DR. TEMPLE: I'm asking the ethicists.

DR. GOLDKIND: Well, since I repeatedly talked about minimization of risk versus the ability to maximize scientific validity. I think that before you enroll a population of that sort, you'd need to ask the question of how do you come to know that they truly are antibiotic resistant? Are you enrolling populations that are pathogen -- that has a pathogen-directed diagnosis?

And you cannot, I don't really think you could ethically support a trial where you were putting a population -- a portion of the study population at increased risk in the trial. So you'd have to answer a lot of other questions that would be based on the information you have that's either clinically relevant information or supportive information from the preclinical setting before you could do that.

DR. TEMPLE: suppose you didn't

ee9f750f-bf3b-4186-a0c3-d763750b38ce

know at the time of randomization that they were going to have a problem organism? It is I guess practice in many antibiotic trials to exclude people who are resistant to the drug being tested. Maybe that's not always necessary if you randomize them appropriately and ethically, because you don't know they have a resistant organism, but simply discover along the way that the people who have resistant organisms don't do as well on the previous therapy and leave them in the trial.

DR. NELSON: I guess two thoughts, Bob. One would be, to some extent you could figure out a way to include that as an endpoint, and similar to a randomized withdrawal design, or some design where that would become a treatment failure.

I'm not sure you would want to continue someone who you know it's resistant on, a drug that doesn't work. But how you as a statistician or a trialist would combine that data I think would be the question.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

I might just add one thing in thinking about it which impacts on pediatrics to a large extent more perhaps than adults is the issue of standard of care that lacks sufficient evidence to justify it other than expert opinion. You heard some of that in the previous presentation.

One of the big questions is the extent to which there is an evidence base for the comparator arm, and at least in pediatrics, often that is a challenge if pediatricians are using a drug that may not have good evidence, based on -- other than their expert opinion, but it's being done in an off-label setting. And then what do you then choose as your comparator I think is a challenge. And that challenge was raised I think in the presentation about the guidelines that are being used.

ACTING CHAIR TOWNSEND: One last question from. Dr. Follmann.

DR. FOLLMANN: Yes, I'd like to

ee9f750f-bf3b-4186-a0c3-d763750b38ce

talk a little more about the delayed start design issue you alluded to and I think was brought up over there.

So I had some in thinking that there were attractive features to that design if we looked at some kind of time to event at a point like defervescence or something, and that maybe with a couple of hundred or 600 patients you could do a superiority trial using defervescence as an outcome if you delayed therapy by four hours or perhaps eight hours or something like that.

So I was very interested in your comment about the study of 14,000 patients where they looked at the rate of mortality of patients who had been delayed four hours.

And I had a couple of specific questions on that. One is, for the people who had been delayed more than four hours -- I guess it was people who got therapy within four hours, they showed a decreased mortality rate compared to who. Would that include

ee9f750f-bf3b-4186-a0c3-d763750b38ce

people who didn't therapy and died before they go therapy? So that's one question.

And a related question is, what severity of CAP did those patients have where that difference in mortality was observed? Because it seems to me that if you're looking at a population which has less than a one in a 100 rate of death, a delayed start option of four hours or so might still be something to consider.

DR. GOLDKIND: I would open this also to the folks who are representing the ATC, and IDSA if they'd like to comment further as well.

I encourage you to go to JAMA as well as the Archives of Internal Medicine, which is where I pulled those articles. And I'd have to actually go back and look at them very carefully to answer your questions.

But it was a multicenter retrospective cohort study of 14,000 hospitalized elderly patients with pneumonia.

So there is a certain -- they're not -- you know it was a retrospective study of a cohort of Medicare database information across 14,000 subjects.

DR. GILBERT: Dr. Gilbert, just to respond as far as time -- is it okay? IDSA representative.

ACTING CHAIR TOWNSEND: Actually, can we save your comments for later?

DR. GILBERT: Oh, I thought she asked for a comment?

ACTING CHAIR TOWNSEND: We're running a little bit later right now. Thank you.

Thanks, Dr. Gilbert.

Dr. Fleming will talk about noninferiority issues in trials of community-acquired pneumonia.

DR. FLEMING: Thank you. So what I'd like to do is really two parts. I'd like to revisit some of what we've already touched on today at some length, and that is the

ee9f750f-bf3b-4186-a0c3-d763750b38ce

complexities and the criteria that we need to address to do valid noninferiority.

And then to move on and talk about the application of those principles to the setting of CAP.

And I'd like to return the thanks to my cochair Dave Gilbert for his efforts in the workshop recently.

So pressing ahead, the classical setting for noninferiority is where we have an effective standard intervention, and we are interested in an alternative experimental that would have some improvement in profile, and if the experimental then gave comparable results to the standard, this improvement in profile would make it an attractive alternative option.

So invasive Aspergillosis Voriconazole providing better side effect profile than Amphotericin B would be of interest, and CAP, a new Quinolone would be a more convenient administration than

ee9f750f-bf3b-4186-a0c3-d763750b38ce

penicillin. In mother-to-child transmission of HIV where intensive therapies do reduce the risk fo mother-to-child transmission. Unfortunately because of cost and convenience, they are not available where we need them most in developing country settings, and we find more cost and convenient alternatives.

So I'll focus going forward in the example here, in the CAP setting, where let's say we are comparing a new Quinolone to penicillin.

So there are dual goals in noninferiority. The first is to have a direct evaluation of the new Quinolone against the penicillin, and also to be able to establish the new Quinolone is effective against placebo. But there is no placebo.

And so we can only get indirect insight about the new Quinolone against placebo by looking at the noninferiority comparison against penicillin, and then the historical evidence for penicillin's effect

ee9f750f-bf3b-4186-a0c3-d763750b38ce

against the active comparator, or against placebo.

And herein lies the challenges to being able to do noninferiority. ICHE-9 says the standard, penicillin in our example, should have efficacy that is of substantial magnitude, precisely estimated. And where those estimates from those historical trials of penicillin need to be relevant to the setting of our noninferiority trial of the new Quinolone against penicillin.

And to just provide a little more insight into the importance of this constancy assumption, suppose we are comparing an experimental against Vancomycin, and Vancoymcin-resistant enterococci patients, if this experimental is similar to Vancomycin, the question is, is that similar effective, or similar ineffective? And you would say, Fleming, it's similar effect, because we know in pristine patients historically Vancomycin was shown to be highly effective.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

Ah, but if in fact Vancomycin is much less effective in VRE patients, then comparable effectiveness could be relatively ineffective, unless we are confident, unless we can reassure than Vancomycin in the setting of a noninferiority trial is as effective as it originally was in the historical trials in pristine patients.

So if we are in fact looking at a noninferiority trial to compare the new Quinolone against penicillin, and we have historical estimates from historical trials about the effect of penicillin, what are the factors that could lead to penicillin having a different effect in this noninferiority trial than it had historically?

Well, first in a noninferiority trial disease could be caused by pathogens that are resistant to penicillin.

Secondly, in the noninferiority trial there could be enhanced and competent treatment that attenuates the effect of

ee9f750f-bf3b-4186-a0c3-d763750b38ce

penicillin. Or in the noninferiority trial penicillin might be given with a dose schedule where there would be less adherence, or there could be different endpoints.

So the essence of why this is important is if we have historical data that does in fact establish an effective standard, an effective penicillin, and if in fact the experimental truly is ineffective, a result where the experimental could actually look similar to the standard in a setting where the constancy assumption fails.

And similarity here, even though it's an ineffective intervention, would be interpreted as similar effective, because we are falsely assuming the validity of constancy assumption.

This issue of the constancy assumption is not a new concept. Those who have been involved in lab experiments have long known the importance of maintaining laboratory conditions in order to extend or

ee9f750f-bf3b-4186-a0c3-d763750b38ce

extrapolate beyond previous laboratory experiments.

All right, so let's talk about a hypothetical illustration of how we might be looking at community-acquired pneumonia in a noninferiority trial. And as we've been discussing at some length today, what is integral is to be able to come up with a margin, so that if we can rule out that we are worse by more than this margin, we can conclude reasonably reliably that we truly have a beneficial intervention.

So let's talk in a setting where it's pneumococcal pneumonia. The standard again let's say is penicillin against a new Quinolone. Let's say the endpoint is failure in mild disease. This could be a composite that would include death, persistence of symptoms, or worsening of symptoms or breakthrough infections.

In a more severe patient it might be mortality. So for sake of illustration

ee9f750f-bf3b-4186-a0c3-d763750b38ce

let's suppose the endpoint is mortality.

And let's suppose in our noninferiority trial, the new Quinolone has a 5 percent higher mortality than penicillin, with 150 per arm, two standard errors, or plus or minus 10 percent.

So let's graphically look at this. We are plotting along this axis. The failure rate, in this example, the relative mortality on the new Quinolone against the standard penicillin. So you'd like to be over here to the left, where the new Quinolone has a lower mortality.

