Evidence Review Group's Report

Title: Dronedarone for atrial fibrillation and atrial flutter

Produced by *Centre for Reviews and Dissemination (CRD), Centre for Health Economics (CHE)*

Authors

Claire McKenna, Research Fellow, CHE
Emma Maund, Research Fellow, CRD
Muhammad Sarowar, Research Fellow, CHE
David Fox, Information Specialist, CRD
Matt Stevenson, ScHARR, University of Sheffield
Chris Pepper, Consultant Cardiologist, Yorkshire Heart Centre, Leeds
Stephen Palmer, Senior Research Fellow, CHE
Nerys Woolacott, Senior Research Fellow, CRD

Correspondence to *Nerys Woolacott, Senior Research Fellow, CRD*

Date completed *28th October 2009*

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 08/214

Declared competing interests of the authors
Mark Sculpher (CHE) has a minority shareholding in a consulting company which has undertaken work for Sanofi including a project on quality of life associated with atrial fibrillation. He did not participate personally in this consultancy nor has be been involved in the Evidence Review Group for Dronedarone

Acknowledgements
Steve Price, Research Fellow, CRD

Rider on responsibility for report
The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:
McKenna C, Maund E, Sarowar M, Fox D, Stevenson M, Pepper C, Palmer S, Woolacott N. Dronedarone for atrial fibrillation and atrial flutter. A Single Technology Appraisal. Centre for Reviews and Dissemination and Centre for Health Economics, 2009.

Table of contents

List of tables and figures

Glossary and list of abbreviations

List of abbreviations and definition of terms

AAD	Anti-arrhythmic drugs
ACS	Acute coronary syndrome
AE	Adverse event
AF	Atrial fibrillation
AFL	Atrial flutter
AV	Atrioventricular
Baseline therapy	Standard therapy for AF according to guidelines: e.g. anticoagulants and beta-blockers.
BID	Twice daily
CAD	Coronary artery disease
$CHADS_2$ score	Clinical prediction rule for estimating risk of stroke in patients with AF
CHF	Congestive heart failure
CI	Confidence interval
CV	Cardiovascular
DES	Discrete event simulation

ECG	Electrocardiogram
ERG	Evidence review group
HRQoL	Health-related quality of life
ICER	Incremental cost-effectiveness ratio
LVD	Left ventricular dysfunction
LVEF	Left ventricular ejection fraction
LVF	Left ventricular function
MI	Myocardial infarction
MTC	Mixed treatment comparison
NYHA	New York Heart Association
RCT	Randomised controlled trial
SAE	Serious adverse event
SHD	Structural heart disease
TEAE(s)	Treatment-emergent adverse event(s)
TIA	Transient ischaemic attack

Note on use of page numbers

All page numbers given in parentheses in this ERG report refer to the manufacturer's original submission, unless otherwise stated. References to the ERG report are given in terms of section number (e.g. "see section 4.2.2 for details").

1 Summary

1.1 Scope of the submission

This report presents the ERG's assessment of the manufacturer's (Sanofi Aventis Limited) submission to NICE on the use of Multaq ®, dronedarone for the treatment of atrial fibrillation (AF) and atrial flutter (AFL). The report includes an assessment of both the clinical and cost-effectiveness evidence submitted by the company. The ERG report identifies the submission's strengths and weaknesses, supplemented, where appropriate, with additional analyses. A clinical expert was asked to advise the ERG to help inform the review.

The manufacturer's submission largely adhered to the scope for the appraisal issued by NICE in that it evaluated the use of dronedarone both as a first line adjunctive treatment to standard baseline therapy (with or without beta-blockers and anticoagulation therapy) and as a second line treatment compared to other anti-arrhythmic drugs (AADs) in accordance with their respective indications and

recommendations from current NICE clinical guidelines. However, the manufacturer's submission differed slightly to the NICE scope for the appraisal in terms of the population and subgroups. Firstly, the population addressed in the submission was restricted to the anticipated licensed indication *i.e. "stable adult patients with a recent history of, or current non-permanent atrial fibrillation (AF). This population is likely to exclude patients with NYHA CHF Class IV and also NYHA Class III CHF with a recent haemodynamic instability"* (p7 of manufacturer's submission*)*. Secondly, the evaluation of dronedarone as first line adjunctive treatment to standard baseline therapy was restricted further to patients with multiple CV risk factors, corresponding to a $CHADS_2$ score ≥ 4[a] In addition, the wording of the decision problem was altered so that the comparison of dronedarone as a second line therapy versus other AADs was changed to "as an alternative 1st line to current anti-arrhythmic agents when it is considered appropriate to introduce an AAD" (p7 of manufacturer's submission). Subgroups were considered in the submission based on $CHADS_2$ scores, as opposed to measures of cardiovascular risk. Finally, patients with AFL were not considered separately in the submission or the economic model. In part this appears to reflect the proposed EMEA wording for the license for dronedarone which does not specifically mention AFL. However, the manufacturer assumed that for those groups of patients whose AFL is clinically indistinguishable from AF, the assumptions and results of the economic model would also apply.

1.2 Summary of submitted clinical effectiveness evidence

The main clinical effectiveness data for dronedarone and other AADs were derived from Randomised Controlled Trials (RCTs) (both against control and head-to-head RCTs comparing alternative AADs), meta-analysis (presenting direct and indirect comparisons) and a synthesis of the direct and indirect RCT evidence using a mixed treatment comparison (MTC).

The manufacturer's submission identified seven phase II and III trials of dronedarone and subsequently focused on the four phase III RCTs that enrolled patients with persistent or paroxsysmal AF/AFL. Three of these RCTs compared dronedarone to placebo as an adjunctive treatment to standard baseline therapy (e.g. beta blockers

[a] $CHADS_2$ is a stroke risk stratification scheme which is based on specific risk factors including congestive heart failure, hypertension, age >75, diabetes mellitus, and prior stroke or transient ischemic attack.

and anticoagulation therapy). The baseline therapy in these RCTs (EURIDIS, ADONIS and ATHENA) was considered by the manufacturer to be representative of standard baseline therapy in the UK. The remaining RCT (DIONYSOS) directly compared dronedarone with the AAD amiodarone.

