National Institute for Health and Care Excellence

Final

Atrial fibrillation: diagnosis and management

Network meta-analysis J2: ablation

NICE guideline NG196

Network meta-analysis report

April 2021

Final

Developed by the National Guideline Centre, Royal College of Physicians

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and, where appropriate, their carer or guardian.

Local commissioners and providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN **978-1-4731-4043-1**

Contents

1 Introduction

Network meta-analysis (NMA) is a statistical technique that allows simultaneous pooling of data for three or more interventions when the available evidence forms a connected network of intervention comparisons from RCTs (for example: evidence from trials comparing interventions A vs B, trials of B vs C and trials of C vs A). This enables both direct evidence (for example A vs B trials for the AvB comparison) and indirect evidence (for example A vs C and B vs C trials provide an indirect estimate of AvB) to be pooled.[8, 12, 30] NMA combines all the available data simultaneously into a single set of treatment effects that provide a unique ordering of intervention effectiveness, whilst respecting the randomisation in the included RCTs.[8, 30] The resulting estimates are therefore easier to interpret than a series of pairwise comparisons, and because both direct and indirect evidence is pooled these are more precisely estimated (have greater statistical power).

NMA assumes that the included studies are similar in terms of factors that might interact with the intervention effects (effect modifiers). So, the relative effect of intervention B vs intervention A would be expected to be similar in all of the studies (if they had included A and B interventions). This assumption is the same as that made in conventional pairwise meta-analysis, but we have to be particularly careful that the studies making different comparisons do not differ in effect modifiers (the data are consistent).[11] We can assess this assumption by measuring statistical heterogeneity, and also by checking if the direct and indirect estimates are in agreement when there are loops of evidence in the network (eg an ABC triangle of evidence).[6]

The analysis provides estimates of relative effects (with 95% credible intervals) for each intervention compared to a reference intervention (in this case the reference intervention was medical care with antiarrhythmic drugs) as well as estimates of all pairwise comparisons. In addition, for a given assumed “baseline effect” on the reference intervention, we can obtain absolute effects for all interventions. These estimates provide a useful clinical summary of the results and facilitate the formation of recommendations based on the best available evidence. Having a single set of treatment effects that takes into account all the available evidence also facilitates cost effectiveness analysis.

The ablation review for this guideline update (comparing radiofrequency [RF] point by point ablation, RF multielectrode ablation, cryoballoon ablation, laser ablation, thoracoscopy ablation, hybrid ablation (combination of thoracoscopy and catheter ablation) and medical care in people with atrial fibrillation) formed a connected network of RCT evidence for the paroxysmal AF stratum and so an NMA was considered for this stratum. For the other AF-type strata there were insufficient data to allow an NMA (see section 1.1).

This topic was considered a high clinical priority for the guideline due to variations in practice and uncertainty about the most clinically and cost effective strategy in the paroxysmal AF population. It was also given the highest priority for new economic modelling. Given this, the committee agreed that network meta-analysis was warranted to facilitate cost effectiveness analysis and help decision making in this area.

1.1 Study selection

A systematic review of RCTs comparing RF point by point, RF multielectrode, cryoballoon ablation, laser ablation, thoracoscopy, hybrid ablation/thoracoscopy, open surgery and medical care in people with atrial fibrillation was undertaken for the guideline, although no eligible studies were found for open surgery. Studies identified in this review were considered for inclusion in the NMA. The full details for the pairwise ablation evidence review can be found in review J1.

We performed NMAs that simultaneously used all the relevant RCT evidence from the clinical evidence review. As with conventional meta-analyses, this type of analysis does not break the randomisation of the evidence.

1.1.1 Population

The review and pairwise meta-analyses stratified studies according to predominant (>75%) AF type within the study: 1) 'paroxysmal AF', 2) 'persistent AF <1 year', 3) 'persistent AF >1 year' and 4) 'mixed (any type <75%)/unclear'. Data for both the persistent strata were regarded as too sparse for NMA: for the persistent >1 year stratum there was only one comparison, and for the persistent <1 year stratum there were only 2 comparisons. The data for the mixed/unclear stratum were regarded as inappropriate for NMA as any results would not be useful for decision-making because it was unclear to which population group that stratum pertained. Hence the GC agreed that only the data for the paroxysmal AF stratum, which contained a rich network of comparisons, should be subject to an NMA.

The committee discussed the importance of clinical homogeneity between comparisons in the paroxysmal AF NMA, and whether heterogeneity could be caused by the presence of 1) three trials[36, 37, 57] where the patients were undergoing first line treatment (in contrast to most other trials where they had been treated with drugs before), and 2) two trials[44, 45] where the patients had all failed ablation before.

In terms of the first category of potential heterogeneity, the committee decided to keep first line treatments in the proposed NMA on the pragmatic basis that pairwise results showed this made little difference to effect. This was bolstered by the committee's understanding that it was biologically plausible that effect sizes would not be altered. For example, in the between-ablation trials the committee saw no reason why the strength of relative effects would be affected by prior failure of an antiarrhythmic drug (AAD) or not. Similarly, in the ablation versus medical care trials where treatment was not first line, the medical care group were given an alternative AAD drug to that which they had previously failed, so again the committee did not feel this would lead to different strength of relative effects in comparison to trials on patients receiving first line treatment.

In terms of the second category of potential heterogeneity, however, the committee decide to remove the trials where patients had previously failed ablation, on the basis that this constituted a very different population of patients; patients failing ablation once would be at a higher probability of failing again, which would create a source of potential heterogeneity.

1.1.2 Outcome measures

Four outcomes were selected for the NMA. All of the four outcomes were deemed as critical outcomes for decision-making by the committee and/or important for incorporation in the cost effectiveness analysis:

- Recurrence at longest available follow up
- Stroke/TIA at longest available follow up
- Mortality at longest available follow-up
- Serious adverse events at longest available follow-up (not including stroke and mortality)

Study follow-up durations were usually 12 months, but there was some variation across studies for all 4 outcomes (Appendix B). For binary outcomes reported as the number of events for a given follow-up time, the most appropriate NMA model is to use a Binomial likelihood with a cumulative-log-log (cloglog) link to obtain relative treatment effects as hazard ratios[12, 13]. However, for the mortality, stroke, and serious adverse events outcomes, the events were rare and an NMA model with a Binomial likelihood and a logit link was deemed appropriate despite the variation in follow-up time[51]. For the recurrence outcome, a

clog-log link model was used to allow for the variable follow up times. The logit model yielded estimates of odds ratios, which were transformed to risk ratios based on an assumed baseline risk, and the clog-log model yielded hazard ratios.

1.1.3 Comparability of interventions

The interventions compared in the model were those found in the randomised controlled trials and included in the clinical evidence review already presented in review J1 of the full guideline. If an intervention was evaluated in a study that met the inclusion criteria for the network (that is if it reported at least one of the outcomes of interest and matched the inclusion criteria for the meta-analysis) then it was included in the network meta-analysis, otherwise it was excluded.

For the NMA relating to the outcomes of 'recurrence' and 'serious adverse events' the following interventions were included (the code for each intervention used in the WinBUGS models is also given to facilitate understanding of the scripts in the appendices):

Intervention	**WinBugs code for intervention**
Medical care (antiarrhythmic drugs [AADs])	1
RF point by point	2
cryoballoon	3
laser	4
thoracoscopy	5
Hybrid	6
RF multielectrode	7

For the NMA relating to the outcomes of stroke, the following interventions were included:

Intervention	**WinBugs code for intervention**
Medical care	1
RF point by point	2
cryoballoon	3
laser	4
RF multielectrode	5

For the NMA relating to the outcomes of mortality, the following interventions were included:

Intervention	**WinBugs code for intervention**

Medical care	1
RF point by point	2
cryoablation	3
laser	4

2 Statistical methods

2.1 Synthesis methods

A Bayesian framework is used to estimate all parameters, using Markov chain Monte Carlo simulation methods implemented in WinBUGS 1.4.3.[33, 34] A generalised linear model with a binomial likelihood and logit link was fitted for the mortality, serious adverse events and stroke outcomes, and a cloglog model was fitted for the recurrence outcome. Detailed reasons why these models were used are given in section 2.2.

Non-informative Normal(0,10000) priors were assigned to the trial-specific baseline and treatments effects (log odds ratios), and normal (0,10) priors were used for log-hazard ratios (which are sufficiently flat on the log-hazard scale) while a Uniform(0,5) prior was assigned to the between-study standard deviation in the random effects models.[13] Convergence was assessed using the Brooks-Gelman-Rubin diagnostic plot[5, 18] and was satisfactory by 60,000 simulations for all outcomes. A further sample of 60,000 iterations per chain post-convergence was obtained on which all reported results were based. Each analysis was run with 3 chains, each with a different set of initial values, to check that the model had converged through the mixing of chains via history plots, and results were not influenced by the initial values.

We assessed the goodness of fit of the model by calculating the mean of the posterior distribution of the residual deviance. If this is close to the number of unconstrained data points (the number of trial arms in the analysis) then the model is explaining the data well.

Studies with zero or 100% events in all arms were excluded from the analysis because these studies provide no evidence on relative effects.[13] For studies with zero or 100% events in one arm only, we used a continuity correction where we added 0.5 to both the number of events and the number of non-events, which has shown to perform well when there is an approximate 1:1 randomisation ratio across intervention arms.[23]

2.1.1 Between study heterogeneity

When considering models for network meta-analysis (NMA), there are several aspects of the data that will impact the choice of parameters included in the model. To assess the validity of an NMA it is essential to assess the extent of heterogeneity and inconsistency. Heterogeneity concerns the differences in treatment effects between trials within each treatment contrast, while consistency concerns the differences between the direct and indirect evidence informing the treatment contrasts.

A fixed effects NMA model is the simplest model available to estimate the effects of interventions separately while simultaneously synthesizing all available evidence. This model assumes no heterogeneity between trials within each treatment contrast. In other words, all trials within a treatment contrast are estimating the same treatment effect, regardless of any differences in the conduct of the trials, populations, or treatments (i.e., administration or dose). A random effects NMA model relaxes this assumption accounting for any differences in treatment effects between trials, within a treatment contrast, that are beyond chance by estimating the between-study standard deviation. The between-study standard deviation is assumed to be the same for each treatment contrast. When critiquing NMA models, it is good practice to assess and compare the fit of both fixed and random effects models, as differences may provide evidence of potential between-study heterogeneity.

2.1.2 Baseline model and data

The baseline risk is defined as the (absolute) risk of achieving the outcome of interest for patients receiving the reference intervention (medical care) in the population of interest.

Relative effects estimated from the NMA can be applied to the absolute baseline risk to obtain absolute risks under each intervention in the population of interest (see section 2.2). This allows us to convert the results of the NMA, which are estimated as odds ratios, into risk ratios for easier interpretation.

For the recurrence outcome, 3 studies provided the baseline data: Jais,[24] Pappone,[40] and Wazni[57]. These were all with a 1 year follow up from European studies and felt to be the most relevant data to the UK population. For the mortality and serious adverse events outcomes only a subset of these studies were included, as not all of them reported each outcome. For the stroke outcome, none of the included studies had relevant events, and so the baseline data were estimated based on three sources:

1. J-Rhythm study.[38] In a group of paroxysmal AF patients in Japan (aged 64.7 years, 80% on warfarin and 78.1% at CHADS2 score of 0-1), 9/419 randomised to rhythm control had suffered a symptomatic stroke after a mean follow up of 578 days. This yielded an annual rate of 1.3%.
2. The Health Economist calculated a baseline stroke risk of 0.7% for the HE model (using FIRE and ICE CHADSVASC distributions, untreated stroke rates from Asperg 2016 and RR from Sterne 2017).
3. Expert opinion from cardiologists in the GC

Based on these data, it was decided that an annual rate of 1% (expressed by nominal data of 1 event from 100 people) would be an appropriate baseline rate.

The baseline data below were analysed with the baseline NMA models of the 4 outcomes, using the best fitting of the fixed or random effects models. This yielded the logarithmic estimates of absolute risk (mean A) and uncertainty (sd A) for the medical treatment in each of the 4 outcomes (recurrence mean A= 0.2822, sd A= 0.09149; stroke mean A= -5.165, sd A= 1.288;mortality mean A= 3.612, sd A=0.816;serious adverse events mean A=-2.457, sd A=0.322). The mean A and precision of A (inverse square of the sd) were then fed into the consistency NMA models to facilitate estimation of absolute effects for the other treatments.

Table 1: Event rates reported in the trials that informed baseline risk for the medical arm in the different outcomes

Outcome	Estimate based on J-Rhythm, HE estimate and expert opinion		Jais[24]		Pappone[41]		Wazni[57]	
	Number events / Total randomised	%	Number events / Total randomised	%	Number events / Total randomised	%	Number of events / Total randomised	%
Recurrence	-	-	42/55	76.4	75/99	75.8	22/35	62.9
Stroke	1/100	1	-	-	-	-	-	-
Mortality	-	-	2/59	3.4	-	-	-	-
Serious AEs	-	-	AEs described but unclear to which group some events belonged	-	10/99	10.1	1/35	2.85

2.2 Summary measures and reference treatment

The results of pair-wise meta-analyses are presented in the clinical evidence review (Chapter J1).

The number of people who experienced recurrence of atrial fibrillation before a specific time were reported by studies with different follow-up times. The probability of recurrence is expected to increase with follow-up time, and this is likely to mean that the odds ratio depends on follow-up time too. An alternative approach is to model the *rate* of recurrence (i.e. the number of people experiencing recurrence per unit time). The assumption that the rate ratio is constant over time may be more reasonable than the assumption that the odds ratio (or relative risk) is constant over time. If we further assume that the rate of events is constant over time (following an Exponential distribution), then although the probability of an event depends on follow-up time, the complementary-log-log (cloglog) of the probability of an event is the sum of the log of follow-up time and the log of the event rate. Treatment effects are put on the log of the event rate in the NMA to obtain log rate ratios. Since the rate ratio is assumed to be constant over time, the proportional hazards assumption is made, and the rate ratios are equivalent to hazard ratios.

For the recurrence outcome therefore, data were pooled using a clog-log model, which produced hazard ratios robust to variations in follow-up time. We calculated the overall ranking of interventions according to their relative hazard compared to control group. Due to the skewness of the data, the NMA hazard ratios and rank results are reported as posterior medians rather than means to give a more accurate representation of the 'most likely' value.

However, if events are rare then the results from modelling rates will be very similar to modelling odds ratios. Therefore, for the mortality, stroke/TIA and serious adverse events outcomes, data were pooled as log odds ratios. To facilitate comparison with the results of the pairwise MA, we converted the log odds ratios into relative risks as follows. Assuming a baseline probability of effect in the population of interest P[b] (as described above in Section 2.1.2), the relative risks were calculated as RR[k] = P[k]/P[b], where logit(P[k]) = log(OR[k]) + logit(P[k]) for treatment k.

We also calculated the overall ranking of interventions according to their relative risk compared to control group. Due to the skewness of the data, the NMA relative risks and rank results are reported as posterior medians rather than means to give a more accurate representation of the 'most likely' value.

2.3 Methods of assessing inconsistency

A key assumption behind NMA is that the evidence in the network is consistent. In other words, it is assumed that the direct and indirect treatment effect estimates do not disagree with one another. Discrepancies between direct and indirect estimates of effect may result from several possible causes relating to differences between the trials included in terms of their clinical or methodological characteristics that interact with the relative intervention effects.

This form of heterogeneity is a problem for network meta-analysis but may be dealt with by subgroup analysis, meta-regression or by more narrowly defining inclusion criteria.

Inconsistency was assessed by comparing the chosen consistency model (fixed or random effects) to an "inconsistency", or unrelated mean effects, model.[14, 15] The latter is equivalent to having separate, unrelated, meta-analyses for every pairwise contrast, with a common variance parameter assumed in the case of random effects models. Note that inconsistency can only be assessed when there are closed loops of direct evidence on 3 or more treatments that are informed by at least 3 distinct trials.[53] The contribution of each data point

to the posterior mean deviance was also plotted for the inconsistency model against the consistency model, to assess whether individual data points contribute to inconsistency.

The posterior mean of the residual deviance, which measures the magnitude of the differences between the observed data and the model predictions of the data, was used to assess the goodness of fit of each model.[49] Smaller values are preferred, and in a well-fitting model the posterior mean residual deviance should be close to the number of data points in the network (each study arm contributes one data point).[49] In addition to comparing how well the models fit the data using the posterior mean of the residual deviance, models were compared using the deviance information criterion (DIC). This is equal to the sum of the posterior mean deviance and the effective number of parameters, and thus penalizes model fit with model complexity.[49] Lower values are preferred and typically differences of at least 3 points are considered meaningful.[49]

3 Results

3.1 Recurrence of atrial fibrillation

3.1.1 Network and data

Two studies[44, 45] were excluded where patients had all failed ablation previously, according to the pre-hoc decision made by the GC. In addition, 12 further studies with some kind of recurrence data were excluded because their recurrence data did not meet the protocol definition of recurrence (**Table 2**). The protocol definition of recurrence was the first event of AF (however detected) occurring at any point between the end of the blanking period and the end of follow up. The remaining 18 studies[1, 3, 4, 7, 16, 20, 22, 24, 25, 27, 29, 35, 36, 42, 43, 50, 57, 58] involving the 7 interventions were included in the recurrence network. As for all outcomes, data from studies where any switching of interventions had occurred for individual participants was dealt with using the intention to treat (ITT) principle: that is, events were assigned to the randomised treatment rather than the treatment after switching. The ITT principle was applied because patients switching are often those not responding well to initial treatment, and keeping patients in randomised groups permits capture of this information.

Table 2. Studies providing recurrence data that were excluded from the analysis.

Excluded	Reason
Cosedis Nielsen, 2012[9]	Unclear if cumulative data provided in table includes blanking period
Davtyan, 2018[10]	Unclear if events were counted during the blanking period (which would be incorrect); also unclear if data are cumulative (required) or point data (excluded)
Gal, 2014[17]	Unclear whether the data were cumulative or point data
Giannopoulos, 2019[19]	Unclear whether the data were cumulative or point data
Kece, 2019[26]	Unclear if events occurred in blanking period
Packer, 2013[39]	8 Patients in the usual care group crossed over to ablation in the blanking period, and looks likely these were then classed as treatment failures (recurrence) in the final results, even though recurrence occurring in the blanking period should not be counted. Because of the ambiguity of reporting it is certainly not possible to be confident this was not the case. We don't know if these 8 people would have had recurrences after the blanking period so we could not code them as no recurrence (as they may well have gone on to get recurrence after the blanking period) and we could not code them as recurrent (as they may well not have developed it after the blanking period).
Pappone, 2006[40]	Data in RF point by point group unclear. However the data in the medical care group were clear and have been used in the baseline analysis.
Wang, 2014[55]	Did not exclude events occurring very early after ablation
Watanabe, 2018[56]	Unclear outcome – 'use of AADs' provided, but cannot be used as proxy for recurrence, as stated that patients allowed to use them even if no recurrence. Paper also gives number without AF but this includes patients who are using AADs.
Xu, 2012[59]	Unclear if events occurred in blanking period
Yagishita, 2020[60]	Events occurred during early period (4 weeks) after ablation
You, 2019[61]	Unclear if events occurred in blanking period

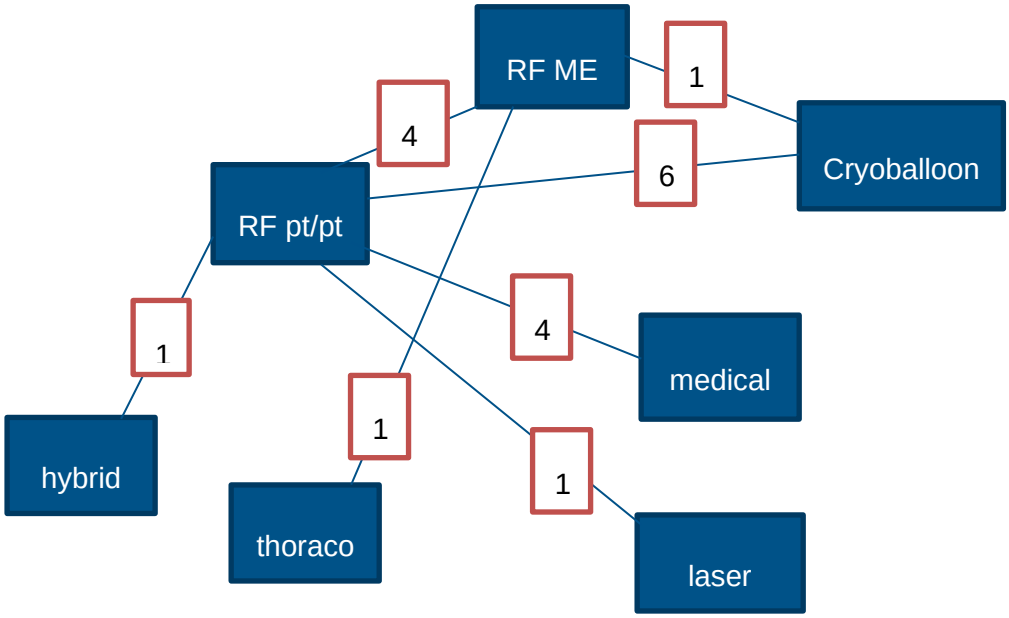
The original outcome in the pairwise review had been designated '*symptomatic* AF recurrence', but few studies had looked at this. Instead they mostly looked at AF recurrence as picked up by ECG/Holter/ILR, which would include both symptomatic and asymptomatic AF ('mixed' symptomatic / asymptomatic). Thus, in the original pairwise review, we accepted

any recurrence (pure symptomatic or mixed) for meta-analysis, but downgraded the mixed evidence for indirectness. There were only 4 studies previously with symptomatic recurrence data and the other 19 had mixed asymptomatic/symptomatic recurrence data.