You certainly don't want to be here to the right where it would have a higher mortality than the standard or penicillin regimen.

Well, in our example it is slightly to the right of zero at 5 percent higher mortality. That could be as much as 15 percent higher.

Is this upper limit sufficiently

ee9f750f-bf3b-4186-a0c3-d763750b38ce

low to allow us to conclude that we have adequate efficacy of the new Quinolone? One of the issues is, we have to know where does placebo reside. So what is the actual efficacy, or what is the effect of the standard, the penicillin, against a placebo.

And let's suppose we in fact did have a comparison here that looked at no specific treatment, or placebo against penicillin, and let's suppose that the placebo had a 30 percent higher mortality in 350 patients, that would also be estimated at plus or minus 10 percent.

With that estimate then we would know that placebo lies here at 30 percent higher risk of mortality, or rate of mortality, than the standard penicillin intervention.

But this isn't a known fact. This is estimated with random variability, with variability. And there is also uncertainty about the validity of the constancy

ee9f750f-bf3b-4186-a0c3-d763750b38ce

assumption. Is this historical data telling us that the placebo has a 30 percent higher mortality rate than standard, relative to what the effect of the standard is in the noninferiority trial.

And so a traditional approach here is to use the lower limit of the 95 percent confidence interval, as where we would place -- where we would place placebo in this scenario.

And in fact in this example, because the upper limit of the confidence interval for the new Quinolone against placebo, against penicillin, does rule out 20 percent, we'd conclude that this is evidence of efficacy.

It is better than a placebo.

But the issue is, if in fact you have a very effective standard in penicillin, why is it good to introduce something else that is just better than placebo?

And so the tradition is, we need

ee9f750f-bf3b-4186-a0c3-d763750b38ce

to preserve an important fraction of the effect of the active comparator, a fraction that is often taken as 50 percent.

So if we want to preserve 50 percent of the effect of penicillin, the margin then that we would be using is 10 percent.

Now there is also another important criterion that must be considered. The margin must be sufficiently small that we can accept that it'd be clinically relevant for the experimental therapy to be worse than penicillin or the standard by at least 10 percent.

We're inclined to say, sure, okay, it's all right if we can rule out that mortality is not higher or greater than 10 percent. We want a margin that's as big as we can make it to keep trial sizes small to make it more likely to get a positive result.

On the other hand, we have to be able to justify the patients would be

ee9f750f-bf3b-4186-a0c3-d763750b38ce

comfortable using this new intervention simply by knowing that we're ruling out that it has more than a 10 percent higher mortality.

So to justify that, I always turn it around and say, suppose you had with an experimental intervention the ability to reduce 30 percent mortality to 20 percent; that's a 10 percent improvement. Would you be marching off claiming that's a major advance, even if there were inconveniences with that new therapy? Absolutely.

Well, if a 10 percent improvement would be highly clinically important on mortality, why are we willing to give up 10 percent mortality in the other direction?

So in essence the clinical relevance of the reduction in efficacy has a major component in defining what the appropriate margin is, and if we are going to allow a 10 percent increase in mortality, there needs to be substantial other benefits that counterbalance that from the perspective

ee9f750f-bf3b-4186-a0c3-d763750b38ce

of the patient.

So in essence then the margin that we arrive at has to be sufficiently rigorous that it allows us to establish efficacy; to in particular allow us to know that we are preserving a substantial fraction of the effect of an active comparator, an active comparator that we're saying is so effective it would be unethical to use a placebo.

Well, if it's unethical to use a placebo, it's unethical to lose an important fraction of that active comparator effect. And also it has to be sufficiently small that we can justify that it is acceptable to have loss of efficacy of up to that amount.

Well, if you do a noninferiority trial, and the new Quinolone is in fact established to be noninferior to penicillin, what can we say? It's at least as good as penicillin? Or it's not worse than penicillin?

Well, let's suppose that the

ee9f750f-bf3b-4186-a0c3-d763750b38ce

noninferiority trial was actually 10 times larger than what I showed in the previous example. But you are still estimating it has a 5 percent higher mortality. Now with 3,000 people that's 5 percent plus or minus only 3 percent.

So now when we look on the plot, in this particular scenario, we still see that the mortality rate on the new Quinolone is 5 percent higher than it is in penicillin, but with an upper limit of 8 percent that rules out our margin, so we've successfully established noninferiority.

On the other hand, the other limit of the confidence interval rules out equality, approving inferiority. So in this noninferiority trial we have established that the new Quinolone is noninferior to penicillin, while proving its inferior.

All right? Actually I'm okay with that paradox. I'm okay with the paradox because noninferiority doesn't mean that I've

ee9f750f-bf3b-4186-a0c3-d763750b38ce

established that the experimental therapy is not worse than. Noninferiority simply says that I've ruled out that it's unacceptably worse than. It may be worse; I've just ruled out that it's unacceptably worse. That's all we conclude with noninferiority.

As a result this 10 percent margin better be ruling out anything that is, in fact, unacceptable. So if you tried to use a 15 or 20 percent margin you'd have to be justifying that any increase in mortality less than 15 or 20 percent wouldn't be clinically important.

So no, those aren't the conclusions of -- noninferiority simply allows you to conclude that the new Quinolone is not meaningfully worse than the standard, in this case penicillin.

All right, what I'd like to do in the second part here, then, is to go forward with these principles to apply them in the setting of CAP trials.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

And we have seen, we have had much discussion about the fact that standard antibiotics in CAP are effective. The question we have to address isn't a simple one as that. The question is, what is the effect? What is the magnitude of the effect? On what clinically important outcome measures, in what population, and in what experimental conditions?

Because we need the answers to all these questions that Ed Cox laid out at the beginning today in order to be able to define noninferiority margins, and to design noninferiority trials.

So choice of endpoint. There is a temptation to use our biological understanding to say, we intend to achieve, we hope to achieve, clinical benefit on mortality, on symptoms, on breakthrough infections, mediated through biological effect, such as effects on microbiological endpoints.

So in bacteremic patients, one

ee9f750f-bf3b-4186-a0c3-d763750b38ce

such endpoint might be getting clearance or negative blood cultures. Brad Finland showed that in a considerable number of bacteremic patients, that we had deaths occurring with purulent focal complications in spite of repeatedly negative blood cultures.

And so this is the classic example of a false positive conclusion that we often get with surrogates. And the reasons that surrogates are often misleading isn't that we aren't partially right about how it is we are trying to achieve benefit. But the totality of the mechanisms through which the disease process influences outcome, and the mechanisms through which the treatment affects those processes, is highly multidimensional. And the ultimate effect on clinical measures is often not adequately captured by just looking at the effect on one specific biomarker.

We talked about defervescence. The argonaut reported an experience in patients where temperature became normal at

ee9f750f-bf3b-4186-a0c3-d763750b38ce

three days in half the treated patients. It did become normal in a substantial number of the controls, about 25 percent, but this represented a doubling, a substantial improvement with therapy on defervescence. And yet, secondary pyrexia was fairly common. And he reports the overall average duration of pyrexia in the hospital was little affected.

So simply getting clearance or normalization of the temperature isn't capturing the totality of the effect on pyrexia, because even a more significant issue with defervescence, is defervescence in essence capturing the essence of the benefit and risk of what we're hoping to achieve in a CAP setting?

If so, if lowering the body temperature is the goal, why not just use antipyretics, if that's the goal. But in fact it's not the goal. In fact if that were the goal we could do placebo control trials. Because what serious harm would be induced if

ee9f750f-bf3b-4186-a0c3-d763750b38ce

the worst scenario here is that we would have a slight delay.

It is a clinically relevant measure, but it's not capturing nearly the totality of what we really most care about in terms of mortality and symptoms.

So while defervescence, if that was the essence of what we're trying to do would be a proper endpoint. If we're actually thinking of it as a surrogate for a more comprehensive benefit, for what really matters to a greater extent to patients, it's actually a very bad surrogate.

Because antipyretics do lower body temperature, but don't affect mortality and symptoms. And yet serum therapy that induces febrile reactions in a third of the patients decreases mortality. We get the wrong answer when we compare antipyretics to serum therapy when we are using defervescence when in reality it's the serum therapy that provides the more global benefit.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

Here is a listing, kind of a prioritization of what the endpoints could be, from the most clinically compelling, to what could also be used as measures if we have proper instruments.

If we have properly validated patient reported outcomes, PRO measures, we could be using these types of measures to establish benefit.

However, as we know from the broad science of measuring such outcomes, the devil is in the details to really validating PROs. The outcomes can depend very significantly on the outcome measures. Metley referred to the fact that in treated patients, medium time to resolution of cough is 14 days. Twenty percent of patients reported substantial fatigue three months after the initial time to diagnosis in contrast to the single article by Kingston that says you have resolution of fatigue.