The EURIDIS/ADONIS trials demonstrated that dronedarone was statistically significantly more effective than placebo for maintenance of sinus rhythm and in reducing the ventricular rate during recurrence of AF/AF. [1]

The ATHENA study was by far the largest RCTs (n=4628) investigating the effectiveness of dronedarone (400 mg BID) versus placebo in addition to baseline therapy on the combined endpoint of CV hospitalisation and all cause mortality. Unlike the other RCTs the population recruited to the ATHENA study represents a moderate to high-risk elderly AF population, 75% of whom were in sinus rhythm. Over a mean follow-up period of 21±5 months, dronedarone resulted in a significant reduction in the primary composite endpoint of time to first CV hospitalisation or death from any cause (Hazard ratio (HR) 0.76, 95% confidence interval (CI): 0.69, 0.84, p<0.001). The primary endpoint appeared to be mainly driven by a reduction in time to first cardiovascular hospitalisation due to a significant reduction in hospitalisation for AF (HR 0.63, 95% CI: 0.55, 0.72, p<0.001).[redacted]

A statistically significant difference in all-cause mortality between patients receiving dronedarone and those receiving placebo was not reported (HR: 0.84, 95% CI: 0.66, 1.08; p=0.18). A post-hoc analysis showed that there was a statistically significant reduction in the risk of stroke in patients receiving dronedarone compared to those receiving placebo (HR: 0.66, 95% CI: 0.46, 0.96, p=0.027). [2]

The DIONYSOS trial (n=504) was the only RCT identified that directly compared dronedarone (400mg BID) with another AAD (amiodarone 600 mg daily for 28 days, then 200mg daily). This was a short term study with patients followed up for at least 6 months. The primary endpoint was recurrence of AF or premature study drug discontinuation. The incidence of the primary endpoint was 73.9% for dronedarone and 55.3% for amiodarone (p-value<0.0001).

The manufacturer's submission also compared the efficacy and safety of dronedarone with four other AADs commonly used in the UK (flecainide, propafenone, sotalol and amiodarone). The RCTs for flecainide and propafenone were subsequently combined to represent class 1c agents. A range of alternative synthesis approaches were employed: (i) a direct meta-analysis based on the RCTs reporting a comparison of each AAD against placebo/control; (ii) an indirect meta-analysis comparing the different drugs using the placebo/control groups as a common comparator and (iii) a MTC combining the direct and indirect RCT evidence.

Outcomes assessed in the different syntheses were AF recurrence, all-cause mortality, treatment discontinuations, stroke and serious adverse events (SAEs). A total of 39 studies were considered eligible for inclusion, although the studies included in the direct meta-analysis and the MTC were subject to different inclusion criteria. The most notable difference was that the MTC analysis imposed additional restrictions on inclusion in order to achieve convergence. This had an important effect on the evidence base considered in the different approaches. For example, the direct meta-analysis for all-cause mortality incorporated data from 33 studies, which only 7 trials subsequently met the inclusion criteria for the same outcome in the MTC.

Overall, the results from the different synthesis approaches showed that the odds of AF recurrence appears statistically significantly lower with all AADs compared to non-active control. However, the results of the direct, indirect and MTC analyses consistently showed that the odds of AF recurrence are statistically significantly higher for dronedarone compared to amiodarone. No direct data were available to compare dronedarone versus class 1c or sotalol. However, the indirect comparison [redacted].

Whilst the mean risk of all-cause mortality with dronedarone from the ATHENA trial suggested a beneficial effect, this was not statistically significant in the moderate to high risk ATHENA population[b]. None of the differences estimated between AADs for all-cause mortality based on the head-to-head RCTs (DIONYSOS) or the results from the indirect comparison were statistically significant. However, in the MTC, dronedarone was reported to have a statistically significant reduction in the odds of all-cause mortality compared to both sotalol and amiodarone.

[b] Correction inserted after manufacturer identified a factual error

Only limited data from 4 studies was available for stroke, including 2 RCTs of dronedarone versus control, 1 RCT of dronedarone versus amiodarone and 1 RCT comparing sotalol versus amiodarone [c]versus control. Notably, none of these studies included stroke as a pre-specified endpoint. Only the results from the MTC were reported in the manufacturer's submission for this outcome. Dronedarone was associated with a statistically significant reduction in stroke compared to control, although this result was driven largely from the ATHENA trial based on a moderate to high-risk elderly AF population. No significant difference was reported between dronedarone and either amiodarone or sotalol based on the results from the MTC.

The adverse effect profile of dronedarone suggests it is well tolerated. However the results of various approaches to synthesis the evidence did not demonstrate a statistically significant reduction in treatment discontinuations with dronedarone compared with other AADs. Inconsistency was also apparent between results from the direct, indirect and MTC suggesting that there appears considerable uncertainty surrounding the findings for this outcome. An additional comparison of treatment discontinuation due to adverse events also failed to demonstrate a statistically significant difference between dronedarone and other AADs. However, the manufacturer noted that, while the odds ratio for dronedarone compared to amiodarone was not statistically different, the reported difference might still be considered clinically relevant.

1.3 Summary of submitted cost effectiveness evidence

No previous published cost-effectiveness studies of dronedarone in patients with AF/AFL were identified by the manufacturer. The submission included a discrete event simulation model which was used to estimate the cost-effectiveness of dronedarone with other licensed AADs and standard therapy alone for AF. The comparison with standard therapy alone was restricted to high-risk elderly AF patients with a $CHADS_2$ score ≥4. This model was used to evaluate the cost-effectiveness over five main patient groups in accordance with the clinical pathways for these populations presented in current UK clinical guidelines. The results from the manufacturer's submission demonstrated that dronedarone appeared highly cost-effective in each of the populations compared to using standard baseline therapy

[c] Correction inserted after manufacturer identified a factual error

alone as first line treatment, or compared to sotalol or amiodarone as a first line anti-arrhythmic. The results for dronedarone, relative to class 1c agents, showed that dronedarone was borderline in terms of cost-effectiveness with an incremental cost-effectiveness ratio (ICER) just above £20,000 per Quality-Adjusted Life Year (QALY) and a 50% probability of being cost-effective at this threshold. The findings were reported to be robust across a wide range of alternative assumptions. The results appeared most sensitive to the time horizon of the model and assumptions regarding the benefits from AADs on mortality.

