Evidence based medicine without a statistician?

#DantripIsDead

Thomas Alexander Gerds

https://chat.openai.com/chat

Is it a good idea to do a randomized clinical trial without a statistician?

No, it is not a good idea to conduct a randomized clinical trial (RCT) without the involvement of a statistician. Statistical expertise is essential for the design, analysis, and interpretation of RCTs.

A statistician can help ensure that the study is properly designed with appropriate sample size calculations, randomization procedures, and blinding protocols to minimize bias and increase the precision of the results.

Additionally, a statistician can help select appropriate statistical tests and models to analyze the data, as well as identify and address any potential confounding factors or sources of variability.

Without a statistician, there is a greater risk of making errors in the design, analysis, and interpretation of the study, which could lead to inaccurate or misleading conclusions.

This could have serious consequences for both the study participants and the broader population if the results are used to inform clinical practice.

Therefore, it is highly recommended to involve a statistician in the planning and execution of an RCT to ensure that the study is conducted rigorously and that the results are reliable and meaningful.

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JANUARY 31, 2019 VOL. 380 NO. 5

Partial Oral versus Intravenous Antibiotic Treatment of Endocarditis

Kasper Iversen, M.D., D.M.Sc., Nikolaj Ihlemann, M.D., Ph.D., Sabine U. Gill, M.D., Ph.D., Trine Madsen, M.D., Ph.D., Hanne Elming, M.D., Ph.D., Kaare T. Jensen, M.D., Ph.D., Niels E. Bruun, M.D., D.M.Sc., Dan E. Høfsten, M.D., Ph.D., Kurt Fursted, M.D., D.M.Sc., Jens J. Christensen, M.D., D.M.Sc., Martin Schultz, M.D., Christine F. Klein, M.D., Emil L. Fosbøll, M.D., Ph.D., Flemming Rosenvinge, M.D., Henrik C. Schønheyder, M.D., D.M.Sc., Lars Køber, M.D., D.M.Sc., Christian Torp-Pedersen, M.D., D.M.Sc., Jannik Helweg-Larsen, M.D., D.M.Sc., Niels Tønder, M.D., D.M.Sc., Claus Moser, M.D., Ph.D., and Henning Bundgaard, M.D., D.M.Sc.

ABSTRACT

BACKGROUND

Patients with infective endocarditis on the left side of the heart are typically treated with intravenous antibiotic agents for up to 6 weeks. Whether a shift from intravenous to oral antibiotics once the patient is in stable condition would result in efficacy and safety similar to those with continued intravenous treatment is unknown.

METHODS

In a randomized, noninferiority, multicenter trial, we assigned 400 adults in stable condition who had endocarditis on the left side of the heart caused by streptococcus, *Enterococcus faecalis*, *Staphylococcus aureus*, or coagulase-negative staphylococci and who were being treated with intravenous antibiotics to continue intravenous treatment (199 patients) or to switch to oral antibiotic treatment (201 patients). In all patients, antibiotic treatment was administered intravenously for at least 10 days. If feasible, patients in the orally treated group were discharged to outpatient treatment. The primary outcome was

The authors' affiliations are listed in the Appendix. Address reprint requests to Dr. Bundgaard at the Department of Cardiology B 2141, the Heart Center, Rigshospitalet, Copenhagen University Hospital, Blegdamsvej 9, 2100 Copenhagen, Denmark, or at henning.bundgaard@regionh.dk.

Drs. Iversen, Ihlemann, Høfsten, Fosbøll, Køber, and Bundgaard are members of Copenhagen Health Science Partners.

This article was published on August 28, 2018, at NEJM.org.

N Engl J Med 2019;380:415-24.

From the abstract

METHODS

In a randomized, noninferiority, multicenter trial, we assigned 400 adults in stable condition who had endocarditis on the left side of the heart caused by streptococcus, *Enterococcus faecalis*, *Staphylococcus aureus*, or coagulase-negative staphylococci and who were being treated with intravenous antibiotics to continue intravenous treatment (199 patients) or to switch to oral antibiotic treatment (201 patients). In all patients, antibiotic treatment was administered intravenously for at least 10 days. If feasible, patients in the orally treated group were discharged to outpatient treatment. The primary outcome was a composite of all-cause mortality, unplanned cardiac surgery, embolic events, or relapse of bacteremia with the primary pathogen, from the time of randomization until 6 months after antibiotic treatment was completed.

From the abstract

METHODS

In a randomized, noninferiority, multicenter trial, we assigned 400 adults in stable condition who had endocarditis on the left side of the heart caused by streptococcus, *Enterococcus faecalis*, *Staphylococcus aureus*, or coagulase-negative staphylococci and who were being treated with intravenous antibiotics to continue intravenous treatment (199 patients) or to switch to oral antibiotic treatment (201 patients). In all patients, antibiotic treatment was administered intravenously for at least 10 days. If feasible, patients in the orally treated group were discharged to outpatient treatment. The primary outcome was a composite of all-cause mortality, unplanned cardiac surgery, embolic events, or relapse of bacteremia with the primary pathogen, from the time of randomization until 6 months after antibiotic treatment was completed.

- Standard care: intravenous antibiotic treatment
- Active arm: oral antibiotic treatment

Noninferiority trial!

From the abstract

RESULTS

After randomization, antibiotic treatment was completed after a median of 19 days (interquartile range, 14 to 25) in the intravenously treated group and 17 days (interquartile range, 14 to 25) in the orally treated group (P=0.48). The primary composite outcome occurred in 24 patients (12.1%) in the intravenously treated group and in 18 (9.0%) in the orally treated group (between-group difference, 3.1 percentage points; 95% confidence interval, –3.4 to 9.6; P=0.40), which met noninferiority criteria.