It is likely that this variability in actual outcome might contribute to inconsistency in an NMA, so we originally thought we should include the type of recurrence (symptomatic vs mixed) as a covariate in a meta-regression. However on further examination of the papers we found we could gather mixed asymptomatic/symptomatic recurrence data from three of the 4 papers from which we had originally only collected symptomatic recurrence data. This seemed more sensible than adjusting for it – if we could make the outcome as homogeneous as possible across comparisons this might lead to better coherence overall. We were aware that we were not using the ideal clinical outcome of symptomatic recurrence, but since that was only available for a minority of studies it did not seem too much of a loss to be unable to include it. Certainly the gains from reduced inconsistency were deemed to outweigh the disadvantages of not using symptomatic AF recurrence when available. The only alternative options were to use only those 4 symptomatic AF studies for the NMA (which would not have yielded a network) or to have tried to adjust for type of AF recurrence (we were doubtful that we could have gained any valid adjustment from only those 4 studies). The NMA has therefore been run using the outcome of symptomatic/asymptomatic recurrence for almost all studies; in one study[57] there was no mixed recurrence data so the pure symptomatic recurrence data were used.

The network can be seen in Figure 1 and the trial data for each of the studies included in the NMA are presented in Table 3.

Figure 1: Network diagram for recurrence



RF=radiofrequency; pt/pt=point by point; ME=multielectrode; thoraco=thoracoscopy; numbers in red boxes refer to the number of studies in a direct comparison

Table 3: Study data for recurrence network meta-analysis

Study	Intervention	Comparison	Intervention		Comparison	
			Events	n	Events	n
Andrade, 2019[1]	RF pt/pt	Cryoballoon	53	115	111	231
Bin Waleed, 2019[3]	RF pt/pt	Cryoballoon	3	29	4	28
Gunawardene[20]	RF pt/pt	Cryoballoon	3	30	6	30

Study	Intervention	Comparison	Intervention		Comparison	
			Events	n	Events	n
Hunter[22]	RF pt/pt	Cryoballoon	41	77	26	78
Kuck[29]	RF pt/pt	Cryoballoon	143	376	138	374
Perez Castellano[42]	RF pt/pt	Cryoballoon	8	25	13	25
Jan[25]	RF pt/pt	hybrid	17	26	10	24
Dukkipati[16]	RF pt/pt	laser	60	166	61	167
Boersema[4]	RF pt/pt	RF ME	11	58	14	59
Bulava[7]	RF pt/pt	RF ME	15	51	12	51
McCready[35]	RF pt/pt	RF ME	40	91	37	92
Podd[43]	RF pt/pt	RF ME	12	25	11	25
Jais[24]	Medical	RF pt/pt	42	55	7	53
Morillo[36]	Medical	RF pt/pt	44	61	36	66
Wazni[57]	Medical	RF pt/pt	22	35	4	32
Wilber[58]	Medical	RF pt/pt	46	56	38	103
Koch[27]	Cryoballoon	RF ME	13	22	10	15
Sugihara[50]	thoraco	RF ME	3	20	20	49

RF=radiofrequency; pt/pt=point by point; ME=multielectrode; thoraco=thoracoscopy

3.1.2 Inconsistency and goodness of fit

Both fixed effects and random effects baseline models were fitted to the medical data from the Jais,[24] Pappone,[41] and Wazni[57] studies. As seen in Table 4, the fixed effects baseline model had a DIC of 17.25 compared to 18.32 for the random effects baseline model, and so the fixed effect baseline model was preferred, and used to combine with the relative effects from the NMA to obtain absolute probabilities and relative risks outputs.

There was no evidence of heterogeneity in the NMA model, but there was a better fit for the Random Effects NMA model than for the Fixed Effects model. There was a lower DIC and significantly lower ResDev.

An inconsistency model was run and the model fit statistics were as seen in Table 4. The NMA has a similar DIC suggesting that there is no evidence of inconsistency, supported by the similar direct and indirect estimates in Table 5. In addition, the posterior median standard deviation, a measure of the between study variability, is lower for the RE consistency NMA than RE inconsistency model, further confirming the lack of inconsistency (Table 4).

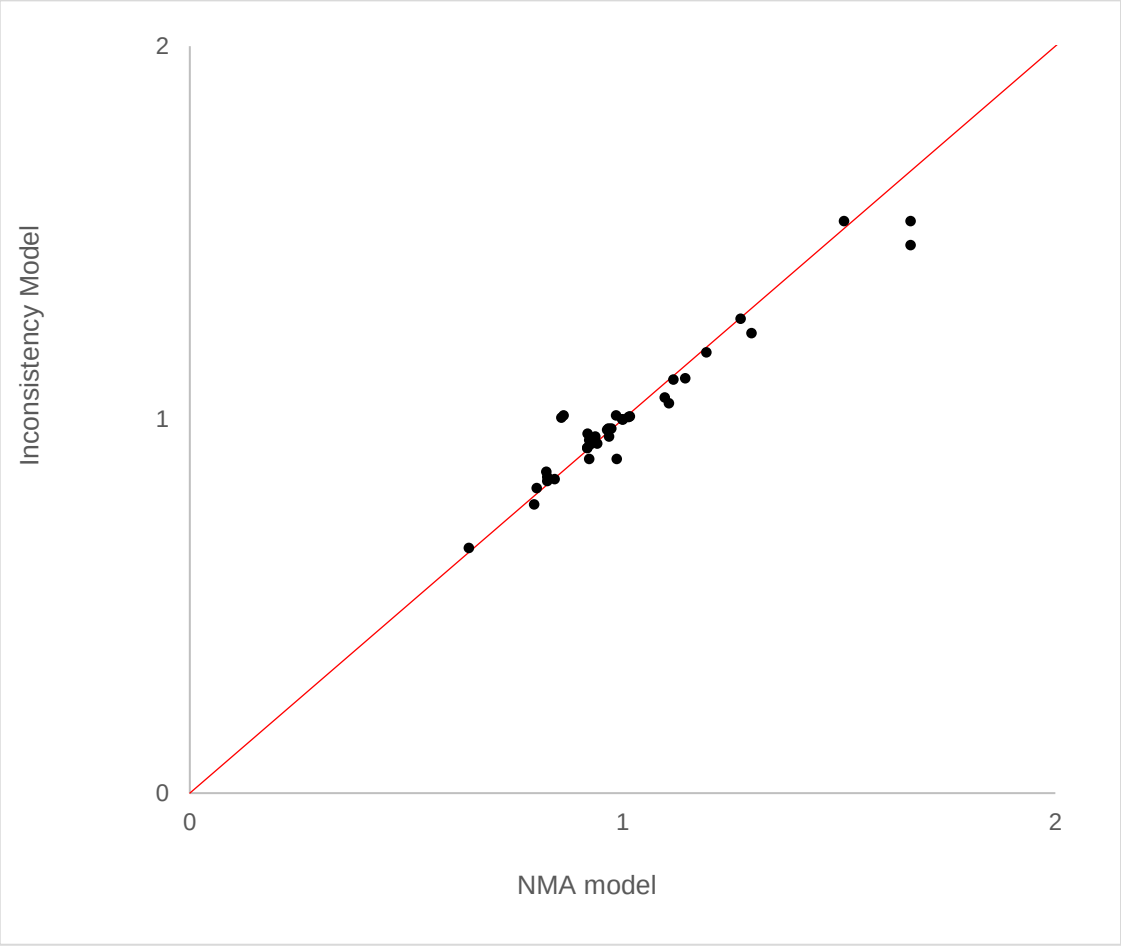
Figure 2 presents the contributions to the posterior mean of the deviances for each data-point for the inconsistency model against that for the consistency NMA model. There is no evidence of inconsistency, as there are no points notably below the line of equality, which would be indicative of data better predicted by the inconsistency model.

Table 4: Model fit statistics – recurrence

	Deviance information criterion (DIC)	Mean of the residual deviance (ResDev)*	Posterior median sd (95% CrIs)
Baseline models			
Fixed effects	17.251	3.401	NA
Random effects	18.325	2.912	2.473 (0.2509-8.432)
Relative effect models			
NMA Fixed effects	231.216	55.97	NA
NMA Random effects	219.046	35.98	0.461 (0.198-0.899)
Inconsistency model [RE]	219.694	35.93	0.493 (0.21-0.978)

Number of data points: baseline 3, NMA 36

Figure 2: Posterior mean of the contribution to the posterior mean residual deviance of the inconsistency model vs. the consistency model – recurrence



3.1.3 Results of estimation

Table 5 summarises the final results of the NMA in terms of hazard ratios for every possible treatment comparison.

Table 6 presents summary statistics for the 7 interventions included in the network, including the rank of the intervention, probability of the intervention being the best and mean absolute probability of an event. The mean absolute probability of the event in the medical treatment was based on the results of the baseline analysis, and the absolute probabilities for the other treatments are based upon application of the NMA relative effects to the baseline probability for the medical treatment.

Table 5: Hazard ratios for recurrence; direct pairwise meta-analysis results and NMA results

Intervention	Comparison	Random Effects Direct effects - median (95% credible intervals)	Random Effects NMA - median (95% credible intervals)
RF pt pt	Medical	0.243(0.121 to 0.446)	0.2652 (0.1456-0.4762)
Cryo	Medical	-	0.2707(0.1289-0.5952)
laser	Medical	-	0.2775(0.08254-0.9607)
thoraco	Medical	-	0.08638(0.01485-0.4699)
Hybrid	Medical	-	0.1425(0.03562-0.5904)
RF ME	Medical	-	0.2664(0.1192-0.6145)
cryoballoon	RF pt pt	1.039(0.623 to 1.873)	1.021(0.6461-1.726)
laser	RF pt pt	1.013(0.312 to 3.290)	1.047(0.3626-3.177)
thoraco	RF pt pt	-	0.328(0.06044-1.624)
hybrid	RF pt pt	0.500 (0.126 to1.954)	0.5405(0.1519-1.984)
RF ME	RF pt pt	0.927 (0.481 to 1.800)	1.007(0.5792-1.83)
laser	cryo	-	1.027(0.3087-3.317)
thoraco	cryo	-	0.3197(0.05574-1.611)
hybrid	cryo	-	0.5274(0.1327-2.06)
RFME	cryo	1.215 (0.296 to 4.943)	0.9849(0.4877-1.956)
thoraco	laser	-	0.3097(0.04248-2.126)
hybrid	Laser	-	0.5141(0.09611-2.774)
RF ME	Laser	-	0.9614(0.2837-3.28)
Hybrid	thoraco	-	1.658(0.2165-13.74)
RF ME	thoraco	3.317 (0.698 to 19.375)	3.063(0.6966-3.063)
RFME	hybrid	-	1.867(0.4599-7.595)

**Random effects model was used as this gave a better fit to the data (lower total residual deviance than the fixed effects model)*

Table 6: Intervention rank and mean probability of event – recurrence

	Probability of recurrence at one year – posterior median (and credible intervals)	Intervention rank - median (95% CrIs)	Probability intervention is best (%)
medical	0.7344(0.6697-0.7949)	7 (6-7)	0.0011%
RF pt pt	0.2962(0.1717-0.477)	4 (2-6)	0.408%
cryo	0.3018(0.154-0.5527)	4 (2-6)	0.938%

	Probability of recurrence at one year – posterior median (and credible intervals)	Intervention rank - median (95% CrIs)	Probability intervention is best (%)
laser	0.308(0.1023-0.7238)	5 (1-6)	4.177%
thoraco	0.108(0.01929-0.4668)	1 (1-6)	66.05%
hybrid	0.1724((0.04562-0.5488)	2 (1-6)	27.82%
RF ME	0.2974(0.1439-0.5637)	4 (2-6)	0.611%

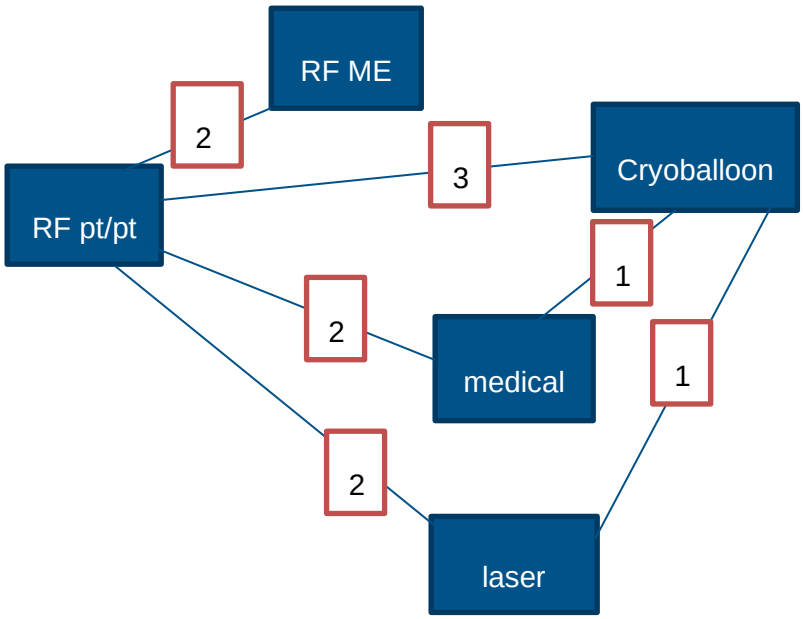
3.2 Stroke/TIA

3.2.1 Network and data

After excluding studies that reported zero events in all arms, since they do not contribute evidence to the NMA [Dias, S., et al., NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials, in Technical Support Document. 2011], 9 studies[1, 9, 16, 26, 29, 35, 39, 40, 48] involving 5 interventions were included in the stroke network. As for all outcomes, data from studies where any switching of interventions had occurred for individual participants was dealt with using the intention to treat (ITT) principle: that is, events were assigned to the randomised treatment rather than the treatment after switching. The ITT principle was applied because patients switching are often those not responding well to initial treatment, and keeping patients in randomised groups permits capture of this information.

The network can be seen in Figure 3 and the trial data for each of the studies included in the NMA are presented in Table 7.

Figure 3: Network diagram for stroke



RF=radiofrequency; pt/pt=point by point; ME=multielectrode; Note that there was a three arm trial between RF pt/pt, laser and cryoballoon. Numbers in red squares denote numbers of studies.

Table 7: Study data for stroke/TIA network meta-analysis

Study	Intervention	Comp 1	Comp 2	Intervention		Comp 1		Comp 2	
				E	n	E	n	E	n
Andrade[1]	RF pt/pt	Cryo	NA	0.5	116	2.5	232	NA	NA
Kuck[29]	RF pt/pt	Cryo	NA	2	376	2	374	NA	NA
Schmidt[48]	RF pt/pt	Cryo	laser	8	33	6	33	8	33
Dukkipatti[16]	RF pt/pt	laser	NA	1	172	2	170	NA	NA
Kece[26]	RF pt/pt	RF ME		2	35	8	35	NA	NA

Study	Intervention	Comp 1	Comp 2	Intervention		Comp 1		Comp 2	
				E	n	E	n	E	n
McCready[35]	RF pt/pt	RF ME	NA	0.5	92	2.5	93	NA	NA
Cosedis Neilsen[9]	medical	RF pt/pt	NA	1	148	2	146	NA	NA
Pappone[40]	medical	RF pt pt		0.5	100	1.5	100	NA	NA
Packer[39]	medical	cryo	NA	0.5	83	7.5	164	NA	NA

Comp= comparison; E = number of events; n= total number in group; NA = not applicable; RF = radiofrequency; pt/pt=point by point; ME=multielectrode; cryo=cryoballoon

3.2.2 Inconsistency and goodness of fit

Both fixed effects and random effects baseline models were fitted to data based on a consensus agreement of the likely baseline risk. As seen in Table 8 there was no noticeable difference in DIC between the fixed and random effects baseline models, and so the fixed effect baseline model was preferred, and used to combine with the relative effects from the NMA to obtain absolute probabilities and relative risks outputs.

There was no evidence of heterogeneity in the NMA model, but there was a slightly better fit for the Fixed effects NMA model than for the random effects model, with a slightly lower DIC and ResDev.

A fixed effect inconsistency model was run and the model fit statistics were as seen in Table 8. The Fixed effect NMA has a slightly smaller DIC suggesting that there is no evidence of inconsistency, a conclusion which is supported by comparing risk ratios from the pairwise and NMA models (Table 9).

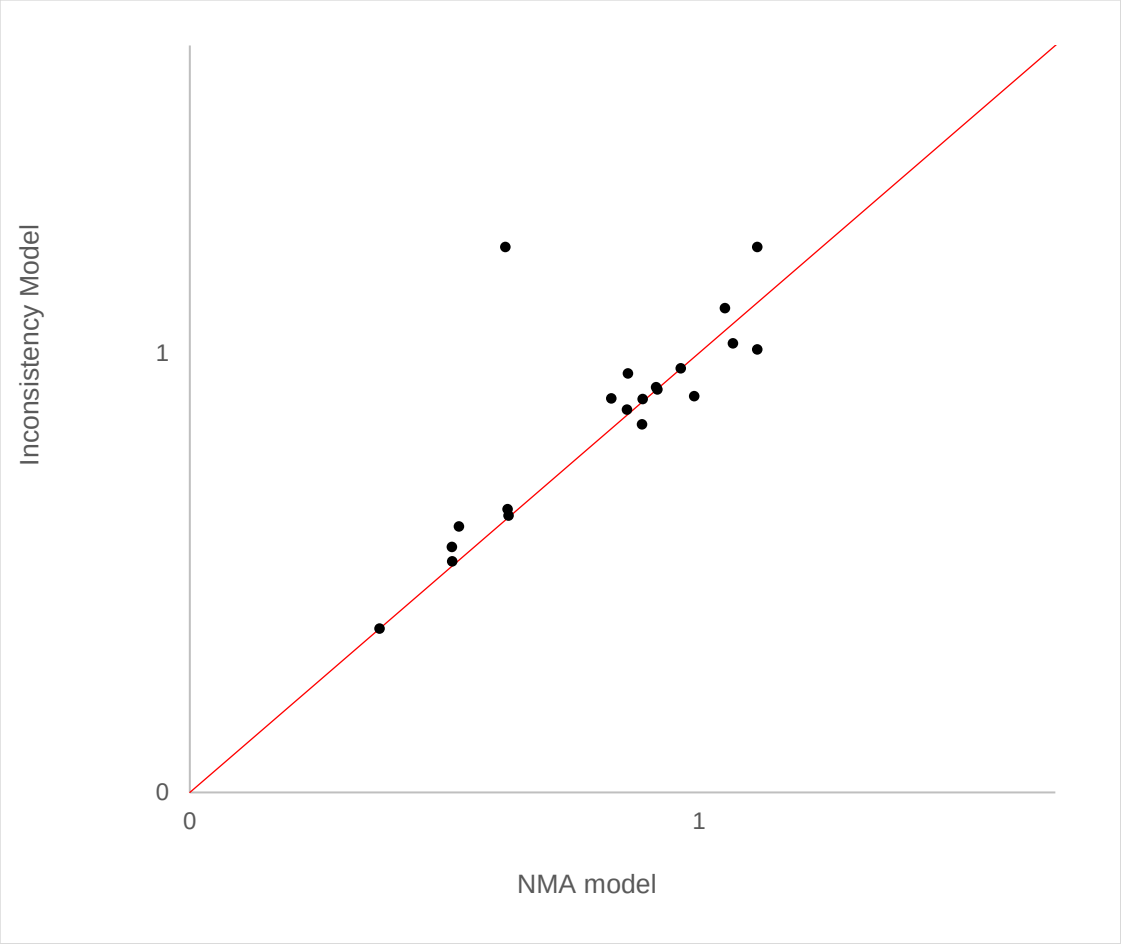
Figure 4 presents the contributions to the posterior mean of the deviances for each data-point for the inconsistency model against that for the consistency NMA model. There is no evidence of inconsistency, as there are no points notably below the line of equality, which would be indicative of data better predicted by the inconsistency model.

Table 8: Model fit statistics – stroke/TIA

	Deviance information criterion (DIC)	Mean of the residual deviance (ResDev)	Posterior median sd (95% CrIs)
Baseline models			
Fixed effects	4.036	1.158	NA
Random effects	4.014	1.154	2.49 (0.1277 - 4.875)
Relative effect models			
NMA Fixed effects	76.706	15.15	NA
NMA Random effects	78.504	15.8	0.4669 (0.02075 – 2.128)
Inconsistency model [FE]	78.107	15.87	NA

Number of data points: baseline 1, NMA 19

Figure 4: Posterior mean of the contribution to the posterior mean residual deviance of the inconsistency model vs. the consistency model – stroke/TIA



3.2.3 Results of estimation

Table 9 summarises the final results of the pairwise meta-analyses in terms of risk ratios generated from studies directly comparing different interventions, together with the results of the NMA in terms of risk ratios for every possible treatment comparison.