The main driver of cost-effectiveness for the comparisons of dronedarone versus standard therapy as first line treatment, and sotalol or amiodarone as first line anti-arrhythmics, is the additional mortality benefit attributed to dronedarone. Stroke benefits and differences in treatment-related adverse events have only a very limited impact on cost-effectiveness for these comparisons. In contrast, the main drivers of cost-effectiveness for the comparison of dronedarone versus class 1c agents are a combination of the benefits assumed from stroke and a reduction in adverse events. The ERG noted that if only the potential benefits of AF recurrence are included in the model then dronedarone does not appear cost-effective for any of the populations considered.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

The ERG checked the systematic review search strategy and considered the search to be comprehensive in that all relevant trials of dronedarone appeared to be included in the submission. The full clinical trial programme for dronedarone appears to have been the largest to date for any AAD with more than 6500 patients with AF/AFL recruited, although the majority of the data comes from the ATHENA study (n=4628) which was specifically designed to evaluate the prevention of hospitalisation due to cardiovascular events or death rather than antiarrhythmic effects.

Since only one of the RCTs compared dronedarone head-to-head versus another AAD (amiodarone), the manufacturer conducted a more comprehensive systematic review to identify additional RCTs for other relevant comparator AADs, including

class 1c agents, sotalol and amiodarone. A range of alternative synthesis approaches were employed by the manufacturer in order to assess the relative effectiveness of dronedarone compared to other AADs that are currently used in the NHS. The results of these separate comparisons were reported for each of the main clinical outcomes

In general, the ERG considered the economic submission to be of high-quality meeting the requirements of the NICE reference case. The economic model structure was considered appropriate for the decision problem and the detailed sensitivity analyses were thorough and informative in exploring the robustness of the results.

1.4.2 Weaknesses

Although the systematic review produced by the manufacturer appeared comprehensive, the inclusion/exclusion criteria applied to studies to be included in direct and indirect analysis were not explicitly stated in the supplementary report of the main submission. Furthermore, although it was stated in the protocol of the systematic review that study quality would be assessed, no details of study quality were provided and there was no evidence that study quality was used to inform the analysis.

In the data synthesis reported by the manufacturer there were different inclusion/exclusion criteria applied to studies for the direct meta-analysis and the MTC. The additional filter applied to the MTC, due to apparent problems of convergence, resulted in a substantial reduction in the number of studies entering the MTC compared to the direct meta-analysis. Generally, a MTC analysis is used to increase the network of evidence available so that the relative effectiveness of the comparators can be simultaneously estimated by borrowing strength from the wider range of trials, but in the manufacturer's submission the network of evidence has been reduced considerably relative to that available for the direct and indirect analysis. This represents a potential weakness given that the MTC results are subsequently used to inform the relative effectiveness estimate in the cost-effectiveness analysis. Although it is unclear whether the additional filter applied to the MTC would introduce bias, the exclusion of potentially relevant evidence will certainly increase the overall uncertainty surrounding these outcomes. Furthermore, the methods and results of the MTC were not reported in sufficient detail in the

original submission to allow a detailed evaluation of the approach employed by the manufacturer. Finally, there was inconsistent use of continuity corrected data and inconsistency in the use of time points.

Issues of clinical and statistical heterogeneity between the different studies were not considered to have been sufficiently reported or explored by the manufacturer. Consequently, the validity of pooling the individual studies in the different synthesis approaches represents a potential weakness. This is an important issue given that the majority of evidence for dronedarone is derived from the ATHENA study which represents a moderate to high-risk elderly AF population. The exchangeability of this study with lower risk and younger AF populations has not been fully considered within the submission[d] and nor has the generalisability of the ATHENA population to the overall AF population managed in the NHS.

The ERG identified a number of potential weaknesses related to the economic submission and electronic model which were considered to impact on the validity of the cost-effectiveness results. These included: (i) the treatment pathways evaluated by the manufacturer may not represent the full range of relevant strategies or sequences; (iii) the use of baseline data from the ATHENA trial may not be generalisable to the UK AF population; (iv) the use of a restricted set of studies, and the assumptions used for class 1c agents, to inform the relative effectiveness estimates applied in the model; (v) uncertainty surrounding the HRQoL data used in the model; (vi) uncertainty in relation to the acquisition costs, initiation and monitoring costs of dronedarone.
The ERG explored the robustness to a number of these uncertainties. The ICER of dronedarone remained relatively robust throughout (< £20,000 per QALY) except for the following assumptions: (i) amiodarone and sotalol have the same effect on all-cause mortality as dronedarone; and (ii) class 1c has the same effect on stroke as dronedarone. In these situations, the ICER of dronedarone was well above £30,000 per QALY.

Finally, the submission does not explicitly consider the potential clinical or cost-effectiveness of dronedarone for patients with AFL. Although the manufacturer reports that AFL is not mentioned specifically within the proposed licensed indication,

[d] Correction inserted after manufacturer identified a factual error

some AFL patients are assumed to be clinically similar to AF and hence the manufacturer assumes that the results of the economic model will apply to these patients. However, there is no discussion of the validity of this assumption or the proportion of patients with AFL who would be considered to be clinically similar to AF.

1.4.3 Areas of uncertainty

The relative effectiveness and cost-effectiveness of dronedarone versus other AADs remains subject to a number of areas of uncertainty in terms of informing current NHS practice. These uncertainties include:

- The majority of evidence for dronedarone is derived from the ATHENA study which reflects a moderate to high-risk elderly AF population. The generalisability of this evidence to inform the management of a lower risk and younger AF population remains uncertain.

