant treatment effect, and number two,

that a 10 or 15 percent margin would preserve sufficient treatment effect to conclude non-inferiority. Is that correct?

DR. COX: I think one of the reasons we're here today and talking about this is because we are taking, you know, a look at clinical trial designs for community-acquired pneumonia, and you can tell we are asking the question now of what is an informative study. So, you know, those studies were done at a time where we were selecting margins, in part, based upon convention or sample size issues. Now what we are here talking about today is non-inferiority margins selected upon, or based upon data, and understanding what the available treatment effect is. So that is really the question that we are talking about here today.

DR. VENITZ: No, I understand, but my point is then you do have some experience with things other than mortality, which is

what the whole discussion has been, you know, comparing retrospectively to what happened decades ago, whether we have assay sensitivity, what the margins should be. You have experience with endpoints other than mortality, based on eight years worth of registration trials.

DR. COX: Right.

DR. VENITZ: Your question is to us whether we think that is acceptable or not.

DR. COX: And you are correct. Those studies did look at whether the patient had clinically responded, whether there was a need for further antibiotic therapy, and also included in there, you know, patients who died would be considered failures. But the reason that we are asking the question here today about the appropriate endpoint is to get to this issue of the treatment effect, and what is the appropriate endpoint based upon what we know from information that's out there on treatment effects. So, you know --

DR. TEMPLE: What population?

DR. COX: Yes, and what is the correct population. So we're really trying to get to an understanding of what the treatment effect is in this group in order to be able to appropriately pick a margin, or set a margin, when appropriate to do so.

DR. VENITZ: But the 10 or 15 percent that was chosen presumably then already reflects some expectation that the treatment effect on clinically cure is similar to what it supposedly is on mortality. Is it not?

DR. COX: I don't think so. I think that really reflects convention at that point in time, you know, selection of a margin. And at one point, we were selecting margins largely based upon what the expected cure rate would be, and then what sample size that would be that would fall in the range of two to three hundred patients per treatment arm.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

ACTING CHAIR TOWNSEND: Dr. Calhoun.

DR. CALHOUN: I'd like to follow up on something that Dr. Rex and Dr. Musher were talking about. I think this really is maybe the crux of the matter. It's clear that antibiotics work in pneumonia. I think the evidence is pretty compelling, and that underlies Dr. Musher's assertion that a placebo controlled trial is unethical, because we know that antibiotics work.

The evidence that has been presented to us so far has really focused on mortality, but that's simply because mortality was the thing that was measured. It doesn't mean that there were not also effects on some of the clinical outcomes that Dr. Musher was talking about. So the fact of the matter is that a non-inferiority trial, using some of those other measures, would not necessarily need to have already established what the effect of standard antibiotic therapy is on

resolution of fever, for example. It would not necessarily have to demonstrate, we wouldn't necessarily need to know that a priori.

I think, not to speak for Dr. Musher, but I think the point is that there are outcomes other than mortality that would necessarily have had to improve with treatment. And then showing that a new drug is not worse than, or perhaps, if it shows superiority, better yet. But I think if we are completely focused on having to demonstrate your M1 first, and then pick a small piece of that as an M2, we are going to be stuck in this mortality hole.

ACTING CHAIR TOWNSEND: Dr. Fleming.

DR. FLEMING: The reality is, we have spoken greatly about mortality because there, it's on that endpoint that we have substantial evidence to be able assess what the effect is of standard interventions. In

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

a non-inferiority trial, if you want to, in essence, show you are similar to the active comparator, and to be able to conclude that you have reliable evidence of benefit, we need to have valid documentation of what the effect of the active comparator is on that endpoint. We've had a lot of discussion in lesser severe, in mild cases, that showing non-inferiority on mortality would require very large numbers, and that there are other measures that are important, and that would be more frequently occurring.

So in a mild patient's resolution of symptoms, cough, dyspnea, chest pain, fatigue, et cetera, would be, in fact, also important measures. Of course, hoping or needing some reassurance that, when you achieve those benefits, it's not at a negative or inferior effect on mortality, but there could, in fact, be more attention given to those other measures in less severe patients.

But to do so, we would have to

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

follow the traditional approach. We would have to, first of all, identify what are the measures that are most clinically relevant to the patients at this point, and what would be an instrument that would be reliably able to assess what the effects are on those measures. And you get into issues of content validity, criterion validity, construct validity.

And, by the way, a lot of work is being done on that, Lamping, the Dial-Lamping article that Dr. Musher referred to in Chest is one such example. And, interestingly, when patients are asked from a content validity what are those measures, you do see things like cough, dyspnea, chest pain, fatigue. You don't see fever. Fever is, appropriately, a measure physicians use in guiding management. That doesn't mean it's the predominant measure that patients use to characterize what it is that they feel, or how they want to improve. That's what content and criterion validity is all about. And when that work has been done,

you see things showing up like cough, dyspnea, chest pain, and fatigue.

Once you have those measures, once you have a PRO, this is not unique to this setting. The science of PRO is something that has been pursued across all clinical areas, and it is a very difficult science. What we're trying to do is easily justified. We're trying to establish benefits on what matter to patients from a symptom perspective. That's where PROs come in. But the devil is in the details to be able to do that in a rigorous and reliable way. So you establish the validity of the PROs.

Once that's done, you can now use those measures, but you can't use them in an non-inferiority trial unless you have the ability to determine, what was the effect of the active comparator on those measures? We can use those in superiority. And so it's very appropriate to build on the work of Lamping and others with PROs on these measures

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

that will give us valid assessments of the effects on symptoms, and to compare to an active comparator showing non-inferiority on mortality, or superiority on the PROs.

ACTING CHAIR TOWNSEND: Dr. Musher.

DR. MUSHER: Listen guys, you are asking. Dr. Cox, Dr. Temple, Dr. Fleming, you are asking, but you are simply not listening.