From the abstract

RESULTS

After randomization, antibiotic treatment was completed after a median of 19 days (interquartile range, 14 to 25) in the intravenously treated group and 17 days (interquartile range, 14 to 25) in the orally treated group (P=0.48). The primary composite outcome occurred in 24 patients (12.1%) in the intravenously treated group and in 18 (9.0%) in the orally treated group (between-group difference, 3.1 percentage points; 95% confidence interval, –3.4 to 9.6; P=0.40), which met noninferiority criteria.

Phew!

Enrollment

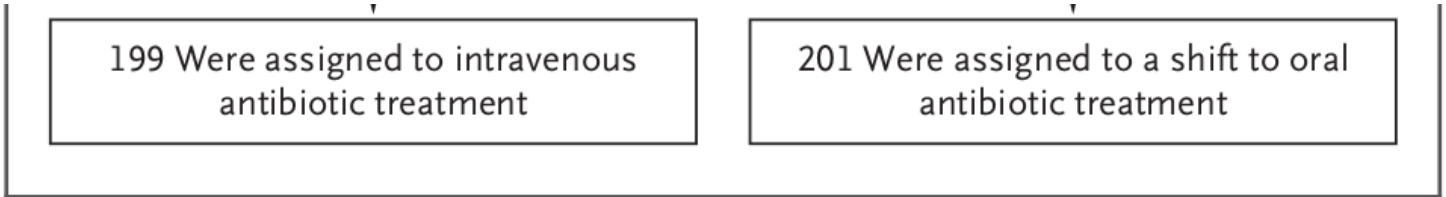


Figure 1. Enrollment and Randomization of Patients.

Inclusion and exclusion criteria are listed in Table S1, and additional details on reasons for exclusion are provided in Table S4, in the Supplementary Appendix. Signs of abscess formation were identified by transesophageal echocardiography (TEE) immediately before randomization. No patients were lost to follow-up. The body-mass index (BMI) is the weight in kilograms divided by the square of the height in meters.

Non-inferiority analysis

STATISTICAL ANALYSIS

The trial was designed as a noninferiority trial; that is, it was designed to determine, with the use of a noninferiority margin, whether partial oral treatment was noninferior to conventional intravenous treatment. We estimated event rates for the four components of the primary composite outcome from the literature[17]; we estimated the risk of all-cause mortality to be 2 to 5%, the risk of unplanned surgery to be 1 to 3%, the risk of embolic events to be 1 to 2%, and the risk of relapse of bacteremia to be 1 to 3%. Thus, the overall risk of the primary outcome was 5 to 13%. A risk difference (i.e., a noninferiority margin) of 10 percentage points was chosen (see the Supplementary Appendix). Under the assumption of a 10% event rate and a 5% loss to follow-up, we determined that inclusion of 400 patients would be required to provide a power of 90% to confirm noninferiority, with a one-sided confidence interval of 97.5%. Continuous variables are presented

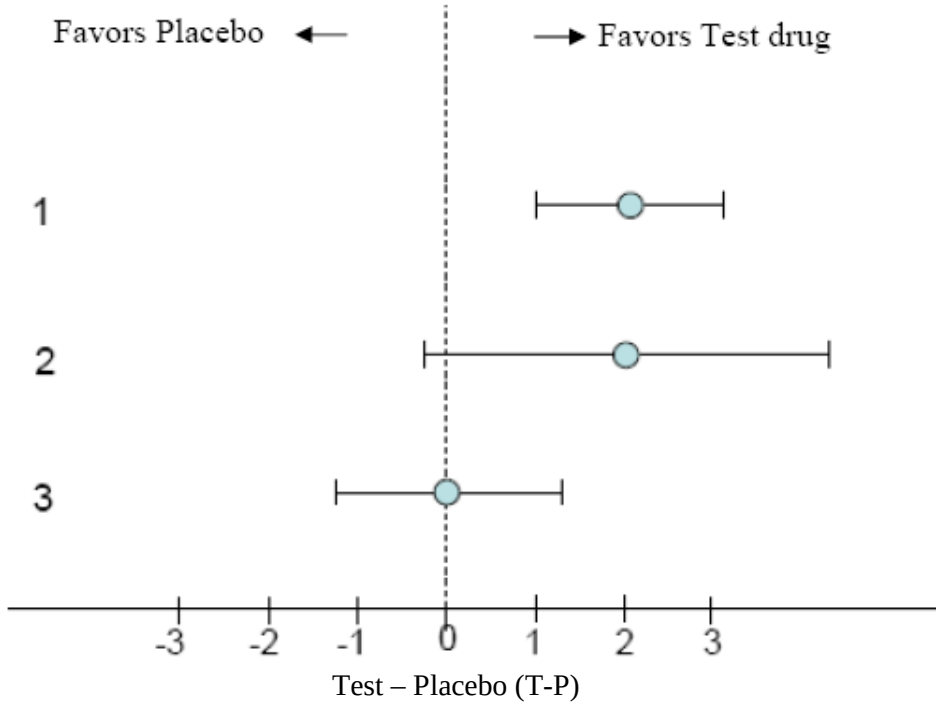
Choice of non-inferiority margin

The non-inferiority margin was chosen based recommendation from the FDA;
https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf
First, we considered the clinical acceptable difference in treatment effect. Considering the estimated large gain of oral treatment (hospital resources, economical and patient related) and estimated low risk of fatal outcome due to treatment failure in the orally treated group (inclusion *after* the initial phase, where the vast majority of complications is seen combined with close control of out-patients) we considered a difference of 10% to be clinically acceptable. Secondly, we considered the overall treatment effect. As mortality in untreated patients with infectious endocarditis is close to 100% the treatment effect is very high, which supports the choice of the non-inferiority margin. Finally, practical aspects of realistic recruitment of patients and completion of the trial were also taken into consideration.

So, oral antibiotic treatment is still non-inferior even if the risk of having the composite event (all-cause death is the most important) is 10% higher than what we expect with standard care!