- The evaluation of dronedarone as first line adjunctive treatment to standard baseline therapy was restricted to patients with multiple CV risk factors, corresponding to a $CHADS_2$ score ≥ 4. $CHADS_2$ is a clinical prediction score for estimating the risk of stroke in patients with AF and its validity as a prediction score for CVD risk more generally and for all-cause mortality remains uncertain. Furthermore, the all-cause mortality effect for this subgroup was derived from a post-hoc analysis.

- The relative efficacy of dronedarone compared to other AADs remains highly uncertain. The short-term (6 months) DIONYSOS trial is the only head-to-head RCT identified comparing dronedarone with another AAD.

- The comparison of dronedarone and the other AADs was based on a range of alternative synthesis approaches incorporating direct and indirect evidence. Given the lack of consideration of clinical and statistical heterogeneity across the different studies, the validity of pooling the individual studies in the different synthesis approaches remains uncertain.

- Existing clinical evidence across the AADs appears most robust for AF recurrences and appears considerably more uncertain for the other major clinical endpoints identified as major drivers of cost-effectiveness (e.g. all-

cause mortality and stroke). Although dronedarone was reported to have a statistically significant reduction in the odds of all-cause mortality compared to both sotalol and amiodarone based on the MTC, neither the results from the head-to-head RCT (DIONYSOS) nor the results from the indirect comparisons reported a statistically significant difference. The existing evidence for stroke is also highly uncertain and only a small number of studies have reported this outcome. This was not a pre-specified outcome in any of these studies and no significant difference was reported between dronedarone and amiodarone and sotalol based on the results from the MTC. In the absence of data on stroke for class 1c agents the manufacturer assumed that these had no effect on stroke compared to standard baseline therapy. Consequently, the additional benefits attributed in the economic model to dronedarone compared to other AADs and to all-cause mortality and stroke remain highly uncertain.

- Although dronedarone is also licensed to lower ventricular rate, rate control was not included as an outcome measure of the scope or in the submission.

- HRQoL has not been directly assessed in any of the existing dronedarone RCTs. The presence and potential magnitude of any quality of life benefits attributed to dronedarone are thus uncertain.

There remain a number of additional sources of uncertainty related to the cost-effectiveness of dronedarone which the ERG has been unable to adequately address. These include establishing the most appropriate source of data to inform the baseline event rates applied in the model; the position for dronedarone in the pathway of treatment sequences; HRQoL benefits of dronedarone and the maintenance of benefits over the longer term.

Finally, it is important to note that the final acquisition price for dronedarone has not yet been confirmed by the manufacturer. A daily cost of £2.30 is assumed in the base case analysis and a range between £2.20 - £2.50 is considered in the sensitivity analysis. The ERG made a final request for confirmation of the acquisition price from the manufacturer on 22/10/09 but failed to receive confirmation prior to submission. Should the final acquisition price exceed these estimates, then the cost-effectiveness results may no longer be valid.

1.5 Key issues

Further trials or the implementation of registries would be helpful to establish the efficacy and safety of dronedarone relative to other AAD treatments that are regularly used in this indication within UK clinical practice. Additional evidence related to the effectiveness of AADs for patients with AFL would also be valuable. Longer-term follow-up of trials, with prespecified outcome measures and analyses, are required to better establish the longer term efficacy or safety of dronedarone compared to other AADs. This is of particular importance in regard to outcomes of all-cause mortality and stroke since these appear to be the key drivers of the cost-effectiveness results. Given the lack of existing health related quality of life data, future RCTs of dronedarone should also consider using a relevant HRQOL measure.

Focusing on the existing dronedarone RCTs the key issues relate to the following: the generalisability of the ATHENA study which reflects a moderate to high-risk elderly AF population to a younger and less risky population; the use of a post-hoc subgroup analysis for the 1st line comparison of dronedarone as an adjunctive treatment to standard care and the lack of HRQoL data.

In terms of the broader comparison made using existing RCTs for the comparator AADs, the ERG considers that potential clinical and statistical heterogeneity has not been adequately considered. In particular, the exchangeability of the ATHENA study with these studies remains subject to a number of uncertainties. Furthermore, the additional restrictions imposed on the inclusion of RCTs in the MTC are likely to increase the overall decision uncertainty compared to a fuller use of this evidence. Finally, the additional benefits assigned to dronedarone in the cost-effectiveness analysis attributed to a reduction in all-cause mortality and stroke compared to other AADs are subject to considerable uncertainty and are clearly not assumed to be directly related to the impact on AF recurrence, since dronedarone is assumed to be the least effective of the AADs for this outcome. Since these additional benefits are mediated through AF recurrence, a key issue that remains is the need to more precisely identify the mechanism of effect by which these wider cardiovascular and mortality benefits attributed to dronedarone are potentially achieved.

Key issues specifically relevant to the economic evaluation, include the following: establishing the most appropriate source of data to inform the baseline event rates applied in the model; the potential cost-effectiveness of dronedarone in a range of alternative and feasible treatment sequences, the potential HRQoL benefits of

dronedarone and the maintenance of benefits over the longer term; and the absence of a final confirmed acquisition price at the time of the submission of the ERG report. Finally, the lower initiation costs assumed for dronedarone and differential monitoring costs between treatments are uncertain, although these do not appear to have a significant impact on the final ICER results[e].

2 Background

2.1 Critique of manufacturer's description of underlying health problem

The ERG considers that the description of atrial fibrillation (AF) presented in the manufacturer's submission is adequate and appropriate. The description of the clinical outcomes associated with AF places much emphasis upon the risk of stroke, with only a brief mention of the adverse effect that symptomatic AF can have on a patient's quality of life.

2.2 Critique of manufacturer's overview of current service provision

The ERG considers that the manufacturer's overview of current service provision for atrial fibrillation (AF) is adequate and appropriate, and appears to be structured around the NICE guidelines in terms of the treatment strategies as illustrated in figures 1 and 2, which are taken from the manufacturers submission (pages 71 and 72).

[e] Correction inserted after manufacturer identified a factual error

Figure 1: Comparator treatment pathway and considered dronedarone positions for paroxysmal AF patients.