If your non-inferiority model, the way you're defining it, doesn't work, then I am proposing to you we figure out some totally different way. I hate to use a cliche, but then think out of the box. I'm not a statistician, but there are ways to compare drugs, and you don't go back to makeup data from the pre-antibiotic era to do it.

If you are comparing drug B to drug A, I don't know, look at the -- if the patient is treated with drug B, don't defervesce for 7.4 days on average, and the patients on drug A defervescing in 3.2 days, you do a statistical analysis, and they are

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

significantly different based on the size of the group you have, and then you have got a difference. And if it's 7.3, if you think it's worse to have fever for 7.3 days then it is for 3.2 days, then that is worse. And that is significantly worse.

Figure out a way to do it. But to say all you can use is mortality, the outpatient mortality, it has been shown, guys, the mortality is 0.3 percent. You have got to 15,000 patients in your studies. Nobody is going to do it. Just drop it, and think of something else. You are not listening to the answers.

DR. TEMPLE: Yes, who is not listening I think could be debated. Look --

DR. MUSHER: Bob --

DR. TEMPLE: I'm sorry, but --

DR. MUSHER: -- you are a statistician, and I am a doctor.

DR. TEMPLE: I am most assuredly not a statistician.

(Laughter.)

DR. MUSHER: Well, I am most assuredly a doctor, and I am the one taking care of the patient.

DR. TEMPLE: Me, too.

DR. MUSHER: Well, but then you have got, and we have got to say, this is the way we evaluate patients, please design for us a statistical model that will help us objectify our data. You can't keep saying, I got to have a comparison with people who weren't treated in the 1930s. I haven't got the data from the 1930s, and you haven't got it, and it doesn't exist.

ACTING CHAIR TOWNSEND: Dr. Musher, let Dr. Temple respond.

DR. TEMPLE: Let's be clear on what the problem is. It's our obligation under law to reach a conclusion that a drug will do what it's cracked up to do. And you have an interest in that, too. You don't want an ineffective antibiotic, either.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

If somebody shows that a new drug causes more rapid defervescence than an old drug, that's a trial, that's a superiority trial. Nobody has any trouble with that. That's the old way.

DR. MUSHER: That's a label.

DR. TEMPLE: We're very happy with that.

DR. MUSHER: I don't know what it means, that's a superiority trial.

DR. TEMPLE: I mean --

DR. MUSHER: I don't know what that means.

DR. TEMPLE: Sorry. That is an interpretable result. Everybody would be very happy with that result. That's no problem.

DR. MUSHER: Thank you.

DR. TEMPLE: The problem is when what your goal is is to show that a new drug works, because it is not inferior to another drug. And to do that, you have to know what the effect of the drug you are comparing it is

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

to.

DR. MUSHER: Well, that is what you keep saying, and I just don't think so.

DR. TEMPLE: Well, I mean, it is logically necessary. You don't have to be a statistician to understand it, because if you did, I wouldn't be able to understand it. It's fairly straight forward.

Now, it's also true that people have not always recognized this. People used to do trials of A versus B, show no significant difference, and declare that A works. That's not correct. It's illogical. It's very common practice, but it isn't true. Before you can reach that conclusion, you have to know what the effect of the control was.

Now, you don't want to be over rigid about it. We were allowed to use some intuition. We recognize this is a sort of qualitative thing, but you do have to know that, or your trial is not interpretable. We call it assay sensitivity. I like to think of

whether there is a bulb in the colorimeter. This has been recognized by people for 30 years. Lou Lasagna used to write about this.

DR. MUSHER: But we do think that the antibiotics --

DR. TEMPLE: So it's not something we just --

DR. MUSHER: We think that the antibiotics work, and that's not what the question is. So the question is whether a new antibiotic works as well as, or is not worse than the old one. You don't have to go back to the controls every time.

DR. TEMPLE: Now, before you leave that --

ACTING CHAIR TOWNSEND: Gentlemen, is it possible to have this discussion maybe later? There are a couple more questions. We are almost out of time. Is that okay? Okay.

Dr. Patterson.

DR. PATTERSON: Well, I appreciate Dr. Temple and Dr. Cox seeking the input and

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

appropriate answers here, and I also appreciate the statistician's viewpoint, and the numbers in black and white. I guess, you know, in clinical medicine, we have a sense that things are not always black and white, and so I think that's why we are all here.

I did want to comment on the aspect of fever. I think fever is a valid clinical sign to interpret in terms of response. It is a very prominent patient complaint, and it is a very valid monitor by the patient of how they are doing. Patients are quite aware of when they are having high fevers.

I don't buy this argument about serum therapy in the 1930s that caused febrile reactions, and people having lower mortality compared to untreated. I don't buy that as a valid argument for why we can't use fever because, number one, serum was a treatment, and fever itself was a treatment in the pre-antibiotic era. We used to use, or physicians used to use induction of fever to treat things

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

like syphilis, and other bacterial infections. Fortunately, we have better therapies these days.

But all that aside, I just think that we should not toss out fever as a very valid response to therapy.

ACTING CHAIR TOWNSEND: Thank you. Dr. Rex?

DR. REX: I am going to follow up right where Dr. Patterson left off about fever.

Fever is really both overanalyzed and underappreciated. The overanalyzed bit is to say that it's just fever that we are treating. Well, it's actually not. Fever is -- you can't get better if your fever doesn't go away. I mean, at the end of the day, nobody goes away improved still with a fever.

Now, of course, there are exceptions. They might have a drug fever. Their pneumonia might have gotten better.