These data have to be taken with

ee9f750f-bf3b-4186-a0c3-d763750b38ce

great caution looking at the totality of evidence on such measures, and looking at whether or not we have reliable instruments to assess the effects on such measures.

Metley pointed out that there are no well controlled studies. We understand that; we have no randomized trials. But his quote, it's nicely pointing out what observational studies -- and that's we have, we have historical observational cohorts -- what are they highly suited to do? And what are they weakly suited to do?

Observational cohorts are well suited to defining what is the expected rate of complications of certain kinds of events, describing how patients are managed.

What they are poorly suited to do is to estimate the impact of an intervention.

In essence where we have our best opportunity from the historical literature to look at what the effects are of interventions and CAP would be for mortality endpoints.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

Dowling, this is one of the overall summary papers that Dowling presents, and it points out that we have a significant amount of literature looking at no specific therapy against serum therapy, sulfonomides and penicillin on overall mortality.

Two limitations or two issues: these columns provided the most substantive evidence used in the IDSA report. The sulfonamide report, though, certainly the sulfonamide column shows lesser efficacy, or at least in these data, a higher level of mortality than on penicillin.

The other significant issue here is this is only allowing us -- it's important to assess effects by age, and to adjust for confounding by age, but it only allows us to look at age.

So how do we address the confounding that exists in the lack of randomization? Finland and Brown noted, any presentation concerning mortality from

ee9f750f-bf3b-4186-a0c3-d763750b38ce

pneumonia is incomplete if it fails to take into account bacteremia age and presence of systemic complications.

And in a significant overview article by Finland, Finland is looking specifically in these white bars, in those patients that had no specific therapy.

And when you look in bacteremia patients by age, you see age is strongly predictive of mortality. And in non-bacteremia patients by age, age is strongly predictive of mortality.

Even more predictive of mortality is bacteremic status, when you compare age groups across bacteremic-nonbacteremic. Bacteremic status is highly predictive of mortality.

Why is this important? Well, Finland and Brown in their 1939 article show two groups, one group that had nine deaths in 18 patients, one group that had one death in 18 patients. Of course we don't know, there

ee9f750f-bf3b-4186-a0c3-d763750b38ce

wasn't randomization, so there is probably a lack of comparability.

But you say, come on, Fleming. Yes, there is no randomization here. But this is a 44 percent difference, P of .003. This is pretty compelling.

Well, this is sulfanilimide that had the 50 percent, and no specific treatment that had the 6 percent.

So before we take the conclusion that sulfanilimide is harmful, we looked at the data to see that in fact in the sulfanilimide category more than half the patients were bacteremic; more than half were multi-lobar, and almost three-quarters were age greater than 30, whereas in the no specific treatment there were no bacteremic patients, and no patients that were over the age of 30. High confounding. Therefore, it's critical to sort out treatment effects from these highly predictive prognostic factors.

So essentially what I'd like to

ee9f750f-bf3b-4186-a0c3-d763750b38ce

talk about is an independent literature review to derive margins in the noninferiority setting on a mortality endpoint that addresses as best possible this confounding.

So essentially what we did was, an extensive literature review of original source articles. We did not use the review articles for two reasons. The review articles led to double counting, because review articles, two separate review articles, often had the same source articles.

Secondly, when you went to the source articles you found more data on the characteristics of patients, particularly age and bacteremia.

We also looked predominantly at the experience on sulfamide derivatives and penicillin, and didn't include the data on serum therapy because we wanted to have mechanisms of action here that were actually similar to Beta-lactams and Quinolones that would be used today.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

And the majority of subjects had pneumococcal pneumonia, and had microbiological confirmation of disease.

It would be difficult then to extrapolate to some of the other types of pneumonia.

We were able to find, among many many articles, 16 that did characterize results by age and by bacteremia, and eight in particular that characterized by age and by bacteremia, in a search led by John Powers and assisted by me.

And this extensive effort, I'm going to summarize them in a single slide. So in this single slide, what we characterized then is the death rate, mortality over number of patients on no specific treatment versus antibiotics in a bacteremic setting; and no specific treatment against antibiotics in a nonbacteremic setting, subdivided again by the same age categories that Finland had used, less than 30, 30 to 49, and greater than 50.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

So we have six separate cells in which to make the comparison.

The interesting phenomenon, the interesting results that we can see is that age is highly predictive of mortality. As you go down the column, as you go down the rows, you have more - much higher mortality, and that's also certainly the case in bacteremic patients.

And bacteremia is highly predictive of mortality as well. So bacteremia and age are highly predictive of outcome. As a result they are confounders. This is the aspect that is fundamentally different, as you well know. If we had randomized trials, we would have randomization to account for this high predictiveness of these covariants.

Separately, in addition to bacteremia and age being predicters of outcome, they have a completely separate property: they are also effect modifiers. And

ee9f750f-bf3b-4186-a0c3-d763750b38ce

effect modifiers are the critical issue that deal with the issue of constancy assumption.

So when you have covariants like bacteremia and age, if you had randomization, you don't have to worry so much about them being predicters; you have to worry about them being effect modifiers. But here we have to worry about both. We have to worry about the fact that they are predicters, confounding our analysis, and they are effect modifiers. The conclusions can be specific to the category.

And what we find here is the estimated differences here are very different across these cells. There certainly is effect modification that is occurring here.

And what we also find is that while bacteremia status is an effect modifier, even when you are adjusting for age, it's not so much an effect modifier in the older patients. It's much more profoundly an effect modifier in younger patients.

Conclusion: if you want to develop

ee9f750f-bf3b-4186-a0c3-d763750b38ce

a noninferiority margin in younger people, you must take into account bacteremia status in addition to age. Age is absolutely critical; so is bacteremia status.

So what we have done is just a partial adjustment for the lack of randomization. We have looked at age and bacteremia status. There are many other -- there are at least five other key predicters and potential effect modifiers, as identified in the literature.

The PORT scoring certainly allows for and takes into account age and comorbid illness. It doesn't take into account bacteremia status. And in fact Musher did a very nice summary here that points out that when you look at the PORT scores in nonbacteremic patients versus bacteremic patients, much of the predictiveness of bacteremia status is not captured in the PORT score.

Bottom line is, you need to look

ee9f750f-bf3b-4186-a0c3-d763750b38ce

at both. When you look at both, what you find are 10 percent margins that can be justified in any of these bacteremia patients, even in younger patients.

So this in fact is a higher margin than IDSA was talking about, if you look at younger people who are bacteremic. And if you're not bacteremic, if you are age greater than 50, PORT 4 or 5, you also have a substantial margin.

But when you are non-bacteremic and younger, we are not saying that antibiotics don't work, but it's a much more complicated issue to drill down on exactly what that level of efficacy is.

And in all likelihood, even if you could refine the amount of information you have in antibiotic treated patients who are non-bacteremic in these other categories, you are probably looking at margins that could be as high as 2 percent to 5 percent, in this range; much less than the 10 percent margins

ee9f750f-bf3b-4186-a0c3-d763750b38ce

in the other categories.

So conclusions: when we do a noninferiority trial, it doesn't establish that the new Quinolone is as effective as the standard. It establishes that the new Quinolone is not unacceptably worse, and that puts significant restrictions on how rigorous we need to have that margin.

It's not adequate to do a simple statistical calculation working backwards, saying, I'd like a trial of 300 people. My success rate in the control arm is 80 percent. Therefore what is the margin I can rule out with 90 percent probability? The answer is 15 percent. That doesn't justify the margin.

You justify the margin first based on clinical criteria as to what is a small enough excess that if it occurred would be clinically acceptable.

But that's not it, that's not the end of the story. The margin isn't based simply on a consensus and clinical judgment.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

It also must be evidence based, and the evidence-based aspect of the margin is that it must be sufficiently rigorous that we know we are preserving a substantial fraction of the effect of the active comparator, if for no other reason because we can't deprive people of that active comparator by doing a placebo-controlled trial. So how can we deprive the clinical community of the effects of that active comparator if the alternative therapies that we'd be approving are in fact meaningfully less effective?

So the bottom line here is, indeed, antibiotics are effective, and they are effective on mortality, but that's not the end of the story. To come up with margins, we have to understand as clearly as possible the magnitude of the effect of our comparator antibiotics, and that magnitude, and hence the margin is specific to the endpoint; you have a different margin for every endpoint. It's specific to the population. You have a

ee9f750f-bf3b-4186-a0c3-d763750b38ce

different margin for each population.

And it's specific to the experimental conditions, including the exact intervention that you are using. You could have a different margin for each specific intervention that you are using as the active comparator.

So in conclusion, as has already been stated, noninferiority trial designs have generally been recognized as a necessary approach in certain settings, but to be avoided whenever possible. They share many of the inherent dangers of historically controlled trials.

In our setting, though, they have the complication that we not only have these highly predictive factors that are effect modifiers, but they are also predicters, and we don't have randomized controlled trials to look at the estimate of the effect of those historical interventions.

