Published in final edited form as:

J Clin Epidemiol. 2015 December ; 68(12): 1504–1511. doi:10.1016/j.jclinepi.2015.03.016.

Many multicentre trials had few events per centre, requiring analysis via random-effects models or GEEs

Brennan C Kahan^{1,*} and Michael O Harhay^{2,*}

¹Pragmatic Clinical Trials Unit, Queen Mary University of London, 58 Turner Street, London, UK, E1 2AB

²Division of Epidemiology, Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA, 19146

Abstract

Objectives—Adjustment for center in multicenter trials is recommended when there are between-center differences or when randomization has been stratified by center. However, common methods of analysis (such as fixed-effects, Mantel-Haenszel, or stratified Cox models) often require a large number of patients or events per center to perform well.

Study design and setting—We reviewed 206 multicenter randomized trials published in four general medical journals to assess the average number of patients and events per centre, and determine whether appropriate methods of analysis were used in trials with few patients or events per centre.

Results—The median number of events per center/treatment arm combination for trials using a binary or survival outcome was 3 (IQR 1–10). Sixteen percent of trials had less than 1 event per center/treatment combination, 50% fewer than 3, and 63% fewer than 5. Of the trials which adjusted for center using a method of analysis which requires a large number of events per center, 6% had less than 1 event per center-treatment combination, 25% fewer than 3, and 50% fewer than 5. Methods of analysis which allow for few events per center, such as random-effects models or GEEs, were rarely used.

Conclusions—Many multicenter trials contain few events per center. Adjustment for center using random-effects models or GEE with model-based (non-robust) standard-errors may be beneficial in these scenarios.

*Correspondence to b.kahan@qmul.ac.uk.

*BCK and MOH contributed equally to this manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Contributor: BCK and MOH jointly came up with the concept, and co-wrote the manuscript. BCK developed the data extraction form, and extracted the majority of the data. Both authors gave final approval for the manuscript to be submitted.

Competing interests: The authors declare they have no conflicts of interest.

Background

Randomized controlled trials (RCT) often recruit participants from multiple centers to increase accrual rate and enhance the generalizability of results by allowing assessment of the intervention in a variety of settings. Patient outcomes sometimes vary by center, due to factors such as differing patient characteristics, methods of measuring or recording data, processes of care, or training of staff [1–10].

When between-center differences are large, accounting for center in the trial analysis can offer benefits over an unadjusted analysis. In particular, adjustment for center will (a) account for chance imbalances in the number of patients assigned to each treatment within a center [1, 2, 11]; (b) increase power [1, 2, 10–14]; and (c) lead to correct results when randomization has been stratified according to center [1, 2, 11, 13, 15–17].

However, adjusting for center is often more complex than adjusting for other covariates, such as age or gender. This is because center typically has a higher number of categories than other covariates, meaning there are fewer patients within each category. Most methods of analysis make assumptions regarding the number of patients or events per center, and if these assumptions are violated then results may be adversely affected [1–3]. In particular, a poorly chosen method of analysis can lead to problems such as some patients or centers being excluded from the analysis [1, 2], biased estimates of treatment effect [3], or inflated type I error rates [1, 3]. We therefore undertook a review of published trials to determine how often the assumptions regarding the number of patients or events per center are met in practice, and whether trials are using appropriate methods of analysis when these assumptions are violated.

Methods

We begin by discussing when adjustment for centre-effects in multicenter trials is useful, and we then compare different methods of performing an adjusted analysis.

Adjustment for centre-effects: rationale

The primary benefit of adjusting for centre-effects is that it can increase power when there are large differences in patient outcomes between different centres [1, 2]. This may occur when different types of patients present to different centres, when outcome assessment varies across centres, or when some centres are more effective at preventing adverse outcomes than others. The extent to which power can be increased depends on the difference between centres; the larger the differences, the larger the increase in power.