But if you look at the big pattern

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

of things, fever is merely the most prominent and easily measured for a physician of the complex of symptoms that represents community-acquired pneumonia. And it's one that we write down. It's one that everybody in all the old papers would show a little graph of. It would also show respiratory rate and heart rate, but fever has, as a quality of standing out from a physician's perspective, in that we can graph it in the interim. Every morning we'll say, and Mrs. Smith's temperature over the last 24 hours, her Tmax was --

So, it is important to recognize the value of it. And then let me point out that it is quite underappreciated. We have talked about the idea of time to fever response as if it's something we've never studied before. Actually, every study we have ever done has incorporated a time to fever response.

You say, what? Where did this come from? Where is this? It's right there in the

way that you get to the end of the case report form. Think about a case report form, the page that says final outcome, improved, yes or no. Did we get to that final outcome by starting the patient on drug, walking away for two weeks, coming back and saying, hi, Mrs. Smith, haven't seen you for two weeks, how did you do? No. We actually typically will see the patient, or talk to the patient every day, sometimes even twice, or three, or four times a day, depending on how sick they are. But you will have lots of data along the way, and the patient can fail at any time along the way.

And indeed, in every study I have ever done, somewhere around day three or four, as a physician, if this person is not getting better, I am pretty unhappy, and I may very well wash them out of the trial. And indeed, that is often part of the way the protocol is written. If the patient is not better by day three or four, you are out of here.

And so the fact that we say that the success is defined at end of therapy plus seven days does not mean that there wasn't a time to event measure in there. You actually could fail at any point along the way, and people were failing at points along the way earlier on. So we have always incorporated, at least at a basic level, a test that, somewhere around day three, four, or five, in most infections, if you are not getting better, pretty much every physician is going to wash you out. That's no good. By day three or four, you better be showing me something that says, I'm improving. And there are lots of ways for that to occur.

And Dr. Musher and I might see the same patient, and we might disagree on whether or not today Mrs. Smith has passed over that boundary. But between today, tomorrow, and the next day, we are probably going to agree that she is either headed up, or headed down. So it may not be as precise as we might wish,

and a PRO might be more precise, but it certainly is an aggregate accurate. And I think that's the theme to be picked up on here, is we've got more time to resolution data than you've actually realized. It's built into all of these studies.

So, the idea of fever. Think of it not by itself. And Dr. Fleming, it is not the fever that we are going after, and the fact that fever is not the reason that we can't do a placebo study. If fever was the disease, if that was the entire disease, then you're right, we could do a placebo study. But it's not. The concern is about the disease that can rapidly progress. Even a young person with pneumococcus can go downhill in a big hurry.

And we've got data. Go back to like the Agranat. I love the Agranat report from 1939, because there is this beautiful demonstration in that report of the effect of antibiotics on fever by cohort. I mean, it's

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

not perfectly randomized, but clearly, it moves the peak of the resolution from about a week, to about three days. Very encouraging demonstrations. And is there something that we can do with that statistically? That's really what I'm asking.

So, that's my comment on fever, and I thank Dr. Patterson for bringing up the subject.

ACTING CHAIR TOWNSEND: Dr. Fleming has the last comment, and I think that will be it.

DR. FLEMING: The Agranat data does address this issue. And the issue, what is not at issue here is whether care givers use fever as a guide to assessing the patient's condition, and as a guide to use of interventions. That's a separate issue from what the actual outcome is, and what the patient values as an actual outcome.

And this is a phenomenon that's true across all clinical areas. If I had more

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

time, I would give many examples. But in that Agranat study, he did talk about how fever resolved more quickly. But he also went on to say secondary pyrexia was very common, and the ultimate average time in reduction wasn't impressive.

Then Flippin talked about, most clinical reports have stressed the frequency of initiation of drug treatment, and it's followed by, within 24 to 36 hours of a drop in temperature. So he was referring to the recognition of the drop in temperature. It goes on to say, resolution of pneumonia then follows within a variable period of days, although we cannot say that it was hastened or retarded by the fall in temperature.

The evidence that that fall in temperature is causally leading to the benefits that patients care about is the weak point. And when we look at what patients care about from a symptoms perspective, and patients are asked, which is what the Lamping

survey was done from a content and criterion validity, fever isn't on the list of what was listed there.

So the distinction is, this isn't challenging the appropriateness of clinicians monitoring fever, and a recognition that this is one of the measures that clinicians use in guiding interventions. That is a separate issue, though, from what is it that patients ultimately are trying to achieve?

ACTING CHAIR TOWNSEND: Thank you, Dr. Fleming. Real quick.

DR. REX: Very quickly. I don't disagree with you that, with the older, less effective, and somewhat toxic agents, the fever patterns, double pyrexias, late empyemas, all of that could certainly occur, but it's kind of like my discussion earlier about mortality. In the current era, things are cleaner. The drugs, actually the drug fevers are less common. We weren't dealing with the sulfa that wipes out your bone

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

marrow. So, there is an improvement that has occurred along the way. And I am not saying -- fever is not the disease. I agree with you 100 percent. It is not the disease. It is not the disease. It is not the disease.

But with modern, relatively clean drugs, if your fever doesn't ago away, you don't ultimately get better. And so they are very closely linked. And that is my point.

DR. FLEMING: But the Lamping experience is from current day, where people are asked current day, what is it that they are looking to achieve.

ACTING CHAIR TOWNSEND: Excellent questions. We will have lots more time to discuss this after the break. And then this afternoon, clearly, there are many things to resolve. We'll take a break for 15 minutes. Be back here at 10:35.

Panel members, there is going to be a list going around for anybody who needs a taxi.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

(Whereupon, the meeting went off the record at 10:22 a.m. and went back on the record at 10:45 a.m.)

ACTING CHAIR TOWNSEND: It has become apparent, I think, to most of the panel members and probably members of the audience, that we have a lot of questions that need to be answered, many of which we are actually planning to answer in a formal question session later this afternoon. In the interest of trying to get that session accomplishing its goals, we are going to move things up and shorten things a bit on the schedule before then, so we can start that part earlier.