FDA guidelines

Figure 1. Possible Results of a Placebo-Controlled Superiority Study (Point Estimate and 95% Confidence Interval (CI))

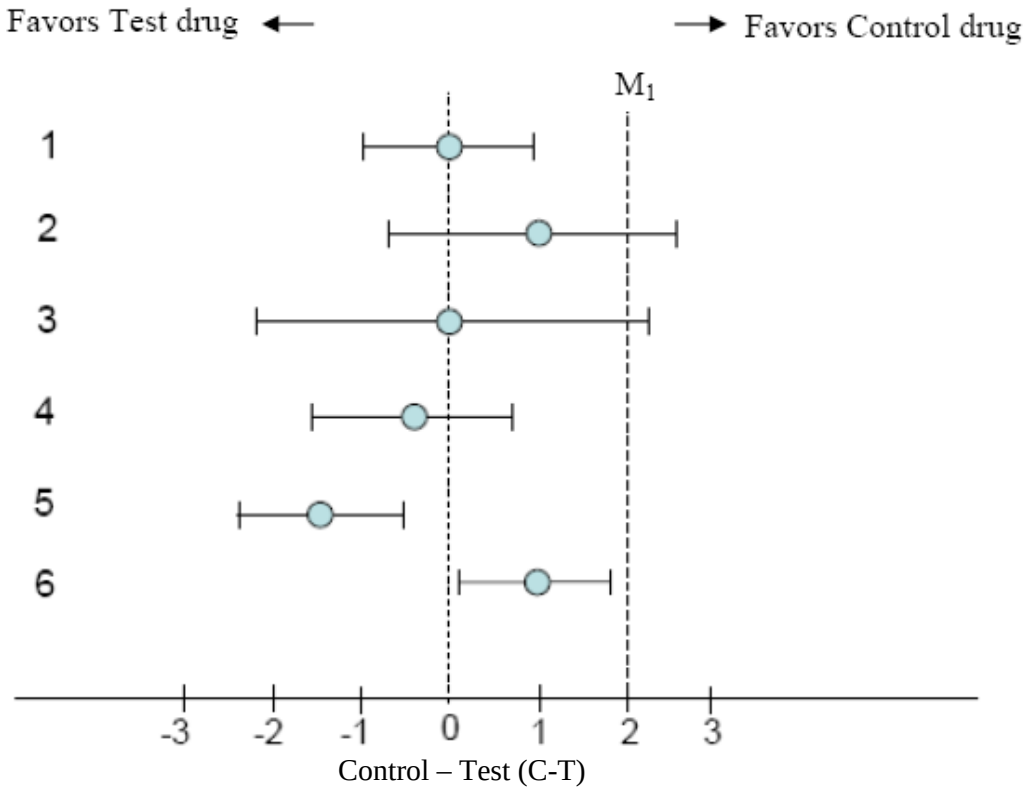


1. Point estimate of effect is 2; 95% CI lower bound is 1. Conclusion: Drug is effective and has an effect of at least 1.
2. Point estimate of effect is 2; 95% CI lower bound is <0. Conclusion: Drug is not shown to be effective.
3. Point estimate of effect is 0; 95% CI lower bound is well below 0. Conclusion: Drug is not shown to be effective.

Although there is no difference in the conclusions of scenarios 2 and 3, the magnitude of the treatment difference and width of the confidence interval in scenario 2 may encourage the conduct of another study (possibly larger) before deciding that the test drug has no effect

FDA guidelines

Figure 2. Possible Results of an NI Study Showing Control Drug–Test Drug Differences (Point Estimate and 95% CI)



1. Point estimate of C-T is 0, suggesting equal effect of C and T; upper bound of the 95% CI for C-T is 1, well below M_1; NI is demonstrated.
2. Point estimate of C-T favors C; the upper bound of the 95% CI for C-T is >2, above M_1; NI is not demonstrated.
3. Point estimate of C-T is zero, which suggests an equal effect; but the upper bound of the 95% CI for C-T is >2 (i.e., above M_1), so that NI is not demonstrated.
4. Point estimate favors T; NI is demonstrated, but superiority is not demonstrated.

FDA guidelines

IV. CHOOSING THE NON-INFERIORITY MARGIN AND TESTING THE NON-INFERIORITY HYPOTHESIS

A. Introduction

In this section, various sources of uncertainty that come into play for NI trials are discussed, with particular emphasis on those affecting the choice of margin. We also discuss in more detail the choice of statistical test for the NI hypothesis. As described briefly in section III, there are two different approaches to analysis of the NI study, one called the *fixed margin method* and the other called the *synthesis method*. Both use the same data from the historical studies and NI study, but in different ways.

● With the fixed margin method, the margin M_1 is based upon estimates of the effect of the active comparator in previously conducted studies. The NI margin to be ruled out in the NI study is then prespecified, and it is usually chosen as a quantity smaller than M_1 (i.e., M_2) to ensure that a reasonable fraction of the effect of the control is preserved. The NI study is successful if the results rule out with a sufficient level of confidence (e.g., 97.5%) inferiority of the test drug to the control by an amount equal to the NI margin or more. It is referred to as a fixed margin method because the past studies comparing the control drug with placebo are used to derive a single fixed value for M_1. The value typically chosen is the lower bound of the 95% confidence interval about the treatment effect of a single placebo-controlled trial or meta-analysis of such trials and represents a conservative estimate of the effect the active control drug is expected to have in the NI study. If M_1 is ruled out by the 95% confidence interval upper bound for C-T, then the conclusion is made that the test drug has an effect. If M_2 is ruled out, then the effect of the test drug has been shown to preserve a clinically important fraction of the effect of the control drug.