Adjustment for centre-effects will not benefit all trials; if there are very small differences between centres then an adjusted analysis will give similar results to an unadjusted analysis. Therefore, it is recommended to adjust for centre-effects if there are expected to be large differences between centres, or if centre was used as a stratification factor during randomization (as in this case, not adjusting for centre can lead to incorrect estimates of the standard error for treatment effect, adversely affecting the type I error rate) [1, 2, 13, 15–17].

The decision on whether (and how) to adjust for center should be pre-specified in the trial protocol or statistical analysis plan, before investigators have access to the trial data.

Methods of adjusting for centre-effects

We describe some common methods of adjusting for center in multicenter trials, and then discuss the assumptions made by each. These points are summarized in table 1. More comprehensive descriptions of each method is available elsewhere [1–4, 6, 7]. Additionally, we do not discuss issues such as conditional vs. marginal treatment effects, or treatment by center interactions, both of which have been discussed previously [1–3, 6].

Commonly used (or recommended) methods of analysis include fixed-effects regression models, random-effects regression models, generalized estimating equations (GEEs), Mantel-Haenszel (for binary outcomes), and stratified Cox models (for time-to-event outcomes).

Within-center methods: fixed-effects, Mantel-Haenszel, and Stratified Cox models—Fixed-effects regression models, Mantel-Haenszel methods, and stratified Cox models are all analysis methods which rely on within-center comparisons; that is, they estimate the treatment effect separately within each center, and combine the estimates to produce an overall result.

A fixed-effects regression analysis is performed by incorporating an indicator (or dummy) variable for all centers but one into the regression model, similar to how a variable such as gender or ethnicity would be included. Likewise, Mantel-Haenszel and stratified Cox models can be easily performed in most standard statistical software packages.

Random-effects models—Random-effects regression models (also called ‘frailty’ models for survival outcomes) include center as a random intercept, which is generally assumed to follow a normal distribution. Unlike the methods above, random-effects models combine within- and between-center estimates to produce an overall treatment effect. The within-center estimate is produced in a similar manner to above (results are estimated separately within each center and then combined); the between-center estimate is produced from a regression analysis of the center-level summaries [2, 18].

Generalized estimating equations (GEE)—GEEs produce marginal estimates of treatment effect [1], similar to a regression model that has not been adjusted for center. They then attempt to correct the standard errors (SEs) to account for the clustering [19]. This is done by selecting a correlation structure for the data (for example, that patients within a center are all equally correlated). Two different types of SEs can then be used; (a) robust ‘sandwich-estimator’ SEs, which ensure that even if the specified correlation structure is incorrect, the type I error rate will be valid; or (b) nonrobust model-based SEs, which assume the correlation structure has been correctly specified.

Assumptions regarding the number of patients or events per center

A within-center comparison, where treatment effects are estimated separately within each center, relies heavily on the assumption that there is an adequate number of patients or

events within each center/treatment combination (a trial with 10 centers and 3 treatment arms would have 30 center/treatment combinations). A low number of patients or events per center/treatment combinations will often lead to some centers and patients being excluded from the analysis.

This occurs because centers for which a treatment effect cannot be calculated are excluded from the analysis of all comparisons involving that treatment arm. A treatment effect cannot be calculated when (a) there are no patients allocated to one of the treatment groups involved, within that center [1, 2, 8]; (b) all patients on one of the treatment arms within a center experience the same outcome (for binary outcomes calculated using an odds ratio) [1, 2, 8]; or (c) no patients on one of the treatment arms within the center experience an outcome (for survival outcomes).

This will reduce both precision and power, leading to less reliable results. It also raises some ethical concerns about arbitrarily excluding patients from the analysis (thus denying them the opportunity to allow their experience to benefit future patients) after exposing them to the risks of the trial [2].

Key assumptions for each method of analysis

Fixed-effects regression models—As fixed-effects models are based solely on a within-center comparison, they assume an adequate number of patients or events within each center/treatment combination; when this assumption is violated, they may lead to reduced power [1, 2]. Furthermore, for a binary or survival outcome, they require a relatively small number of centers, as including too many centers can lead to biased estimates of treatment effect and an increased type I error rate [1, 6, 20].