So, what we are going to do is we are going to run the remainder of this question clarification session until 11:15. We are then going to have the open public hearing from 11:15 to 12:15. And then we will have lunch from 12:15 to 1:00. So we are going to shorten the lunch session a bit. There is vending machines upstairs. And then

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

we will start the advisory committee question session at 1:00. So, we are going to give ourselves an extra hour to try to get through some of these questions. Okay?

Dr. Rex, you have a question.

DR. REX: I wanted to, Dr. Fleming and I were having a conversation just before the break about fever in pneumonia PRO efforts.

There have been five studies that I was able to find in the literature where somebody tried to put together a PRO-ish tool for community-acquired pneumonia. Lamping, and I have put up a slide. Lamping, Metlay, I can't pronounce this person's name, Dean and Marrie. I'm sorry, I clipped this slide out really quickly so it doesn't actually have the cites, but if anybody wants them, I can provide them readily. They are easy to find.

So these are the five where people have attempted to develop a scheme. And across the top are the symptoms. The symptoms

that got into their scheme. So, if you look part way across, you will see sweating and chills in the Lamping scheme. Metlay and Dean didn't use sweating and chills. They actually called it fever, whatever that was. Again, this is patient-based stuff. These are the symptoms, these are the words they used for the symptoms. So, in fact, the Lamping CAP-Sym score does include an element that represents fever.

Now, I should point out that sweating and chills don't necessarily equal fever. There are times when I sweat, like right now, that have nothing to do, I hope, with having a fever. But in these general scheme of things, in this setting, I will argue that most of the time sweating and chills reflect actually the physiologic abnormality of having a significant variation in your body temperature, i.e., a fever.

So, most of the time, though I will admit the caveat. So that is the comment.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

That is their data.

ACTING CHAIR TOWNSEND: Dr. Musher.

DR. MUSHER: I would like to add just to that and then I am going to make some other comments afterwards.

Let it be noted that Dr. Lamping is a psychologist. Doctors Metlay and Dean are practicing physicians. My feeling remains that there is too much emphasis on death or what patients report.

And doctors are sitting here and saying, guys, there is a whole bunch of things that we look at. This is what we look at every day when we take care of patients and I am asking you to tell me how you would evaluate the rapidity of the response in six or seven parameters.

And every time I ask, I am told, well you have got to know what it was like in the pre-antibiotic era and I don't think the data are available. And if I don't know what it was like in the pre-antibiotic era, then we

have got to do something about it. You can't just dismiss it or say well, we have got to talk about death or now there is some data on defervescence. There have got to be better ways. It was already stated. And guys, it is absolutely true.

Defervescence in the pre-antibiotic era is just a totally different thing. It actually biases it away from what my point is. But I will tell you, that if a complication appeared, which it did because the original infection didn't get treated, the fever would go on for weeks and months until a patient died of wasting away or empyema. That is what happened. So the fever never went away.

We don't have that nowadays. The rate of empyema used to be somewhere between 10 and 15 percent pneumococcal pneumonia. Now it is two percent. And the reason is we have got antibiotics.

I am telling you, this, ladies and gentlemen, is what we as doctors use to

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

evaluate patients. And you can be a nihilist and you can say, if you don't think antibiotic A is any good, it is not doing anything, then your comparison with drug B is not valid. You have got to go compare with controls.

Well, I am not a nihilist. I am just an ordinary doctor and drug A is working. So along comes drug B and I just want to know that drug B is not significantly worse. I think that there are ways to do that without going back to the 1930s and getting data. And I think that the statisticians can help me. I have already said just tell me if there is a statistically significant difference for five parameters between the response in drug B and drug A. There have got to be ways to do it.

ACTING CHAIR TOWNSEND: Dr. Cox.

DR. COX: Dr. Townsend. If it would be helpful, I mean, I can go and show a couple of slides on this that will try and graphically illustrate this point. Is that --

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

ACTING CHAIR TOWNSEND: That would be great.

DR. COX: -- okay?

ACTING CHAIR TOWNSEND: Thank you.

DR. WHITNEY: While he is setting that up, I just wondered if I could follow up on the slide that was just up there. Were the blue squares the things that were statistically significant?

DR. MUSHER: No, they were what were looked at.

DR. WHITNEY: What was the blue versus the white?

DR. REX: No, the blue is what each system used.

DR. MUSHER: What they looked at.

DR. REX: So, Lamping has 18, you know, Lamping has this series of 12 to 18 things. But the white just means that it didn't appear in the other person's system.

And I thought this chart was interesting because I was sort of interested

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

in what appeared in everybody's scheme, or at least in three or four of the schemes. And so that is actually, I had a health economist pull this together for me so that I could see. And it is interesting how strong fatigue is, how strong dyspnea is, how strong coughing is. But in addition, this feverish thing appears in three of the five schemes.

DR. WHITNEY: Well, I am still confused. Does that mean that the investigators just asked if the patients had those symptoms and not how many patients actually reported having?

DR. REX: No. Okay, so thank you for asking. So let me just talk about a couple of them and give you an idea of how they occurred.

So, what did Lamping do? Lamping interviewed 33 CAP patients from the UK and France and recorded verbatim the conversations and then had somebody go through and kind of pick out the things that recurred. And that

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

is how they cooked down to their list of 12 to 18 items. Okay? So they did some psychometric work on the text.

Metlay looked at the symptoms from the original Fine patient outcome research team, the original Fine PORT study. And the experts selected questions based on symptoms in two previous CAP cohort studies. They created a questionnaire. The questionnaire was then modified based on another little investigation with people. So, in each case, the groups are basically, they are starting with some data from patients, kind of picking out some things based on a little bit of intuition, a little bit of summarizing of patient comments and creating a scheme represented by the blue boxes that they then thought was interesting.