Interesting, when we talk about

ethics, there is a recent article in Lancet that says that noninferiority themselves are unethical. And the argument that's given is, if you have a highly effective standard therapy, and you do a noninferiority trial, you randomize half the patients to that highly effective standard therapy and half of them to an intervention that you hope is as effective but could be clinically meaningfully less effective, why is it in the patient's best interest to go into such a trial and have half a chance to get something that you hope is effective but could be clinically less effective?

Hence the ethical issue raised by the Lancet article.

So in conclusion we believe that valid noninferiority trials can be done in CAP. They can be done specifically using standard control regimen. So regimens that would - for which we could argue that the effect of sulfanomides and penicillin

ee9f750f-bf3b-4186-a0c3-d763750b38ce

historically would apply.

They should be done in populations that would have significant mortality. In particular, patients who are either bacteremic or PORT 4 or 5, there is substantial evidence of efficacy in those settings. With microbiological confirmation of pneumococcal like disease in those settings, we could then be using all cause mortality on a seven-day or a 14-day mortality endpoint, using setting-specific margins that could be as large as 10 percent.

In closing I'd like to thank my colleague John Powers for his significant efforts in the literature review and for coauthoring the manuscript that was the source of these comments.

Thanks.

ACTING CHAIR TOWNSEND: Thank you, Dr. Fleming.

We're running just a little bit late. So what I'd like to do is if there are

ee9f750f-bf3b-4186-a0c3-d763750b38ce

any questions for Dr. Fleming to take a couple for them. Dr. Gilbert, if you are still willing to discuss Dr. Goodkind's question, and then Dr. Rex, you had a question from the last speaker. So if you still have that. Okay. All right.

So a couple of questions for Dr. Fleming if you have any, and then we'll go to Dr. Gilbert.

DR. GILBERT: Well, I would like to comment, Dr. Gilbert, I'd like to comment on the question that was answered in the previous discussion, and then ask Tom three simple questions.

Very simple. First of all, as far as timing of starting therapy and effect on mortality, probably the best data albeit imperfect is the Austrian and gold data showing that the earlier you started the antibiotic the higher the survivorship rate, and that data has been widely quoted, and we can provide the reference if needed. And it's

ee9f750f-bf3b-4186-a0c3-d763750b38ce

in our position paper.

I appreciate Dr. Fleming's eloquence and expertise. I want him to run for political office. But I have three very simple questions.

You've answered the first one: you agree that there is a drug effect in the treatment of community-acquired pneumonia.

DR. FLEMING: Yes.

DR. GILBERT: Do pneumonia patients become afebrile and normalize their white count?

DR. FLEMING: Do pneumonia patients become afebrile and normalize their white count on their own?

DR. GILBERT: If they survive either - if they survive and/or they are in a treatment.

DR. FLEMING: The argonaut data was referencing a cohort that was treated and had no specific treatment, and in terms of febrile status, in terms of defervescence, 25 percent

ee9f750f-bf3b-4186-a0c3-d763750b38ce

were achieving normalization of temperature by three days.

DR. GILBERT: Well, I'm sharing with the audience a discussion that Tom and I have had now over the last four or five months trying to get clinical reality into valid and acceptable clinical trial designs. Because all of us use the responsive or lack of responsive to fever, normalization of the white count, as an endpoint, as well as whether the patients are feeling better or not, which are the patient reported observations.

So I love the analysis that was done, but somehow it has to be comprised with clinical reality.

DR. FLEMING: So in many disease settings, we use biomarkers for purposes of directing clinical therapy, and that's completely appropriate.

The biomarkers have five distinct purposes. And while they are appropriate for

ee9f750f-bf3b-4186-a0c3-d763750b38ce

some purposes, it doesn't mean they are appropriate for others. So they are useful for assessing prognosis. They are useful for detecting disease. They are useful as you just said for defining how to change regimens, how to change clinical care. They are useful for enrichment of the patient population as we just saw here; bacteremic patients enriches your population, and they are useful as biomarkers.

And because a measure is useful in clinical care to direct how you are administering a therapy does not at all establish that it is useful as a biomarker for enrichment or a biomarker to be used as a surrogate endpoint.

And the argument that is given here is, if the goal of therapy is defervescence, it's the right endpoint If the goal however, and this is what I believe I'm hearing consistently is, no, we can't do placebo-controlled trials because the risk to

ee9f750f-bf3b-4186-a0c3-d763750b38ce

these patients is far more than a delay in defervescence.

The risks are these much more significant mortality and symptom consequences. What I'm saying is, defervescence, then, could be a very bad measure for discerning far less appropriate versus more appropriate strategies.

ACTING CHAIR TOWNSEND: Dr. Rex.

DR. REX: My comment is not specifically for Dr. Fleming. It's really for Dr. Temple who asked the question earlier about enrolling groups of patients who have resistant isolates.

And my observation is that resistance is not a binary observation. You get a range of MICs. And the very interesting and instructive area under the curve ratio to the MIC analysis is actually a form of exploring that question in a setting where it is not really resistant. It just has a little bit higher MIC.

So I would observe that that kind of a result is actually an internal measure of assay sensitivity. If you've done a study in which even though I understand, MICs are not randomly distributed, nor are AUCs. Patients have them because that's their isolate, and that's their pharmacokinetics. But if you take, in large aggregates, the patients who have differing exposure MIC ratios, that begins to point at that question.

It's not a complete answer, I will observe, but it certainly does point in the direction.

I just wanted to offer that as a bit more of a closing of the circle on your question.

ACTING CHAIR TOWNSEND: Dr. Patterson.

DR. PATTERSON: Well, I saw that Dr. Spellberg had got up to make a comment on this, and he wasn't allowed to speak. So if he is not going to speak to the group, I'd

ee9f750f-bf3b-4186-a0c3-d763750b38ce

like to either hear him after lunch and talk to him during the break, and have permission to do that. Because I'd like to hear what he had to say.

ACTING CHAIR TOWNSEND: Dr. Temple.

DR. TEMPLE: My assumption is if you could find a somewhat resistant population and show that the new drug was better than the other one, we wouldn't have to have this problem, because superiority is usually interpretable.

Tom, suppose there were well documented effects of effective therapy on fever, and people decided that defervescence was in fact a valuable thing, even if it wasn't mortality. But it was still valuable.

If such data existed, that could be an endpoint, couldn't it? I'm not sure that the evidence is that there is such data, but if there were --

DR. FLEMING: It is. It is an endpoint, and it is a valid clinical

ee9f750f-bf3b-4186-a0c3-d763750b38ce

consequence to patients.

If however we said, this is the essence of what we're doing is, we're treating fever, we are trying to achieve defervescence, then I think a lot of us would say it's not such a problem to do a placebo controlled trial with proper informed consent.

The problem is, that's not my position. I agree with people who are saying it is a problem to be doing placebo controlled trials in such settings where even if the mortality rate is low, if you are reducing that mortality rate then the effect on mortality, the effect on preventing significant breakthrough infections and in achieving benefit more globally on symptoms are more globally what I'm hoping to achieve, and what this example is saying, you may get the wrong picture of the better choice by simply looking at defervescence alone.

DR. TEMPLE: Well, let's suppose you establish that the drug in question is not

ee9f750f-bf3b-4186-a0c3-d763750b38ce

an anti-fever drug. It's not acetominofen. That's easy enough to find out.

And I'm not sure what the reasons might be why you'd want to pick an endpoint; maybe more people have it or something. There are more endpoints.

We -- it's common for example when your goal is to reduce cardiovascular mortality to live with reducing heart attacks which are a surrogate after all.

DR. FLEMING: Well, I would find that no problem as well. Heart attacks are themselves a very significant component of the mortality effect and morbidity effects.

DR. TEMPLE: Actually not the heart attacks we measure. We measure minor enzyme elevations. People don't recognize that surrogate.

DR. FLEMING: I would even take that measure as you refer to as MI even on enzyme elevation as a more profound clinical event than defervescence at day three versus

ee9f750f-bf3b-4186-a0c3-d763750b38ce

day five.

DR. TEMPLE: Okay, just as a thought, I'm not sure a major consequence of being infected like fever couldn't conceivably be an endpoint. I don't think there is any reason to think about that, at least in the severely ill people, because you have a clear mortality effect for the untreated people; you don't have to get there.

But I do wonder about the less severe people, the under 30s or something, where you've displayed quite clearly that you are not going to be able to do that.

So maybe you want to include everybody in the trial because you like to get information on younger people too. And that could be an endpoint for that population. I wouldn't rule that out yet.

DR. FLEMING: So if you could find a cohort sufficiently young, sufficient low risk, non-bacteremic, don't have any of these other complications that put them at a

ee9f750f-bf3b-4186-a0c3-d763750b38ce

detectible risk of mortality, then I understand your point.