Mantel-Haenszel and Stratified Cox models—Both these methods rely on within-center comparisons, and so assume an adequate number of patients or events for each center/treatment combination; as above, these analysis methods may lead to reduced power when this assumption is violated [1, 6]. However, unlike fixed-effects, they make no assumptions regarding the total number of centers used.

Generalized estimating equations (GEE)—As GEEs are not based on within-center comparison, they make no assumptions regarding the number of patients or events per center/treatment combination. However, if using robust SEs, they require a large number of centers; when this is not the case, the risk of a false-positive will be increased. The number of centers required to ensure adequate results may be prohibitive; it has been shown that even with 100 centers, error rates can still be too high [1]. Using model-based SEs does not require as many centers, and can be safely used with as few as 5 [1]. However, model-based SEs should only be used with an exchangeable correlation structure within centers (rather than an independent correlation structure, for example). Alternatively, robust SEs could be used provided a small sample correction is implemented [21].

Random-effects—Similarly to GEEs, random-effects models do not rely on within-center comparisons, and so make no assumptions regarding the number of patients or events per center/treatment combination. They do assume an adequate number of centers, and that the

center-effects follow a normal distribution. However, in practice, 5 or more centers has been shown to be adequate, and random-effects models are quite robust to misspecification of the randomeffects distribution [1, 2, 6]. In some cases where there is a small number of centers, or a small number of patients, it may be useful to incorporate a degree-of-freedom correction to ensure valid type I error rates [2, 22]

Recommended methods of analysis for multicenter trials

A large amount of research has compared the different methods of analysis in a variety of settings [1, 2, 4, 6–8]. For continuous outcomes, it is generally agreed that random-effects models will give good results in almost all scenarios [2, 4, 7]. Fixed-effects models can be used provided the number of patients per center is not too small [1, 2].

For binary or time-to-event outcomes, random-effects models have been widely recommended as they perform well in most scenarios [2, 6]. GEEs with robust SEs are not recommended as they require a restrictively large number of centers (unless a small sample correction is incorporated); however, GEEs with model-based SEs (using an exchangeable correlation structure) perform as well as random-effects in most scenarios [1]. Fixed-effects models should only be used with a small number of centers, and a large number of patients per center; otherwise, they can lead to biased estimates of treatment effect, and inflated type I error rates [1]. Stratified Cox models and Mantel-Haenszel methods can be used provided there is an adequate number of patients and events in each center, but can lead to reduced power otherwise [1, 6].

What impact does our choice of analysis method have in practice?

Figure 1 shows the impact that using a within-center method of analysis (fixed-effects, Mantel-Haenszel, stratified Cox model) can have on results when there are a small number of patients or events per centre. For a trial with 1 event per centre on average (based on 5 patients per centre, and an event rate of 20%) that has an equal number of patients in each centre, using a within-center approach to the analysis would lead to 83% of patients being excluded from the analysis on average. Increasing the average number of events per centre to 5 still leads to 12% of patients being excluded, and 10 events per centre leads to a 1% exclusion rate. This can have a large impact on power, and may affect both the internal and external validity of the trial.

Recent research has shown that the choice of analysis method can have a large impact on results. A re-analysis of the MIST2 trial [23] for the outcome of surgery up to 90 days (a binary outcome) found that adjustment for centre using fixed-effects regression or Mantel-Haenszel led to 15% and 52% of patients being excluded from the analysis respectively [1] (results are shown in table 2). This resulted in the odds ratios (ORs) for fixed-effects and Mantel-Haenszel being larger than the ORs for random-effects or GEEs (OR=3.34 and 3.59 for fixed-effects and Mantel-Haenszel respectively, compared to 2.72 and 2.57 for randomeffects and GEEs). This led to results being statistically significant for fixed-effects and Mantel-Haenszel, but not for random-effects or GEEs.