● The synthesis method, derived from the same data, combines (or synthesizes) the estimate of treatment effect relative to the control from the NI trial with the estimate of the control effect from a meta-analysis of historical trials. This method treats both sources of data as if they came from the same randomized trial, to project what the placebo effect would have been had the placebo been present in the NI trial. The process makes use of the variability from both the NI trial and the historical trials and yields one confidence interval for testing the NI hypothesis that the treatment rules out loss of a prespecified fixed fraction of the control effect without actually

FDA guidelines

B. Statistical Uncertainties and Quantification of the Active Control Effect (M_1)

1. What Are the Sources of Uncertainty in an NI Study?

The interpretation of an NI study depends on three linked critical conclusions:

1. That there is reliable information about the effect the active control drug had in past studies
2. That there is reason to believe the effect the active control drug has in the current NI study is similar to the effect observed in past studies
3. That the NI study provides reliable information about the effect of the test drug relative to the comparator

All three sources of information are subject to uncertainty. For the first and third, the uncertainty is largely of a statistical nature, measured by standard errors (for the synthesis method) and confidence intervals (for the fixed margin method). The second conclusion is subject to scientific uncertainty that is largely unquantifiable.

The first source of uncertainty is the estimation of the effect of the active control in past studies. Particular problems arise when there is only a single historical study of the active control versus placebo because there is then no information about study-to-study variability; when there are multiple studies but substantial inconsistency in the estimates of the size of the effect among them; and when data from several pharmacologically related drugs are used to develop the estimate for the effect of the active control. All three of these potential problems need to be considered when choosing M_1.

The second source of uncertainty is not statistically based but rather arises from the concern that the magnitude of effect estimated from past studies will be larger than the effect of the active control in the current NI study. Assuming that the effect will be unchanged is often referred to as the "constancy assumption." If the assumption is incorrect and the magnitude of the effect in the current NI study is smaller than the estimated effect from historical studies, M_1 will have been incorrectly chosen (too large) and an apparently successful study showing NI could have given an erroneous result. Lack of constancy can occur for many reasons, including advances in adjunctive medical care, differences in the patient populations, or changes in the assessment of

FDA guidelines

As previously discussed, the assumption of constancy of the effect of the active comparator between the historical studies and the NI study is crucial to a conclusion of efficacy based on non-inferiority. There are, however, technical difficulties in defining what is meant by a constant effect across studies in different populations and at different times. The mathematical way in which the treatment effect is formulated will affect the meaning and plausibility of the constancy assumption.

Consider, for example, studies of fairly infrequent, binary outcomes, typical of most cardiovascular studies, a situation in which NI trials have been successfully employed. Suppose the chance of an event in the placebo group of a historical study was p_1, and in the active treatment group p_2. Then the treatment effect in the historical study can be expressed in several ways:

1. The ratio p_2/p_1, called the relative risk or risk ratio
2. The relative risk reduction $1 - p_2/p_1$
3. The odds ratio $[p_2/(1 - p_2)]/[p_1/(1 - p_1)]$
4. The risk difference $p_1 - p_2$
5. The number needed to treat (NNT), $1/(p_1 - p_2)$

If p_1 and p_2 vary across studies, they might nevertheless vary in such a way that p_2/p_1 is constant. For example, if p_1 and p_2 are 0.10 and 0.05 in one study, and 0.06 and 0.03 in another study, the ratio p_2/p_1 is 0.5 in both studies. The relative risk reduction will then also be the same in both studies, and the odds ratio will be approximately the same. The risk difference, however, is 0.05 in one study and 0.03 in the other, and the numbers needed to treat are 20 and 33.

In contrast, if two studies have different p_1 and p_2 but the same risk difference, they will have the same NNT but cannot have the same relative risk, relative risk reduction, or odds ratio. Thus, the very definition of *constancy* depends on how the effect is formulated mathematically. Because the interpretation of a NI study relies crucially on the assumption of constancy, careful consideration must be given to the question of what, if anything, is likely to be constant across historical studies and between the historical studies and the NI study. As noted, based on both experience and expectation, the constancy assumption for outcome studies has generally been based on expected constancy of relative and not absolute effects.

Protocol: power calculation

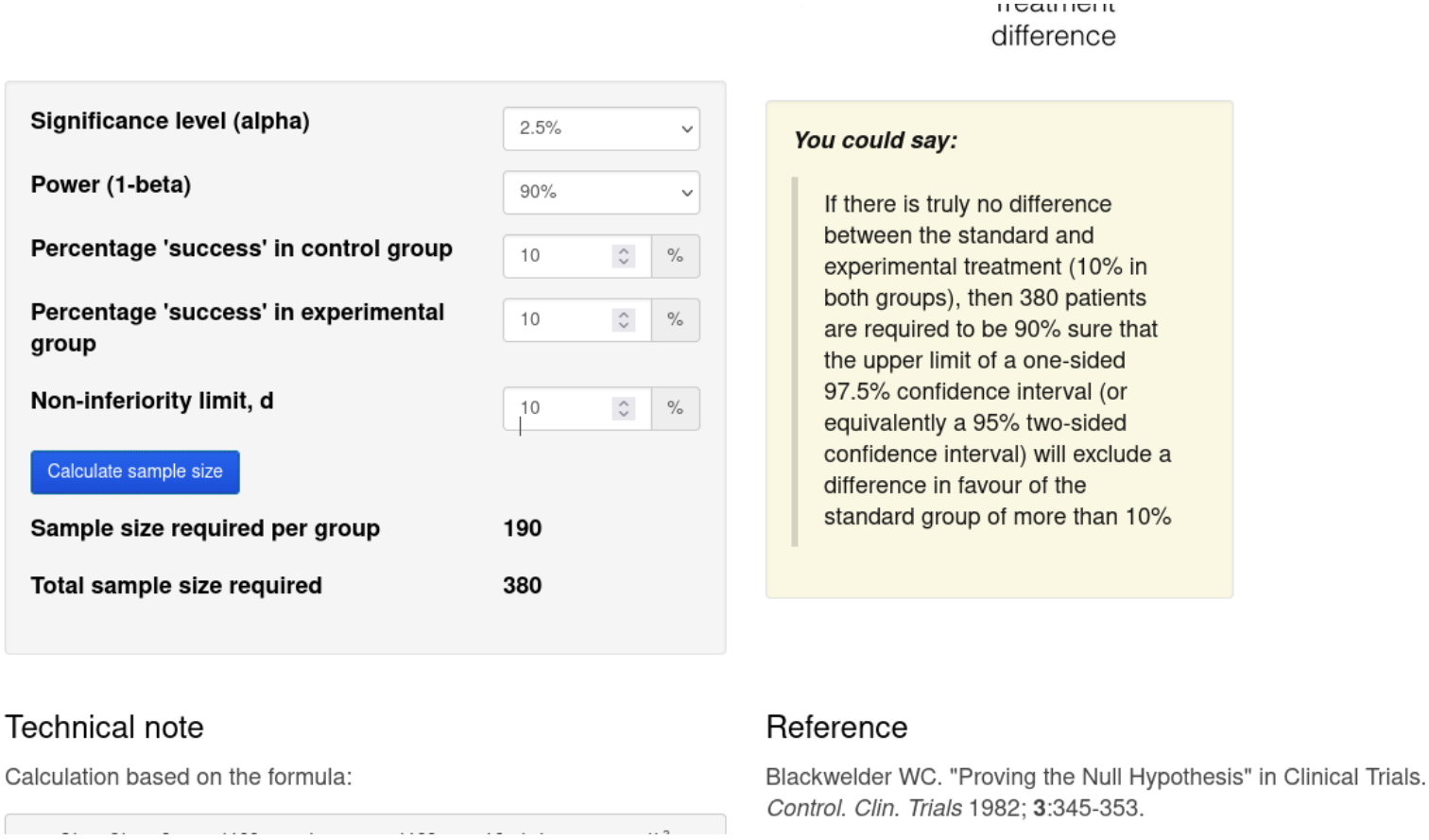
Due to the relatively low incidence of endpoints with the standard treatment on one hand, and the potentially large economic and psychological benefits of oral treatment on the other, we have chosen a Δ value of 10% risk difference. The power to confirm non-inferiority is set at 90%, with a one-sided confidence interval of 97.5%, which would require a sample size of