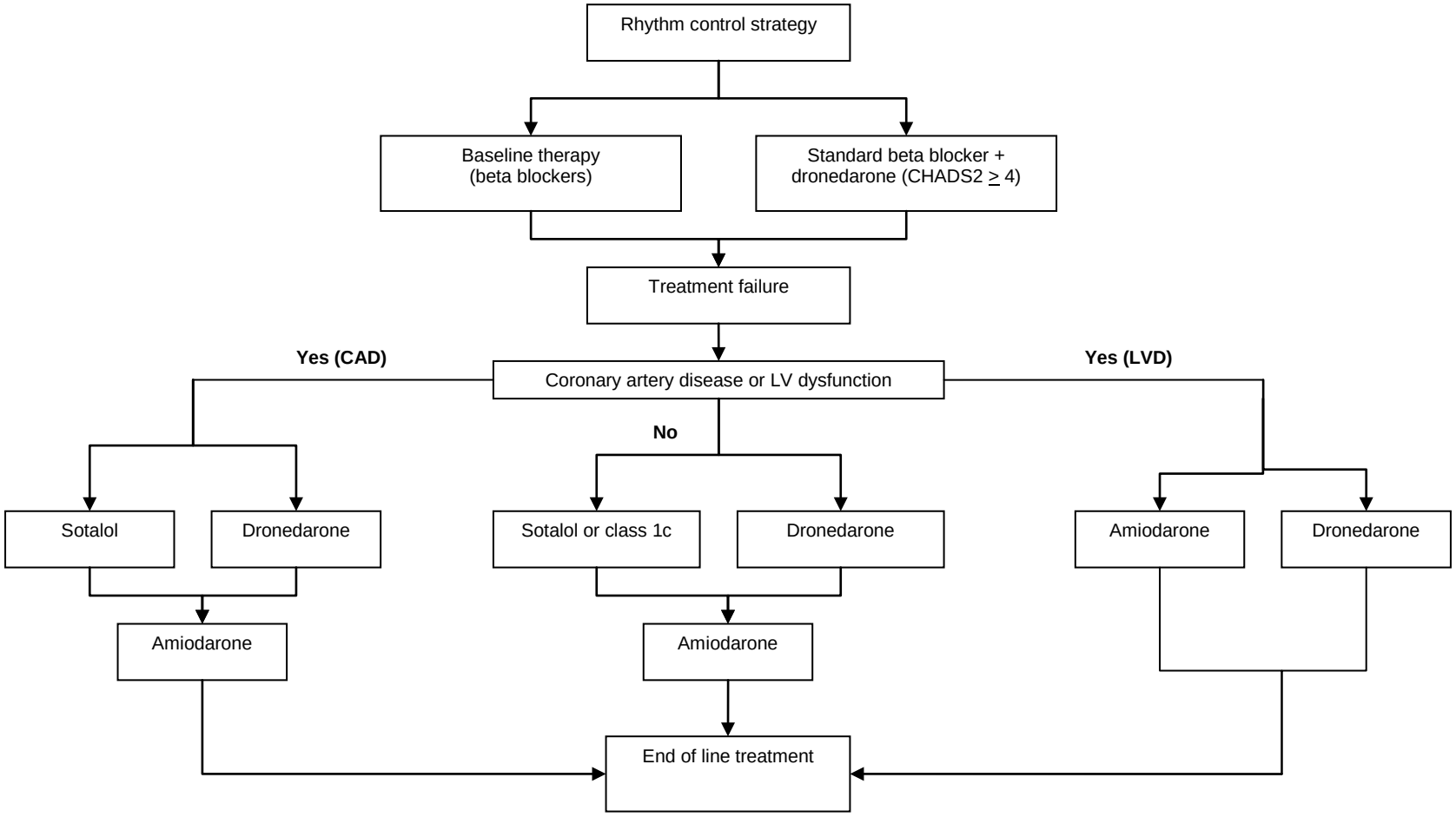
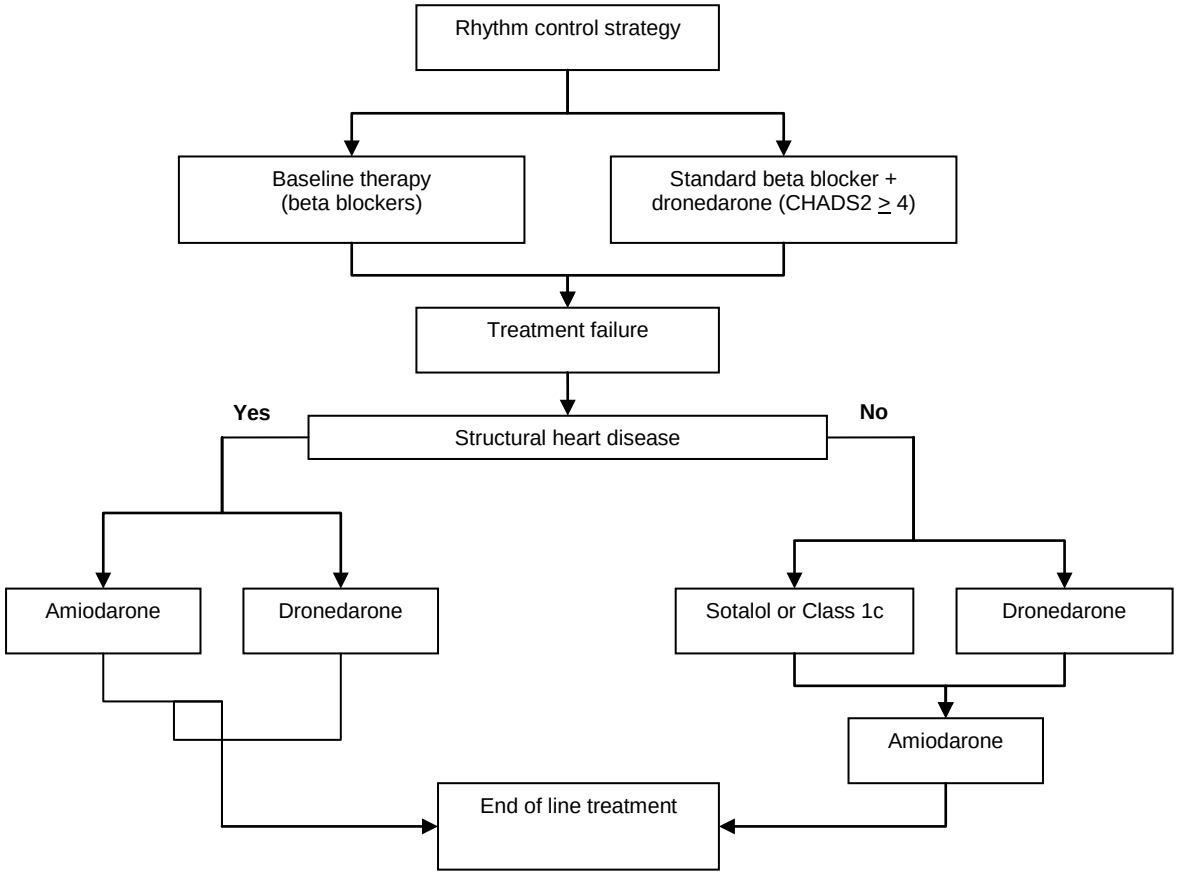