So, it is not, you know, I don't know how else to say it. They were proceeding kind of in combination, they were proceeding iteratively to come up with a list of symptoms

that they thought were relevant to community-acquired pneumonia.

Moussaoui questionnaire developed using text books, literature, and expert opinions. Dean used a symptom assessment developed from the literature. So, does that answer your question, how they got to it?

DR. WHITNEY: So, in other words, these are what patients said they had, or doctors picked up from the patients at the onset of pneumonia, and therefore we might be able to use them to follow the resolution of pneumonia.

DR. REX: Right. That was kind of the idea.

And Lamping was probably the best one because they found, they got 33 CAP patients and had some trained interviewers sit down with them and talk to him. Talk to me about your experience of having had community-acquired pneumonia, that is kind of how I read the paper. And you know, you spend an hour

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

talking about what did it feel like, what bothered you? And then, as you were getting better, what still bothered you?

And they took all these transcripts and analyzed them. And out of that, distilled a set of symptoms. And they have actually done a little, you know, they did some iterative validation work on the symptoms to be sure that like the wording that you would use to ask somebody about that, would be a wording that most people would understand the words being used. You know, I am not a PRO developer, but you are trying to make the language clean.

DR. WHITNEY: And is there a similar thing? These are patient reported. Is there a similar analysis that could be done, those things that physicians monitor, the respiratory rate, the white count, the fever? Because I think that is Dan's point. This is the patient side. Is there something from the physician side that also could be --

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

DR. MUSHER: The paper by Halm, et al. I gave the reference in my slide. I think that there are just two sides to the same coin and I think it all should be used.

DR. REX: And we saw at the January workshop, perhaps somebody here who may have presented the slide, I don't remember who did, but there was a slide that looked at an experience from the `40s or `50s where at least a couple of physicians were sent in to interview the same person back to back and looked at the correlation between what the physicians got out of that patient in back to back interviews.

And the correlation wasn't awful but it wasn't great either. So it is telling you that, you know, if Dan and I go interview the same patient, we may walk away with a summary of that patient that is a little bit different on any given day and I fully recognize that.

So physicians have variability and

bias. But I think the reason people got interested in PROs, if you think about patient to physician to case report form, if you put a physician in the middle, it is sort of intuitively better to go directly from patient to case report form, if you can, even though we know that at each step there is some bias. You know, the less of it we have got in there, the better.

But I still think that things like fever, which is relatively objective, and some of the physicals signs the physicians focus on, plus the aggregate sense of I got better, I am ready to get out of bed, I actually feel like having dinner. Those things do capture a lot of the patient's sense of the disease. They are not the same and I fully recognize that. So they are not as precise but they certainly, in aggregate over time, I think, are an accurate representation of this person got better.

So, you know, that is the theme

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

that I want to build on.

ACTING CHAIR TOWNSEND: Dr. Fleming.

DR. MUSHER: But is on the same point. Just to finish up, the Halm paper takes the objective data the physicians use, which is oxygen saturation, respiratory rate, pulse, and does the same thing and looks at the rate of improvement of those. So that is why I said it is the other half.

It is not how the physician interprets information given to him or her by the patient, it is the objective data the physicians use. And I think that they should all be used together.

So, what the patient says about his own symptoms is going to be more reliable than through the doctors and interpreter in some ways, not always. But certainly the objective data need to be included and can be.

DR. FLEMING: So just for a little more amplification clarification, so there are

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

the domains, the components, the signs, the physicians will use to guide management of patients. The focus of this is to say, ultimately, we are assessing the affect of treatment on what patients value. And so, in the concept of saying what is it that patients value? Certainly death but much more than death.

So, essentially, in the Chest article for Lamping, what she is saying here is that trained interviewers conducted telephone and face-to-face interviews asking patients about their daily life with CAP, their symptoms and circumstances in which they were most bothered and limited because of CAP, basically focusing on patients' views about the bothersomeness of their symptoms and this is what emerged.

And indeed, chills and sweating are here. Those are symptoms. Those are symptoms that are related to fever but they are not fever, identically fever. I.e., you could

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

have a raised temperature and not have chills and sweating. And you could have sweating without fever.

The reality though is yes, they are the symptoms that are related to fever. But it is also relevant to say that what comes forward here are a dozen. So, when you are looking at this, the comprehensive aspect of this is much more than just the chills and sweating, if you were using chills and sweating.

And it is interesting, I agree with Dr. Rex, to look to see. It is the coughing, the shortness of breath, dyspnea, and fatigue are the ones that really show up consistently across all of those measures.

DR. MUSHER: Dr. Fleming, you said something along the way, excuse me, but it is very important and not correct. You said we all agree what is really important is how the patient feels. Look, I am a doctor, of course it is how the patient feels.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

The patient says, Doctor, I have got shortness of breath today. It is not much better. And you see the patient is lying there comfortably with a respiratory rate of 16, which is normal, then the doctor's observation needs to be included, along with the patient's observation. That is all I am saying. And, therefore, I am trying to tell you that the doctor's observation, the medical observations on pulse, and respiratory rate, and on temperature are very important. Please don't dismiss them.

DR. FLEMING: They are indeed. This is such a key point. We aren't dismissing them.

DR. MUSHER: But you did. You said, we agree that what is important is how the patient feels.

DR. FLEMING: We aren't dismissing them in terms of what guides a physician. A physician is guided by --

DR. MUSHER: And the physician

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

judges the outcome. The physician largely judges the success of the therapy.

DR. FLEMING: The ultimate judge of the success of the therapy and the ultimate goal of what we are trying to do is to benefit patients in prognosis and in quality of life. And patients, ultimately, are weighing in on what aspects they most care about. We do appropriately use a lot of signs to address how to manage a patient.