Table 10 presents summary statistics for the 5 interventions included in the network, including the rank of the intervention, probability of the intervention being the best and mean absolute probability of an event. The mean absolute probability of the event in the medical treatment was based on the results of the baseline analysis, and the absolute probabilities for the other treatments are based upon application of the NMA relative effects to the baseline probability for the medical treatment.

Table 9: Risk ratios for stroke/TIA; direct pairwise meta-analysis results and NMA results

Intervention	Comparison	Fixed Effects Direct (95% confidence intervals)	Fixed Effects* NMA - median (95% credible intervals)
RF pt pt	Medical	2.35(0.35-15.82)	4.277(0.9741-27.35)
Cryo	Medical	7.59(0.44-131.31)	4.413(1.024-28.3)
laser	Medical	-	5.602(1.035-38.38)
RF ME	Medical	-	19.8(3.024-144.4)
Cryo	RF pt pt	0.91(0.40-2.04)	1.032(0.4406-2.434)
laser	RF pt pt	1.11(0.50-2.48)	1.282(0.4988-3.282)
RF ME	RF pt pt	4.19(1.11-15.82)	4.277(1.321-19.79)
laser	cryo	1.33(0.52-3.42)	1.242(0.4575-3.358)
RF ME	cryo	-	4.166(1.035-22.8)
RF ME	laser	-	3.33(0.7968-19.55)

**Fixed effects model was used as this gave a better fit to the data (lower total residual deviance than the random effects model)*

Table 10: Intervention rank and mean probability of event – stroke/TIA

	Probability of recurrence – posterior median (and credible intervals)	Intervention rank - median (95% CrIs)	Probability intervention is best (%)
medical	0.005652 (0.00045-0.06665)	1 (1-2)	95.13%
RF pt pt	0.02608(0.001322-0.3952)	3 (2-4)	1.829%
cryo	0.02693(0.001376-0.4036)	3 (2-4)	1.452%
laser	0.03417(0.001542-0.4876)	4 (2-5)	1.568%
RF ME	0.1344(0.00541-0.8478)	5 (4-5)	0.0278%

3.2.4 Sensitivity analysis – removal of Schmidt, 2013[48] and Kece[26]

Two studies[26, 48] were felt to be somewhat different to the others, because they did not look at clinical strokes but instead asymptomatic cerebral lesions identified by magnetic resonance imaging (MRI). Although these lesions are still 'strokes', it was felt important to examine results without these studies included. A further analysis was therefore conducted with exclusion of the data from Schmidt, 2013[48] and Kece[26]. This was not based on an *apriori* plan but the committee felt that it should be carried out on a post-hoc basis given their feeling that the inclusion of these studies might influence results.

3.2.4.1 Inconsistency and goodness of fit

There was no evidence of heterogeneity in the NMA model, but there was a slightly better fit for the Fixed effects NMA model than for the random effects model. There was a slightly lower DIC and ResDev.

An inconsistency model was run and the model fit statistics were as seen in Table 11. The NMA has a slightly smaller DIC suggesting that there is no evidence of inconsistency, a conclusion which is supported by comparing risk ratios from the pairwise and NMA models (Table 12).

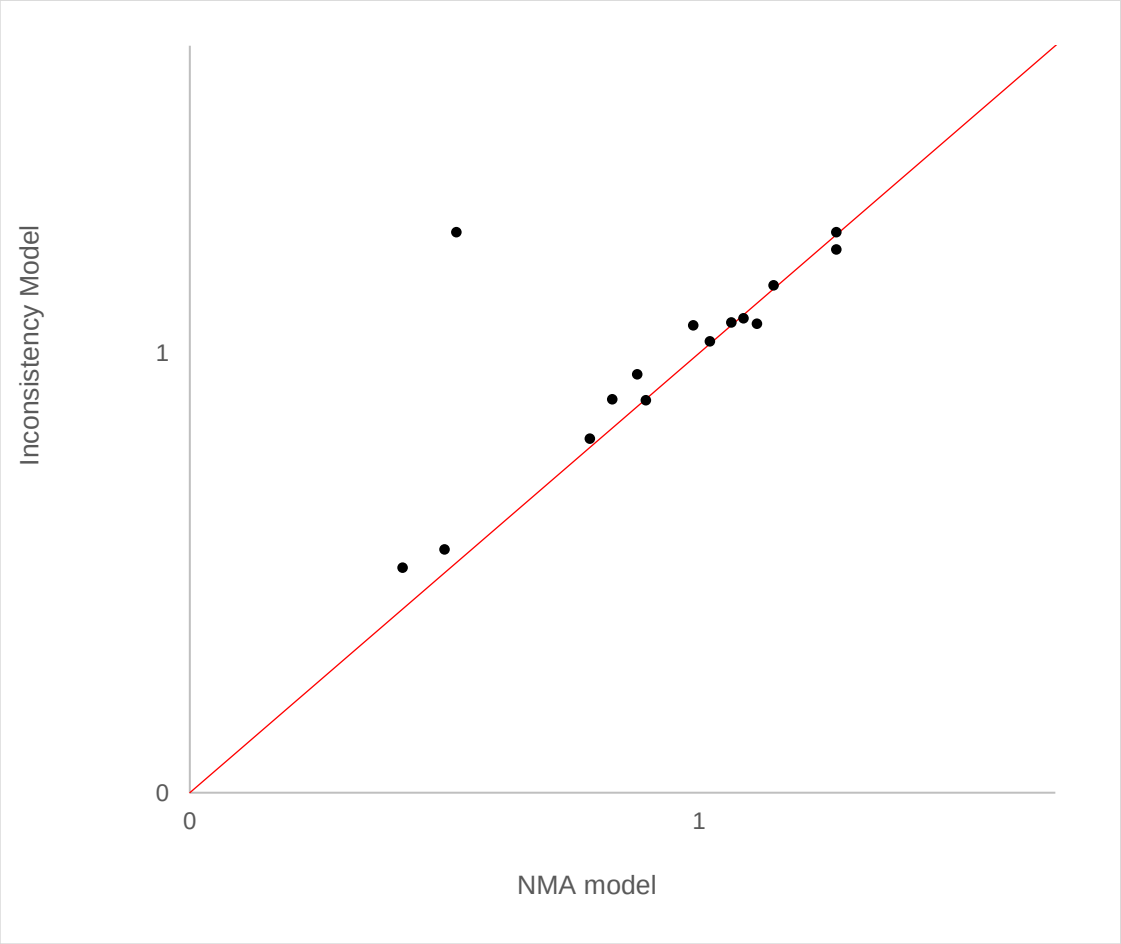
Figure 5 presents the contributions to the posterior mean of the deviances for each data-point for the inconsistency model against that for the consistency NMA model. There is no evidence of inconsistency, as there are no points notably below the line of equality, which would be indicative of data better predicted by the inconsistency model.

Table 11: Model fit statistics – stroke

	Deviance information criterion (DIC)	Mean of the residual deviance (ResDev)	Posterior median sd (95% CrIs)
Baseline models			
Fixed effects	4.036	1.158	NA
Random effects	4.014	1.154	2.49 (0.1277 - 4.875)
Relative effect models			
NMA Fixed effects	54.593	12.53	NA
NMA Random effects	56.409	13.42	0.934 (0.0354 – 4.174)
Inconsistency model [FE]	56.378	13.58	NA

Number of data points: baseline 1, NMA 12

Figure 5: Posterior mean of the contribution to the posterior mean residual deviance of the inconsistency model vs. the consistency model – stroke



3.2.4.2 Results of estimation

This sensitivity analysis showed qualitatively similar results. There were some quantitative differences in effects (for example, for laser versus medical the RR was 5.602 for the original model but 8.519 with the 2 studies removed, and for cryo versus RF pt/pt the RR was 1.032 for the original model but 1.67 with the 2 studies removed. However, no other directions of effect changed, and rankings remained the same. The probabilities of being the best were also very similar. The results are given below (Table 12 to Table 13).

Table 12: Risk ratios for stroke/TIA; direct pairwise meta-analysis results and NMA results

Intervention	Comparison	Fixed Effects Direct (95% confidence intervals)	Fixed Effects* NMA - median (95% credible intervals)
RF pt pt	Medical	2.35(0.35-15.82)	3.508 (0.7218-24.12)
Cryo	Medical	7.59(0.44-131.31)	6.004(1.234-43.85)
laser	Medical	-	8.519(0.4228-146.1)
RF ME	Medical	-	25.49(1.147-467.3)
Cryo	RF pt pt	1.38(0.27-6.93)	1.67(0.4442-7.834)
laser	RF pt pt	2.02(0.19-22.11)	2.197(0.1965-30.38)
RF ME	RF pt pt	4.95(0.24-101.62)	6.146(0.5081-134.1)
laser	cryo	-	1.326(0.07518-22.05)
RF ME	cryo	-	3.714(0.2062-87.88)
RF ME	laser	-	2.681(0.08233-134.1)

**Fixed effects model was used as this gave a better fit to the data (lower total residual deviance than the random effects model)*

Table 13: Intervention rank and mean probability of event – stroke/TIA

	Probability of stroke/TIA – posterior median (and credible intervals)	Intervention rank - median (95% CrIs)	Probability intervention is best (%)
medical	0.005688(0.000458-0.06617)	1 (1-3)	86.18%
RF pt pt	0.02111(0.001016-0.3577)	2 (1-4)	4.205%
cryo	0.03737(0.001755-0.5203)	3 (2-5)	0.7417%
laser	0.05347(0.000971-0.8492)	4 (1-5)	7.287%
RF ME	0.2001(0.002813-0.9946)	5 (2-5)	1.588%

3.3 Mortality

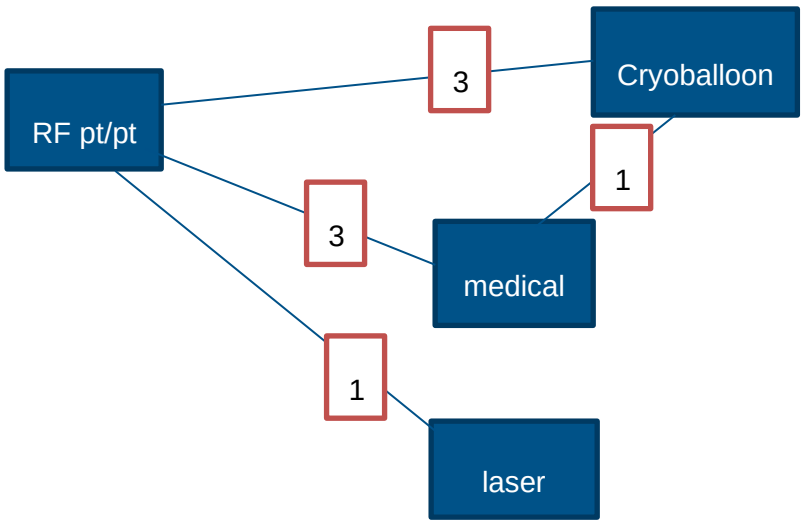
3.3.1 Network and data

After excluding studies that reported zero events in all arms, since they do not contribute evidence to the NMA [Dias, S., et al., NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials, in Technical Support Document. 2011], 8 studies[1, 2, 9, 16, 24, 29, 39, 58] involving 4 interventions were included in the mortality network. As for all outcomes, data from studies where any switching of interventions had occurred for individual participants was dealt with using the intention to treat (ITT) principle: that is, events were assigned to the randomised treatment rather than the treatment after switching. The ITT principle was applied because patients switching are often those not responding well to initial treatment, and keeping patients in randomised groups permits capture of this information.

The network can be seen in

Figure **6** and the trial data for each of the studies included in the NMA are presented in Table 14.

Figure 6: Network diagram for mortality



RF=radiofrequency; pt/pt=point by point; ME=multielectrode; numbers in red squares refer to numbers of studies

Table 14: Study data for mortality network meta-analysis

Study	Intervention	Comparator	Intervention		Comparator	
			Events	n	Events	n
Andrade[1]	RF pt/pt	cryo	0.5	116	1.5	232
Kuck[29]	RF pt/pt	cryo	0.5	377	2.5	375
Hunter[2]	RF pt/pt	cryo	1	67	2	67
Dukkipatti[16]	RF pt/pt	laser	0.5	173	1.5	171
Jais[24]	medical	RF pt/pt	2.5	60	0.5	54
Cosedis Neilsen[9]	medical	RF pt/pt	4	148	3	146

Study	Intervention	Comparator	Intervention		Comparator	
			Events	n	Events	n
Wilber[58]	medical	RF pt/pt	0.5	58	1.5	104
Packer[39]	medical	cryo	0.5	83	1.5	164

n= total number in group; RF = radiofrequency; pt/pt=point by point; cryo=cryoballoon

3.3.2 Inconsistency and goodness of fit

Both fixed effects and random effects baseline models were fitted to the data from the Jais[24] study. As seen in Table 15, the fixed effects baseline model had a DIC of 4.629 compared to 4.626 for the random effects baseline model. Because the DIC values were very similar, and only 1 study had informed the baseline estimate, the fixed effects baseline model was the preferred model and used to combine with the relative effects from the NMA to obtain absolute probabilities and relative risks outputs.

There was no evidence of heterogeneity in the NMA model, and there was a slightly better fit for the Fixed effects NMA model than for the random effects model, with a slightly lower DIC and ResDev.

A fixed effects inconsistency model was run and the model fit statistics were as seen in Table 15. The consistency NMA has a slightly smaller DIC suggesting that there is no evidence of inconsistency, a conclusion which is supported by comparing risk ratios from the pairwise and NMA models (Table 16).

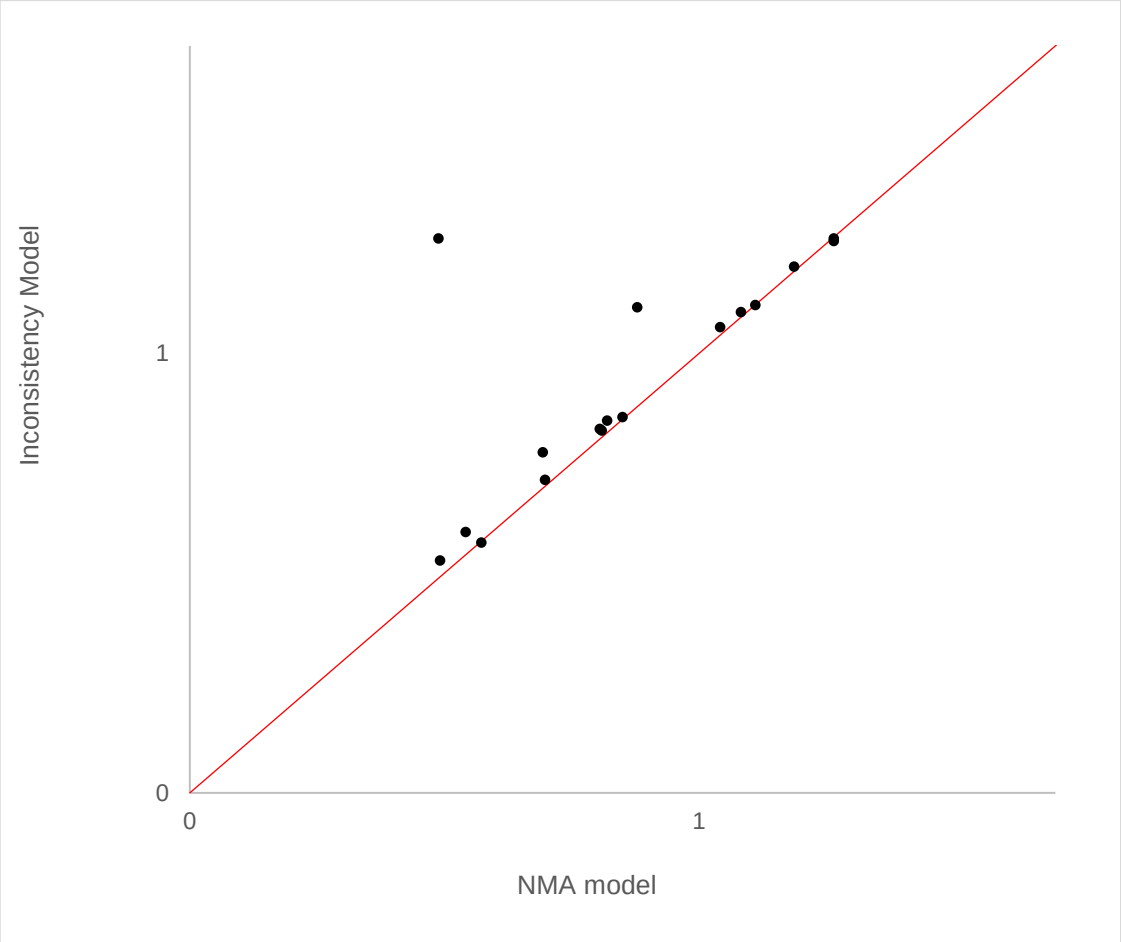
Figure 7 presents the contributions to the posterior mean of the deviances for each data-point for the inconsistency model against that for the consistency NMA model. There is no evidence of inconsistency, as there are no points below the line of equality, which would be indicative of data better predicted by the inconsistency model.

Table 15: Model fit statistics – mortality

	Deviance information criterion (DIC)	Mean of the residual deviance (ResDev)	Posterior median sd (95% CrIs)
Baseline models			
Fixed effects	4.629	1.086	NA
Random effects	4.626	1.085	2.514 (0.129-4.875)
Relative effect models			
NMA Fixed effects	57.9	13.34	NA
NMA Random effects	59.715	14.11	0.6349 (0.029 – 3.088)
Inconsistency model [FE]	59.904	14.62	NA

Number of data points: baseline 1, NMA 16

Figure 7: Posterior mean of the contribution to the posterior mean residual deviance of the inconsistency model vs. the consistency model – mortality



3.3.1 Results of estimation

Table 16 summarises the results of the pairwise meta-analyses in terms of risk ratios generated from studies directly comparing different interventions, together with the results of the NMA in terms of risk ratios for every possible treatment comparison.

Table 17 presents summary statistics for the 4 interventions included in the network, including the rank of the intervention, probability of the intervention being the best and mean absolute probability of an event. The mean absolute probability of the event in the medical treatment was based on the results of the baseline analysis, and the absolute probabilities for the other treatments are based upon application of the NMA relative effects to the baseline probability for the medical treatment.

Table 16: Risk ratios for mortality; direct pairwise meta-analysis results and NMA results

Intervention	Comparison	Fixed Effects Direct (95% confidence intervals)	Fixed Effects* NMA - median (95% credible intervals)
RFptpt	Medical	0.66(0.20-2.14)	0.6472(0.1985-1.938)
cryo	Medical	1.52 (0.06-26.87)	1.771(0.3464-9.821)
laser	Medical	-	3.112(0.09159-56.62)
cryo	RF ptpt	2.54(0.53-12.29)	2.709(0.6985-13.3)
laser	RF ptpt	3.04(0.12-73.98)	4.635(0.1748-95.46)
laser	cryo	-	1.691(0.0463-45.31)

**Fixed effects model was used as this gave a better fit to the data (lower total residual deviance than the random effects model)*

Table 17: Intervention rank and mean probability of event – mortality

	Probability of recurrence – posterior median (and credible intervals)	Intervention rank - median (95% CrIs)	Probability intervention is best (%)
medical	0.02616 (0.00541-0.1182)	2 (1-4)	18.05%
RFpt pt	0.01678 (0.00233-0.1089)	1 (1-3)	59.69%
cryo	0.04745 (0.004753-0.3835)	3 (1-4)	5.49%
laser	0.08649(0.001668-0.983)	4 (1-4)	16.77%

3.4 Serious adverse events (not including mortality and stroke)

3.4.1 Network and data

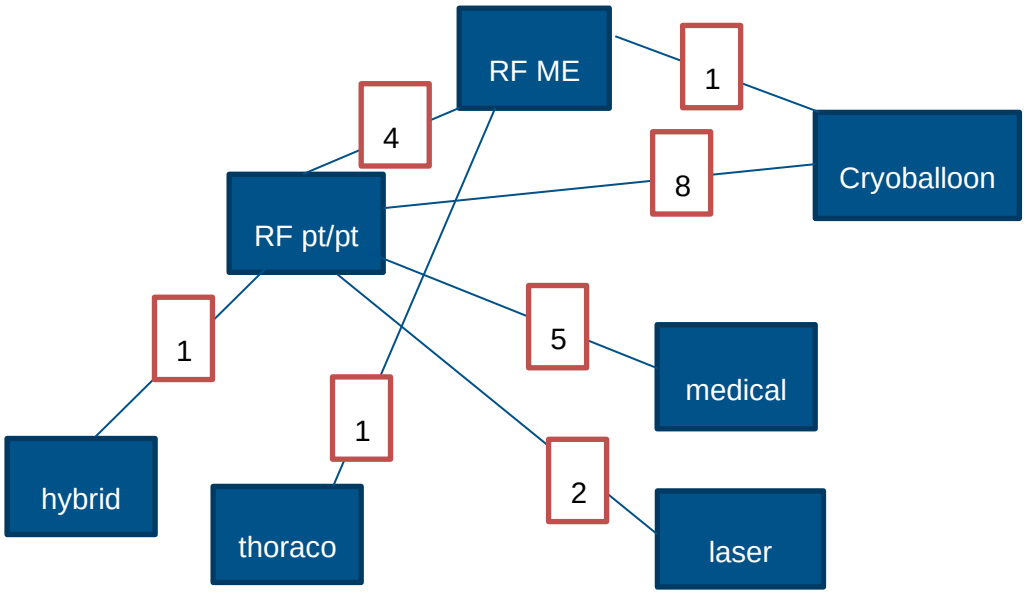
Two studies providing serious adverse outcome data were excluded. The data from Jais, 2008[24] were excluded because data were reported unclearly, and it was not possible to gain a response to our query from the authors. The data from Packer, 2013[39] were excluded because the pre-defined adverse events were strongly biased towards those experienced with cryoablation – hence adverse events of medical care would not be adequately captured.