Figure 2: Comparator treatment pathway and considered dronedarone positions for persistent AF patients.



3 Critique of manufacturer's definition of decision problem

3.1 Population

The wording describing the population addressed in the submission differed slightly from that described in the scope of this appraisal.

"People with either a recent history of, or current paroxysmal or persistent atrial fibrillation or atrial flutter, who are current receiving standard baseline treatment with or without beta blockers"

was replaced with:

"*As per the anticipated licensed indication: for stable adult patients with a recent history of, or current non-permanent atrial fibrillation (AF).*

Although they are essentially the same population, the wording in the submission more closely reflects that of the draft SmPC. The important difference is that the submission decision problem specifies that "*This population is likely to exclude patients with NYHA CHF Class IV and also NYHA Class III CHF with a recent haemodynamic instability."* This exclusion is one that anticipates the product licence and reflects concerns over the use of dronedarone in such patients given the results of the terminated ANDROMEDA trial in which dronedarone was found to be statistically significantly associated with increased early mortality related to worsening heart failure in patients who were hospitalised with symptomatic heart failure or severe left ventricular systolic dysfunction (HR: 2.13. 95% CI: 1.07, 4.25).[3]

The evaluation of dronedarone as a first line adjunctive treatment to standard baseline therapy was restricted further to patients with multiple CV risk factors, corresponding to a $CHADS_2$ score ≥ 4. Patients with atrial flutter (AFL) were not considered separately in the submission or the economic model. In part this appears to reflect the proposed EMEA wording for the license for dronedarone which does not specifically mention AFL. However, the manufacturer also assumed that for there would be groups of patients for whom their AFL is clinically indistinguishable from AF and that the assumptions and results of the economic model for AF would also apply to this group.

3.2 Intervention

The intervention considered in the submission is dronedarone (Multaq ®). Dronedarone has properties belonging to all four Vaughan-Williams classes. It is a multi-channel blocker inhibiting the potassium currents (including IK (Ach), IKur, IKr, IKs) and thus prolonging cardiac action potential and refractory periods (Class III). It also inhibits the sodium currents (Class Ib) and the calcium currents (Class IV). Furthermore it non-competitively antagonises adrenergic activities (Class II). [4] In their statement of the decision problem the manufacturer specifies just "dronedarone". However, the submission is more specific in the section on cost-effectiveness (see section 7.2.1.1 for details), stating that "Dronedarone is assumed to be used in the model as per its licence i.e. 400 mg bid." This reflects the anticipated product licence.

3.3 Comparators

In the final NICE scope comparators were divided into first and second line therapy, "As a first line treatment or as an adjunct to standard baseline therapy, dronedarone will be compared with standard baseline therapy with or without beta blockers. As a second line therapy, dronedarone will be compared to the following drugs according to their indications: class 1c anti arrhythmic agents (flecainide); and sotalol"

In the manufacturer's submission the above terminology has been slightly altered although the interpretation appears to remain the same. Within the manufacturer's decision problem dronedarone is compared as first line treatment as an adjunct to standard baseline therapy and as an alternative 1st line to current anti-arrhythmic drugs (AADs), such as class 1c antiarrhythmic agents (flecainide and propafenone)[f], sotalol, and amiodarone.

This proposed treatment pathway is described in more detail in Section 7 of the manufacturer's submission (p 70-72 and Figures 7.2a and 7.2b). This reflects the main treatment sequences based on UK clinical guidelines. However, it is not clear from the manufacturer's description of the comparator technology why dronedarone should be introduced at an earlier stage than current anti-arrhythmic agents for moderate to high-risk elderly AF patients. Also the manufacturer only considers the use of dronedarone at specific points within the treatment pathway, thus precluding a full assessment of the effectiveness and cost-effectiveness of using dronedarone at

[f] Correction inserted after manufacturer identified a factual error

different points in the pathway. For example, for paroxysmal patients with CAD, the manufacturer compares sotalol and dronedarone as alternative 1st line AADs. Patients who withdraw from these treatments are then assumed to receive amiodarone as a 2nd line AAD. However, no assessment is made of the potential effectiveness and cost-effectiveness of alternative management approaches which could involve using dronedarone as an alternative 2nd line AAD to amiodarone. This approach would require additional strategies to be included in the economic model (e.g. a strategy of sotalol followed by dronedarone or amiodarone).

Missing from the decision problem and the submission is a full investigation of the outcome 'rate control' i.e. the effect of dronedarone and comparators. The effect of dronedarone on heart rate during AF is specified in the (draft) SPC. If dronedarone is administered primarily to control ventricular rate during AF rather than as an AAD, then comparators should include purely rate limiting drugs..

3.4 Outcomes

The outcomes considered in the manufacturer's submission include all-cause mortality, AF recurrence, stroke, cardiac events, treatment withdrawals, adverse events of treatment, and health-related quality of life. The outcomes listed in the decision problem reflect those specified in the final NICE scope, although time to recurrence of AF/AFL is not considered in the broader synthesis. However, some dronedarone RCTs provide time to AF recurrence data.