Reitsma *et al* [8] found similar, although less dramatic results. For the outcome of preterm birth (a binary outcome), they found that fixed-effects regression and Mantel-Haenszel led to

35 centres (330 patients, 19%) and 3 centres (3 patients, <1%) being excluded from the analysis respectively. This led to larger ORs for fixed-effects and Mantel-Haenszel (OR=1.57 and 1.55 respectively) compared to random-effects and GEEs (1.49 and 1.46 respectively). As above, results were statistically significant for fixed-effects and Mantel-Haenszel, but not for random-effects and GEEs.

Review of published trials

We used the same set of articles used in a previous review that one of the authors (BCK) was involved in [15]. Full details of the search strategy and eligibility criteria are available elsewhere [15]. Briefly, we included parallel group, individually randomized, controlled trials which were published in one of four major medical journals in 2010 (*BMJ*, *Journal of the American Medical Association*, *Lancet*, and *New England Journal of Medicine*). For this current review we only included trials that identified themselves as multicenter.

For eligible trials, one author (BCK) extracted data on the number of centers, the number of patients, and, for trials with a binary or survival primary outcome, the number of events. For trials which stratified the randomization procedure according to center, we also determined which method of adjustment they used (if an adjusted analysis was indeed performed). A second author (MOH) extracted data on a subset of trials (n=10) to assess agreement. Agreement was determined for the number of centers included, the number of patients, the number of events (if applicable), and the method of analysis. Overall agreement between authors was 91%.

Some trials had multiple types of centers (for example, both hospitals and doctor's offices). In these scenarios, we use whichever type of center had been used as a stratification factor in the randomization process. If center had not been used in the randomization process, we used the center from which patients were recruited.

Results

Reporting

Of the 258 trials which were eligible for inclusion, 206 (80%) were multicenter, 22 (9%) were single center, and 30 (12%) did not state whether they were single or multicenter. Only 178 of the 206 multicenter trials (86%) gave information on the number of centers included. Of 151 trials whose primary outcome was binary or survival, 13 (9%) did not give information on the number of events that occurred, meaning we could not calculate the average number of events per center.

Number of centers, patients, and events

Results are shown in table 3. Of the 178 trials which provided information on the number of centers included, the median number of centers used was 20 (IQR 8 to 70). 56 trials (31%) used less than 10 centers, whereas 55 (30%) used more than 50. The median number of patients per center was 24 (10 to 76). 13 trials (7%) had fewer than 5 patients per center, and 48 (27%) fewer than 10. The median number of events per center for trials with a binary or

survival outcome was 6 (IQR 3 to 21). 35 trials (28%) had fewer than 3 events per center, and 54 (43%) fewer than 5.

Of these 178 trials, 147 (83%) used two treatment arms, 22 (12%) used 3 arms, and the remainder used between 4 and 9 arms. The median number of patients per center/treatment combination was 12 (IQR 4 to 37), and 52 trials (29%) had fewer than 5 patients for each center/treatment combination. The median number of events per center/treatment combination was 3 (IQR 1 to 10); 20 trials (16%) had less than 1 event for each center/treatment combination, 62 (50%) fewer than 3, and 78 (63%) fewer than 5.

Adjustment for center

Results are shown in table 4. 120 trials stratified by center during the randomization process (49 binary, 21 continuous, 44 survival, 6 other). Only 35 (29%) adjusted for center in their analysis. This differed according to outcome type; trials with a binary (11/49, 22%) or survival outcome (9/44, 20%) were much less likely to adjust for center than trials with a continuous outcome (13/21, 62%) ($p=0.001$ for both comparisons).