But in that scenario, you could take it to the next step and say, I can do superiority, because you are saying, I'm finding a scenario where defervescence is the essence of what I'm trying to do.

DR. TEMPLE: No, I don't think that's true. What I hear people - one last thing, what I hear people saying is, we don't want to accept any risk of mortality here.

So if the risk is 1 in 500 or something, you are not going to be able to study that. But nobody thinks leaving people with bad pneumonia untreated, we'll accept that risk.

So you are not going to have mortality endpoints in that population, but maybe you can have a relevant endpoint in that population. I just wouldn't discard it too soon.

DR. FLEMING: And that's what we

ee9f750f-bf3b-4186-a0c3-d763750b38ce

have to discuss.

ACTING CHAIR TOWNSEND: What I'd like to do is, since we didn't have time for questions for the speakers who spoke before the break, is to have any other questions directed to those if possible, before we go to lunch.

DR. WEIDERMANN: Let me -- and this may come up later too, but since I see myself apparently as the token pediatrician on the panel here -- oh, we have one more, thanks. I hope we understand -- I mean we are talking about age under 30, and I don't, Tom, in that collection of studies how many were truly in the pediatric age range.

But I think the numbers we're talking about apply to adults with pneumonia, and the biologic variability, once you -- under 30 compared to 30 to 49 is tremendous variability.

So I hope, especially when we're talking about specific margins or

ee9f750f-bf3b-4186-a0c3-d763750b38ce

categorization of risk factors or things like that, it starts to break down if we really include children.

ACTING CHAIR TOWNSEND: Dr. Patterson?

DR. PATTERSON: Could we hear Dr. Spellberg's comments now?

ACTING CHAIR TOWNSEND: If you could keep them brief, Dr. Spellberg. Thank you.

DR. SPELLBERG: Well, first let me respond to the pediatric issue.

As I mentioned in my talk, there are multiple historical studies done on the pediatric population. And when I say, pediatric population, under 12.

And some of the studies break them apart to under two and between two and 12. Bradley and McCracken have summarized that data and that manuscript will be coming out. And the effect size is very similar.

My comments with respect to

ee9f750f-bf3b-4186-a0c3-d763750b38ce

endpoints are, one, as somebody who takes care of patients I profoundly disagree that fever is not a useful endpoint.

We always rate PRN Tylenol, and we - the nurses are always flooding the patients with Tylenol. They can't give them the Tylenol fast enough.

Nevertheless, it is PRN. And when the Tylenol effect wears off, if the patient's pneumonia is not responding, they become febrile again. If the patient's pneumonia is responding they defervesce. And there are clinical data in multiple studies that show, and Rich Wunderink alluded to these clinical composites, that if you are clinical responding, and fever is a component of that, your risk of bad outcomes or relapse is extremely low.

So we could use that data to help guide us to know that we are effectively treating the pneumonia.

The second issue is with respect

ee9f750f-bf3b-4186-a0c3-d763750b38ce

to bacteremia. There are two problems with bacteremia in the modern era, trying to link them to the historical data sets.

One is, there is a much smaller incidence of bacteremia in patients with pneumonia nowadays as compared to historical datasets. And there are probably a variety of reasons for that.

But unlike age, which is constant - a 30-year-old is a 30-year-old - there is just a much smaller fraction of patients nowadays that are bacteremic, so it's harder to make a constancy assumption argument.

And the second issue is that there are data in the modern era that if you are bacteremic, you don't have a higher mortality if you get treated with antibiotics than if you're non-bacteremic.

That was an effect of the historical pre-antibiotic era, and the very early antibiotic era with sulfa drugs which are not very effective at treating bacteremia.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

In the modern era the clinical data show that whether you are bacteremic or not the mortality is minimally effective as long as you are on effective antibiotic therapy.

If you are not on effective antibiotic therapy your mortality is much higher.

ACTING CHAIR TOWNSEND: Thank you.

DR. FLEMING: I'll take that as a question. Just very quickly in response, we are using the historical data in which bacteremia is very prevalent to arrive at the margins. And therefore we have to take into account that confounding or that property, how predictive bacteremia is.

And the second issue is, the issue on the table is not, is defervescence of clinical revelance. The issue is, when you look at an effect on defervescence, are you capturing adequately what the totality of the intended benefit is to patients?

ee9f750f-bf3b-4186-a0c3-d763750b38ce

DR. MUSHER: And that is exactly the point, Dr. Fleming, that I don't understand. The totality of the benefit is not what was under discussion. There are a number of different things that you can look at. So it wasn't a totality.

And in your response to Dr. Temple I didn't understand it. Because it sounded as if he says we should use temperature as an endpoint, and you went back to the placebo argument, and it really just didn't follow. And just before your comment, I would like to add, as was just pointed out, we clinicians use multiple endpoints. And if you are taking a disease that is not so serious, where you don't really expect mortality, and there is not going to be enough mortality that you are going to be able to tell from it, unless you got a huge study, then you have to use multiple clinical endpoints. And they would include the time it takes until you feel better, and the time it takes until you stop

ee9f750f-bf3b-4186-a0c3-d763750b38ce

coughing, and the time you defervesce, which of course is all the need for a placebo control for a controlled study with comparator, and to be double blinded.

DR. FLEMING: Yes, indeed. And there is concurrence on the fact that you need to take into account the multidimensional aspect of what we're trying to do. And the strongest arguments that are being put forward about the reason we have to use effective active controls is in particular because some of those elements of what we're providing are of profound irreversible morbidity and mortality.

Some others are also important. Persistence of symptoms, breakthrough infections, et cetera. And then symptoms such as defervescence. They are all part of the whole story.

If we're however putting forward a criterion for registrational trials, is defervescence as the primary endpoint

ee9f750f-bf3b-4186-a0c3-d763750b38ce

adequately comprehensive to the totality of what I agree with you, you have to think about clinically.

DR. MUSHER: So it's multiple endpoints is what's being proposed, not a primary endpoint?

DR. FLEMING: Well, the strategies that could go forward certainly could be based on, for example, a mortality endpoint. If you looked in a severe patient at a mortality endpoint, and you are seeing benefit there, that is such a profound measure that even though you are not looking at some of the other elements, we would e persuaded of the benefit.

So the opposite extreme, when you take something at the lower end of the spectrum of relative clinical importance and rely on that alone, then there is the uncertainty as to how effects on that measure are predictive of effects on other measures. And being a predicator doesn't validate a

ee9f750f-bf3b-4186-a0c3-d763750b38ce

surrogate endpoint. Simply saying that historical data says that when you have lower temperatures -- in fact I think there is evidence to say when you have particularly low temperatures you could be at higher risk.

But just being a predicator doesn't validate getting a change on that measure will achieve the intended clinical benefit on the clinical endpoint.

DR. MUSHER: Could I just one more very quick comment.

Again, it seems to me, and you realize I'm struggling to understand the statistics, although I think you presented them very clearly -- it's just hard for me. But if the patient -- if the patients do better in five or six clinical regards, but somehow there is a much higher mortality among the ones who did better, then we're going to say that mortality trumps all those things, and that's not going to make a superior drug.

But if the patients do better in

ee9f750f-bf3b-4186-a0c3-d763750b38ce

all those other regards, and there is no mortality difference, then you would say you have very good drug, a better drug, whatever it is. I think you can't dismiss all the -- and I think that may be what you were saying.

I've already asked several questions, and the guy said, that's exactly what they were saying. So I'm just sort of dense.

But I am trying -- I always figure if I'm the only one in the room, I doubt I'm the only one in the room who doesn't understand something.

ACTING CHAIR TOWNSEND: Real quick.

DR. FLEMING: So real quick, the goal in any therapy is benefit to risk. And favorable benefit to risk. And we would ideally like to have our measure of benefit be as representative of what it is you as a care giver and as patient truly want to achieve.

And we hope there are concordances in how those effects occur. In the literature

ee9f750f-bf3b-4186-a0c3-d763750b38ce

there is in clinical practice innumerable examples of where if you take a COX-2 and you can achieve lower GI ulceration than a nonselective end set, and have analgesic effects, you still may be achieving increased risk of cardiovascular death, stroke and MI. And there are just numerous examples of this.

If you knew that, yes, it would trump the symptom relief, if you were causing deaths. We hope that's not the case here. But the bottom line is, we need to make an adequately reliable assessment of what we are doing, that you as a care giver are trying to achieve when you treat your patient. And is defervescence an adequate measure of that totality.

And I'm giving some examples where I think it's very -- it ought to be concerning if that's all we're measuring.

So I like your idea of saying, I want to see something more comprehensively, so that I have a more reliable sense of what I'm

ee9f750f-bf3b-4186-a0c3-d763750b38ce

doing in totality.

ACTING CHAIR TOWNSEND: Thank you. We'll have time to discuss this more. Thank you very much.

I think people are probably people are getting antsy for lunch. I know I'm getting hungry.