After excluding studies that reported zero events in all arms, since they do not contribute evidence to the NMA, 22 studies[1, 9, 10, 16, 17, 21, 25-27, 29, 31, 35, 36, 40, 42, 43, 50, 52, 57, 58, 60, 61] involving 7 interventions were included in the serious adverse events network. As for all outcomes, data from studies where any switching of interventions had occurred for individual participants was dealt with using the intention to treat (ITT) principle: that is, events were assigned to the randomised treatment rather than the treatment after switching. The ITT principle was applied because patients switching are often those not responding well to initial treatment, and keeping patients in randomised groups permits capture of this information.

To avoid double counting of data the serious adverse events outcome does not include stroke or mortality events. Serious adverse events were any adverse event reported in any of the included papers that were defined by 2 cardiologists (one was the topic expert on the guideline) as ‘serious’. See Appendix C for more information.

The network can be seen in Figure 8 and the trial data for each of the studies included in the NMA are presented in Table 18.

Figure 8: Network diagram for serious adverse events



RF=radiofrequency; pt/pt=point by point; ME=multielectrode; thoraco = thoracoscopy; Numbers in red squares refer to numbers of studies

Table 18: Study data for serious adverse events network meta-analysis

Study	Intervention	Comparator	Intervention		Comparator	
			Events	n	Events	n
Andrade[1]	RF pt/pt	cryo	3	115	13	231
Davytyan[10]	RF pt/pt	cryo	2.5	45	0.5	46
Hunter[21]	RF pt/pt	cryo	2	77	4	78
Kuck[29]	RF pt/pt	cryo	29	376	25	374
Luik[31]	RF pt/pt	cryo	3	159	11	156
Perez[42]	RF pt/pt	cryo	1	25	1	25
Yagishita[60]	RF pt/pt	cryo	3	125	1	125
You, 2019[61]	RF pt/pt	cryo	2	70	3	140
Jan[25]	RF pt/pt	hybrid	0.5	27	3.5	25
Dukkipatti[16]	RF pt/pt	laser	5	172	8	170
Ucer[52]	RF pt/pt	laser	1	25	1	25
Gal[17]	RF pt/pt	RF ME	6	230	3	230
Kece, 2019 [26]	RF pt/pt	RF ME	1	35	1	35
Mccready[35]	RF pt/pt	RF ME	4	91	1	92
Podd[43]	RF pt/pt	RF ME	0.5	26	1.5	26
Morillo[36]	Medical	RF pt/pt	3	61	6	66
Cosedis Nielsen[9]	Medical	RF pt/pt	12	148	15	146
Pappone[40]	Medical	RF pt/pt	10	99	3	99
Wazni[57]	Medical	RF pt/pt	1	35	2	32
Wilber[58]	Medical	RF pt/pt	2	57	4	103
Koch[27]	cryo	RF ME	2	17	2	15
Sugihara[50]	thoraco	RF ME	6.5	21	0.5	50

n= total number in group; RF = radiofrequency; pt/pt=point by point; ME=multielectrode; cryo=cryoballoon

3.4.2 Inconsistency and goodness of fit

Both fixed effects and random effects baseline models were fitted to the data from the Pappone[40] and Wazni[57] studies. As seen in Table 19, the fixed and random effects baseline models had similar DICs, and so the fixed effect baseline model was preferred, and used to combine with the relative effects from the NMA to obtain absolute probabilities and relative risks outputs.

The fixed and random effects NMA models also had similar DICs, and so the fixed effect NMA model was preferred. A fixed effects inconsistency model was run and the model fit statistics were as seen in Table 19. The consistency NMA has a slightly smaller DIC suggesting that there is no evidence of inconsistency, a conclusion which is supported by comparing risk ratios from the pairwise and NMA models (Table 20). *Number of data points: baseline 2, NMA 44*

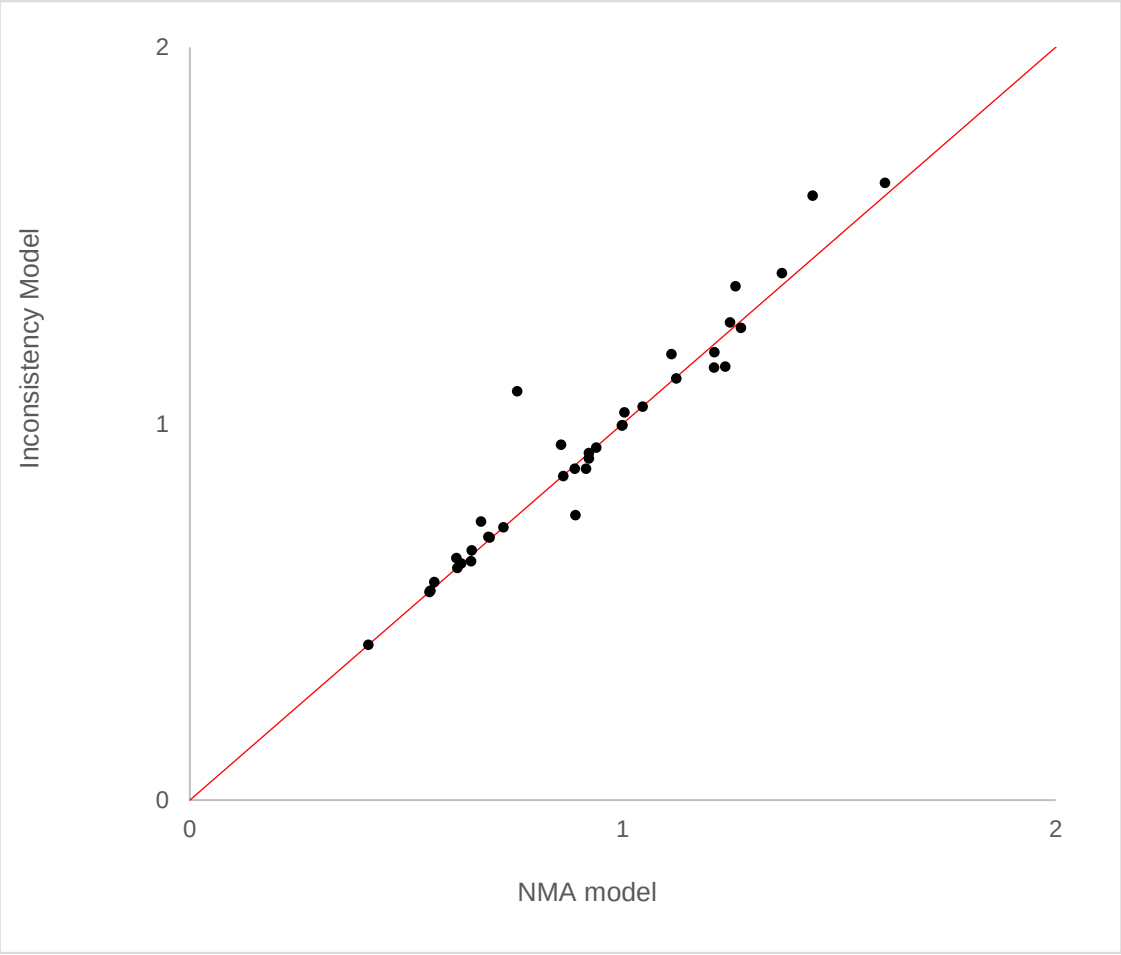
Figure 9 presents the contributions to the posterior mean of the deviances for each data-point for the inconsistency model against that for the consistency NMA model. There is no evidence of inconsistency, as there are no points significantly below the line of equality, which would be indicative of data better predicted by the inconsistency model.

Table 19: Model fit statistics – serious adverse events

	Deviance information criterion (DIC)	Mean of the residual deviance (ResDev)	Posterior median sd (95% CrIs)
Baseline models			
Fixed effects	10.226	3.203	NA
Random effects	9.988	2.023	1.859 (0.1133 – 4.773)
Relative effect models			
NMA Fixed effects	201.988	48.45	NA
NMA Random effects	201.605	43.33	0.482(0.037 – 1.22)
Inconsistency model [FE]	203.601	49.07	NA

Number of data points: baseline 2, NMA 44

Figure 9: Posterior mean of the contribution to the posterior mean residual deviance of the inconsistency model vs. the consistency model – serious adverse events



Number of data points: baseline 2, NMA 44

3.4.1 Results of estimation

Table 20 summarises the results of the pairwise meta-analyses in terms of risk ratios generated from studies directly comparing different interventions, together with the results of the NMA in terms of risk ratios for every possible treatment comparison.

Table 21 presents summary statistics for the 4 interventions included in the network, including the rank of the intervention, probability of the intervention being the best and mean absolute probability of an event. The mean absolute probability of the event in the medical treatment was based on the results of the baseline analysis, and the absolute probabilities for the other treatments are based upon application of the NMA relative effects to the baseline probability for the medical treatment.

Table 20: Risk ratios for serious adverse events; direct pairwise meta-analysis results and NMA results

Intervention	Comparison	Fixed Effects Direct (95% confidence intervals)	Fixed Effects* NMA - median (95% credible intervals)
RF pt pt	Medical	1.01(0.61-1.66)	1.006(0.6072-1.645)
Cryo	Medical	-	1.108(0.5965-2.02)
laser	Medical	-	1.515(0.5036-4.108)
thoraco	Medical	-	9.633(2.717-19.29)
Hybrid	Medical	-	7.076(0.9586-17.66)
RF ME	Medical	-	0.6506(0.2317-1.672)
Cryo	RF pt pt	1.136(0.769-1.66)	1.103(0.7671-1.573)
laser	RF pt pt	1.52(0.55-4.18)	1.500(0.567-3.73)
thoraco	RF pt pt	-	9.298(2.802-22.12)
hybrid	RF pt pt	7.56(0.41-139.17)	6.743(1.004-19.74)
RF ME	RF pt pt	0.56(0.22-1.46)	0.6486(0.2636-1.447)
laser	cryo	-	1.361(0.485-3.626)
thoraco	cryo	-	8.404(2.466-21.59)
hybrid	cryo	-	6.066(0.8819-19.08)
RF ME	cryo	1.13(0.18-7.09)	0.5878(0.2287-1.385)
thoraco	laser	-	6.08(1.449-22.76)
hybrid	Laser	-	4.299(0.5617-19.43)
RF ME	Laser	-	0.4303(0.1198-1.534)
Hybrid	thoraco	-	0.7868(0.09125-3.115)
RF ME	thoraco	0.03(0.00-0.55)	0.07213(0.02157-0.2288)
RF ME	hybrid	-	0.0998(0.02302-0.7517)

**Fixed effects model was used as this gave a better fit to the data*

Table 21: Intervention rank and mean probability of event – serious adverse events

	Probability of adverse events – posterior median (and credible intervals)	Intervention rank - median (95% CrIs)	Probability intervention is best (%)
medical	0.079(0.04351-0.1388)	3 (1-5)	13.35%
RF pt/pt	0.07954(0.0362-0.1663)	3(1-5)	4.184%
cryoballoon	0.08769(0.03706-0.1956)	4(1-5)	3.596%
laser	0.1203(0.03416-0.3567)	5(1-6)	6.102%
thoraco	0.8424(0.2046-0.9996)	7(6-7)	0.0027%
Hybrid	0.5969(0.06979-0.9988)	6(3-7)	1.00%
RF ME	0.05125(0.01543-0.155)	1(1-5)	71.77%

4 Risk of bias

An overall risk of bias assessment was conducted for the studies and outcomes included in the NMA. Overall risk of bias for each study-outcome was determined by consideration of the independent domains of bias: selection bias, performance bias, attrition bias, outcome reporting bias and detection bias. Limitations in each domain were summed, and overall risk of bias was deemed 'very serious' if there were 2 or more serious limitations overall, 'serious' if there was one serious limitation overall, and not serious if there were no limitations overall. Details are provided in review J.

As seen in Table 22, the majority of the relevant evidence for the NMAs had a very serious risk of bias and this was mainly due to concerns about selection and performance bias. Full risk of bias details can be found in Chapter J1 of the guideline

Table 22: Pairwise meta-analysis risk of bias (RoB) assessment per NMA outcome

Study	recurrence	stroke	mortality	Serious AES
Andrade[1]	serious	serious	serious	serious
Bin Waleed[3]	Very serious	-	-	-
Boersema[4]	serious	-	-	-
Bulava[7]	Very serious	-	-	-
Davytyan[10]	-	-	-	Very serious
Dukkipati[16]	Very serious	Very serious	Very serious	Very serious
Gal[17]	Very serious	-	-	Very serious
Giannopoulos[19]	Very serious	-	-	-
Gunawardine[20]	Very serious	-	-	-
Hunter[21]	Very serious	-	-	Very serious
Jais[24]	Very serious	-	Very serious	-
Jan[25]	Very serious	-	-	Very serious
Kece [26]	-	Very serious	-	Very serious
koch[27]	Very serious	-	-	Very serious
Kuck[29]	Very serious	Very serious	Very serious	Very serious
Luik[31]	-	-	-	Very serious
McCready[35]	serious	serious	-	serious
Morillo[36]	Very serious	-	-	Very serious
Nielsen[37]	serious	serious	serious	serious
Packer[39]	-	Very serious	Very serious	-
Pappone[41]	Very serious	-	-	Very serious
Perez castellano[42]	Very serious	-	-	Very serious
Podd[43]	Very serious	-	-	Very serious
Schmidt[48]	-	Very serious	-	-
Sugihara[50]	Very serious	-	-	Very serious
Ucer[52]	-	-	-	Very serious
Wang[55]	-	-	-	-
Watanabe[56]	-	-	-	-
Wazni[57]	Very serious	-	-	Very serious
Wilber[58]	Very serious	-	-	Very serious
Xu[59]	-	-	-	-

Study	recurrence	stroke	mortality	Serious AES
Yagishita, 2020[60]	-	-	-	Very serious
You, 2019[61]	-	-	-	Very serious

5 Discussion

Recurrence

Evidence shows thoracoscopy is more effective than medical treatment, with the 95% credible (CrIs) of the hazard ratios not including the null effect. There is also some evidence suggesting thoracoscopy is more effective than cryoballoon, RF ME, RF pt/pt, laser ablation and hybrid, although this is not conclusive. In terms of the point estimates, thoracoscopy led to about a third of the recurrence observed with the catheter ablation treatments, and to about a tenth of the recurrence seen with medical treatment. The difference with hybrid was smaller, with thoracoscopy leading to about 2/3 of the recurrences seen with hybrid. Thoracoscopy was ranked as best treatment, with a 66% probability of being the best treatment to avoid recurrence of AF. However there was a high level of uncertainty due to direct evidence being derived from only one small study, and the 95% credible intervals of ranking therefore ranged from 1st to 6th.

The hybrid approach had a median ranking of 2nd, with a 28% probability of being the best treatment, but there was again high uncertainty, with 95% credible intervals from 1st-6th. Hybrid was significantly better than medical treatment but was not significantly different to the catheter ablation treatments.

Conversely, evidence shows that medical treatment is inferior to thoracoscopy, hybrid, RF point by point, RF multielectrode, laser and cryoballoon, with the credible intervals not crossing the null line. This inferiority of medical treatment was reflected in its ranking, where it ranked the worst [7th (95% credible intervals 6th to 7th)], and by its 0% probability of being the best treatment to avoid recurrence of AF.

The other four ablation treatments (RF point by point, RF ME, cryoablation and laser ablation) had very similar levels of efficacy in terms of recurrence, as all NMA comparisons between them had point estimates very close to 1. All were ranked in 4th or 5th place with probabilities of being the best treatment of 0.4% (RF point by point), 0.6% (RF ME), 0.9% (cryoballoon) and 4.2% (laser).

On the basis of these results it can be stated with reasonable confidence that thoracoscopy is the best treatment for avoiding recurrence, followed by the hybrid approach. Medical treatment is the worst treatment choice, and the other 4 catheter ablation treatments have similar effects to each other.

Stroke/TIA

Evidence on this outcome did not encompass thoracoscopy and hybrid, and so a full appraisal of the benefits and harms of thoracoscopy and hybrid were unfortunately not possible. Nevertheless, medical treatment was uniformly better than the four other ablation treatments in terms of the risk of stroke/TIA, with relative risks of stroke/TIA from the other 4 ablation treatments being between 4 and 20 times greater than using medical treatment. This evidence was conclusive based on comparisons against laser, cryoballoon and RE ME, but not conclusive against RF point by point, as the 95% CrIs of the risk ratios included the null effect. Medical treatment was ranked 1st (95% CrIs 1st to 2nd) and it had a 95% probability of being the best treatment in terms of reducing the risk of stroke/TIA.

RF multielectrode ablation appeared to carry the greatest risk of stroke/TIA, with a 20 fold increased risk compared to medical treatment, and a 3.3 to 4.3 fold increase in risk compared to the other ablation treatments. Although there was high certainty for its inferiority compared to medical treatment, RF pt pt and cryo, there was some uncertainty about the true direction of effect in the comparisons with laser, as the 95% credible intervals for the risk ratio included the null effect. Unsurprisingly, RE ME ranked the worst (5th) of all treatments in terms of a patient's risk of stroke, with tight 95% CrIs which ranged from 4 to 5.

Cryoballoon, laser and RF point by point were all quite similar to each other in terms of stroke/TIA risk, with relative risks quite close to 1. Very similar results were obtained in the sensitivity analysis where the two studies reporting asymptomatic cerebral lesions were excluded. We are therefore fairly confident that the inclusion of these studies has not unduly influenced the findings.

On the basis of these results, it can be stated that medical care may be the best treatment for avoiding stroke, RF multielectrode is the worst, and cryoballoon, RF point by point and laser may, with some uncertainty, have similar effects to each other. However, because of the lack of data for thoracoscopy and hybrid these assertions are not made with confidence.

Mortality

Evidence on this outcome did not include thoracoscopy, hybrid or RF multielectrode, and so this limits the ability to make an overall appraisal of benefits and harms across all 6 treatments in the NMA.

In terms of point estimates, RF point by point was superior to the other two ablation treatments and medical treatment, with about 2/3 the risk of death compared to medical treatment, and about one quarter to a fifth of the risk of death compared to cryoballoon and laser. This led to RF point by point ranking as the best treatment in terms of risk of mortality. However there was high uncertainty reflected by the wide credible intervals of both the risk ratios and rank, and this contributed to RF point by point having a more modest probability of being the best treatment (60%) than would be expected from the point estimates.

Of the other three treatments, cryotherapy and laser had the worst performance in terms of point estimates, with both having a double to a threefold risk compared to medical treatment and 3-5 times the risk compared to RF point by point. However as there was considerable uncertainty in the effect estimates, the probabilities of being the best treatment were similar between medical care (18%), cryoballoon (5.5%) and laser (16.8%).

Serious adverse events

Evidence on this outcome included all 7 treatments, providing some scope for a weighing up the benefits and harms between all treatments.

The point estimates suggested that RF ME had the lowest risk of serious adverse events, with a 0.43 to 0.65 risk compared to the other catheter ablation techniques, a 2/3 risk compared to medical care and about a 1/14 risk compared to thoracoscopy and a 1/10 risk compared to hybrid (though it should be remembered that serious adverse events did not include stroke or mortality). RF ME therefore ranked as the best treatment in terms of serious adverse events, with a probability of being the best of 72%. However these relative effects were very imprecise, reflected by the considerable uncertainty in the rank of RF ME (95% credible intervals of 1st-5th). The three remaining catheter ablation treatments (RF point by point, cryoballoon and laser) had similar effects to each other, ranked 3rd,4th and 5th respectively just behind ME. However, there was not enough evidence to draw firm conclusions on the superiority or inferiority of the catheter ablation treatments in terms of risk of SAEs, as again there was considerable uncertainty in the estimated risk ratios (the 95% CrIs for most comparisons included the null effect).

Conversely, most comparisons involving thoracoscopy were precise, clearly demonstrating that it was worse than medical care, cryoballoon, laser, RF ME, and RF pt/pt, with point estimates demonstrating a 6 to 14-fold increased risk of serious adverse events compared to these treatments. Evidence also suggested it was worse than the hybrid approach, although this was not conclusive. Thoracostomy was ranked the worst treatment, 7th, with tight 95% credible intervals between 6th and 7th.

Hybrid was ranked second worst treatment, with point estimates indicating 4 to 10 fold increases in risk over the catheter ablation treatments and medical care. However the

precision of these estimates was very low, making definite conclusions difficult. Medical treatment, meanwhile, was ranked as third best.