200-380, based on an event rate of 5 to 10%. Assuming a 10% event rate in the standard arm,
58

---

power of 90%, delta of 10% and a one-sided confidence interval of 97.5%, with 5% of subjects expected to be lost to follow-up or withdrawal, a total of 400 patients with 200 patients in each arm is required.

Power calculation for non-inferiority [1]



[1] https://www.sealedenvelope.com/power/binary-noninferior/

Confidence interval for proportion: which method?

intervention and control groups. Patients will be followed as one cohort and data will be analyzed as randomized. Patients will be included at randomization and are followed until 6 months after end of endocarditis treatment. Any patient lost to follow-up will be censored at the last time known to be alive. Patients will be followed until occurrence of first primary endpoint or 6 months.. The primary analysis is an intention-to-treat analysis. Results from a per-protocol analysis will also be reported,

Non-inferiority will be considered to have been confirmed if the 97.5% one-sided confidence interval for the primary endpoint does not include Δ. The primary endpoint will be treated as a binary variable, with statistical analysis based on difference in proportions. Confidence in- tervals will be calculated using Newcombes method. We do not expect patients to be lost to follow-up regarding the primary outcome. If patients are lost to follow-up differences in out-

Newcombe's method: with or without continuity correction? [2]

10. Newcombe (Score) without continuity correction

The Wilson (Score) confidence limits without continuity correction for each individual binomial proportion, $p_i = x_i/n_i$, (i = 1, 2) is given by

$$\frac{1}{2(n_i + z^2)}\left((2n_i\hat{p}_i + z^2) \pm z\sqrt{4n_i\hat{p}_i(1-\hat{p}_i) + z^2}\right)$$

Denote the lower and upper Wilson (Score) confidence limits for p_i as L_i and U_i. The Newcombe (Score) confidence limits are given by

Lower limit: $(\hat{p}_1 - \hat{p}_2) - \sqrt{(\hat{p}_1 - L_1)^2 + (U_2 - \hat{p}_2)^2}$
Upper limit: $(\hat{p}_1 - \hat{p}_2) + \sqrt{(U_1 - \hat{p}_1)^2 + (\hat{p}_2 - L_2)^2}$

[2] Constructing Confidence Intervals for the Differences of Binomial Proportions in SAS ® Will Garner, Gilead Sciences, Inc., Foster City, CA

Newcombe's method: with or without continuity correction?

Constructing Confidence Intervals for the Differences of Binomial Proportions in SAS®, Continued

11. Newcombe (Score) with continuity correction

    The method is the same as the Score method (Method 10) above, but the confidence intervals for each individual binomial proportion are obtained using the Wilson (Score) confidence limits with continuity correction, given by the following

$$\frac{2n_i\hat{p}_i + z^2}{2(n_i + z^2)} \pm \frac{\left(1 + z\sqrt{z^2 - 2 - \frac{1}{n_i} + 4\hat{p}_i(n_i(1-\hat{p}_i) + 1)}\right)}{2(n_i + z^2)}$$

Note: The profile likelihood methods (Methods 8 and 9) are not discussed further here. Please see Newcombe (1998) for more information.

Methods 1, 2, 10, and 11 are available in SAS® 9.3, and Methods 5 and 6 were newly added in SAS® 9.4. Methods 3 and 4 can be computed using the formulas provided. Appendix A contains SAS® code that can be run in earlier versions (prior to 9.4) that can compute Methods 5 and 6. This code also computes Method 7. (Note: Method 7 is not available via SAS® 9.4 procedures.)

Survival analysis

Five-years Outcome of Partial Oral Treatment of Endocarditis - The POET Trial

Supplements

This supplement contains the following items:

1. Original protocol, final protocol, summary of changes.
2. Final statistical analysis plan, summary of changes (short-term study).
3. Final statistical analysis plan (long-term study).
4. Final statistical analysis plan (five-years study).

Statistical plan for main outcome paper

Statistical analyses will be performed using SPSS, SAS and R.