Dr. Patterson, I know you had a question. If we can wait until later that would be great. Thank you.

So we'll now take a break for lunch. We'll reconvene in this room again in an hour. So at 1:25, if you have any personal belongings please take them with you. Panel members, please remember, what happened in this room stays in this room. Don't discuss this out there.

Thank you.

(Whereupon at 12:25 p.m. the proceeding in the above-entitled matter went off the record to return on the record at 1:28 p.m.)

ee9f750f-bf3b-4186-a0c3-d763750b38ce

(Whereupon at 12:25 p.m. the proceeding in the above-entitled matter went off the record to return on the record at 1:28 p.m.)

ACTING CHAIR TOWNSEND: We are going to go ahead and start without them.

So we are going to start off this afternoon with a presentation by Dr. Mary Singer about the treatment effect of anti-bacterial drugs, and community-acquired pneumonia, historical perspective.

TREATMENT EFFECT OF ANTIBACTERIAL DRUGS IN CAP: A HISTORICAL PERSPECTIVE

DR. SINGER: Good afternoon, everybody.

Today I'm going to discuss what we know about the treatment effect of antibacterial drugs in CAP from the historical data.

My objectives are to review evidence for treatment effect of antibacterial drugs in CAP, and to estimate the magnitude of

ee9f750f-bf3b-4186-a0c3-d763750b38ce

that treatment effect.

Usually generally treatment effect is determined from placebo-controlled studies. But there have been no placebo controlled studies, or studies with no specific therapy as a control since the late 1930s for CAP.

So we searched through the literature for data which might allow for an estimation of the treatment effect, looking for information on natural history of untreated pneumonia, and we focused on published studies performed in the pre-antibiotic era, and shortly after antibiotics were introduced.

Most were studies of either pneumococcal or lobar pneumonia, which were considered synonymous at the time. And most were in hospitalized patients.

We found a number of observational studies which looked at patients treated with antibacterials, or those who received no specific therapy except for symptomatic

ee9f750f-bf3b-4186-a0c3-d763750b38ce

treatment.

They all use mortality as an endpoint. We also found a number of - a few controlled trials of antibacterial drugs versus untreated controls. Again most were in pneumococcal or lobar pneumonia, and we also looked at a couple of studies that studied mycoplasma pneumonia.

Dr. Nambiar is going to talk some more about some alternative approaches we also took to looking for an antibacterial effect.

This slide shows a brief summary, excuse me, a brief history of effective treatment for pneumococcal pneumonia.

Strep pneumoniae, or Diplococcus pneumoniae was first identified as the cause of pneumonia in around 1881. At the same time by Sternberg in the U.S. and Pasteur in France.

Serum therapy was first used for pneumonia, treatment of pneumococcal pneumonia, starting around 1913 with some

ee9f750f-bf3b-4186-a0c3-d763750b38ce

success. And sulfapyridine was introduced into clinical practice around 1938 and 1939. Sulfapyridine is a sulfaminide derivative, and was shown in a few early clinical studies to have possible efficacy in pneumonia, and was very quickly introduced into clinical practice.

It wasn't until the `40s that penicillin and other true antibiotics were introduced into practice.

Before I show you the data I'd like to show a quote from Sir William Osler, who himself died of Haemophilus influenza pneumonia in 1919.

He described the natural history of CAP before antibiotics. Recovery followed the crisis, an abrupt decrease in temperature over 12 hours accompanied by a passage from a condition of extreme distress and anxiety to one of comparative comfort, and occurred in a large proportion of cases. A fatal outcome was noted in 20 - 35 percent.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

He went on to say that worst prognosis was evident in drunkards, in the elderly, with fatality increasing to 50 - 65 percent in the elderly and those in their sixth and seventh decades.

Other than a few observational studies and anecdotes like I just showed you, we know really very little about the natural history of untreated community-acquired pneumonia.

This figure was taken from a chapter by Bullowa in 1937 in his textbook on the management of pneumonia for medical students and physicians.

He described the proportion of payments with untreated pneumococcal pneumonia who recovered, or terminated by recovery, describing the proportion of patients who recovered by day for different pneumococcal types.

So there was a total of 662 cases here of eight different pneumococcal types,

ee9f750f-bf3b-4186-a0c3-d763750b38ce

each shown by a different pattern in the bar, over a period of days from day one to day 19, for the percent of patients ranging from zero to 20 percent here.

So he first described a bell-shaped curve for - and this was fever resolution. Note that most patients had resolution of fever by days seven and eight, while only a small percentage had a resolution of fever in the first few days.

Bullowa also described mortality in untreated pneumococcal pneumonia in infants and children, and this table shows mortality by age, in those under two years old, and those who were over two, depending on whether or not they had bacteremia or not.

Overall mortality was highest in children under two years old, so for all cases, whether they were bacteremic or not, 20 percent, compared to 4 percent in those who were over two years old.

In both cases mortality was higher

ee9f750f-bf3b-4186-a0c3-d763750b38ce

in bacteremic patients than those who had negative blood cultures.

First I'll discuss the data from the observational studies. This was a study from Tilghman and Finland in 1937 who reported a percent mortality as a function of age and the presence of bacteremia in this prospective study of over 1,500 patients with pneumococcal pneumonia at Boston City Hospital between 1929 and 1935.

So most of -- the large majority, 82 percent of the cases, of the 1,500 cases, receive no specific therapy. So that was 82 percent who received no therapy here; 18 percent did receive pneumococcal antiserum.

So this figure shows, is combined for both, those who received treatment and those who did not.

It shows mortality on the Y axis by age group, and for all cases, whether they were bacteremic or not, mortality increased with age.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

Mortality was higher in those who were bacteremic than those who were not; and the proportion of patients who were bacteremic also increased with age, except for the oldest group in this study.

Note that the proportion of those who were bacteremic in this particular study far exceeds what we would see in clinical trials today. Also note that in bacteremic patients mortality was almost 30 percent in the youngest group, increased to about 100 percent in the oldest group, while in those who were non-bacteremic, mortality was still in the range of 10 percent in the youngest group, and quickly increased to about 80 percent.

Tilghman and Finland in this study also looked at outcomes, that is mortality depending on treatment. So whether they were treated with serum or no serum. So mortality was highest in bacteremic patients who did not receive any serum treatment compared to those

ee9f750f-bf3b-4186-a0c3-d763750b38ce

who did, 84 versus 45 percent. And in non-bacteremic patients, untreated, mortality was 32 percent compared to 12 percent for treated.

In the same study the authors also looked at duration of acute illness by presence of bacteremia, age - excuse me, not age - and whether or not they were treated with serum. So this slide is a little bit difficult to read, so I'll take some time to explain it.

The upper chart refers to those who received no specific therapy. So that was really the bulk of the patients in the study. The lower chart refers to those who received serum treatment. The X axis is duration of acute illness and days, and the Y axis is the number of cases.

You'll see that for each time period here there are two bars, the first one represents bacteremic cases; the second one represents non-bacteremic cases. And each bar is divided into two showing survivors at the

ee9f750f-bf3b-4186-a0c3-d763750b38ce

top and those who died at the bottom.

So in patients who received no specific therapy duration of the acute illness was most often about seven or eight days in comparison to those who received serum treatment duration of illness was shortened to about four to six days.

You've seen this slide before. This is a summary slide from Max Finland published in 1943, and then again in 1960. He reported mortality rates in patients with pneumococcal pneumonia at Boston City Hospital, comparing those who received no specific therapy to those treated with serum or sulfonamide derivatives.

So percent mortality is on the Y axis. Age group is on the X axis. The top left shows all cases, while the other two, the bottom left and right, show bacteremic subset and non-bacteremic cases.

So for all patients, okay, one other thing, the white bars or open bars

ee9f750f-bf3b-4186-a0c3-d763750b38ce

represent those who receive no specific therapy. Striped bars show those who received serum treatment, and black bars represent those who received sulfonamide derivatives. A small proportion of those also received serum.

And you probably can't read these numbers. There were about 2,800 who received no specific therapy; over 1,000 with serum alone; and about 1,200 cases who received sulfonamides.

So overall mortality was 41 percent in those who were not treated; so that's for all ages. And about 17 or 18 percent in those who received sulfonamides.

As already pointed out, the mortality increased with age regardless of treatment. Mortality in bacteremic patients was much higher overall, about 78 percent compared to about 30 percent in those who received sulfonamides; again, increased with age.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

In the non-bacteremic group, mortality was considerably lower, but there was still a small treatment difference. For the untreated group mortality was about 28 percent compared to about 10 percent or 11 percent in those who received sulfonamides; much larger in those that were over 50 years old.

So this was an observational study by Max Finland's group at Boston City Hospital in 1944 in which patients had moderate to severe pneumococcal pneumonia. The study doesn't really provide any direct information about treatment effect over placebo, but does provide some other useful information.