DR. MUSHER: But we are evaluating the effect of an antimicrobial agent and the patient may not be able to distinguish the effect of the antimicrobial agent from all sorts of other things in his or her life. So you cannot just focus on the patients' symptoms.

ACTING CHAIR TOWNSEND: I think we are probably agreed that we need to be comprehensive on our evaluation. We are going to talk about endpoints when we get to the question discussion session.

Dr. Cox still needs to talk. And then if you have got a follow-up, then you can go after Dr. Cox.

DR. COX: I thought it might be helpful just to take a step back for a moment and just, you know, look at sort of a simplified depiction here of outcomes in a study. And this really could be any endpoint. I know we have had a lot of discussions about mortality, fever, other clinical symptoms that patients may exhibit.

And let's just put that aside for a minute and just say that this is the primary endpoint. And maybe it is a constellation of clinical symptoms such as fever. And I think the key point here, that we are trying to get at, in a circumstance like this, where the test drug is behaving the same as the active control, in essence, in this study, against this primary endpoint. And let's say for instance, this is a constellation of clinical findings and there is no difference in

mortality, this study is informative because although we don't have it from this particular study what the placebo rate would be, it is that difference that we are seeing between the test and the active and what a placebo would have done if it were included in such a study. And that is the key point that makes this study informative.

When we get to this situation where we may be looking at clinical symptoms, mortality, I mean, whatever the endpoint is, if there is really no difference between the test active and what a placebo would have done if it were included, and it is not going to be included because we are not, you know, we are doing an active controlled study in patients who are sick, in this setting, it is not informative. So this, I think, is a key point.

And how do we know where the placebo is? Is it over here, close to where the test and active are or is it on the other

side? And that's why we keep trying to find information that informs us about that. And one of the pieces of information that we do have is all this old historical data. And if there is other information that helps us to understand where that placebo rate is, whether it be current day, whether it be on any of a variety of meaningful endpoints or a constellation of findings that would constitute an endpoint, is really what we need to get at.

So, it is this critical point of, you know, what makes the comparison of the test and the active informative.

DR. MUSHER: Well, if you think that your test drug, if you think that your existing drugs aren't effective, then we have a whole other problem. I mean, if you think that -- I'll just take an example.

If you think that moxifloxicin is not effective in treating pneumonia so, therefore, when you come along with a new drug

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

which I will call drug B and moxifloxicin is drug A, if you don't think the moxifloxicin is effective, then you are absolutely right, you have got to have some kind of control.

So, then the question is, are you a nihilist and say that the moxifloxicin is not effective or do you say well, yes, it is effective and we are going to compare drug B to drug A?

We just haven't got the data from the pre-antibiotic era. And things were totally different then. It was different patients, a different era, different expectations. It was all different.

DR. COX: Right. And I think the question that we are asking is actually different than that. The question is, how do we design a clinical trial that is informative? And I think, you know, obviously this is a very difficult topic. I mean, it has not been easy for any of us to figure this out but that is what we are trying to get at

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

today and that is what we are hoping to understand. What are the options for a trial that will be informative? And that is what we are hoping to hear from folks today.

ACTING CHAIR TOWNSEND: Dr. Dowell.

DR. MUSHER: Well, I still say if drug B, which is a new one, is about as good as drug A, which is a old one, and because I am not a nihilist, then I am going to accept that drug B is effective. It just seems like common sense to me.

ACTING CHAIR TOWNSEND: Dr. Dowell?

DR. DOWELL: Yes, the problem is, you just don't have that many examples in the modern era where it has been clearly shown that a drug is not effective. So, where is that placebo?

But I have to come back to that daptomycin trial and really compliment the people, I don't know if they are here, who worked on that trial.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

DR. MUSHER: It was a beautiful study.

DR. DOWELL: But it seems like that gives us a lot of information for what we are going to be discussing this afternoon. Because, in fact, they did show that one of the drugs was not as good as the other. And so, what did they do? We talked yesterday. They enriched the trial for patients who are more severely ill. They didn't just look at mortality as an outcome. They looked at a variety of outcomes. So I don't actually think this whole thing is hopeless. I think there are some things that we can learn from that trial and others that will help you guys to set out the parameters that other companies can do the same thing.

ACTING CHAIR TOWNSEND: Thank you. Dr. Whitney?

DR. WHITNEY: Yes, I guess just getting back to the discussion of the patient reported outcomes. I think I guess that is

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

fine for the mild disease, where you could ask the patient, well, how are you doing. Do you have, you know, are you fatigued today?

But for something that is a little more serious, you know, what can we use that is short of death, where you are trying to get at, you know, in Dr. Gitterman's talk, he talked about some sort of composite clinical endpoint for clinical failure. And what goes in that category, I guess is my question. Do we have some consensus on that?

ACTING CHAIR TOWNSEND: Hold on a second. Just, I was given a reminder. The daptomycin information is not -- some of it not yet public information? So we actually -- okay.

UNIDENTIFIED SPEAKER: If it is public, it is okay.

ACTING CHAIR TOWNSEND: Okay. I guess we are just reminded to stay on task.

Okay, where are we now? Dr. Temple, I think, had a question.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

DR. TEMPLE: Let me just drop back and get away from antibiotics for a second so it won't bother everybody who knows all about antibiotics.

If someone wanted to develop a new antidepressant, okay, they would have to come to grips with the fact that about 50 percent of all trials of depression drugs that we know work, can't distinguish drug from placebo. Okay? That is long-standing, well known.

So, if someone was nervous about leaving people who are depressed on a placebo and said, I want to do an active control trial, they would have to then say, okay, what is the effect of the drug I know works in this trial? And we know from enormous experience, that they can give us no reassurance at all that the drug that they know works worked in any given trial because half of them fail. You get the same results with antihistamines. There is a lot of symptomatic treatments where that is true.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

I don't know how many people have participated in it, but we are having the same problem in otitis and sinusitis where it is not so clear that the drugs that have been used really work or can regularly be shown to work, even though everybody sort of believes they must work in some people.