This may have been because investigators were concerned about adjustment with a low number of events for trials with binary or survival outcomes; trials which did not adjust had fewer events per center (median 5, IQR 3 to 20) compared with trials that did adjust (median 17, IQR 6 to 43) ($p=0.05$ Wilcoxon rank-sum test). A similar trend was seen for events per center/treatment combination (unadjusted 3 (1 to 10) vs. adjusted 5 (3 to 22)), although was not statistically significant ($p=0.08$ by Wilcoxon rank-sum test).

Methods of adjustment

Of the 35 trials which adjusted for center, the majority ($n=22$, 63%) used a fixed-effects analysis, and only 3 (9%) used a random-effects model. For trials with a continuous outcome, most used fixed-effects (11/13, 85%), and only one (8%) used random-effects. Most trials with a binary outcome used either fixed-effects (6/11, 55%) or Mantel-Haenszel (3/11, 27%); only 2/11 (18%) used random-effects, and none used GEEs. Most trials with a survival outcome used either fixed-effects (3/9, 33%) or a stratified Cox model (5/9, 56%); none used random-effects.

Overall, 17 trials with a binary or survival outcome adjusted using a 'within-center' method of analysis (either fixed-effects, Mantel-Haenszel, or a stratified Cox model). Of the 16 trials for which we were able to calculate the number of events per center-treatment combination, 1 (6%) had less than 1 event per center-treatment combination, 4 (25%) had fewer than 3, and 8 (50%) fewer than 5.

Discussion

Adjustment for center in the analysis of multicenter trials can be beneficial when there are between-center differences, or when center has been used as a stratification factor during randomization. However, many methods of analysis make strong assumptions regarding the number of patients or events per center, and can produce misleading results when these assumptions are violated. We therefore undertook a review of published trials to determine

whether these assumptions are often met in practice, and whether trials are using appropriate methods of analysis.

We found that multicenter trials are common, however the reporting of key aspects was often poor. For example, 12% of trials did not state whether they were single or multicenter; of the trials that did state they were multicenter, 14% did not give the number of centers; and of the trials using a binary or survival endpoint, 9% did not give adequate information on the number of events that occurred. Given that many methods of analysis rely on key assumptions regarding the number of centers, and number of patients or events per center, it is important that this information is clearly reported to allow readers to assess the validity of the methods used.

The number of events per center in trials with a binary or survival outcome was often quite low. The median was 6, and almost a third of trials had fewer than 3 events per center. When broken down by center/treatment combination, the median was 3, and half of all trials had fewer than 3 events per center/treatment combination, and 16% had less than one. Despite the low number of events per center in many trials, over 80% of studies which adjusted for center in the analysis used a within-center method of analysis (such as fixedeffects, Mantel-Haenszel, stratified Cox model), which can lead to poor results in these scenarios. Methods of analysis which can cope with a small number of events per center, such as random-effects models or GEEs were rarely used; only 3 trials reported using random-effects, and none used GEEs.

Re-analysis from previous trials has found that methods such as fixed-effects and Mantel-Haenszel can lead to a large proportion of patients being excluded from the analysis, which resulted in larger treatment effect estimates and different conclusions from analyses based on random-effects models or GEEs. Given the frequency of trials with a small number of events per center, we would recommend that random-effects models and GEEs with modelbased SEs (with an exchangeable correlation structure) are used much more frequently in practice.

Conclusions

Many multicenter trials contain few events per center. Adjustment for center using randomeffects models or GEE with model-based (non-robust) standard-errors may be beneficial in these scenarios.

Acknowledgments

The authors would like to thank two anonymous reviewers, whose thoughtful comments led to an improved manuscript. Grant numbers R01CA159932 from the National Cancer Institute and 1F31HL127947-01 from the National, Heart, Lung and Blood Institute of the United States National Institutes of Health support MOH.