The emphasis placed on certain outcomes in the manufacturer's submission reflects the clinical trial program of dronedarone, during which the emphasis moved away from demonstrating that dronedarone was an effective AAD at preventing AF/AFL recurrence to demonstrating it could have an overall beneficial effect on cardiovascular morbidity and all cause mortality. These outcomes are appropriate for an AAD. However, consideration must be given to how the effects on broader outcomes such as all-cause mortality and stroke are mediated.

Missing from the decision problem and the submission is a full investigation of the outcome 'rate control' i.e. the effect of dronedarone and comparators on heart rate during AF. This effect of dronedarone is specified in the (draft) SPC.

3.5 Time frame

The submission proposes that patients demonstrating a good response to dronedarone therapy should continue to take the drug indefinitely. This is appropriate for an AAD, given the nature of AF as a chronic progressive disorder. Most trials of dronedarone are relatively short term with 6 to 12 months follow up, with the exception of the ATHENA study that has a mean follow up of 21 months. However, long term effectiveness of dronedarone remains uncertain.

3.6 Other relevant factors

In the submission subgroups of the population to be treated with dronedarone are defined using the $CHADS_2$ risk score (1 point for each of the following: recent congestive heart failure, hypertension, age 75 or over, diabetes mellitus, and 2 points for a history of stroke of transient ischaemic attack). However, $CHADS_2$ is a clinical prediction score for estimating the risk of stroke in patients with AF and its validity as a prediction score for CVD risk more generally and for all-cause mortality remains uncertain. Furthermore, this score has a poor ability to both predict thromboembolism in AF patients and to separate these patients into risk categories that correspond to different rates of thromboembolism [5]. Therefore, the use of $CHADS_2$ to define patients groups at risk of CVD events generally and to stratify treatment effectiveness for all-cause mortality may not be clinically meaningful. The ERG requested that the manufacturer provide details of any previous published studies in which AF patients had been stratified according to $CHADS_2$ score. The company performed the search as requested and were unable to locate any studies. Thus, this appears the first instance of the $CHADS_2$ score being used to stratify treatment effectiveness for all-cause mortality in an AAD study.

4 Clinical Effectiveness

4.1 Critique of manufacturer's approach

In the manufacturer's submission, the clinical evidence supporting the use of dronedarone as a treatment for AF is presented and synthesised in a number of ways. Firstly, the relevant dronedarone RCTs are presented and discussed. These are then compared with a broader set of RCTs for the comparator AADs. The

effectiveness of dronedarone and the comparator AADs are compared based on the results of a direct meta-analysis comparing each AAD with the control/placebo group. Further analyses are then conducted to generate results for indirect comparisons between individual AADs using the control/placebo group as a common comparator. Finally, the submission includes results from a mixed treatment comparison (MTC) combining evidence from the direct meta-analysis with the results from head-to-head studies.

4.1.1 Description of manufacturers search strategy and comment on whether the search strategy was appropriate.

The manufacturer's submission described the search strategies used to identify relevant studies of dronedarone for AF/AFL, and full details of the search strategies used in each section were reported in the appendices or in supplementary material provided. Overall, the search strategies employed for each of the sections of the submission were appropriate. A detailed commentary of the search strategies employed is given in Appendix 1.

4.1.2 Statement of the inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate.

Within the submission the trials of dronedarone are presented separately from the review of dronedarone and comparators. The detailed presentation of the dronedarone trials focuses on 4 phase III randomised controlled trials (RCTs) of dronedarone (EURIDIS/ADONIS, ATHENA and DIONYSOS), These four trials are in the licensed indication and provide data on the efficacy and safety of dronedarone. Additionally, one phase II dose ranging study (DAFNE) and one phase III trial of dronedarone in patients with permanent AF (ERATO), which are not discussed in detail in the submission, provide data on treatment discontinuations and safety of dronedarone.

The comparative review of dronedarone and other AADs presented in the submission is based on a broader systematic review and meta-analysis commissioned by the manufacturer. This broader review encompassed class 1a, class 1c, class III, class II, class IV drugs and digoxin, and ablation. Within this broader review, dronedarone,

amiodarone, sotalol, flecanide and propafenone were identified as 'priority drugs' and the review was based on 72 RCTs that involved at least one priority drug treatment arm. Although this is not clearly stated in the submission, only comparisons of these priority drugs with a non-active control or another priority drug were included in the meta-analysis presented in the submission (a total of 39 trials). Thus, only trials of dronedarone, amiodarone, sotalol, flecanide and propafenone versus placebo, no treatment, or each other are included.

The manufacturer's submission states that the systematic review of clinical effectiveness included RCTs and controlled trials, although the systematic review submitted as supplementary material to the ERG itself states that only RCTs were included, with [redacted] Outcomes of interest included mortality, stroke, AF recurrence, treatment withdrawals (due to any cause and due to adverse effects) [g]and serious adverse effects (SAEs).

The intervention criteria were appropriate given current NICE guidelines for the treatment of AF. The use of data from RCTs only was appropriate as this is likely to provide the most reliable estimates of efficacy and safety. However, the inclusion of participants with any type of AF was inappropriate given that the anticipated licensed indication for dronedarone was for non-permanent AF, i.e. paroxysmal and persistent AF, rather than permanent AF. The omission of health-related quality of life as an outcome deviates from the inclusion criteria stipulated in systematic review protocol, and the appraisal scope.

The protocol for the systematic review stated that [redacted]. This method of study selection and data extraction would reduce the potential for reviewer error and bias. However, it is not stated within the systematic review or the company submission whether this method of study selection or data extraction was adhered to, and therefore it can not be determined if these review processes were subject to reviewer error or bias.

4.1.3 Table of identified studies. What studies were included in the submission and what were excluded.

Direct evidence of the clinical effectiveness and safety of dronedarone comprised four Sanofi-Aventis funded studies: EURIDIS/ADONIS, ATHENA and DIONYSOS, which were all performed in a population of the licensed indication. One phase II dose ranging study (DAFNE) and one phase III trial of dronedarone in patients with permanent AF (ERATO) are not discussed in detail in the submission but illustrate dronedarone's efficacy in terms of AF recurrence and ventricular rate and provide data on treatment discontinuations and safety of dronedarone. Another major dronedarone trial that is not included in the submission is the ANDROMEDA trial. The exclusion of this trial is justified given that its population of patients without AF, and with congestive heart failure (CHF) is contra-indicated for dronedarone.