6 Conclusions

Whilst thoracoscopy was the best treatment in terms of reducing the risk of AF recurrence, most evidence was very imprecise because it was based on one small study. Thoracoscopy also carried the highest risk of serious adverse events. Although some data were found that showed zero events for mortality and stroke with this treatment, for technical reasons these could not be included in the NMA.

The hybrid approach showed promise as a means of reducing recurrence, being ranked just below thoracoscopy, but because data were based on one small trial the precision of estimates were again insufficient to allow firm conclusions. Hybrid was also associated with a relatively high rate of serious adverse effects, though to a lesser extent than thoracoscopy.

Conversely, medical care was relatively free from harms, but it was not effective for recurrence, with almost 3/4 of people having medical treatment experiencing a recurrence. Consequently, all ablation approaches were significantly better at reducing recurrence than medical treatment. Importantly, however, medical care carried the lowest risk of stroke of all *evaluated* treatments, which is very relevant if it is considered that avoidance of stroke is of paramount importance for people with AF. However because thoracoscopy and hybrid were not evaluated for stroke there is the possibility they may have shown lower stroke rates than medical care.

The remaining treatments were catheter ablation treatments: RF point by point, RF multielectrode, cryoballoon and laser. Of these, laser seemed to have the best efficacy in terms of recurrence, though this is uncertain. Bearing in mind the harms of treatment, RF ME is disadvantaged greatly by its high risk of stroke, despite conferring a low risk of other serious adverse events. The lower, albeit fairly uncertain, risk of mortality from RF point by point does give it some advantage over the others, as death is the most critical measure of harm.

Finally, mention should be made that quality of data was impaired by serious or very serious risk of bias in all four outcomes, mainly due to issues around selection, attrition and performance bias. This should be borne in mind when interpreting results, as there is a risk that estimates may be inflated.

In conclusion, medical care is relatively ineffective for preventing AF recurrence. Whilst thoracoscopy, and possibly the hybrid approach, are the most effective ways of reducing the risk of AF recurrence, the high rates of adverse events in these modalities suggest that the catheter ablation treatments, with the exception possibly of RF ME, are a safer option.

References

1. Andrade JG, Champagne J, Dubuc M, Deyell MW, Verma A, Macle L et al. Cryoballoon or radiofrequency ablation for atrial fibrillation assessed by continuous monitoring: a randomized clinical trial. Circulation. 2019; 140:1779–1788

2. Ang R, Hunter RJ, Lim WY, Opel A, Ullah W, Providencia R et al. Long term outcome and pulmonary vein reconnection of patients undergoing cryoablation and/or radiofrequency ablation: results from the Cryo versus RF trial. Journal of Atrial Fibrillation. 2018; 11(3):2072

3. Bin Waleed K, Yin X, Yang X, Dai B, Liu Y, Wang Z et al. Short and long-term changes in platelet and inflammatory biomarkers after cryoballoon and radiofrequency ablation. International Journal of Cardiology. 2019; 285:128-132

4. Boersma LV, van der Voort P, Debruyne P, Dekker L, Simmers T, Rossenbacker T et al. Multielectrode pulmonary vein isolation versus single tip wide area catheter ablation for paroxysmal atrial fibrillation: a multinational multicenter randomized clinical trial. Circulation: Arrhythmia and Electrophysiology. 2016; 9(4):e003151

5. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics. 1998; 7(4):434-455

6. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of Clinical Epidemiology. 1997; 50(6):683-691

7. Bulava A, Hanis J, Sitek D, Osmera O, Karpianus D, Snorek M et al. Catheter ablation for paroxysmal atrial fibrillation: a randomized comparison between multielectrode catheter and point-by-point ablation. Pacing and Clinical Electrophysiology. 2010; 33(9):1039-1046

8. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005; 331(7521):897-900

9. Cosedis Nielsen J, Johannessen A, Raatikainen P, Hindricks G, Walfridsson H, Kongstad O et al. Radiofrequency ablation as initial therapy in paroxysmal atrial fibrillation. New England Journal of Medicine. 2012; 367(17):1587-1595

10. Davtyan K, Shatakhtsyan V, Poghosyan H, Deev A, Tarasov A, Kharlap M et al. Radiofrequency versus cryoballoon ablation of atrial fibrillation: an evaluation using ECG, Holter monitoring, and implantable loop recorders to monitor absolute and clinical effectiveness. BioMed Research International. 2018; 2018:3629384

11. Dias D, Ades AE, Welton NJ, Jansen A, Sutton AJ. Network meta-analysis for decision-making. Wiley. 2018.

12. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Medical Decision Making. 2013; 33(5):607-617

13. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. 2011. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/A-general-linear-modelling-framework-for-pair-wise-and-network-meta-analysis-of-randomised-controlled-trials..pdf

14. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Medical Decision Making. 2013; 33(5):641-656

15. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. NICE DSU technical support document 4: inconsistency in networks of evidence based on randomised controlled trials. 2011. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD4-Inconsistency.final_.15April2014.pdf

16. Dukkipati SR, Cuoco F, Kutinsky I, Aryana A, Bahnson TD, Lakkireddy D et al. Pulmonary vein isolation using the visually guided laser balloon: a prospective, multicenter, and randomized comparison to standard radiofrequency ablation. Journal of the American College of Cardiology. 2015; 66(12):1350-1360

17. Gal P, Aarntzen AE, Smit JJ, Adiyaman A, Misier AR, Delnoy PP et al. Conventional radiofrequency catheter ablation compared to multi-electrode ablation for atrial fibrillation. International Journal of Cardiology. 2014; 176(3):891-895

18. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science. 1992; 7(4):457-472

19. Giannopoulos G, Kossyvakis C, Vrachatis D, Aggeli C, Tsitsinakis G, Letsas K et al. Effect of cryoballoon and radiofrequency ablation for pulmonary vein isolation on left atrial function in patients with nonvalvular paroxysmal atrial fibrillation: a prospective randomized study (Cryo-LAEF study). Journal of Cardiovascular Electrophysiology. 2019; 30(7):991-998

20. Gunawardene MA, Hoffmann BA, Schaeffer B, Chung DU, Moser J, Akbulak RO et al. Influence of energy source on early atrial fibrillation recurrences: a comparison of cryoballoon vs. radiofrequency current energy ablation with the endpoint of unexcitability in pulmonary vein isolation. Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology. 2018; 20(1):43-49

21. Hunter RJ, Baker V, Finlay MC, Duncan ER, Lovell MJ, Tayebjee MH et al. Point-by-point radiofrequency ablation versus the cryoballoon or a novel combined approach: A randomized trial comparing 3 methods of pulmonary vein isolation for paroxysmal atrial fibrillation (The Cryo Versus RF Trial). Journal of Cardiovascular Electrophysiology. 2015; 26(12):1307-1314

22. Hunter RJ, Berriman TJ, Diab I, Kamdar R, Richmond L, Baker V et al. A randomized controlled trial of catheter ablation versus medical treatment of

atrial fibrillation in heart failure (the CAMTAF trial). Circulation: Arrhythmia and Electrophysiology. 2014; 7(1):31-38

23. J. Sweeting M, J. Sutton A, C. Lambert P. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Statistics in Medicine. 2004; 23(9):1351-1375

24. Jais P, Cauchemez B, Macle L, Daoud E, Khairy P, Subbiah R et al. Catheter ablation versus antiarrhythmic drugs for atrial fibrillation: the A4 study. Circulation. 2008; 118(24):2498-2505

25. Jan M, Zizek D, Gersak ZM, Gersak B. Comparison of treatment outcomes between convergent procedure and catheter ablation for paroxysmal atrial fibrillation evaluated with implantable loop recorder monitoring. Journal of Cardiovascular Electrophysiology. 2018; 29(8):1073-1080

26. Kece F, Bruggemans EF, de Riva M, Alizadeh Dehnavi R, Wijnmaalen AP, Meulman TJ et al. Incidence and clinical significance of cerebral embolism during atrial fibrillation ablation with duty-cycled phased-radiofrequency versus cooled-radiofrequency: A randomized controlled trial. JACC: Clinical Electrophysiology. 2019; 5(3):318-326

27. Koch L, Haeusler KG, Herm J, Safak E, Fischer R, Malzahn U et al. Mesh ablator vs. cryoballoon pulmonary vein ablation of symptomatic paroxysmal atrial fibrillation: results of the MACPAF study. Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology. 2012; 14(10):1441-1449

28. Kuck KH, Brugada J, Furnkranz A, Metzner A, Ouyang F, Chun KR et al. Cryoballoon or radiofrequency ablation for paroxysmal atrial fibrillation. New England Journal of Medicine. 2016; 374(23):2235-2245

29. Kuck KH, Furnkranz A, Chun KR, Metzner A, Ouyang F, Schluter M et al. Cryoballoon or radiofrequency ablation for symptomatic paroxysmal atrial fibrillation: reintervention, rehospitalization, and quality-of-life outcomes in the FIRE AND ICE trial. European Heart Journal. 2016; 37(38):2858-2865

30. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in Medicine. 2004; 23(20):3105-3124

31. Luik A, Kunzmann K, Hormann P, Schmidt K, Radzewitz A, Bramlage P et al. Cryoballoon vs. open irrigated radiofrequency ablation for paroxysmal atrial fibrillation: long-term FreezeAF outcomes. BMC Cardiovascular Disorders. 2017; 17(1):135

32. Luik A, Radzewitz A, Kieser M, Walter M, Bramlage P, Hormann P et al. Cryoballoon versus open irrigated radiofrequency ablation in patients with paroxysmal atrial fibrillation: the prospective, randomized, controlled, noninferiority FreezeAF study. Circulation. 2015; 132(14):1311-1319

33. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. The BUGS book – a practical introduction to Bayesian analysis. CRC Press. 2012.

34. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing. 2000; 10:325-337

35. McCready J, Chow AW, Lowe MD, Segal OR, Ahsan S, de Bono J et al. Safety and efficacy of multipolar pulmonary vein ablation catheter vs. irrigated radiofrequency ablation for paroxysmal atrial fibrillation: a randomized multicentre trial. Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology. 2014; 16(8):1145-1153

36. Morillo CA, Verma A, Connolly SJ, Kuck KH, Nair GM, Champagne J et al. Radiofrequency ablation vs antiarrhythmic drugs as first-line treatment of paroxysmal atrial fibrillation (RAAFT-2): a randomized trial. JAMA. 2014; 311(7):692-700

37. Nielsen JC, Johannessen A, Raatikainen P, Hindricks G, Walfridsson H, Pehrson SM et al. Long-term efficacy of catheter ablation as first-line therapy for paroxysmal atrial fibrillation: 5-year outcome in a randomised clinical trial. Heart. 2017; 103(5):368-376

38. Ogawa S, Yamashita T, Yamazaki T, Aizawa Y, Atarashi H, Inoue H et al. Optimal treatment strategy for patients with paroxysmal atrial fibrillation: J-RHYTHM Study. Circulation Journal. 2009; 73(2):242-248

39. Packer DL, Kowal RC, Wheelan KR, Irwin JM, Champagne J, Guerra PG et al. Cryoballoon ablation of pulmonary veins for paroxysmal atrial fibrillation: first results of the North American Arctic Front (STOP AF) pivotal trial. Journal of the American College of Cardiology. 2013; 61(16):1713-1723

40. Pappone C, Augello G, Sala S, Gugliotta F, Vicedomini G, Gulletta S et al. A randomized trial of circumferential pulmonary vein ablation versus antiarrhythmic drug therapy in paroxysmal atrial fibrillation: the APAF Study. Journal of the American College of Cardiology. 2006; 48(11):2340-2347

41. Pappone C, Vicedomini G, Augello G, Manguso F, Saviano M, Baldi M et al. Radiofrequency catheter ablation and antiarrhythmic drug therapy: a prospective, randomized, 4-year follow-up trial: the APAF study. Circulation: Arrhythmia and Electrophysiology. 2011; 4(6):808-814

42. Perez-Castellano N, Fernandez-Cavazos R, Moreno J, Canadas V, Conde A, Gonzalez-Ferrer JJ et al. The COR trial: a randomized study with continuous rhythm monitoring to compare the efficacy of cryoenergy and radiofrequency for pulmonary vein isolation. Heart Rhythm. 2014; 11(1):8-14

43. Podd SJ, Sulke AN, Sugihara C, Furniss SS. Phased multipolar radiofrequency pulmonary vein isolation is as effective and safe as conventional irrigated point-to-point ablation. A prospective randomised 1-year implantable cardiac monitoring device follow-up trial. Journal of Interventional Cardiac Electrophysiology. 2015; 44(3):257-264

44. Pokushalov E, Romanov A, Artyomenko S, Baranova V, Losik D, Bairamova S et al. Cryoballoon versus radiofrequency for pulmonary vein re-isolation after a failed initial ablation procedure in patients with paroxysmal atrial fibrillation. Journal of Cardiovascular Electrophysiology. 2013; 24(3):274-279

45. Pokushalov E, Romanov A, De Melis M, Artyomenko S, Baranova V, Losik D et al. Progression of atrial fibrillation after a failed initial ablation procedure in patients with paroxysmal atrial fibrillation: a randomized comparison of drug

therapy versus reablation. Circulation: Arrhythmia and Electrophysiology. 2013; 6(4):754-760

46. Reynolds MR, Walczak J, White SA, Cohen DJ, Wilber DJ. Improvements in symptoms and quality of life in patients with paroxysmal atrial fibrillation treated with radiofrequency catheter ablation versus antiarrhythmic drugs. Circulation: Cardiovascular Quality and Outcomes. 2010; 3(6):615-623

47. Schirdewan A, Herm J, Roser M, Landmesser U, Endres M, Koch L et al. Loop recorder detected high rate of atrial fibrillation recurrence after a single balloon- or basket-based ablation of paroxysmal atrial fibrillation: results of the MACPAF study. Frontiers in Cardiovascular Medicine. 2017; 4(4):1-8

48. Schmidt B, Gunawardene M, Krieg D, Bordignon S, Furnkranz A, Kulikoglu M et al. A prospective randomized single-center study on the risk of asymptomatic cerebral lesions comparing irrigated radiofrequency current ablation with the cryoballoon and the laser balloon. Journal of Cardiovascular Electrophysiology. 2013; 24(8):869-874

49. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B. 2002; 64(4):583-639

50. Sugihara C, Furniss S, Hyde J, Lewis M, Sulke N. Results of the first investigator-initiated randomized clinical trial of nMARQTM, PVACTM, and thoracoscopic ablation for paroxysmal atrial fibrillation. Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology. 2018; 20(FI_3):F384-F391

51. Thom H, López-López JA, Welton NJ. Shared parameter model for competing risks and different data summaries in meta-analysis: Implications for common and rare outcomes. Research Synthesis Methods. 2020; 11(1):91-104

52. Ucer E, Janeczko Y, Seegers J, Fredersdorf S, Friemel S, Poschenrieder F et al. A RAndomized Trial to compare the acute reconnection after pulmonary vein ISolation with Laser-BalloON versus radiofrequency Ablation: RATISBONA trial. Journal of Cardiovascular Electrophysiology. 2018; 29(5):733-739

53. van Valkenhoef G, Dias S, Ades AE, Welton NJ. Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. Research Syntheis Methods. 2016; 7(1):80-93

54. Walfridsson H, Walfridsson U, Nielsen JC, Johannessen A, Raatikainen P, Janzon M et al. Radiofrequency ablation as initial therapy in paroxysmal atrial fibrillation: results on health-related quality of life and symptom burden. The MANTRA-PAF trial. Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology. 2015; 17(2):215-221

55. Wang S, Liu L, Zou C. Comparative study of video-assisted thoracoscopic surgery ablation and radiofrequency catheter ablation on treating paroxysmal atrial fibrillation: a randomized, controlled short-term trial. Chinese Medical Journal. 2014; 127(14):2567-2570

56. Watanabe R, Sairaku A, Yoshida Y, Nanasato M, Kamiya H, Suzuki H et al. Head-to-head comparison of acute and chronic pulmonary vein stenosis for cryoballoon versus radiofrequency ablation. Pacing and Clinical Electrophysiology. 2018; 41(4):376-382

57. Wazni OM, Marrouche NF, Martin DO, Verma A, Bhargava M, Saliba W et al. Radiofrequency ablation vs antiarrhythmic drugs as first-line treatment of symptomatic atrial fibrillation: a randomized trial. JAMA. 2005; 293(21):2634-2640

58. Wilber DJ, Pappone C, Neuzil P, De Paola A, Marchlinski F, Natale A et al. Comparison of antiarrhythmic drug therapy and radiofrequency catheter ablation in patients with paroxysmal atrial fibrillation: a randomized controlled trial. JAMA. 2010; 303(4):333-340

59. Xu Y, Sharma D, Du F, Li G, Xu G. Comparison of circumferential pulmonary vein isolation and antiarrhythmic drug therapy in patients with atrial fibrillation. Cardiology and Therapy. 2012; 1(3):1-7

60. Yagishita A, Goya M, Iesaka Y, Nitta J, Takahashi A, Nagata Y et al. A prospective multicenter study of direct comparison of feasibility and safety of pulmonary vein isolation using the minimally interrupted apixaban between second-generation cryoballoon and radiofrequency ablation of paroxysmal atrial fibrillation: J-HIT apixaban. Journal of Arrhythmia. 2020; 36(4):617-623

61. You L, Yao L, Zhou B, Jin L, Yin H, Wu J et al. Effects of different ablation strategies on long-term left atrial function in patients with paroxysmal atrial fibrillation: a single-blind randomized controlled trial. Scientific Reports. 2019; 9(1):7695

Appendices

Appendix A: WinBUGS Code

A.1 recurrence

A.1.1 Main code

A.1.1.1 Random effects

```
# Binomial likelihood, cloglog link
# Random effects model for multi-arm trials
model{                          # *** PROGRAM STARTS
for(i in 1:ns){                 # LOOP THROUGH STUDIES
   w[i,1] <- 0    # adjustment for multi-arm trials is zero for control arm
   delta[i,1] <- 0            # treatment effect is zero for control arm
   mu[i] ~ dnorm(0,.01)          # vague priors for all trial baselines
   for (k in 1:na[i]) {           # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
# model for linear predictor
      cloglog(p[i,k]) <- mu[i] + delta[i,k]
      rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators
#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))        }
#  summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
   for (k in 2:na[i]) {           # LOOP THROUGH ARMS
# trial-specific LHR distributions
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LHR distributions (with multi-arm trial correction)
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LHR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k
# adjustment for multi-arm RCTs
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
    }
 }
totresdev <- sum(resdev[])          # Total Residual Deviance

d[1]<-0       # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.1) }
sd ~ dunif(0,5)     # vague prior for between-trial SD
tau <- pow(sd,-2)   # between-trial precision = (1/between-trial variance)

A ~ dnorm(meanA,precA)
for (k in 1:nt) { cloglog(T[k]) <- A + d[k]  } # Note log(1)=0, so not needed when time = 1 year

# Ranking and prob{treatment k is best}
 for (k in 1:nt) {
            rk[k]<-rank(d[],k)
```

```
                        best[k]<-equals(rank(d[],k),1)
                        }

# pairwise HRs
for (c in 1:(nt-1)) {
        for (k in (c+1):nt)  {
                        lhr[c,k] <- d[k] - d[c]
          log(hr[c,k]) <- lhr[c,k]
          }
        }
}                        # *** PROGRAM ENDS
```

Data

ns= number of studies; nt=number of treatments #1=medical,2=RF pt pt,3=cryo,4=laser,5=thoraco,6=hybrid,7=RF ME

Baseline time in years

list(ns=18, nt=7, meanA=0.2822, precA=119.468)

r[,1]	r[,2]	n[,1]	n[,2]	t[,1]	t[,2]	na[]	
53	111	115	231	2	3	2	#andrade
3	4	29	28	2	3	2	#bin waleed
3	6	30	30	2	3	2	#gunawardine
41	26	77	78	2	3	2	#hunter
143	138	376	374	2	3	2	#kuck
8	13	25	25	2	3	2	#perez
17	10	26	24	2	6	2	#jan
60	61	166	167	2	4	2	#dukkipatti
11	14	58	59	2	7	2	#boersma
15	12	51	51	2	7	2	#bulava
40	37	91	92	2	7	2	#mcready
12	11	25	25	2	7	2	#podd
42	7	55	53	1	2	2	#jais
44	36	61	66	1	2	2	#morrillo
22	4	35	32	1	2	2	#wazni
46	38	56	103	1	2	2	#wilber
13	10	22	15	3	7	2	#koch
3	20	20	49	5	7	2	#sugihara

END

Initial Values

#chain 1

list(d=c(NA, 0,0,0,0,0,0), sd=1, mu=c(0, 0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0))

#chain 2

list(d=c(NA, -1,-2,-1,-1,-3,-1), sd=4, mu=c(-1, -1, -1, -1, -1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,1,-1,1))

#chain 3

list(d=c(NA, 1,0,2,0,3,0), sd=2, mu=c(-1, 1, -1, -1, 1, -1, 1,1,1,0,-1,-1,0,0,1,-1,1,-1))