References

1. Kahan BC. Accounting for centre-effects in multicentre trials with a binary outcome - when, why, and how? *BMC medical research methodology*. 2014; 14(1):20. [PubMed: 24512175]

2. Kahan BC, Morris TP. Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects? *Statistics in medicine*. 2013; 32(7):1136–1149. [PubMed: 23112128]
3. Agresti A, Hartzel J. Strategies for comparing treatments on a binary response with multi-centre data. *Statistics in medicine*. 2000; 19(8):1115–1139. [PubMed: 10790684]
4. Chu R, Thabane L, Ma J, Holbrook A, Pullenayegum E, Devereaux PJ. Comparing methods to estimate treatment effects on a continuous outcome in multicentre randomized controlled trials: a simulation study. *BMC medical research methodology*. 2011; 11:21. [PubMed: 21338524]
5. Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicentre studies: an overview. *Annals of internal medicine*. 2001; 135(2):112–123. [PubMed: 11453711]
6. Munda M, Legrand C. Adjusting for centre heterogeneity in multicentre clinical trials with a time-to-event outcome. *Pharmaceutical statistics*. 2014; 13(2):145–152. [PubMed: 24523155]
7. Pickering RM, Weatherall M. The analysis of continuous outcomes in multi-centre trials with small centre sizes. *Statistics in medicine*. 2007; 26(30):5445–5456. [PubMed: 17924360]
8. Reitsma A, Chu R, Thorpe J, McDonald S, Thabane L, Hutton E. Accounting for center in the Early External Cephalic Version trials: an empirical comparison of statistical methods to adjust for center in a multicenter trial with binary outcomes. *Trials*. 2014; 15:377. [PubMed: 25257928]
9. Biau DJ, Porcher R, Boutron I. The account for provider and center effects in multicenter interventional and surgical randomized controlled trials is in need of improvement: a review. *Journal of clinical epidemiology*. 2008; 61(5):435–439. [PubMed: 18394535]
10. Vierron E, Giraudeau B. Design effect in multicenter studies: gain or loss of power? *BMC medical research methodology*. 2009; 9:39. [PubMed: 19538744]
11. Kahan BC, Jairath V, Dore CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials*. 2014; 15:139. [PubMed: 24755011]
12. Hernandez AV, Steyerberg EW, Habbema JD. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *Journal of clinical epidemiology*. 2004; 57(5):454–460. [PubMed: 15196615]
13. Kahan BC, Morris TP. Assessing potential sources of clustering in individually randomised trials. *BMC medical research methodology*. 2013; 13:58. [PubMed: 23590245]
14. Hernandez AV, Eijkemans MJ, Steyerberg EW. Randomized controlled trials with time-to-event outcomes: how much does prespecified covariate adjustment increase power? *Annals of epidemiology*. 2006; 16(1):41–48. [PubMed: 16275011]
15. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. *BMJ*. 2012; 345:e5840. [PubMed: 22983531]
16. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Statistics in medicine*. 2012; 31(4):328–340. [PubMed: 22139891]
17. Parzen M, Lipsitz S, Dear K. Does clustering affect the usual test statistics of no treatment effect in a randomized clinical trial? *Biometrical Journal*. 1998; 40(4):385–402.
18. Rabe-Hesketh, S.; Skrondal, A. *Multilevel and Longitudinal Modeling Using Stata*. College Station, Texas: Stata Press; 2012.
19. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73(1):13–22.
20. Kahan BC, Morris TP. Adjusting for multiple prognostic factors in the analysis of randomised trials. *BMC medical research methodology*. 2013; 13:99. [PubMed: 23898993]
21. Pan W, Wall MM. Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations. *Statistics in medicine*. 2002; 21(10):1429–1441. [PubMed: 12185894]
22. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997; 53(3):983–997. [PubMed: 9333350]
23. Rahman NM, Maskell NA, West A, Teoh R, Arnold A, Mackinlay C, et al. Intrapleural use of tissue plasminogen activator and DNase in pleural infection. *The New England journal of medicine*. 2011; 365(6):518–526. [PubMed: 21830966]

Figure 1. Number of patients excluded from the analysis based on a within-center approach

This figure shows the proportion of patients who would be excluded from the analysis if a within-center approach was used (e.g. fixed-effects models, Mantel-Haenszel, or a stratified Cox model). We assumed the same number of patients in each centre, the same number of patients on control and intervention within each centre, and a 20% event rate for each patient (sensitivity analyses using different event rates found similar results). We considered patients would be excluded if all the outcomes were the same (either all successes or all failures) in either one of the treatment groups. We calculated these proportions using binomial probabilities.