Baseline description of groups:

Table of summary statistics for each group (age, sex, results of routine blood tests, comorbidities, microbiological findings, valve involved, prior valve disease, cardiac devices).
Continuous variables will be summarized with: n (non-missing sample size), mean, standard deviation, median, interquartile range, number of missing values as appropriate. Categorical variables will be reported as frequency and percentages (based on non-missing sample size) and number of missing values.

Primary outcome analysis:

The primary outcome assessment will be performed as an **analysis of difference in the combined endpoint between the intervention and control groups, i.e. similarly to the analysis performed and published for the original (short-term) study**. Patients will be followed as one cohort and data will be analyzed as randomized. Patients will be included at randomization and are followed until 10th of July 2020. Any patient lost to follow-up will be censored at the last time known to be alive. Patients will be followed until occurrence of first primary endpoint or until 10th July 2020. The primary analysis is an intention-to-treat analysis. Results from a per-protocol analysis will also be reported,

Due to the long follow-up time the primary endpoint will be not treated as a binary variable, but time to event will be compared using cox regression analysis. The proportional-hazard assumption

Protocol page 88

Secondary endpoints will be analyzed similar to the primary endpoint. Sensitivity analyses will not be performed for the secondary endpoints. Hazard ratios will be reported for the secondary endpoints using cox regression due to competing risk (mortality).

Further analysis

Causes of death will be reported and difference in distribution of death between groups will be assessed using Chi square.

Handling of missing data:

The primary outcome analysis should be subject to no or little missing data as it is based on objective data from the patient records.

Statistical analysis section (continued)

as absolute numbers and frequencies and were compared with the chi-square test, including Yates' correction for continuity. Logistic-regression analysis was used to calculate odds ratios for the primary outcome in prespecified subgroups. Cox regression analysis was used to assess the components of the primary composite outcome to address competing risks (e.g., death). The proportional-hazard assumption was assessed with Schoenfeld residuals. All analyses were performed according

Protocol: Kaplan-Meier

regarding the primary outcome. If patients are lost to follow-up differences in out-

come will be estimated by the Kaplan-Meier method and confidence limits will be calculated using the method, suggested by Altman et al (1999)

Kaplan-Meier plots will be used to illustrate event free survival.

Lost without a statistician

PRIMARY OUTCOME

All enrolled patients were followed for 6 months after the antibiotic treatment was completed or until death. No patients were lost to follow-up. The primary composite outcome occurred in a total of 42 patients (10.5%) — in 24 patients (12.1%) in the intravenously treated group and in 18 (9.0%) in the orally treated group (odds ratio, 0.72; 95% confidence interval [CI], 0.37 to 1.36). The between-group difference was 3.1 percentage points (95% CI, –3.4 to 9.6; P=0.40) in favor of oral treatment, and the criterion for noninferiority was therefore met. In the per-protocol

Kaplan-Meier?

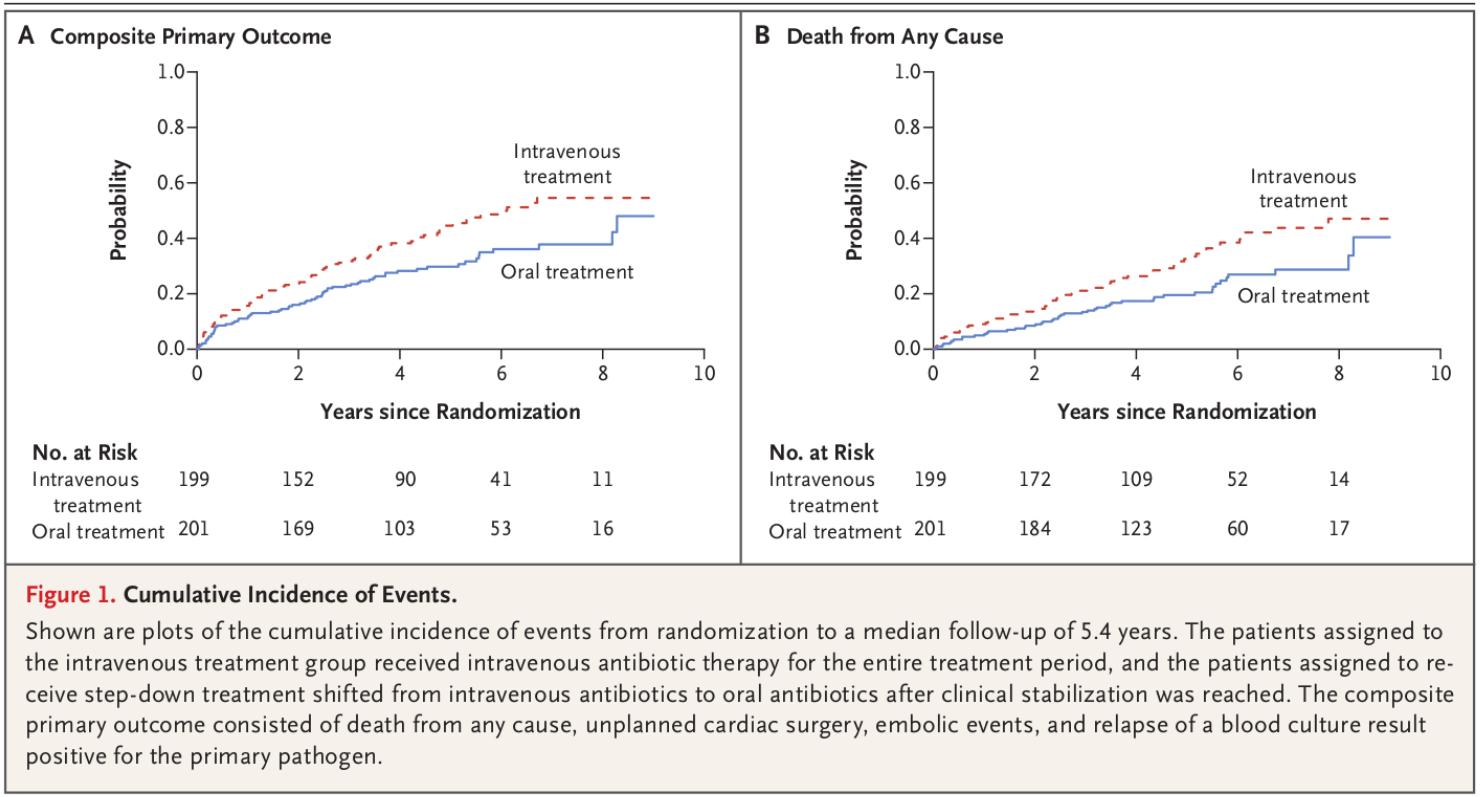


Figure 1. Cumulative Incidence of Events.