In contrast to other studies of the time period, patients were classified by severity at baseline. They were characterized as grade two, or moderate; grade three, meaning that they were acutely ill or irrational; grade four meaning they had shock and/or heart failure.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

So grades three and four would be considered severe even by today's standards. And you can see that the treatment arms, which were penicillin alone or penicillin after failing or developing intolerance to sulfa treatment.

The treatment arms are not balanced by severity. In this case 16 out of 17, and in the second treatment group to be considered to have severe pneumonia, compared to 21 out of 37 in the penicillin alone group.

Outcomes were not reported by severity in this study, but we can see that overall mortality was 18 to 19 percent in both groups.

Another important point about the study was that they looked at other endpoints other than death, including relapse, complications, bacteremia, duration of acute symptoms and fever.

So for example in the penicillin group there were two relapses; no evidence of

ee9f750f-bf3b-4186-a0c3-d763750b38ce

bacteremia in the patients who did have blood cultures once they started penicillin; and in the large majority of patients, 80 to 90 percent, there was resolution of fever and/or acute symptoms in under 48 hours in those who received penicillin.

So there are some differences between the two groups, but the point of the slide was really not not show those differences but what we might expect if we were looking at other alternative endpoints to mortality.

Dowling and Lepper in 1951 compared case fatality rates in patients with pneumococcal pneumonia who received no specific treatment, which is the solid line, to those who received serum therapy, sulfonamides, or penicillin and other antibiotics.

So the case fatality rate is shown on the Y axis; age and years is shown on the X axis. And important points in this study

ee9f750f-bf3b-4186-a0c3-d763750b38ce

were that mortality increased with age in each of the treatment groups. There appeared to be a lower mortality in the serum treatment group, but especially noticeable in younger patients rather than older.

Mortality was reduced even further in those who received sulfonamides, but the largest treatment difference between treatment and untreated control was seen in the penicillin and other antibiotic-treatment group.

In 1964 Austrian and Gold described a prospective observational study of patients with pneumococcal pneumonia with bacteremia. They followed patients who were hospitalized between 1952 and 1962, so a total of 455 patients, with bacteremic pneumococcal pneumonia without extra-pulmonary complications, for example, meningitis.

In this study about half were under 50 years old, and 55 percent had preexisting comorbidity, such as cardiac

ee9f750f-bf3b-4186-a0c3-d763750b38ce

disease.

Most were treated with penicillin or other antibacterial agents, while a small proportion received no specific treatment during this time period. So this could be considered a concurrent control.

This table shows case fatality rate by treatment in Austrian study. So penicillin, those who received penicillin were in the largest group. Mortality was 17 percent. And similar mortalities or case fatality rates were seen in those who received tetracycline or antibacterials.

And if you combine all those groups, overall mortality for any anti-bacterial treatment was 17 percent.

Among the few patients who received no specific treatment, 14 out of 17 died for a case fatality rate of 82 percent.

They also, Austrian and Gold also looked at survival in patients with bacteremic pneumococcal pnuemonia over a 21-day time

ee9f750f-bf3b-4186-a0c3-d763750b38ce

course in patients treated with penicillin -- this is from their prospective study -- compared to historical controls treated with serum or those who received no specific therapy.

So there were let's see, 298 patients in the penicillin treatment group; 93 with serum; and 384 in the untreated group.

So at day 21 the treatment between penicillin as far as survival in those who received penicillin versus no specific therapy was about 65 percent.

Just to summarize the microbiology from these observational studies I just described, all of these studied S. pneumoniae almost exclusively.

And this figure summarizes the treatment effect based on mortality in pneumococcal pneumonia from those observational studies. The blue bars show percent mortality in the untreated group, while the lavender and gold represent

mortality in those who received sulfonamides or penicillin.

So in the Finland study the treatment difference would be considered to be 41 minus 17, or 24 percent, while in those that were bacteremic in that study, 78 minus 30 would be the treatment difference, or 48; in the Dowling and Lepper study, treatment difference for sulfonamides was about let's see 18 percent I believe, compared to 26 percent with penicillin.

The Austrian study, which looked only at bacteremic patients, the treatment difference was much larger, about 65 percent.

So this table summarizes both the point estimates which I just pointed out, as well as the 95 percent confidence intervals among the observational studies.

So the treatment differences, that is, the difference in mortality between treated versus untreated groups ranged from 19 to 25 percent in the Dowling study; 24 percent

ee9f750f-bf3b-4186-a0c3-d763750b38ce

in the Finland paper; and was much higher in bacteremic patients, 48 to 65 percent.

Recognizing potential for bias in this type of post hoc subset analysis, we looked at mortality by age in treated and untreated patients from these observational studies. From Finland's summary we could not determine the number of patients in each of the subgroups. But for the other two studies mortality was higher in patients who were at least 50 years, compared to younger patients.

For example in the Dowling study mortality in treated patients was 20 percent if patients were over 50; 6 percent if they were under 50. In the untreated group mortality was 53 percent, compared to 19 percent in the under 50 group; and similar results were shown in the Austrian study.

Now I'd like to focus on some of the early clinical trials of anti-bacterial drugs for treatment of pneumonia. Note that really none of these studies would be

ee9f750f-bf3b-4186-a0c3-d763750b38ce

considered adequate and well controlled by today's standards for a number of reasons.

In this study by Park and colleagues in 1928 alternate patients with lobar pneumonia who were admitted from Harlem Hospital from 1927 to 1928 were treated with either pneumococcal antiserum or standard treatment.

And standard treatment at that time was fluids, pain relief with an elastic adhesive plaster, restriction of opiates, no drastic catharsis, oxygen for cyanosis or rapid breathing, and digitalization for heart rate greater than 120.

This table shows case fatality rate for the subset of patients with Type I pneumococcal pneumonia. And in this group the authors classified patients condition at baseline.

And this was really the only study that we found that directly reported outcome by severity at baseline. So condition was

ee9f750f-bf3b-4186-a0c3-d763750b38ce

considered good, fair or poor depending on a number of points. But it wasn't described in the paper in any further detail how that was done.

Also the numbers of patients in each of these subgroups was not reported.

But for some perspective here, this shows that regardless of condition overall mortality among those treated with standard therapy was 34 percent compared to 20 percent in those who received serum, for a treatment difference of 14 percent.

Mortality increased with severity across these groups, and the treatment difference increased with severity.

Note that in those who were in good condition at baseline, those mortality was 13 percent in those who received standard therapy, with a slight reduction to 9 percent in those who received serum.

This study by Evans and Gaisford in 1938 evaluated case fatality rates in

ee9f750f-bf3b-4186-a0c3-d763750b38ce

alternate patients with lobar pneumonia, treated with sulfapyridine, also called MMB 693 at that time.

And the untreated versus untreated controls. So these were all hospitalized patients with lobar pneumonia, eight to 68 years old; 86 percent of the population was under 50. And the proportion of bacteremic patients was not reported.

Specific pneumococcal serotypes were reported in only 22 percent, but among the other 78 percent there was no description of the organisms that were present. So they could have been pneumococcus, and they could have been other organisms.

Treatment was determined by enrollment on alternate day, and patients who died within 24 hours were excluded.

So for all patients case fatality rate was 27 percent in those who received no specific therapy, compared to 8 percent in those who received sulfapyridine.

That treatment difference was larger in the subset of patients over 50.

The same authors also describe the effect of sulfapyridine on resolution of fever. Among those treated with sulfapyridine, 60 percent had resolution of fever within 48 hours, compared to 34 percent of the controls.

In the following year, Gaisford published a follow up case study or case series of 400 cases of lobar pneumonia in hospitalized patients, including 100 cases that were described in the previous paper that I just discussed.

Microbiology was not described at baseline, and the proportion of bacteremic patients was not reported. Again, he excluded deaths within 24 hours of admission.

He compared case fatality rates in those treated with sulfapyridine to those now with historical controls in the preceding years at the same hospital.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

So this table shows the outcome in 400 patients treated with sulfapyridine between 1938 and 1939. If you look at the age distribution, about half of the patients overall were between 20 and 50; about 30 percent were in the five to 19 age range, whereas only 20 percent were over 50 years old.

So regardless of treatment group, case fatality rates increased with age.

Treatment differences were about 4-1/2 percent in the youngest group; 21 percent in the 20 to 50 year old group, and 26 percent in the oldest group here.

Okay, in another small study by Graham and colleagues in 1939, alternate patients with pneumococcal pneumonia were treated with sulfapyridine or no specific therapy. Note that 14 percent had other non-pneumococcal pathogens which were not identified.

Patients were hospitalized; 70

ee9f750f-bf3b-4186-a0c3-d763750b38ce

percent were under 50 years old; 29 percent overall had bacteremia. There was some imbalance between the treatment arms as far as proportion of bacteremic; 20 percent in the control group, 34 percent in the sulfapyridine group.