Table 1

Assumptions for different methods of accounting for center-effects

Method of analysis	Assumptions	Impact on results if assumptions are violated	When this method can be used in practice
Fixed-effects	<ul style="list-style-type: none"> - An adequate number of patients or events per center-treatment combination - A small number of centers if used with a binary or survival outcome 	<ul style="list-style-type: none"> - Can lead to patients or centers being excluded from analysis if there are a small number of patients or events per center-treatment combination, leading to reduced power - Can lead to biased estimates of treatment effect when there is a large number of centers with a binary or survival outcome 	When there is a small number of centers, with an adequate number of patients and events per center-treatment combination
Mantel-Haenszel or stratified Cox model	<ul style="list-style-type: none"> - An adequate number of patients or events per center-treatment combination 	<ul style="list-style-type: none"> - Can lead to patients or centers being excluded from analysis if there are a small number of patients or events per center-treatment combination, leading to reduced power 	When there is an adequate number of patients and events per center-treatment combination
GEEs	<ul style="list-style-type: none"> - A large number of centers if used with robust standard errors - If model-based standard errors are used, 5 or more centers will be adequate 	<ul style="list-style-type: none"> - Can lead to inflated type I error rates if robust standard errors are used without a large number of centers 	<ul style="list-style-type: none"> - When there is a large number of centers if used with robust standard errors - When there are at least a few centers if used with model-based standard errors
Random-effects	<ul style="list-style-type: none"> - An adequate number of centers (5 or more will be adequate) 	<ul style="list-style-type: none"> - Could lead to poor results if used with only a few centers (e.g. 2 or 3) 	<ul style="list-style-type: none"> - When there are at least a few centers

* *Note:* GEE, Generalized Estimating Equation

Table 2

Impact of different analysis strategies on results from MIST2 *

Analysis method	Odds ratio (95% CI)	P-value	Percentage of patients excluded from analysis
Fixed-effects	3.34 (1.01 to 11.00)	0.048	15%
Mantel-Haenszel	3.59 (0.99 to 13.03)	0.04	52%
Random-effects	2.72 (0.88 to 8.42)	0.08	0%
GEE	2.57 (0.76 to 8.66)	0.13	0%

* Results are presented for the number of patients who required surgery up to 90 days from randomization (a binary outcome), and pertain to the DNase vs. placebo comparison. The random-effects logistic regression model was implemented using a random-intercept for center. GEEs were implemented using an exchangeable correlation structure within center, and model-based (non-robust) SEs. All analyses adjusted for the stratification factors (amount of pleural fluid in the hemithorax at baseline, whether infection was hospital-acquired, and whether there was evidence of loculation). Full details of this re-analysis and the trial can be found in references (1) and (20) respectively.