So, in all of those cases, doing a non-inferiority study is hard to interpret because you can't say what the effect of the active control is. You can't find information that allows you to say, in this trial what that effect size was, which makes it hard to interpret it.

The essential requirement for the study that you do and for the particular people you put into the trial is to be able to say with reasonable confidence, you never really know because you don't measure it, what the effect of the drug was in those people.

So, everybody has become quite comfortable, Tom included, with a relatively

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

severe population of people with pneumonia because in a fair number of studies, it was very clear that there was always a difference between the treated and untreated. So, everybody is very confident about that.

Whether there are other populations less sick in which you can pull together data that make you confident that you know for a well defined population what the effect would have been, is the crucial question here. We all understand that that is the question. It is not that people oppose putting a broader population into trials. It is how can we say what the effect was so we know whether the failure to find a difference is meaningful. You will never find a difference if the control drug didn't work. Then there will never be a difference. And I just want to remind everybody that we all think we know whether antibiotics are going to be effective because we know what they do in a test tube but it really would be bad if an antibiotic

didn't work as well. So you really want to know that your system will be able to detect that. It is a real worry. People could die if you are wrong. So, it is very important to know.

And it is not that some of these other ways of measuring effectiveness on defervescence or whatever it is are in any way invalid, it is that we don't yet know or haven't been able to figure out yet how to say what the effect of the control drug on those things will be in the trial, so that we will know whether failure to see a difference is meaningful or not.

I mean, I don't think this is all that arcane. It comes up over and over again and has been coming up repeatedly in the antibiotic world. There are several cases where what we used to do, which is define a difference and say, okay, I have ruled that out, doesn't seem valid anymore because we are no longer confident that the drugs we were

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

using as the control have that effect.

And in very mild degrees of pneumonia, that is the worry here. We don't know how to define what the effect of the active drug is. And you must know that before you can use the non-inferiority design. You must know it.

Does that mean you have to have perfect placebo controlled trials from 1930? No. But you have to be able to figure out in some way that has a certain amount of integrity and allows you to say I know the effect was this big. And you somehow have to get to that point.

We all know there are compromises in getting there. We know that. And you go back and look at the old data. It is not perfect, wouldn't meet modern standards in a lot of ways but because the effect seems so large, we are very confident that there is at least some effect in these trials so that you can go forward.

So far, though, the place where it seems best is in the relatively severely ill patients. That poses some difficulties that everybody is worried about. Especially, can you study drugs that come only in oral dosage form. But that is the problem.

It is not all that arcane. It comes up all the time and has been coming up for a long time. It is not Fleming being an obscurant tester or anything. This is something that has been part of the deal for two decades.

ACTING CHAIR TOWNSEND: Dr. Rex. Dr. Musher, Dr. Rex was next.

DR. REX: Thank you. Let me try to summarize something, Dan, because you are close, but you are drifting a little bit off the edge in a direction that is important.

Let me start by saying that the neat thing about antibiotics, and the thing I tell my management all the time, is that we can do so much prior to going into human

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

beings with an antibiotic. We can prove that it kills a bacterium. We can prove that it kills the bacterium in the mouse. We can prove that it cures the mouse's pneumonia, you know, mice aren't people but you know, it is a step in the right direction.

We can do very elegant pharmacodynamics to tell us what kind of an exposure. What is the shape of the exposure curve that would drive an effect? We then go in to man in phase one. We demonstrate that exposure shape. We demonstrate it not just in the plasma, but in the tissue of relevance. We can sample in skin. We can sample epithelial lining fluid.

So, when we go into a trial, our prior probability, and this is the important idea, is very high that this would work microbiologically. Now, there may be things that go wrong and that is why we must do the studies.

But you go into the experiment

thinking, you know, it ought to work. And more than that, I am going to put in enough drugs so that it really ought to work. And unless all the tenants of microbiology are flawed, or unless some bolt from the blue occurs, daptomycin turned out to be inactivated by cerfactin. We're going to come back to that in a second because you are absolutely right, that is a really pivotal bit of data.

So, theme number one is, our prior probability with a well tested, preclinical antibiotic of it doing something against the bacteria is very, very high.

Theme number two, past history tells us that there is a huge effect of treatment. In the olden days, people had the pneumococcus, looked sick, were sick, and even if they lived, they often lived with complications. We have talked about that. The key is not actually severity. They key is really has a bacterial pneumonia.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

So even a young person with pneumococcus can get in deep trouble. Even the 18-year-olds had a substantial mortality in the olden days when you didn't do anything but hold their hand and give them some tea. That's a very important thing. Those people didn't start off -- they might have showed up on the first day. Make the patient under the care of a third year medical student. A young 18-year-old man. Acute onset of a syndrome just like Gaisford described. Perfect case. Was he severely ill? Well, he was acutely ill. But was he hypotensive? No. Was he confused? No. Was he throwing up? No. So, was he severely ill? No. However, had I sat around and just given him tea and crackers, he would have been severely ill by the next day. That is what we know about bacterial pneumonia.

So, I am not discounting severity. Severity is important. But I want us to also to recognize, it is not just severity. It is

also knowing that it is bacterial, Dr. Musher's theme.

So, now, Dan, the issue is not whether or not antibiotics have an effect. I think everybody in this room is going to say moxifloxicin has an effect. That is not the question. As Ed said, the question is how do we measure it? What do we choose to measure that is meaningful? And here my answer is two-fold. The first thing is the thing that we have been using in the modern era is not actually all that bad. It is a composite of mortality and got better. And remember what I said earlier. It is not just -- we have often described this in negative terms. Say, oh, it's just a measure. After a week's worth of therapy and another seven to 21 days later and that is, we have waited so long, that you have blurred all differences.