Case fatality rate was 23 percent in the control group compared to 6 percent in those who received sulfapyridine. And if you look at the subset of bacteremic patients, and there's quite small numbers here, you can see that there is a larger treatment difference, 50 percent minus 18 percent.

In another early controlled trial, Agranat and colleagues evaluated sulfapyridine for treatment of lobar pneumonia. Microbiology at baseline was not described, but at that time it was assumed that, as I mentioned before, lobar pneumonia was synonymous with pneumococcal pneumonia.

In this study there were several study sites in South Africa. The results from

ee9f750f-bf3b-4186-a0c3-d763750b38ce

each site were recorded separately.

Patients were treated with either sulfapyridine or no specific treatment. The method of treatment allocation was by admission to separate wards.

I'm going to talk about the results from the first site, Johannesburg Hospital, because that was the best described.

And actually results were reported by subsets from this site as well, and those were European versus non-European patients, because they were on different wards and reported separately.

In both groups the large majority of patients was under 50 years old.

Patients were classified as having mild, moderate or severe disease at baseline. But there is really no details about how exactly that was done.

There is a little bit of imbalance as far as baseline severity. About 33 percent were severe in the treated groups compared to

about a quarter in the untreated groups.

Proportion of bacteremic patients was not recorded.

So for both of these subsets there was a treatment difference. In the European subset treatment difference was about 15 percent; in the non-European group about 10 percent.

And these investigators also reported the number of days of hospitalization with fever after receipt of sulfa or standard therapy.

In the subset of non-European patients 52 percent had resolution of fever by day three, compared to 27 percent of control patients.

And in the European subset 70 percent had resolution of fever by day three compared to 10 percent of the controls.

And this slide just summarizes the microbiology at baseline in these studies; what type of pneumonia were we looking at.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

In Evans and Gaisford, as I mentioned, 22 percent - in 22 percent of cases pneumococcal types I, II, III and IV were identified, while organism was not reported at all in 78 percent.

So we don't know whether they were pneumococcal or not.

In the Graham study 86 percent of patients had strep pneumonia; 14 percent had non-pneumococcal isolates.

In the Agranat study there was no baseline microbiology reported. We only know microbiology from two of the deaths, strep pneumoniae in one, and staphylococcus in another.

So this table - excuse me, this figure summarizes the anti-bacterial treatment effect from the controlled clinical trials of CAP which I just described. So the treatment difference ranged from 10 to 15 percent in the Agranat study; to 19 percent in the Evans study; and 17 percent in the Graham study.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

And those were the point estimates I just mentioned, ranging, with a treatment difference or treatment effect, ranging from 10 to 15 percent up to about 19 percent, higher in the subset of bacteremic patients in the one study here that reported bacteremia.

And Dr. Valappil is going to speak a little bit more about the statistical implications of these data. You can see that the confidence intervals for some of these studies are pretty wide.

Now I'm going to talk a little bit about the two mycoplasma studies which you've heard mentioned previously.

There were two randomized double blind placebo controlled trials of mycoplasma pneumonia. The first study, Kingston and colleagues, in 1961, studied military recruits, all or predominantly male, aged 17 through 22. There was a total of 300 patients enrolled, and a subgroup of 103 had mycoplasma, which was previously called the

ee9f750f-bf3b-4186-a0c3-d763750b38ce

Eaton agent.

And that was diagnosed serologically.

Additionally nose and throat cultures were done and shown to be negative for pneumococcus and streptococcus if a patient was to be included.

These patients were randomized to either a tetracycline called demeclocycline or placebo.

And this figure from Kingston's study shows duration of fever in days by cumulative percentage of patients, in those who had mycoplasma pneumonia.

So by day three, about 70 percent of those who were treated with the tetracycline had resolution of fever, compared to only about 5 percent in the placebo group.

So a large treatment difference is demonstrated, but we have to be a little bit careful about interpreting these data. This is not really a - this is not a true analysis

ee9f750f-bf3b-4186-a0c3-d763750b38ce

but rather cumulative percentage over time. And the treatment difference might be magnified somewhat.

Kingston also looked at mean duration of fever and other parameters including cough, fatigue, abnormal chest X-ray, and so forth. In the subset of patients with mycoplasma alone that was 109 patients.

So they showed a treatment difference in days, for example for temperature the treatment difference was six days between treated and placebo groups; resolution of chest X-ray abnormalities, difference was 10-1/2 days for cough, 12.3 days and so forth.

In the second randomized controlled trial of mycoplasma pneumonia, this was also placebo controlled and double blind, again, patients were hospitalized military personnel, mostly male, aged 17 to 23; 130 were enrolled, and only 32 were confirmed to have mycoplasma by serologic methods, and

ee9f750f-bf3b-4186-a0c3-d763750b38ce

found to be bacterial pathogen negative by culture.

There were three treatment groups: clindamycin, tetracycline and placebo, with nine, 15 and eight patients. It's not clear why the differences in - patients - in the number of patients in each of the groups if this was truly randomized.

And antipyretic use was reported in eight of the patients, one in the clindomycin group, four in the tetracycline group, and three in the placebo group.

And this slide shows the results on looking at duration of fever, which is shown in hours on the Y axis for each of the treatment groups, tetracycline, clindomycin and placebo.

And what you are looking at here, the individual circles, closed circles, are individual patients. The heavy horizontal line is the mean. So the mean duration of fever in the tetracycline group was 39,

ee9f750f-bf3b-4186-a0c3-d763750b38ce

compared to 72 in the clindomycin group and 76 in the placebo group.

So the authors concluded from this that there was a significant difference in duration of fever between tetracycline and placebo but no difference between clindomycin and placebo.

So to summarize the data from the studies on mycoplasma pneumonia, a treatment effect of anti-bacterials was shown in young healthy patients on decreased duration of fever and other signs and symptoms, hospitalizations, chest X-ray abnormalities and so forth, with the caveat that reporting of antipyretic use was inconsistent among studies, and there remains considerable uncertainty in these data.

This leaves us with a question, can these data be extrapolated to all outpatients with CAP? And can they be used to determine a noninferiority margin for outpatient CAP.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

There are many limitations to using the data from the studies, all the studies which I described today, in estimating the treatment difference between anti-bacterial drug and placebo, including differences in patient populations, for example, comorbidities, immune status, pneumococcal vaccination status, and so forth.

There are clearly differences in organism and disease between the early studies and those today. The older studies looked mostly at hospitalized patients with pneumococcal pneumonia; severity was not well characterized. While today most CAP is treated in the outpatient setting, and strep pneumonia is isolated less frequently.

We know that in mild CAP now a typical organism such as mycoplasma are common. Clearly there are differences in standard of care between today and the 1930s-1940s. And there are obviously differences in study designs. A number of the studies I

ee9f750f-bf3b-4186-a0c3-d763750b38ce

presented are observational data. And the controlled studies that I did present were not randomized in the way that we would expect today, and were not blinded.

Endpoints are different, mortality was used as the endpoint in virtually all of the studies except for the mycoplasma studies we looked at, versus clinical response is the typical endpoint in clinical trials today. And Dr. Nambiar is going to talk a little bit more in the next presentation about why mortality might be difficult to use as an endpoint.

And then of course there's differences in study drugs, the ones we have information for, from the historical data are penicillin, sulfonamides, and tetracyclines.

So to conclude, we do have evidence for a treatment effect of anti-bacterial drugs in hospitalized patients with CAP, based primarily on a mortality benefit in those with pneumococcal or lobar pneumonia.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

There is also limited data on treatment effect for other endpoints such as time to resolution of signs and symptoms of CAP.

There is additionally some evidence although limited for a treatment effect on resolution of signs and symptoms in mycoplasma pneumonia.

Thank you very much.

ACTING CHAIR TOWNSEND: Thank you Dr. Singer.

We have time for one or two questions. Dr. Calhoun?

DR. CALHOUN: Thanks. I had actually kind of a biostatistical clarification question. In many of your treatment difference slides you were looking at the arithmetic difference in mortality rate, treated versus nontreated, and many of your slides showed an age-dependent increase in the treatment effect.

DR. SINGER: Right.

ee9f750f-bf3b-4186-a0c3-d763750b38ce

DR. CALHOUN: But I'm wondering whether that is the right metric, because if you look proportionately, the reduction in mortality was more like two thirds in all age groups.

And the reason I'm asking is that there is actually kind of a public health issue here. If you consider this to be really a disease that is of old people, then that has one set of public health implications. But in fact if you see a treatment effect of antibiotics across the age spectrum, as I think the data that you showed do show, then it has implications for pediatrics as well.

So what are your thoughts about that?

DR. SINGER: I think this data has implications for all the age groups. At least in our analysis we can show treatment differences across all ages, including pediatric patients.

DR. CALHOUN: But the magnitude of

ee9f750f-bf3b-4186-a0c3-d763750b38ce