Note: GEE, Generalized Estimating Equation

Table 3

Number of patients and events per center

	Binary outcome (n=75)	Survival outcome (n=55)	Continuous outcome (n=43)	Overall (n=177)
Number of patients per center – median (IQR)	46.3 (15.4 to 137.1)	18.2 (6.9 to 71.7)	17.4 (8.7 to 50.4)	24.3 (9.7 to 75.5)
Number of patients per center - no. (%)				
<5	3 (4)	6 (11)	4 (9)	13 (7)
5–10	7 (9)	15 (27)	10 (23)	35 (20)
10–20	12 (16)	10 (18)	9 (21)	32 (18)
20–50	16 (21)	9 (16)	9 (21)	34 (19)
50–100	14 (19)	3 (5)	8 (19)	25 (14)
100+	23 (31)	12 (22)	3 (7)	38 (21)
Number of events per center – median (IQR)	9.2 (3.1 to 30.2)	4.2 (2.5 to 13.8)	-	5.9 (2.8 to 21.2)
Number of events per center - no. (%)				
<1	5 (7)	3 (6)	-	8 (7)
1–3	12 (17)	15 (29)	-	27 (22)
3–5	7 (10)	12 (24)	-	19 (15)
5–10	13 (18)	8 (16)	-	21 (17)
10–20	10 (14)	5 (10)	-	15 (12)
20+	25 (35)	8 (16)	-	33 (27)
Number of patients per center/treatment combination – median (IQR)	22 (8 to 64)	9 (3 to 26)	8 (3 to 20)	11.5 (4.4 to 35.8)
Number of patients per center/treatment combination – no. (%)				
<2	2 (3)	3 (5)	6 (14)	11 (6)
2–5	11 (15)	18 (33)	9 (21)	41 (23)
5–10	9 (12)	10 (18)	8 (19)	28 (16)
10–20	14 (19)	8 (15)	9 (21)	31 (18)
20–50	16 (21)	4 (7)	10 (23)	30 (17)
50+	23 (31)	12 (22)	1 (2)	36 (20)
Number of events per center/treatment combination – median (IQR)	4 (2 to 13)	2 (1 to 6)	-	2.6 (1.4 to 10.3)
Number of events per center/treatment combination – no. (%)				
<1	10 (14)	10 (20)	-	20 (16)
1–3	21 (29)	21 (41)	-	42 (34)
3–5	8 (11)	7 (14)	-	15 (12)
5–10	10 (14)	5 (10)	-	15 (12)
10–20	9 (13)	3 (6)	-	12 (10)
20+	14 (19)	5 (10)	-	19 (15)

Table 4

Methods of analysis for trials using stratified randomization

	Binary outcome (n=49)	Survival outcome (n=44)	Continuous outcome (n=21)	Overall (n=120)
Adjusted for center – no. (%)	11 (22)	9 (20)	13 (62)	35 (29)
Patients per center – median (IQR)				
Adjusted	66 (42 to 480)	243 (11 to 722)	37 (11 to 55)	51 (16 to 243)
Unadjusted	50 (22 to 105)	19 (8 to 50)	20 (8 to 71)	24 (9 to 84)
Events per center – median (IQR)				
Adjusted	21 (6 to 48)	9 (5 to 43)	-	17 (6 to 43)
Unadjusted	8 (3 to 26)	5 (3 to 14)	-	5 (3 to 20)
Patients per center/treatment combination – median (IQR)				
Adjusted	33 (14 to 142)	122 (6 to 241)	14 (5 to 23)	22 (8 to 122)
Unadjusted	22 (11 to 53)	9 (4 to 25)	9 (4 to 27)	12 (4 to 39)
Events per center/treatment combination – median (IQR)				
Adjusted	6 (3 to 24)	4 (2 to 22)	-	5 (3 to 22)
Unadjusted	4 (1 to 13)	2 (1 to 7)	-	3 (1 to 10)
Method of adjustment – no. (%)				
Fixed-effects	6/11 (55)	3/9 (33)	11/13 (85)	22 (63)
Random-effects	2/11 (18)	0 (0)	1/13 (8)	3 (9)
GEE	0 (0)	-	-	0 (0)
Mantel-Haenszel	3/11 (27)	-	-	3 (9)
Stratified Cox model	-	5/9 (56)	-	5 (14)
Other	0 (0)	-	1/11 (8)	1 (3)
Unclear	0 (0)	1/9 (11)	0 (0)	1 (3)

* Patients per center, and patients per center/treatment combination are based on: 44 trials (binary), 35 trials (survival), 20 trials (continuous), and 102 trials (overall). Events per center, and events per center/treatment combination are based on: 44 trials (binary), 34 trials (survival), and 78 trials (overall).

Note: GEE, Generalized Estimating Equation