A.1.1.2 Fixed effects

```
# Binomial likelihood, cloglog link
# Fixed effects model
model{                    # *** PROGRAM STARTS
for(i in 1:ns){           # LOOP THROUGH STUDIES
   mu[i] ~ dnorm(0,.01)     # vague priors for all trial baselines
   for (k in 1:na[i])  {     # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k])    # binomial likelihood
```

```
# model for linear predictor
      cloglog(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]]
# expected value of the numerators
       rhat[i,k] <- p[i,k] * n[i,k]
#Deviance contribution
       dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
           +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
     }
# summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
    }
totresdev <- sum(resdev[])      # Total Residual Deviance

d[1]<-0    # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.1) }

A ~ dnorm(meanA,precA)
for (k in 1:nt) { cloglog(T[k]) <- A + d[k]  } # Note log(1)=0, so not needed when time in years

# Ranking and prob{treatment k is best}
 for (k in 1:nt) {
     rk[k]<-rank(d[],k)
                          best[k]<-equals(rank(d[],k),1)
                          }

# pairwise HRs
for (c in 1:(nt-1)) {
          for (k in (c+1):nt)  {
                          lhr[c,k] <- d[k] - d[c]
            log(hr[c,k]) <- lhr[c,k]
            }
          }

}                                        # *** PROGRAM ENDS
```

Data

```
 # ns= number of studies; nt=number of treatments #1=medical,2=RF pt
pt,3=cryo,4=laser,5=thoraco,6=hybrid,7=RF ME


# Baseline time in years
list(ns=18, nt=7, meanA=0.2822, precA=119.468)
```

r[,1]	r[,2]	n[,1]	n[,2]	t[,1]	t[,2]	na[]	
53	111	115	231	2	3	2	#andrade
3	4	29	28	2	3	2	#bin waleed
3	6	30	30	2	3	2	#gunawardine
41	26	77	78	2	3	2	#hunter
143	138	376	374	2	3	2	#kuck
8	13	25	25	2	3	2	#perez
17	10	26	24	2	6	2	#jan
60	61	166	167	2	4	2	#dukkipatti
11	14	58	59	2	7	2	#boersma
15	12	51	51	2	7	2	#bulava
40	37	91	92	2	7	2	#mcready
12	11	25	25	2	7	2	#podd
42	7	55	53	1	2	2	#jais

```
44      36      61      66      1       2       2       #morrillo
22      4       35      32      1       2       2       #wazni
46      38      56      103     1       2       2       #wilber
13      10      22      15      3       7       2       #koch
3       20      20      49      5       7       2       #sugihara
END
```

Initial Values

```
#chain 1
list(d=c( NA, 0,0,0,0,0,0),  mu=c(0, 0, 0, 0, 0,  0,0,0,0,0,0,0,0,0,0,0,0,0 ))
#chain 2
list(d=c( NA, -1,-2,-1,-1,-3,-1),  mu=c(-1, -1,  -1,    -1, -1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,1,-
1, -1))
#chain 3
list(d=c( NA, 1,0,2,0,3,0),  mu=c(-1, 1, -1, -1, 1,   -1, 1,1,1,0,-1,-1,0,0,1,-1,1,1))
```

A.1.2 Baseline model

A.1.2.1 Random effects

```
# Binomial likelihood, cloglog link
# Baseline random effects model
model{                    # *** PROGRAM STARTS
for (i in 1:ns){          # LOOP THROUGH STUDIES
   r[i] ~ dbin(p[i],n[i])              # Likelihood
   cloglog(p[i]) <- log(time[i]) + mu[i]                    # Log-hazard rate
                  mu[i] ~ dnorm(m,tau.m)     # Random effects model

                  # expected value of the numerators
   rhat[i] <- p[i] * n[i]
                  #Deviance contribution
   dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))
         +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))
 }
totresdev <- sum(dev[])                                  # total residual deviance

mu.new ~ dnorm(m,tau.m)          # predictive dist. (log-odds)
m ~ dnorm(0,.0001)            # vague prior for mean
var.m <- 1/tau.m              # between-trial variance
tau.m <- pow(sd.m,-2)   # between-trial precision = (1/between-trial variance)
sd.m ~ dunif(0,5)             # vague prior for between-trial SD
#tau.m ~ dgamma(0.001,0.001)
#sd.m <- sqrt(var.m)
cloglog(R) <- log(x) + m          # R is posterior probability of response per a unit time
cloglog(R.new) <- log(x) + mu.new   # R.new is predictive probability of response per a unit
time
}
#Time in years
list(ns=3, x=1)  # ns=number of studies, x = specified unit of time

r[]       n[]   time[]
42        55      1                  #jais
22        35       1        #wazni
```

```
75      99      1       #pappone
END
```

Inits

```
list(m=0)

list(m= -1)

list(m = 1)
```

A.1.2.2 Fixed effects

```
# Binomial likelihood, cloglog link
# Baseline fixed effect model

model{                          # *** PROGRAM STARTS
for (i in 1:ns){                # LOOP THROUGH STUDIES
   r[i] ~ dbin(p[i],n[i])               # Likelihood
   cloglog(p[i]) <- log(time[i]) + m                                    # Log-hazard rate

# expected value of the numerators
   rhat[i] <- p[i] * n[i]
#Deviance contribution
   dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))
          +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))
 }
totresdev <- sum(dev[])                                     # total residual deviance
m ~ dnorm(0,.0001)            # vague prior for mean


cloglog(R) <- log(x)+ m                 # posterior probability of response per unit(x) time
}


#Time in years
list(ns=3, x=1)  # ns=number of studies, x = specified unit of time

r[]     n[]   time[]
42      55      1                 #jais
22      35       1        #wazni
75      99       1        #pappone
END
```

Inits

```
list(m=0)
list(m= -1)
list(m = 1)
```

A.1.3 Inconsistency model

```
# Binomial likelihood, cloglog link, inconsistency model
# Random effects model
model{                    # *** PROGRAM STARTS
for(i in 1:ns){           # LOOP THROUGH STUDIES
   delta[i,1]<-0          # treatment effect is zero in control arm
   mu[i] ~ dnorm(0,.1)  # vague priors for trial baselines
   for (k in 1:na[i])  {   # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
      cloglog(p[i,k]) <- mu[i] + delta[i,k]  # model for linear predictor
#Deviance contribution
      rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
        +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
     }
# summed residual deviance contribution for this trial
  resdev[i] <- sum(dev[i,1:na[i]])
  for (k in 2:na[i]) {  # LOOP THROUGH ARMS
# trial-specific LHR distributions
      delta[i,k] ~ dnorm(d[t[i,1],t[i,k]] ,tau)
     }
 }
totresdev <- sum(resdev[])   # Total Residual Deviance

sd ~ dunif(0,5)  # vague prior for between-trial standard deviation
var <- pow(sd,2) # between-trial variance
tau <- 1/var     # between-trial precision

# vague priors for treatment effects
for(c in 1:nt){ d[c,c]<-0 }
for(c in 1:(nt-1)){
        for(k in (c+1):nt){
                d[c,k]~dnorm(0,0.01)
                log(hr[c,k]) <- d[c,k]
                d[k,c] <- -d[c,k]
                }
        }

} # *** PROGRAM ENDS
```

Data

```
 # ns= number of studies; nt=number of treatments

list(ns=18, nt=7)

r[,1]   r[,2]   n[,1]   n[,2]   t[,1]   t[,2]   na[]
53      111     115     231     2       3       2       #andrade
3       4       29      28      2       3       2       #bin waleed
3       6       30      30      2       3       2       #gunawardine
41      26      77      78      2       3       2       #hunter
143     138     376     374     2       3       2       #kuck
8       13      25      25      2       3       2       #perez
17      10      26      24      2       6       2       #jan
60      61      166     167     2       4       2       #dukkipatti
11      14      58      59      2       7       2       #boersma
15      12      51      51      2       7       2       #bulava
40      37      91      92      2       7       2       #mcready
12      11      25      25      2       7       2       #podd
42      7       55      53      1       2       2       #jais
44      36      61      66      1       2       2       #morrillo
22      4       35      32      1       2       2       #wazni
```

```
46      38      56      103     1       2       2       #wilber
13      10      22      15      3       7       2       #koch
3       20      20      49      5       7       2       #sugihara
END
```

Initial Values

```
# chain 1
list(sd=1, mu=c(0,0,0,  0,0,0,0,0,0,  0,0,0,0,0,0,  0,0,0),
d = structure(.Data = c(NA,0,0,0,0,0, 0,
NA,NA,0,0,0,0, 0,
NA,NA,NA,0,0,0, 0,
NA,NA,NA,NA,0,0, 0,
NA,NA,NA,NA,NA,0,0,
NA,NA,NA,NA,NA,NA, 0,
NA,NA,NA,NA,NA,NA,NA), .Dim = c(7,7)))

# chain 2
list(sd=1.5, mu=c(0,1,-1,    2,-2,  0,0,1,-1,    2,-2,  0, 0,1,-1,    2,-2,2),
d = structure(.Data = c(NA,0,1,0,1,0, 0,
NA,NA,1,0,1,0, 1,
NA,NA,NA,0,1,0, 1,
NA,NA,NA,NA,0,1, 1,
NA,NA,NA,NA,NA,0,1,
NA,NA,NA,NA,NA,NA, 0,
NA,NA,NA,NA,NA,NA,NA), .Dim = c(7,7)))

# chain 3
list(sd=3, mu=c(3,2,-2,       0,-1,  3,3,2,-2,       0,-1,  3, 3,2,-2,       0,-1,0),
d = structure(.Data = c(NA,0,1,2,1,0, 0,
NA,NA,1,0,1,2,0,
NA,NA,NA,0,1,2, 1,
NA,NA,NA,NA,2,1,1,
NA,NA,NA,NA,NA,2,0,
NA,NA,NA,NA,NA,NA,0,
NA,NA,NA,NA,NA,NA,NA), .Dim = c(7,7)))
```

A.2 Stroke

A.2.1 Main code

A.2.1.1 Random effects

```
# Binomial likelihood, logit link
# Random effects model for multi-arm trials
model{                          # *** PROGRAM STARTS
for(i in 1:ns){                 # LOOP THROUGH STUDIES
   w[i,1] <- 0    # adjustment for multi-arm trials is zero for control arm
   delta[i,1] <- 0            # treatment effect is zero for control arm
   mu[i] ~ dnorm(0,.0001)           # vague priors for all trial baselines
   for (k in 1:na[i]) {            # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
      logit(p[i,k]) <- mu[i] + delta[i,k]  # model for linear predictor
      rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators
#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))        }
#  summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
```

```
   for (k in 2:na[i]) {            # LOOP THROUGH ARMS
# trial-specific LOR distributions
       delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LOR distributions (with multi-arm trial correction)
       md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
       taud[i,k] <- tau *2*(k-1)/k
# adjustment for multi-arm RCTs
       w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
       sw[i,k] <- sum(w[i,1:k-1])/(k-1)
     }
  }
totresdev <- sum(resdev[])          # Total Residual Deviance
d[1]<-0       # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
sd ~ dunif(0,5)     # vague prior for between-trial SD
tau <- pow(sd,-2)   # between-trial precision = (1/between-trial variance)
# Provide estimates of treatment effects T[k] on the natural (probability) scale
# Given a Mean Effect, meanA, for 'standard' treatment A,
# with precision (1/variance) precA
A ~ dnorm(meanA,precA)
for (k in 1:nt) { logit(T[k]) <- A + d[k]  }

rr[1]<- 1
for (k in 2:nt)  {
rr[k]<- T[k]/T[1]  }                          # calculate relative risk


# Ranking and prob{treatment k is best}
 for (k in 1:nt) {
             rk[k]<-rank(rr[],k)
best[k]<-equals(rank(rr[],k),1)}

# pairwise ORs and RRs
for (c in 1:(nt-1))
        {  for (k in (c+1):nt)
              {  lor[c,k] <- d[k] - d[c]
                 log(or[c,k]) <- lor[c,k]
                 lrr[c,k] <- log(rr[k]) - log(rr[c])
                 log(rrisk[c,k]) <- lrr[c,k]

              }
         }
}

}                              # *** PROGRAM ENDS

 Data
# ns= number of studies; nt=number of treatments
#key1=medical,2=RF pt pt,3=cryo,4=laser,5=RF ME

ist(ns=9 nt=5, meanA=-5.165, precA=0.602793)
r[,1]    r[,2]    r[,3]    n[,1]    n[,2]    n[,3]    t[,1]    t[,2]    t[,3]    na[]
2        2        NA       376      374      NA       2        3        NA       {{{2     #kuck{[
```

```
0.5     2.5     NA      116     232     NA      2       3       NA      2       #andrade
8       6       8       33      33      33      2       3       4       3       #schmidt
1       2       NA      172     170     NA      2       4       NA      2       #dukkipatti
2       8       NA      35      35      NA      2       5       NA      2       #kece
0.5     2.5     NA      92      93      NA      2       5       NA      2       #mcready
1       2       NA      148     146     NA      1       2       NA      2       #nielsen
0.5     1.5     NA      100     100     NA      1       2       NA      2       #pappone
0.5     7.5     NA      83      164     NA      1       3       NA      2       #packer

END
```

Initial Values

```
#chain 1
list(d=c( NA, 0,0,0,0), sd=1, mu=c(0, 0, 0, 0, 0,0,0,0,0 ))
#chain 2
list(d=c( NA, -1,-1,-1,-1), sd=4, mu=c(-3, -3, -3, -3, -3,-3, -3, -3,3))
#chain 3
list(d=c( NA, 2,0,3,1), sd=2, mu=c(-3, 5, -1, -3, 7,2,1,4, 2))
```

A.2.1.2 Fixed effects

```
# Binomial likelihood, logit link
# Fixed effects model
model{                          # *** PROGRAM STARTS
for(i in 1:ns){                 # LOOP THROUGH STUDIES
   mu[i] ~ dnorm(0,.0001)      # vague priors for all trial baselines
   for (k in 1:na[i])  {       # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k])    # binomial likelihood
# model for linear predictor
      logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]]
# expected value of the numerators
      rhat[i,k] <- p[i,k] * n[i,k]
#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
          +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
    }
# summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
    }
totresdev <- sum(resdev[])      # Total Residual Deviance
d[1]<-0    # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
# Provide estimates of treatment effects T[k] on the natural (probability) scale
# Given a Mean Effect, meanA, for 'standard' treatment A,
# with precision (1/variance) precA
A ~ dnorm(meanA,precA)
for (k in 1:nt) { logit(T[k]) <- A + d[k]  }

rr[1]<- 1
for (k in 2:nt)  {
rr[k]<- T[k]/T[1]  }                        # calculate relative risk
```

```
# Ranking and prob{treatment k is best}
 for (k in 1:nt) {
            rk[k]<-rank(rr[],k)
best[k]<-equals(rank(rr[],k),1)}

# pairwise ORs and RRs
for (c in 1:(nt-1))
        {  for (k in (c+1):nt)
              {  lor[c,k] <- d[k] - d[c]
                 log(or[c,k]) <- lor[c,k]
                 lrr[c,k] <- log(rr[k]) - log(rr[c])
                 log(rrisk[c,k]) <- lrr[c,k]

              }
         }
}

}                                        # *** PROGRAM ENDS
```

Data

```
# ns= number of studies; nt=number of treatments
#key1=medical,2=RF pt pt,3=cryo,4=laser,5=RF ME

list(ns=9 nt=5, meanA=-5.165, precA=0.602793)
```

r[,1]	r[,2]	r[,3]	n[,1]	n[,2]	n[,3]	t[,1]	t[,2]	t[,3]	na[]	
2	2	NA	376	374	NA	2	3	NA	2	#kuck
0.5	2.5	NA	116	232	NA	2	3	NA	2	#andrade
8	6	8	33	33	33	2	3	4	3	#schmidt
1	2	NA	172	170	NA	2	4	NA	2	#dukkipatti
2	8	NA	35	35	NA	2	5	NA	2	#kece
0.5	2.5	NA	92	93	NA	2	5	NA	2	#mcready
1	2	NA	148	146	NA	1	2	NA	2	#nielsen
0.5	1.5	NA	100	100	NA	1	2	NA	2	#pappone
0.5	7.5	NA	83	164	NA	1	3	NA	2	#packer

```
END
```

Initial Values

```
#chain 1
list(d=c( NA, 0,0,0,0), mu=c(0, 0, 0, 0, 0,0,0,0,0 ))
#chain 2
list(d=c( NA, -1,-1,-1,-1), mu=c(-3, -3, -3, -3, -3,-3, -3, -3, 3))
#chain 3
list(d=c( NA, 2,0,3,1), mu=c(-3, 5, -1, -3, 7,2,1,4, 2))
```

A.2.2 Baseline model

A.2.2.1 Random effects

```
# Binomial likelihood, logit link
# Baseline random effects model
model{                    # *** PROGRAM STARTS
for (i in 1:ns){          # LOOP THROUGH STUDIES
   r[i] ~ dbin(p[i],n[i])          # Likelihood
   logit(p[i]) <- mu[i]                    # Log-odds of response
                  mu[i] ~ dnorm(m,tau.m)   # Random effects model

                  # expected value of the numerators
   rhat[i] <- p[i] * n[i]
                  #Deviance contribution
   dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))
          +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))
 }
totresdev <- sum(dev[])                          # total residual deviance

mu.new ~ dnorm(m,tau.m)        # predictive dist. (log-odds)
m ~ dnorm(0,.0001)             # vague prior for mean
var.m <- 1/tau.m               # between-trial variance
tau.m <- pow(sd.m,-2)   # between-trial precision = (1/between-trial variance)
sd.m ~ dunif(0,5)              # vague prior for between-trial SD
#tau.m ~ dgamma(0.001,0.001)
#sd.m <- sqrt(var.m)
logit(R) <- m                  # posterior probability of response
logit(R.new) <- mu.new         # predictive probability of response
}
```

Data

```
list(ns=1)  # ns=number of studies

r[]       n[]
1         100     #various sources
END
```

Inits

```
list(mu=c(0), sd.m=1, m=0)

list(mu = c(-1), sd.m=2, m= -1)

list(mu = c(1), sd.m = 0.5, m = 1)
```

A.2.2.2 Fixed effects

```
# Binomial likelihood, logit link
# Baseline fixed effect model
model{                    # *** PROGRAM STARTS
for (i in 1:ns){          # LOOP THROUGH STUDIES
```

```
   r[i] ~ dbin(p[i],n[i])              # Likelihood
   logit(p[i]) <- m                                        # Log-odds of response

                  # expected value of the numerators
   rhat[i] <- p[i] * n[i]
                  #Deviance contribution
   dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))
         +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))
 }
totresdev <- sum(dev[])                                    # total residual deviance
m ~ dnorm(0,.0001)            # vague prior for mean
logit(R) <- m                 # posterior probability of response
}
```

Data

```
list(ns=1)  # ns=number of studies

r[]       n[]
1         100     #various sources
END
```

Inits

```
list(m=0)

list(m= -1)

list(m = 1)
```

A.2.3 Inconsistency model

```
# Binomial likelihood, logit link
# Fixed effects INCONSISTENCY model
model{                        # *** PROGRAM STARTS
for(i in 1:ns){               # LOOP THROUGH STUDIES
   mu[i] ~ dnorm(0,.0001)     # vague priors for all trial baselines
   for (k in 1:na[i])  {      # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k])    # binomial likelihood
# model for linear predictor
      logit(p[i,k]) <- mu[i] + d[t[i,1],t[i,k]]
# expected value of the numerators
      rhat[i,k] <- p[i,k] * n[i,k]
#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
          +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
    }
# summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
    }
totresdev <- sum(resdev[])     # Total Residual Deviance

# vague priors for treatment effects
```

```
for(c in 1:nt){ d[c,c]<-0 }
for(c in 1:(nt-1)){
        for(k in (c+1):nt){
                d[c,k]~dnorm(0,0.0001)
                log(hr[c,k]) <- d[c,k]
                d[k,c] <- -d[c,k]
                }
        }

}                                       # *** PROGRAM ENDS
```

Data

```
# ns= number of studies; nt=number of treatments
#key1=medical,2=RF pt pt,3=cryo,4=laser,5=RF ME
list(ns=9 nt=5)
```

r[,1]	r[,2]	r[,3]	n[,1]	n[,2]	n[,3]	t[,1]	t[,2]	t[,3]	na[]	
2	2	NA	376	374	NA	2	3	NA	2	#kuck
0.5	2.5	NA	116	232	NA	2	3	NA	2	#andrade
8	6	8	33	33	33	2	3	4	3	#schmidt
1	2	NA	172	170	NA	2	4	NA	2	#dukkipatti
2	8	NA	35	35	NA	2	5	NA	2	#kece
0.5	2.5	NA	92	93	NA	2	5	NA	2	#mcready
1	2	NA	148	146	NA	1	2	NA	2	#nielsen
0.5	1.5	NA	100	100	NA	1	2	NA	2	#pappone
0.5	7.5	NA	83	164	NA	1	3	NA	2	#packer

```
END
```