That is not so. You could fail earlier. In order to succeed, you had to make it all through all those other days, but you

could fail at any time you wanted. And by day three or four, any physician worth his or her salt is washing you out of the study if you are not better.

So, it actually is. It includes time. It includes mortality. It includes very relevant patient-based outcomes. If you got better, you didn't develop an empyema. It's not just fever but recognize it is a key symptom.

So, is it perfect? No. Could we make it better? I'm sure we could. Does it have assay sensitivity? Is there a bulb in a colorimeter? Absolutely so, when you have a good population. And this is where the Pertel paper is just superb. They insisted on people having a syndrome that really looked like bacterial pneumonia. And they actually got a bacterial isolate. I'm sorry, I am blanking out on the number. It was about a third -- I'm sorry. I won't say. A pretty good frequency. Somebody can look it up for me and

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

tell me. It is a good frequency. And we know that if you are in that sick of a group, you are actually finding the isolate.

We know there are some more people where you just missed it. You know, there was that study we saw yesterday about doing transthoracic aspirates. I mean, it is there. People that sick, that is bacterial more times than not. When you go into that one, and you use this crummy, I'll use the word I often denigrate it, this outcome of, how was it defined, clinical response. You know, you got to the end of therapy and didn't record anything else and you waited a while. Fine, it worked. It detected the inactive drug. It detected it very nicely.

We also heard a very nice exposure response analysis yesterday. Now, admittedly, the numbers are small but they are really consistent. Remember, what is the value of the statistical test? It is to tell -- a P value is all about the extent to which the

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

data at hand contradict your initial hypothesis. P values aren't truth. Okay? P values do not represent the probability that something is true. A P value represents the degree to which the current data contradict your initial hypothesis.

In the case of exposure response, in the case of clinical response, we have a lot of reasons to believe it ought to be true and what we are seeing is that all signs, all arrows point in the same direction.

So, that is the theme that I want to hear here. And Dan, I hope that answers your question. We are not saying that we think that moxi didn't work. We do. If moxi, levo, ceftriaxone didn't work, we would have hospital wards full of people with empyema. They clearly do work. That is not the debate.

The debate is, what can we agree on to measure reasonably such that we can say, when I studied A versus B and they came out about the same, it is on the basis, the

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

percent response. What is response on that Y axis? What is it? And I am saying that we have actually been smarter than we thought. The clinical response measure that we have been using for the past 15 years is really quite good because it does detect in a good population a disease that is actually where you can fail. It detects it. It detects it by daptomycin. It detects it by exposure response. And it is not just the quinolones. You can also see it for a macrolide and for other things.

So, I actually think we have done better than we think. And what we need to do is be aware of that. It is not perfect, but it is a good start.

DR. MUSHER: I just wanted to say that I agree with Dr. Temple's previous remarks and I agree with -- let the record note that once in a while I do.

And the whole point is, again, if you do have a population that includes

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

patients who are going to respond. So the key really is to have enough patients in there who are not going to be responding to the placebo effect because they have a viral pneumonia that doesn't respond to any antibiotic at all.

So, you need to try to have enough in there with bacterial pneumonia. Then, what we should be doing is just finding criteria that will follow them and death rate isn't going to do it because it is just going to be, thank goodness, three per thousand people treated in that fashion, out-patients being treated.

DR. TEMPLE: So the question is a part -- I mean, various people talked yesterday about how to define a population that the drugs definitely work in. You are, I think, Dr. Rex, suggesting that it isn't only their score on this but it is certain other features that tell you that it is very likely they have a bacterial pneumonia. And that is the population we think it works in.

And I don't think anybody disagrees with that.

I think if you can define that population and get them into the trial, you are in fact, home.

I just want to ask Tom one specific thing. Use of death as an endpoint is not, I don't think, disadvantageous. If nobody dies in the trial, then you rule out your margin. As long as you know this -- I mean, it doesn't really matter, I don't think, whether you call it death or of course, looked like they might die or doing very badly, any of those. But if those rates are very low with the antibiotics, that doesn't mean you need a bigger trial, that means you are likely to be successful and rule out the difference you are worried about.

So, getting other things into it, I don't think helps do the trial.

DR. FLEMING: Well, as you noted earlier, if you have a population of people in which you are assured the death rate would have been substantial and you see no deaths,

that is certainly reassuring. So, if we put everybody on an antibiotic in a population that we were highly persuaded would have been having a high death rate and there were not deaths, that would certainly provide substantial evidence. That is the kind of evidence we had for the sulfonamides and the penicillin.

DR. TEMPLE: But I just wanted to say, having no deaths --

DR. FLEMING: Of course, what that is saying though, that is talking about using an historical control when you have a huge effect.

DR. TEMPLE: But there seemed to be some worry that there wouldn't be a lot of deaths or whatever these endpoints are. That is not a problem here, if you are confident that the population would have done badly without treatment. Having no deaths doesn't mean you need a bigger trial. It's not like endpoints in a different showing trial.

8b70aab3-cc41-4ff9-90d5-6b61e280f91a

ACTING CHAIR TOWNSEND: I think we need to move on.

DR. TEMPLE: Okay.

ACTING CHAIR TOWNSEND: I'm sure we will have plenty of opportunity to discuss this during the question session.

So, we are going to start the open public hearing part of the agenda. I will read a prepared statement.

The FDA and this committee place great importance on the open public hearing process. The insights and comments provided can help the agency and this committee in their consideration of the issues before them.

That said, in many instances and for many topics, there will be a variety of opinions. One of our goals today is for this open public hearing to be conducted in a fair and open way where every participant is listened to carefully and treated with dignity, courtesy and respect. Therefore, please speak only when recognized by the