Shown are plots of the cumulative incidence of events from randomization to a median follow-up of 5.4 years. The patients assigned to the intravenous treatment group received intravenous antibiotic therapy for the entire treatment period, and the patients assigned to receive step-down treatment shifted from intravenous antibiotics to oral antibiotics after clinical stabilization was reached. The composite primary outcome consisted of death from any cause, unplanned cardiac surgery, embolic events, and relapse of a blood culture result positive for the primary pathogen.

How to calculate median follow-up?

Competing risks

to the intention-to-treat principle. A per-protocol analysis is also presented for the primary outcome; in the per-protocol analysis, patients who crossed over from their assigned treatment to the other treatment were excluded. Cumulative incidences were calculated for events with competing risk (death) for the outcomes of unplanned cardiac surgery, embolic events, and relapse of bacteremia with the primary pathogen. Two-sided P values of less than 0.05 were considered to indicate statistical significance. Analyses were performed with the use of SPSS software, version 22.0 (IBM), and R software, version 3.3.3 (R Foundation for Statistical Computing).[22-24]

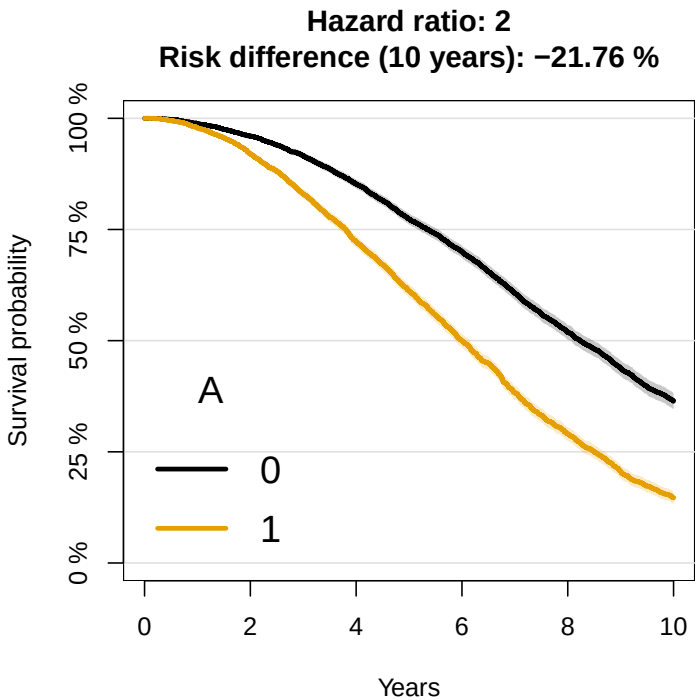
Achieved endpoints

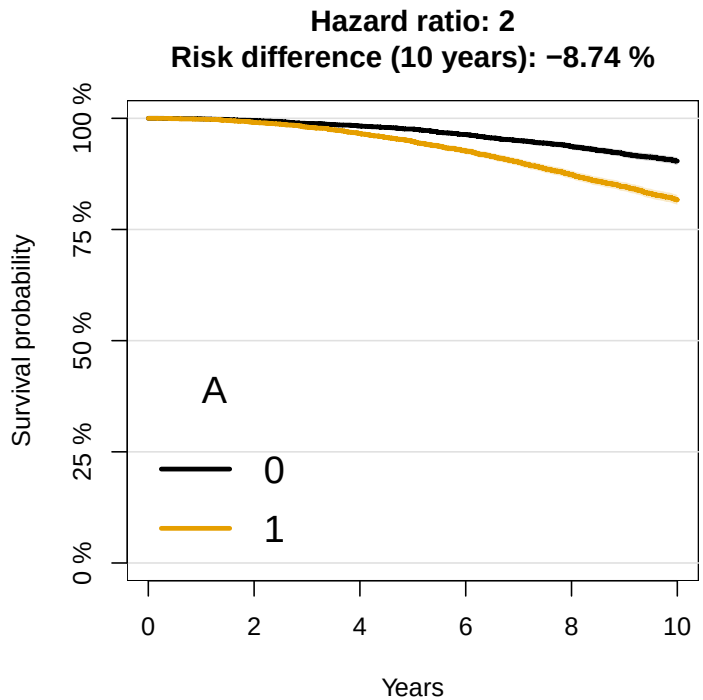
Table S12

Details of the achieved endpoints

	Intravenous treatment	Oral treatment
Causes of death	**12**	**8**
Infection and endocarditis, n (%)	2 (16.7)	2 (25.0)
Infection, not endocarditis, n (%)	2 (16.7)	2 (25.0)
Sudden cardiac death, n (%)	4 (33.3)	0 (0)
Heart failure, n (%)	1 (8.3)	0 (0)
Cerebral haemorrhage, n (%)	1 (8.3)	1 (12.5)
Cancer, n (%)	2 (16.7)	1 (12.5)
Lung disease, (n%)	0 (0)	1 (12.5)
Renal failure, n (%)	0 (0)	1 (12.5)
Reasons for unplanned cardiac surgery	**6**	**6**
Worsening/relapse of infection, n (%)	2 (33.3)	2 (33.3)
Valve dysfunction, no infection, n (%)	3 (50.0)	4 (66.7)
Hematoma in the pericardium, n (%)	1 (16.7)	0 (0)
Type of embolic event	**3**	**3**
Cerebral emboli, n (%)	2 (66.7)	2 (66.7)
Emboli in the eye, n (%)	1 (33.3)	1 (33.3)

Why one cannot interpret effects obtained on the hazard scale as effects directly on the survival scale





No competing risks!