Initial Values

```
# chain 1
list(mu=c(0,0,0,  0,0,0,0,0, 0),
d = structure(.Data = c(NA,0,0,0,0
NA,NA,0,0,0
NA,NA,NA,0,0
NA,NA,NA,NA,0
NA,NA,NA,NA,NA),  .Dim = c(5,5)))

# chain 2
list(mu=c(0,1,-1,    2,-2, 2,-1,2, 1),
d = structure(.Data = c(NA,0,1,0,0
NA,NA,1,0,0
NA,NA,NA,0,0
NA,NA,NA,NA,0
NA,NA,NA,NA,NA), .Dim = c(5,5)))

# chain 3
list(mu=c(3,2,-2,       0,-1, 1,1,-1, 1),
d = structure(.Data = c(NA,0,1,2,0
NA,NA,1,0,0
NA,NA,NA,0,0
NA,NA,NA,NA,0
NA,NA,NA,NA,NA), .Dim = c(5,5)))
```

A.3 Mortality

A.3.1 Main code

A.3.1.1 Random effects

This code is part of
Dias, S., Welton, N.J., Sutton, A.J. & Ades, A.E. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011; last updated September 2016 (available from http://www.nicedsu.org.uk).
This work should be cited whenever the code is used whether in its standard form or adapted.

```
# Binomial likelihood, logit link
# Random effects model for multi-arm trials
model{                                  # *** PROGRAM STARTS
for(i in 1:ns){                         # LOOP THROUGH STUDIES
    w[i,1] <- 0    # adjustment for multi-arm trials is zero for control
arm
    delta[i,1] <- 0             # treatment effect is zero for control arm
    mu[i] ~ dnorm(0,.0001)              # vague priors for all trial baselines
    for (k in 1:na[i]) {                # LOOP THROUGH ARMS
        r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
        logit(p[i,k]) <- mu[i] + delta[i,k]  # model for linear predictor
        rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators
#Deviance contribution
        dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
            +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-
rhat[i,k])))          }
#  summed residual deviance contribution for this trial
    resdev[i] <- sum(dev[i,1:na[i]])
    for (k in 2:na[i]) {                # LOOP THROUGH ARMS
# trial-specific LOR distributions
        delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LOR distributions (with multi-arm trial correction)
        md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
        taud[i,k] <- tau *2*(k-1)/k
# adjustment for multi-arm RCTs
        w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
        sw[i,k] <- sum(w[i,1:k-1])/(k-1)
      }
  }
totresdev <- sum(resdev[])              # Total Residual Deviance
d[1]<-0       # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
sd ~ dunif(0,5)     # vague prior for between-trial SD
tau <- pow(sd,-2)   # between-trial precision = (1/between-trial variance)
# Provide estimates of treatment effects T[k] on the natural (probability)
scale
# Given a Mean Effect, meanA, for 'standard' treatment A,
# with precision (1/variance) precA
A ~ dnorm(meanA,precA)
for (k in 1:nt) { logit(T[k]) <- A + d[k]  }

rr[1]<- 1
for (k in 2:nt)  {
rr[k]<- T[k]/T[1]  }                                  # calculate relative
risk


# Ranking and prob{treatment k is best}
```

```
 for (k in 1:nt) {
            rk[k]<-rank(rr[],k)
best[k]<-equals(rank(rr[],k),1)}

# pairwise ORs and RRs
for (c in 1:(nt-1))
        {  for (k in (c+1):nt)
             {  lor[c,k] <- d[k] - d[c]
                log(or[c,k]) <- lor[c,k]
                lrr[c,k] <- log(rr[k]) - log(rr[c])
                log(rrisk[c,k]) <- lrr[c,k]

             }
        }
}

}                                                  # *** PROGRAM ENDS
```

Data

```
# ns= number of studies; nt=number of treatments
#key1=medical2=RF pt pt3=cryo4=laser

list(ns=8, nt=4, meanA=-3.612, precA=1.503668)
r[,1]   r[,2]   n[,1]   n[,2]   t[,1]   t[,2]   na[]
0.5     1.5     116     232     2       3       2       #andrade
0.5     2.5     377     375     2       3       2       #kuck
1       2       67      67      2       3       2#hunter
0.5     1.5     173     171     2       4       2       #dukkipatti
2.5     0.5     60      54      1       2       2       #jais
4       3       148     146     1       2       2       #nielsen
0.5     1.5     58      104     1       2       2       #wilber
0.5     1.5     83      164     1       3       2       #packer
END


Initial Values
#chain 1
list(d=c( NA, 0,0,0), sd=1, mu=c(0, 0, 0, 0, 0,0, 0,0 ))
#chain 2
list(d=c( NA, -1,-1,-1), sd=4, mu=c(-3, -3, -3, -3, -3, -3, -3,3))
#chain 3
list(d=c( NA, 2,0,3), sd=2, mu=c(-3, 5, -1, -3, 7, 2, 3, 2))
```

A.3.1.2 Fixed effects

```
# Binomial likelihood, logit link
# Fixed effects model
model{                          # *** PROGRAM STARTS
for(i in 1:ns){                 # LOOP THROUGH STUDIES
   mu[i] ~ dnorm(0,.0001)      # vague priors for all trial baselines
   for (k in 1:na[i])  {       # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k])    # binomial likelihood
# model for linear predictor
      logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]]
# expected value of the numerators
      rhat[i,k] <- p[i,k] * n[i,k]
#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
          +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
    }
# summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
    }
totresdev <- sum(resdev[])      # Total Residual Deviance
d[1]<-0    # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
```

```
# Provide estimates of treatment effects T[k] on the natural (probability) scale
# Given a Mean Effect, meanA, for 'standard' treatment A,
# with precision (1/variance) precA
A ~ dnorm(meanA,precA)
for (k in 1:nt) { logit(T[k]) <- A + d[k]  }

rr[1]<- 1
for (k in 2:nt)  {
rr[k]<- T[k]/T[1]  }                          # calculate relative risk


# Ranking and prob{treatment k is best}
 for (k in 1:nt) {
             rk[k]<-rank(rr[],k)
best[k]<-equals(rank(rr[],k),1)}

# pairwise ORs and RRs
for (c in 1:(nt-1))
        {  for (k in (c+1):nt)
              {  lor[c,k] <- d[k] - d[c]
                 log(or[c,k]) <- lor[c,k]
                 lrr[c,k] <- log(rr[k]) - log(rr[c])
                 log(rrisk[c,k]) <- lrr[c,k]

              }
         }
}

}                                          # *** PROGRAM ENDS
```

Data

```
# ns= number of studies; nt=number of treatments
#key1=medical2=RF pt pt3=cryo4=laser

list(ns=8, nt=4, meanA=-3.612, precA=1.503668)
r[,1]   r[,2]   n[,1]   n[,2]   t[,1]   t[,2]   na[]
0.5     1.5     116     232     2       3       2       #andrade
0.5     2.5     377     375     2       3       2       #kuck
1       2       67      67      2       3       2       #hunter
0.5     1.5     173     171     2       4       2       #dukkipatti
2.5     0.5     60      54      1       2       2       #jais
4       3       148     146     1       2       2       #nielsen
0.5     1.5     58      104     1       2       2       #wilber
0.5     1.5     83      164     1       3       2       #packer
END
```

Initial Values

```
#chain 1
list(d=c( NA, 0,0,0), mu=c(0, 0, 0, 0, 0, 0, 0,0 ))
#chain 2
list(d=c( NA, -1,-1,-1), mu=c(-3, -3, -3, -3, -3,-3, 3,-3))
#chain 3
list(d=c( NA, 2,0,3), mu=c(-3, 5, -1, -3, 7,2, 3,2))
```

A.3.2 Baseline model

A.3.2.1 Random effects

```
# Binomial likelihood, logit link
# Baseline random effects model
model{                    # *** PROGRAM STARTS
for (i in 1:ns){          # LOOP THROUGH STUDIES
   r[i] ~ dbin(p[i],n[i])          # Likelihood
   logit(p[i]) <- mu[i]                    # Log-odds of response
                mu[i] ~ dnorm(m,tau.m)     # Random effects model

                # expected value of the numerators
   rhat[i] <- p[i] * n[i]
                #Deviance contribution
   dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))
         +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))
 }
totresdev <- sum(dev[])                          # total residual deviance

mu.new ~ dnorm(m,tau.m)          # predictive dist. (log-odds)
m ~ dnorm(0,.0001)            # vague prior for mean
var.m <- 1/tau.m              # between-trial variance
tau.m <- pow(sd.m,-2)   # between-trial precision = (1/between-trial variance)
sd.m ~ dunif(0,5)             # vague prior for between-trial SD
#tau.m ~ dgamma(0.001,0.001)
#sd.m <- sqrt(var.m)
logit(R) <- m                # posterior probability of response
logit(R.new) <- mu.new        # predictive probability of response
}
```

Data

```
list(ns=1)  # ns=number of studies

r[]       n[]
2         59       #jais
END
```

Inits

```
list(mu=c(0), sd.m=1, m=0)
list(mu = c(-1), sd.m=2, m= -1)
list(mu = c(1), sd.m = 0.5, m = 1)
```

A.3.2.2 Fixed effects

```
# Binomial likelihood, logit link
# Baseline fixed effect model
model{                    # *** PROGRAM STARTS
for (i in 1:ns){          # LOOP THROUGH STUDIES
   r[i] ~ dbin(p[i],n[i])          # Likelihood
   logit(p[i]) <- m                                # Log-odds of response
```

```
                # expected value of the numerators
   rhat[i] <- p[i] * n[i]
                #Deviance contribution
   dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))
         +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))
 }
totresdev <- sum(dev[])                                    # total residual deviance
m ~ dnorm(0,.0001)          # vague prior for mean
logit(R) <- m               # posterior probability of response
}

 Data

list(ns=1)  # ns=number of studies

r[]       n[]
2         59       #jais
END


 Inits
list(m=0)
list(m= -1)
list(m = 1)
```

A.3.3 Inconsistency model

```
# Binomial likelihood, logit link
# Fixed effects INCONSISTENCY model
model{                    # *** PROGRAM STARTS
for(i in 1:ns){           # LOOP THROUGH STUDIES
   mu[i] ~ dnorm(0,.0001)     # vague priors for all trial baselines
   for (k in 1:na[i])  {      # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k])    # binomial likelihood
# model for linear predictor
      logit(p[i,k]) <- mu[i] + d[t[i,1],t[i,k]]
# expected value of the numerators
      rhat[i,k] <- p[i,k] * n[i,k]
#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
          +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
    }
# summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
    }
totresdev <- sum(resdev[])     # Total Residual Deviance

# vague priors for treatment effects
for(c in 1:nt){ d[c,c]<-0 }
for(c in 1:(nt-1)){
         for(k in (c+1):nt){
                  d[c,k]~dnorm(0,0.0001)
                  log(hr[c,k]) <- d[c,k]
                  d[k,c] <- -d[c,k]
                  }
         }
```

```
}                                   # *** PROGRAM ENDS
```

Data

```
# ns= number of studies; nt=number of treatments

list(ns=8 nt=4)

r[,1]   r[,2]   n[,1]   n[,2]   t[,1]   t[,2]   na[]
0.5     1.5     116     232     2       3       2       #andrade
0.5     2.5     377     375     2       3       2       #kuck
1       2       67      67      2       3       2       #hunter
0.5     1.5     173     171     2       4       2       #dukkipatti
2.5     0.5     60      54      1       2       2       #jais
4       3       148     146     1       2       2       #nielsen
0.5     1.5     58      104     1       2       2       #wilber
0.5     1.5     83      164     1       3       2       #packer
END
```

Initial Values

```
# chain 1
list(mu=c(0,0,0,  0,0,0, 0,0),
d = structure(.Data = c(NA,0,0,0,
NA,NA,0,0,
NA,NA,NA,0,
NA,NA,NA,NA),  .Dim = c(4,4)))

# chain 2
list(mu=c(0,1,-1,    2,-2, 2, -2,2),
d = structure(.Data = c(NA,0,1,0,
NA,NA,1,0,
NA,NA,NA,0,
NA,NA,NA,NA), .Dim = c(4,4)))

# chain 3
list(mu=c(3,2,-2,       0,-1, 1, -1,1),
d = structure(.Data = c(NA,0,1,2,
NA,NA,1,0,
NA,NA,NA,0,
NA,NA,NA,NA), .Dim = c(4,4)))
```

A.4 Serious adverse events (not including stroke or mortality)

A.4.1 Main code

A.4.1.1 Random effects

```
# Binomial likelihood, logit link
# Random effects model for multi-arm trials
model{                              # *** PROGRAM STARTS
for(i in 1:ns){                     # LOOP THROUGH STUDIES
   w[i,1] <- 0    # adjustment for multi-arm trials is zero for control arm
   delta[i,1] <- 0            # treatment effect is zero for control arm
   mu[i] ~ dnorm(0,.0001)           # vague priors for all trial baselines
   for (k in 1:na[i]) {             # LOOP THROUGH ARMS
```

```
        r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
        logit(p[i,k]) <- mu[i] + delta[i,k]  # model for linear predictor
        rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators
#Deviance contribution
        dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
           +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))         }
#  summed residual deviance contribution for this trial
    resdev[i] <- sum(dev[i,1:na[i]])
    for (k in 2:na[i]) {            # LOOP THROUGH ARMS
# trial-specific LOR distributions
        delta[i,k] ~ dnorm(md[i,k],taud[i,k])
# mean of LOR distributions (with multi-arm trial correction)
        md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
        taud[i,k] <- tau *2*(k-1)/k
# adjustment for multi-arm RCTs
        w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
# cumulative adjustment for multi-arm trials
        sw[i,k] <- sum(w[i,1:k-1])/(k-1)
      }
  }
totresdev <- sum(resdev[])          # Total Residual Deviance
d[1]<-0       # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
sd ~ dunif(0,5)     # vague prior for between-trial SD
tau <- pow(sd,-2)   # between-trial precision = (1/between-trial variance)
# Provide estimates of treatment effects T[k] on the natural (probability) scale
# Given a Mean Effect, meanA, for 'standard' treatment A,
# with precision (1/variance) precA
A ~ dnorm(meanA,precA)
for (k in 1:nt) { logit(T[k]) <- A + d[k]  }

rr[1]<- 1
for (k in 2:nt)  {
rr[k]<- T[k]/T[1]  }                          # calculate relative risk


# Ranking and prob{treatment k is best}
 for (k in 1:nt) {
              rk[k]<-rank(rr[],k)
best[k]<-equals(rank(rr[],k),1)}

# pairwise ORs and RRs
for (c in 1:(nt-1))
        {  for (k in (c+1):nt)
              {  lor[c,k] <- d[k] - d[c]
                 log(or[c,k]) <- lor[c,k]
                 lrr[c,k] <- log(rr[k]) - log(rr[c])
                 log(rrisk[c,k]) <- lrr[c,k]

              }
         }
}

}                              # *** PROGRAM ENDS
```

Data
ns= number of studies; nt=number of treatments
#1=medical,2=RF pt pt,3=cryo,4=laser,5=thoraco,6=hybrid, 7 =ME

list(ns=22, nt=7, meanA=-2.457, precA=9.644689)

r[,1]	r[,2]	n[,1]	n[,2]	t[,1]	t[,2]	na[]	
2.5	0.5	45	46	2	3	2	#davtyan
3	1	125	125	2	3	2	#yagishita
3	13	115	231	2	3	2	#andrade
2	4	77	78	2	3	2	#hunter
29	25	376	374	2	3	2	#kuck
3	11	159	156	2	3	2	#luik
1	1	25	25	2	3	2	#perez
2	3	70	140	2	3	2	#you
0.5	3.5	27	25	2	6	2	#jan
5	8	172	170	2	4	2	#{dukkipatti
1	1	25	25	2	4	2	#ucer
6	3	230	230	2	7	2	#gal
1	1	35	35	2	7	2	#kece
4	1	91	92	2	7	2	#mcready
0.5	1.5	26	26	2	7	2	#podd
3	6	61	66	1	2	2	#morrillo
12	15	148	146	1	2	2	#neilsen
10	3	99	99	1	2	2	#pappone
1	2	35	32	1	2	2	#wazni
2	4	57	103	1	2	2	#wilber
2	2	17	15	3	7	2	#koch
6.5	0.5	21	50	5	7	2	#sugihara

END

Initial Values
#chain 1
list(d=c(NA, 0,0,0,0,0,0), sd=1, mu=c(0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0))
#chain 2
list(d=c(NA, -1,-1,-1,-1,-1,-1), sd=2, mu=c(-3, -3, -3, -3, -3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3, -3, -3))
#chain 3
list(d=c(NA, 2,0,3,2,1,3), sd=1, mu=c(-3, 5, -1, -3, 7,2, 1, 3, 6, 3, 2, 1, 4, 2, -1, -2, 3, 2, 1, -1, 1, -1))

A.4.1.2 Fixed effects

```
# Binomial likelihood, logit link
# Fixed effects model
model{                              # *** PROGRAM STARTS
for(i in 1:ns){                     # LOOP THROUGH STUDIES
    mu[i] ~ dnorm(0,.0001)          # vague priors for all trial baselines
    for (k in 1:na[i])  {           # LOOP THROUGH ARMS
        r[i,k] ~ dbin(p[i,k],n[i,k])    # binomial likelihood
# model for linear predictor
        logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]]
# expected value of the numerators
        rhat[i,k] <- p[i,k] * n[i,k]
#Deviance contribution
        dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
             +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-
rhat[i,k])))
      }
```

```
# summed residual deviance contribution for this trial
    resdev[i] <- sum(dev[i,1:na[i]])
     }
totresdev <- sum(resdev[])      # Total Residual Deviance
d[1]<-0    # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
# Provide estimates of treatment effects T[k] on the natural (probability)
scale
# Given a Mean Effect, meanA, for 'standard' treatment A,
# with precision (1/variance) precA
A ~ dnorm(meanA,precA)
for (k in 1:nt) { logit(T[k]) <- A + d[k]  }

rr[1]<- 1
for (k in 2:nt)  {
rr[k]<- T[k]/T[1]  }                                  # calculate relative
risk


# Ranking and prob{treatment k is best}
 for (k in 1:nt) {
          rk[k]<-rank(rr[],k)
best[k]<-equals(rank(rr[],k),1)}

# pairwise ORs and RRs
for (c in 1:(nt-1))
      {  for (k in (c+1):nt)
           {  lor[c,k] <- d[k] - d[c]
              log(or[c,k]) <- lor[c,k]
              lrr[c,k] <- log(rr[k]) - log(rr[c])
              log(rrisk[c,k]) <- lrr[c,k]

           }
      }
}

}                                       # *** PROGRAM ENDS
```

Data

```
# ns= number of studies; nt=number of treatments
#1=medical,2=RF pt pt,3=cryo,4=laser,5=thoraco,6=hybrid, 7=ME

list(ns=22, nt=7, meanA=-2.457, precA=9.644689)
r[,1]   r[,2]   n[,1]   n[,2]   t[,1]   t[,2]   na[]
2.5     0.5     45      46      2       3       2       #davtyan
3       1       125     125     2       3       2       #yagishita
3       13      115     231     2       3       2       #andrade
2       4       77      78      2       3       2       #hunter
29      25      376     374     2       3       2       #kuck
3       11      159     156     2       3       2       #luik
1       1       25      25      2       3       2       #perez
2       3       70      140     2       3       2       #you
0.5     3.5     27      25      2       6       2       #jan
5       8       172     170     2       4       2       #dukkipatti
1       1       25      25      2       4       2       #ucer
6       3       230     230     2       7       2       #gal
1       1       35      35      2       7       2       #kece
4       1       91      92      2       7       2       #mcready
0.5     1.5     26      26      2       7       2       #podd
3       6       61      66      1       2       2       #morrillo
12      15      148     146     1       2       2       #neilsen
10      3       99      99      1       2       2       #pappone
1       2       35      32      1       2       2       #wazni
2       4       57      103     1       2       2       #wilber
2       2       17      15      3       7       2       #koch
6.5     0.5     21      50      5       7       2       #sugihara
END
```

Initial Values

```
#chain 1
list(d=c( NA, 0,0,0,0,0,0), mu=c(0, 0, 0, 0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0 ))
#chain 2
list(d=c( NA, -1,-1,-1,-1,-1,-1), mu=c(-3, -3, -3, -3, -3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3,-3, -3))
#chain 3
list(d=c( NA, 2,0,3,2,1,3), mu=c(-3, 5, -1, -3, 7,2, 1, 3, 6, 3, 2, 1, 4, 2, -1, -2, 3, 2, 1, -1, 1, 1))
```

A.4.2 Baseline model

A.4.2.1 Random effects

```
# Binomial likelihood, logit link
# Baseline random effects model
model{                    # *** PROGRAM STARTS
for (i in 1:ns){          # LOOP THROUGH STUDIES
   r[i] ~ dbin(p[i],n[i])          # Likelihood
   logit(p[i]) <- mu[i]                    # Log-odds of response
                mu[i] ~ dnorm(m,tau.m)     # Random effects model

                # expected value of the numerators
   rhat[i] <- p[i] * n[i]
                #Deviance contribution
   dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))
         +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))
 }
totresdev <- sum(dev[])                              # total residual deviance

mu.new ~ dnorm(m,tau.m)          # predictive dist. (log-odds)
m ~ dnorm(0,.0001)            # vague prior for mean
var.m <- 1/tau.m              # between-trial variance
tau.m <- pow(sd.m,-2)   # between-trial precision = (1/between-trial variance)
sd.m ~ dunif(0,5)             # vague prior for between-trial SD
#tau.m ~ dgamma(0.001,0.001)
#sd.m <- sqrt(var.m)
logit(R) <- m                 # posterior probability of response
logit(R.new) <- mu.new        # predictive probability of response
}
```

Data

```
list(ns=2)  # ns=number of studies

r[]       n[]
1         35       #wazni
10        99 #pappone
END
```

Inits

```
list(mu=c(0,0), sd.m=1, m=0)
list(mu = c(-1,-1), sd.m=2, m= -1)
list(mu = c(1,1), sd.m = 0.5, m = 1)
```

A.4.2.2 Fixed effects

```
# Binomial likelihood, logit link
# Baseline fixed effect model
model{                    # *** PROGRAM STARTS
for (i in 1:ns){          # LOOP THROUGH STUDIES
   r[i] ~ dbin(p[i],n[i])          # Likelihood
   logit(p[i]) <- m                                    # Log-odds of response

                  # expected value of the numerators
   rhat[i] <- p[i] * n[i]
                  #Deviance contribution
   dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))
         +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))
 }
totresdev <- sum(dev[])                                # total residual deviance
m ~ dnorm(0,.0001)          # vague prior for mean
logit(R) <- m               # posterior probability of response
}
```

Data

```
list(ns=2)  # ns=number of studies

r[]      n[]
1        35      #wazni
10       99 #pappone
END
```

Inits

```
list(m=0)
list(m= -1)
list(m = 1)
```

A.4.3 Inconsistency model

```
# Binomial likelihood, logit link
# Fixed effects INCONSISTENCY model
model{                    # *** PROGRAM STARTS
for(i in 1:ns){           # LOOP THROUGH STUDIES
   mu[i] ~ dnorm(0,.0001)     # vague priors for all trial baselines
   for (k in 1:na[i])  {      # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k])    # binomial likelihood
# model for linear predictor
      logit(p[i,k]) <- mu[i] + d[t[i,1],t[i,k]]
# expected value of the numerators
      rhat[i,k] <- p[i,k] * n[i,k]
#Deviance contribution
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
```

```
          +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
     }
# summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
    }
totresdev <- sum(resdev[])      # Total Residual Deviance
# vague priors for treatment effects
for (c in 1:(nt-1)){
                    d[c,c]<-0
      for (k in (c+1):nt){
          d[c,k] ~ dnorm(0,.0001)                    # priors for all mean trt effects
          or[c,k] <- exp(d[c,k])                     # all pairwise ORs
      }
                    }

}                                        # *** PROGRAM ENDS
```

Data

```
 # ns= number of studies; nt=number of treatments
#1=medical,2=RF pt pt,3=cryo,4=laser,5=thoraco,6=hybrid, 7=RF me

list(ns=22, nt=7)
r[,1]   r[,2]   n[,1]   n[,2]   t[,1]   t[,2]   na[]
2.5     0.5     45      46      2       3       2       #davtyan
3       1       125     125     2       3       2       #yagishita
3       13      115     231     2       3       2       #andrade
2       4       77      78      2       3       2       #hunter
29      25      376     374     2       3       2       #kuck
3       11      159     156     2       3       2       #luik
1       1       25      25      2       3       2       #perez
2       3       70      140     2       3       2       #you
0.5     3.5     27      25      2       6       2       #jan
5       8       172     170     2       4       2       #dukkipatti
1       1       25      25      2       4       2       #ucer
6       3       230     230     2       7       2       #gal
1       1       35      35      2       7       2       #kece
4       1       91      92      2       7       2       #mcready
0.5     1.5     26      26      2       7       2       #podd
3       6       61      66      1       2       2       #morrillo
12      15      148     146     1       2       2       #neilsen
10      3       99      99      1       2       2       #pappone
1       2       35      32      1       2       2       #wazni
2       4       57      103     1       2       2       #wilber
2       2       17      15      3       7       2       #koch
6.5     0.5     21      50      5       7       2       #sugihara
END
```

chain 1

```
list(mu=c(0,0,0,  0,0,0,0,0,0,  0,0,0,0,0,0,0,  0,0,0,0,0,0,0),
d = structure(.Data = c(NA,0,0,0,0,0,0   NA,NA,0,0,0,0,0,0   NA,NA,NA,0,0,0,0   NA,NA,NA,NA,0,0,0
NA,NA,NA,NA,NA,0,0  NA,NA,NA,NA,NA,NA,0), .Dim = c(6,7)))
```

chain 2

```
list(mu=c(0,0,1,  0,0,1,1,0,0,  1,0,0,0,0,0,1,  0,1,1,0,0,0,0),
d = structure(.Data = c(NA,0,1,0,0,1,0   NA,NA,1,0,0,1,0   NA,NA,NA,0,0,1,0   NA,NA,NA,NA,0,1,0
NA,NA,NA,NA,NA,1,0  NA,NA,NA,NA,NA,NA,0), .Dim = c(6,7)))
```

chain 3

```
list(mu=c(0,0,1,  0,1,0,0,1,0,  0,1,0,1,0,0,  0,0,0,1,0,0,0),
d = structure(.Data = c(NA,0,0,0,0,0,-1   NA,NA,0,0,0,0,-1   NA,NA,NA,0,0,0,-1   NA,NA,NA,NA,0,0,-1
NA,NA,NA,NA,NA,0,-1  NA,NA,NA,NA,NA,NA,-1), .Dim = c(6,7)))
```

Appendix B: Follow up times

Study	Follow up time
RF point by point versus cryoballoon	
Andrade, 2019[1]	12 months
Bin Waleed, 2019[3]	6 months
Davtyan, 2018[10]	12 months
Giannopoulos, 2019[19]	6 months
Gunawardene, 2018[20]	309.7 days
Hunter, 2015[2, 21]	12 months (2 years for mortality)
Kuck, 2016[28] and Kuck, 2016[29] FIRE AND ICE TRIAL	1.5 years
Luik, 2017[31] and Luik, 2015[32] FREEZE AF TRIAL	12months
Perez-Castellano, 2014[42] COR TRIAL	12 months
Watanabe, 2018[56]	12 months
You, 2019[61]	1 year
Yagishita, 2020[60]	4 weeks
RF point by point versus Cryoballoon versus laser	
Schmidt, 2013[48]	1-2 days
RF point by point versus hybrid	
Jan, 2018[25]	30.5 months
RF point by point versus thoraco	
Wang, 2014[55]	12 months
RF point by point versus laser	
Dukkipati, 2015[16]	12 months
Ucer, 2018[52] RATISBONA trial	UNCLEAR
RF point by point versus RF multielectrode	
Boersma 2016[4] MYSTIC-PAF	12 months
Bulava, 2010[7]	202 days
Gal, 2014[17]	12 month sfor recurrence and 43.2 months for other outcomes
Kece, 2019[26]	12 months
McCready, 2014[35].	12 months
Podd, 2015[43]	12 months
RF point by point versus medical care	
Jais, 2008[24] A4 STUDY	12 months
Morillo, 2014[36] RAAFT-2 trial	21 months for recurrence and possibly 24 for SAEs
Nielsen, 2017[37]; Walfridsson, 2015[54] and Cosedis Nielsen, 2012[9] MANTRA-PAF trials	24 months
Pappone, 2011[41] and Pappone, 2006[40] APAF	12 months

Study	Follow up time
Wazni, 2005[57]	12 months
Wilber, 2010[58] and Reynolds, 2010[46]	9 months
Xu, 2012[59]	6 months
RF multielectrode vs cryoballoon	
Koch, 2012[27], Schirdewan, 2017[47] MACPAF trial	12 months for recurrence (Schirdewan).
RF multielectrode vs thoracoscopy	
Sugihara, 2018[50]	12 months
Cryoballoon versus medical	
Packer, 2013[39] STOP AF TRIAL	12 months

Appendix C: Serious adverse events

C.1 Serious adverse events determination

All adverse events listed in the included studies were listed and classified as serious or non-serious by two cardiologists. The list is below. If it was unclear if an adverse effect was serious (because of an ambiguous description) then the adverse event was deemed serious. Only serious adverse events were counted in the analysis.

Complication (all information provided in the papers)	Mark with a YES if deemed 'serious' based on the information provided (err on side of assuming seriousness if unsure)
Aneurysm: Inguinal aneurysm	Yes
Aneurysm: pseudo aneurysm	Yes
Aneurysm: Pseudo aneurysm requiring thrombin injection but no long term sequelae	YES
Arrhythmias: Cardioversion for atrial arrhythmias	NO
Arrhythmias: Life threatening arrhythmias	YES
Arrhythmias: New atrial flutter	NO
Arrhythmias: Post ablation atrial tachycardia requiring ablation	Yes
Arrhythmias: Pro-arrhythmia	YES
Arrhythmias: ventricular tachycardia	YES
Atrial arrhythmias	NO
Atrial flutter or atrial tachycardia	YES
Atrial flutter with 1:1 AV conduction	YES
Bleeding	YES
Bleeding requiring surgery	YES
Bleeding: haematuria	NO
Bleeding: Haemoptysis secondary to haematoma on R inferior PV – resolved spontaneously	YES
Bleeding: haemorrhage requiring transfusion	YES
Bleeding: Major bleeding requiring transfusion	YES
Bleeding: retroperitoneal bleeding, coiling of small artery	YES
Bleeding: Sternotomy for bleeding	YES
Bradycardia	NO
Bradycardia leading to pacemaker insertion	YES
Cardiac tamponade	YES
Cardiac tamponade drained percutaneously	YES
Cardiac tamponade or pericardial effusion	YES

Complication (all information provided in the papers)	Mark with a YES if deemed 'serious' based on the information provided (err on side of assuming seriousness if unsure)
Cardiac tamponade requiring pericardiocentesis	YES
Cardiac: Dressler's syndrome requiring drainage	YES
Cardiac: Heart failure	YES
Cardiac: Major pericardial effusion events requiring drainage	YES
Cardiac: Minor pericardial effusion events – no drainage	NO
Cardiac: Myocardial Infarction	YES
Cardiac: Non-arrhythmia cardiac complication	YES
Cardiac: Pericardial effusion	NO
Cardiac: Pericardial tamponade	YES
Cardiac: Pericardial tamponade requiring drain and 24 hrs extra stay	YES
Cardiac: Pericardial tamponade requiring later (4 week) PVI with RF	YES
Cardiac: perimyocarditis	YES
Cardiac: Small pericardial effusion not requiring pericardiocentesis	NO
Cardiac: suspected perforation at transseptal puncture with no pericardial effusion	NO
Cardiac: Transient ST segment	NO
Drug: Disabling drug intolerance requiring discontinuation	NO
Drug: discomfort due to medication	NO
Fistula: Arteriovenous fistula – managed conservatively without need for further intervention	YES
Fistula: New or worse AV fistula	YES
Fistula: Right femoral AV fistula requiring surgical repair	YES
GI complaints	NO
GI: Gallbladder surgery	NO – unrelated
GI: Oesophageal ulceration	YES
Groin site complications	YES
Groin: Femoral vascular access	NO
Groin: minor groin complications not requiring blood transfusion nor invasive treatment	NO
Hematoma	NO
Hematoma related to anticoagulation	NO
Hematoma: Groin hematoma	NO
Hematoma: Retroperitoneal haematoma	YES
Hematoma: Slight groin haematoma treated conservatively	NO

Complication (all information provided in the papers)	Mark with a YES if deemed 'serious' based on the information provided (err on side of assuming seriousness if unsure)
Hospitalisation for AF	YES
Infection leading to antibiotics and hospitalisation	YES
MSK: knee OA requiring arthroscopy	NO
MSK: Rotator cuff rupture	NO
Neuro: Transient neurological complications (not TIA)	YES
Neuro:Transient global amnesia (not TIA)	YES
Other: Anxiety	NO
Other: Cancer	YES
Other: Chest discomfort	NO
Other: Contrast media reaction	Yes
Other: Contusion	NO
Other: Local oedema	NO
Perforation: Atrial perforation	YES
Perforation: Atrial septal puncture site not occluded requiring atrial septum closure device	YES
PNP: Asymptomatic phrenic nerve injuries	NO
PNP: Persistent phrenic nerve palsy	YES
PNP: Phrenic nerve injury	Yes
PNP: phrenic nerve palsy resolving during 1 year follow up	YES
PNP: Symptomatic phrenic nerve injuries	YES
PNP: transient phrenic nerve palsy resolving before discharge	NO
PNP: Transient phrenic nerve palsy resolving before end of procedure	NO
PNP: Unresolved phrenic nerve injuries	YES
pulmonary complications	YES
Pulmonary oedema	YES
Pulmonary: Acute lung injury	YES
Pulmonary: dyspnoea	NO
Pulmonary: Pneumonia	YES
Pulmonary: Post op lower respiratory tract infection	YES
Pulmonary: Symptomatic pleural effusion	YES
Retinal infarction	YES
Sexual impairment	NO
Stenosis of left superior pulmonary vein requiring dilatation and stent implantation	YES
Stenosis: asymptomatic moderate 50-70% pulmonary vein stenosis	NO
Stenosis: asymptomatic pulmonary vein stenosis	NO
Stenosis: Clinical PV stenosis	YES

Complication (all information provided in the papers)	Mark with a YES if deemed 'serious' based on the information provided (err on side of assuming seriousness if unsure)
Stenosis: Mild <50% pulmonary vein stenosis	NO
Stenosis: pulmonary vein stenosis	NO
Stenosis: PV stenosis	NO
Stenosis: PV stenosis >50%	NO
Stenosis: Severe >70% pulmonary vein stenosis (asymptomatic)	YES
Stenosis: Severe pulmonary vein stenosis	YES
syncope	YES
Thyroid dysfunction	YES
Thyroid: hyperthyroidism	YES
Vascular complication	Yes
Vascular injuries	Yes
Vascular: Major vascular events (no definition)	YES
Vascular: Minor vascular events (no definition)	No

C.2 Serious adverse events by study

Study	Serious adverse events	
	RF pt	Cryo
Andrade, 2020[1]	3/115 3 with one or more of the following: pericardial effusion, pericarditis, hematoma requiring intervention, pseudoaneurysm requiring intervention, esophageal perforation	13/231 Unclear how many people had the following but the following 13 serious AEs were recorded: 1 pericardial effusion, 3 pericarditis, 1 MI, 1 atypical chest pain, 1 HF exacerbation, 1 AV fistula, 3 persistent phrenic nerve palsies, 1 esophageal injury, 1 acute pulmonary infection.
Davtyan, 2018[10]	2/44 2 arteriovenous fistulae – both managed conservatively without need for further intervention	0/45 Deemed non-serious: 2 transient phrenic n palsy resolving before end of procedure
Gunawardene, 2018[20]	0/30 Deemed non-serious: 4 minor groin complications not requiring blood transfusion nor invasive treatment	0/30 Deemed non-serious: 5 minor groin complications not requiring blood transfusion nor invasive treatment 1 transient phrenic nerve palsy resolving before discharge
Hunter, 2015[2, 21]	2/77 1 tamponade drained percutaneously, 1 dresslers syndrome requiring drainage Deemed non-serious: 1 hematoma, 1 asymp PV stenosis	4/78 4 phrenic n palsies resolving in follow up
Kuck, 2016[28] and Kuck, 2016[29] FIRE AND ICE TRIAL	29/376 16 groin site complications, 5 cardiac tamponade or pericardial effusion, 4 pulmonary complications, 3 transient neurological complication (NOT TIA), 1 contrast media reaction Deemed non-serious: 13 atrial arrhythmias, 2 SOB, 2 GI complaint, 1 contusion, 1 haematuria, 1 local oedema	25/374 7 groin site complications, 10 unresolved phrenic injuries, 1 cardiac tamponade/pericardial effusion, 2 pulmonary complication, 1 transient neurological problem (NOT TIA), 3 non arrhythmia cardiac complications, 1 oesophageal ulceration. Deemed non-serious: 8 atrial arrhythmias, 1 SOB, 1 GI complication, 1 anxiety,

Study	Serious adverse events	
Luik, 2017[31] and Luik, 2015[32] FREEZE AF TRIAL	3/159 3 major vascular events Deemed non-serious: 2 minor vascular events – no definition – and 3 minor pericardial effusion - no drainage	11/156 6 major vascular events, 2 major pericardial effusion events (required drainage), 3 symptomatic phrenic nerve palsies. Deemed non-serious: 2 minor vascular events – no def. – and 6 asymptomatic phrenic nerve injuries
Perez-Castellano, 2014[42] COR TRIAL	1/25 1 right femoral arteriovenous fistula requiring surgical repair	1/25 1 haemoptysis secondary to haematoma surrounding R inferior PV and resolved spontaneously Deemed non-serious: 4 temporary phrenic nerve palsies recovering before patient left operating room
You, 2019[61]	2/70 2 vascular injuries	3/140 2 phrenic nerve and 1 vascular injury
Yagishita, 2020[60]	3/125 Pleural effusion requiring drainage Deemed non-serious: Minor bleeding due to groin hematoma: 4/125	1/125 Pleural effusion requiring drainage Deemed non-serious: Minor bleeding due to groin hematoma: 4/125
	RF pt	hybrid
Jan, 2018[25]	0/26	2/24 1 bleeding requiring surgery, 1 acute lung injury, 1 wound infection leading to antibiotics and hospitalisation
	RF pt pt	laser
Dukkipati, 2015[16]	5/172 3 cardiac tamponade, 1 phrenic nerve palsy, 1 major bleeding requiring transfusion Deemed non-serious: 5 PV stenosis >50%, 16 cardioversion for atrial arrhythmias,	8/170 2 cardiac tamponade, 6 phrenic nerve palsy, Deemed non-serious: 14 cardioversion,

Study	Serious adverse events	
Ucer, 2018[52] RATISBONA trial	1/25 1 had pericardial tamponade requiring later (4 week) PVI with RF. Deemed non-serious: 1 slight groin hematoma treated conservatively	1/25 1 Atrial septal puncture site not occluded requiring atrial septum closure device Deemed non-serious: 3 slight groin hematoma treated conservatively
	RF pt pt	RF ME
Gal, 2014[17]	6/230 4 pneumonia, 2 atrial perforation Deemed non-serious: 5 femoral vascular access,	3/230 1 pneumonia, 1 retinal infarction, 1 transient global amnesia (not TIA)
Kece, 2019[26]	1/35 1 tamponade Deemed non-serious: 1 groin hematoma	1/35 1 severe >70% pulm vein stenosis (asymp), Deemed non-serious: 1 UTI
McCready, 2014[35]	4/91 3 cardiac tamponade 1 clinical PV stenosis	1/92 1 pseudo-aneurysm requiring thrombin injection but no long term sequelae
Podd, 2015[43]	0/25	1/25 Pericardial tamponade requiring drain and 24 hr extra stay
	RF pt pt	medical
Morillo, 2014[36] RAAFT-2 trial	6/66 4 tamponade,1 severe pulm vein stenosis, 1 bradycardia leading to pacemaker insertion	3/61 1 atrial flutter with 1:1 AV conduction, 2 syncope
Nielsen, 2017[37]; Walfridsson, 2015[54] and Cosedis Nielsen, 2012[9] MANTRA-PAF trials	15/146 6 cancer, 3 atrial flutter or atrial tachycardia, 1 perimyocarditis, 3 tamponade, 1 ventricular tachycardia, 1 retroperitoneal bleeding, coiling of small artery Deemed non-serious: 1 suspected perforation at transseptal puncture with	12/148 4 cancer, 2 atrial flutter with an AV conduction ratio of 1:1, 3 atrial flutter or atrial tachycardia, , 2 hospitalisation for HF, 1 bradycardia with need for cardiac pacemaker Deemed non-serious: , 1 pericardial effusion 2 discomfort due to medication, 1 rupture of the rotator cuff, 1 gallbladder surgery

Study	Serious adverse events	
	no pericardial effusion, 1 pulmonary vein stenosis, 1 hematoma related to anticoagulation 1 chest discomfort, 1 knee OA requiring arthroscopy	
Pappone, 2011[41] and Pappone, 2006[40] APAF	3/99 3 post-ablation atrial tachycardia requiring ablation Deemed non-serious: 1 small pericardial effusion not requiring pericardiocentesis	10/99 3 pro-arrhythmia, thyroid dysfunction in 7 Deemed non-serious: , sexual impairment in 11; 2 not reported
Wazni, 2005[57]	3/32 2 bleeding Deemed non-serious: 1 asymptomatic moderate 50-70% pulmonary vein stenosis, 1 mild <50% pulmonary vein stenosis	1/35 1 bleeding Deemed non-serious: 3 bradycardia
Wilber, 2010[58] and Reynolds, 2010[46]	4/103 1 pulmonary oedema, 1 vascular complication, 1 HF, 1 pneumonia Deemed non-serious: 1 pericardial effusion	2/57 2 life threatening arrhythmias Deemed non-serious: 3 disabling drug intolerance requiring discontinuation
	RF ME	cryo
Koch, 2012[27], Schirdewan, 2017[47] MACPAF trial	2/15 1 pericardial tamponade, 1 inguinal aneurysm Deemed non-serious: 1 pericardial effusion	2/17 1 retroperotoneal haematoma, 1 inguinal aneurysm Deemed non-serious: 1 transient ST segment
	RF ME	thoraco
Sugihara, 2018[50]	0/49	6/20 2 sternontomy for bleeding, 3 symptomatic pleural effusion, 1 post op lower RTI