Subgroups

Subgroups

Subgroup analyses of primary endpoint:

A forrest plot will be constructed of the primary endpoint including the following subgroups; Gender, diabetes, renal disease, valve prosthesis/native valve, bacteria species (enterococcus, coagulase negative staphylococcus, staphylococcus aureus, streptococci, age above/below median and surgically treated endocarditis. Odds ratios will be reported for the subgroups using logistic regression.

Subgroup analyses

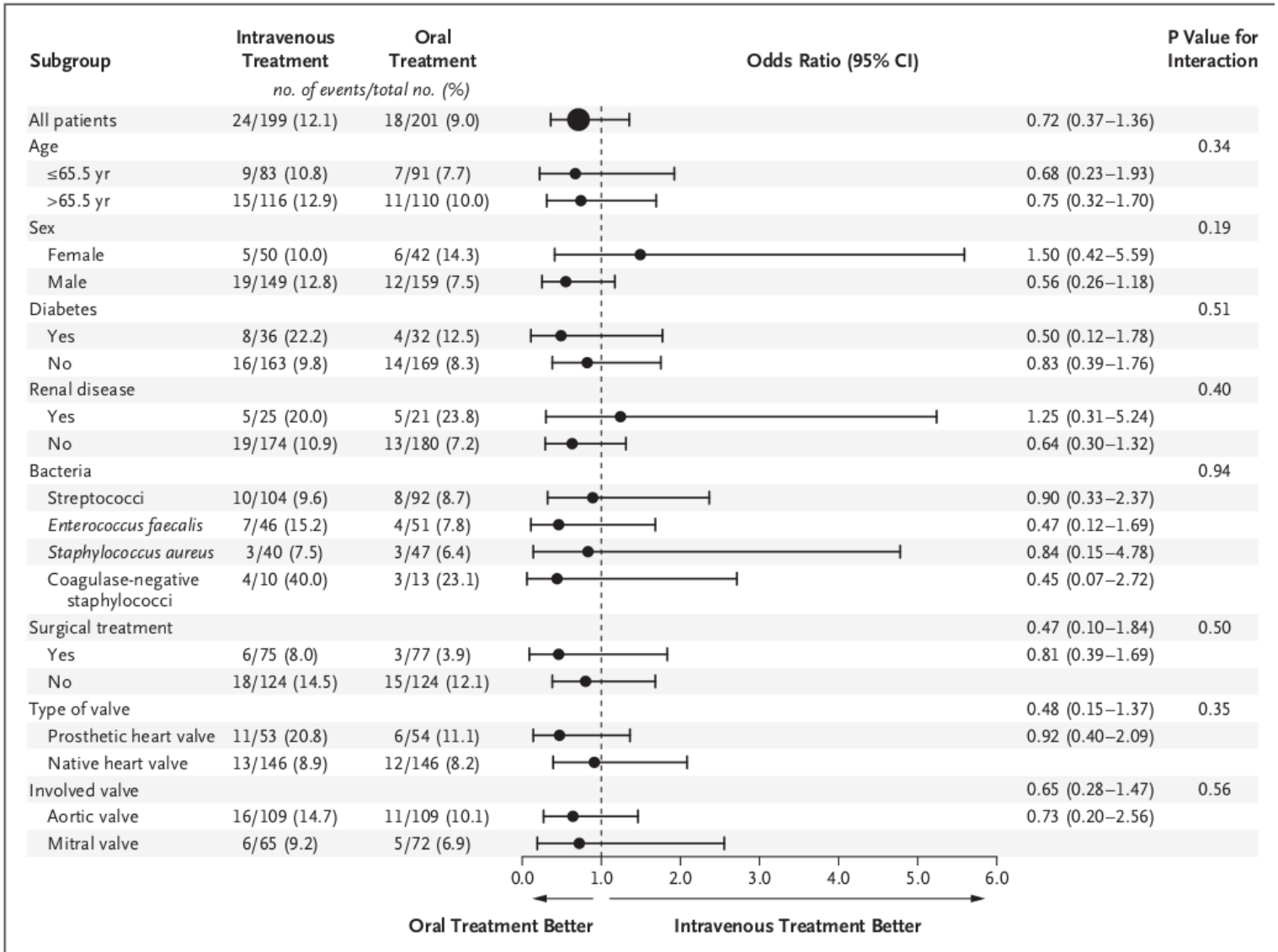


Figure 3. Rates of the Primary Outcome in Prespecified Subgroups.

Discussion

- non-inferiority planned but analyzed as usual?
- multicenter study but center is not a variable in the analyses
- logistic and Cox regression models are not collapsible
  - other variables that affect outcome are ignored
  - attenuation of the true effect
  - loss of power = too large confidence intervals
- Schoenfeld residuals not reported
- subgroup analyses
  - purpose?
  - other variables that affect outcome are ignored: loss of power!
  - different logistic regression models are incompatible with each other
  - multiple testing?
  - center not used?

https://chat.openai.com/chat

Please rewrite as a poem

In planning an RCT,
A statistician is key.
For a trial randomized,
Their expertise cannot be denied.

They'll help design the study right,
With sample size and protocols tight.
Minimizing bias, increasing precision,
Ensuring the data's not missin'.

Analysis is their expertise,
Selecting models with great ease.
Addressing confounding factors galore,
Identifying sources of variability, and more.

Without their aid, errors may arise,
Leading to conclusions that belie
The truth of the study's intent,
A danger for participants and treatment.

So, let the statistician join the fray,
A partner to guide the study's way.
For with their help, a reliable result,
Ensures that in practice, it won't be difficult.