A total of 39 publications that reported data from RCTs were used in the analysis. Details of which trials were included in which analyses are included in appendices to the submission and in the separate report of the systematic review, however, summary details of all trials are not included. The systematic review report supplied to the ERG upon request provides a table of the 72 included studies of the priority drug dataset and the outcomes to which they contribute data, but there is no equivalent table of the smaller number of trials that are actually considered in the submission. In addition, no study details such as sample size, duration or even all treatment arms in each trial are provided making it difficult to understand how trials were selected for inclusion in analyses. A table of excluded studies would have increased the transparency and reproducibility of the results.

[g] Correction inserted after manufacturer identified a factual error

Table 1: Summary of dronedarone trials included in the submission

	ADONIS/EURIDIS[h]	ATHENA	DIONYSOS	DAFNE	ERATO
Population characteristics					
n	1237	4628	504	199	174
Dose	400mg BD	400mg BD	400mg BD	800mg, 1200mg 1600mg	400mg BD
Age range	Dronedarone: Mean 63.5 (10.7) Placebo: Mean 62.2 (11.1)	Mean 71.6 (SD 9.0) <65yr: 18.9% 65 to <75yr 39.5% ≥75yr 41.6%	Mean 64 yrs Range 28-90 52% <65 yrs 19% ≥75 yrs	Mean 64/65 yrs	Mean 65 yrs (range 39-86)
Type of AF	Paroxysmal and persistent AF	Paroxysmal and persistent AF At least one risk factor for cardiovascular hospitalisation	Persistent (cardioversion indicated). (although excluded, some were classed as having paroxysmal or permenant AF)	Persistent (cardioversion indicated).	Permenan
Anti-coagulation used?	Majority of patients were receiving anti-coagulants	44% receiving aspirin	Yes	Yes	
Hypert	Dronedarone: 60% Placebo: 50.1%	86.3%	67%	54%	48%
SHD	Dronedarone: 42.4% Placebo: 39.7%	59.6%	28%		39%
CHF	Dronedarone: 17.3% Placebo: 17.87%	21.2%	22%(not III or IV at time of randomisation)	Approx 20% mean LVEF 55%	40%
Treatment duration	12 months	Min. 12 months	6mths+	6mths	6 mths
Outcome measure					
Primary outcome measure	Recurrence of AF (measured by transtelephonic ECG when symptomatic)	First hospitalisation due to CV events or death	AF recurrence or premature discontinuation due to intolerance or lack of efficacy (AF recurrence measured by unscheduled ECG)	AF recurrence (measured by transtelephonic ECG when symptomatic)	Rate control 24hour Holter monitoring at day 0, 14 and at 4 months
Secondary outcomes	Symptoms related to AF	death from any cause	Occurrence of Major safety endpoint	HR at AF recurrence	Safety and tolerability
	mean ventricular rate during first recurrence of AF	Death from CV causes	Occurrence of drug specific AEs	Side effects	
Post hoc analyses		First hospitalisation due to CV event AF recurrence (measured by scheduled ECGs, hospitalisation for AF/AFL, electrical cardioversion) Stroke [redacted]			

Each of the five dronedarone studies included in the submission are summarised in Table 1 and described briefly below:

[h] These two trials have identical protocols and are often considered as a single trial.

EURIDIS/ ADONIS

EURIDIS and ADONIS were two identical phase III randomised placebo controlled, double blind trials the results of which are usually considered as a single trial. EURIDIS was performed in 12 European countries, whilst ADONIS was performed in the United States of America, Canada, Australia, South Africa and Argentina. Details of these trials are adequately presented in the submission, except it is unclear what proportion of all patients had been previously treated with an AAD. 55%-58% of patients were taking concomitant beta-blockers. EURIDIS/ ADONIS evaluated the effect of dronedarone 400mg twice daily versus placebo on AF recurrence in 1237 adult patients with paroxysmal or persistent AF or AFL, the majority of whom were receiving anticoagulants.

In the 12 month follow up period, both EURIDIS and ADONIS showed that there was a greater time to recurrence of first AF/AFL (primary outcome) in patients who received dronedarone compared to those that received placebo (combined result: median time to first recurrence 116 days versus 53 days), while AF recurrence rate at 12 months was significantly lower with dronedarone (HR 0.75 (95% CI 0.65, 0.87, $p<0.001$). Ventricular rate at first recurrence of AF/AFL was also found to be statistically significantly lower in patients receiving dronedarone compared to those receiving placebo (combined result: mean rate (bpm) 103.4 ± 25.9 versus 117.1 ± 30.4). [1]

Overall this was a good quality trial that demonstrated that dronedarone has clinical anti-arrhythmic and rate control properties in patients with persistent/paroxysmal AF.

ATHENA

The ATHENA trial is discussed at length in the submission and is a very important trial. ATHENA was a very large (n=4628 patients) phase III RCT that evaluated the effect of dronedarone 400mg twice daily versus placebo on hospitalisation due to cardiovascular events or death. It did not re-evaluate the anti-arrhythmic or rate control properties of dronedarone. ATHENA was conducted in patients with a high risk of hospitalisation or death, with paroxysmal or persistent AF/AFL, although 75% of patients were in sinus rhythm at study entry. Patients were considered to be of

moderate to high risk because they had at least one risk factor for cardiovascular hospitalisation. Due to overall mortality rate being lower than expected, the eligibility criteria were changed in the course of the trial so that patients aged less than 70 years of age were no longer eligible for the trial and only patients aged 75 years of age and older could be entered without an additional risk.[6, 7] It is not known what proportion of patients entered into this trial had persistent or paroxysmal AF. However, data provided to the ERG would indicate that most patients had asymptomatic (or at least not troublesome) non-permanent [i]AF (see details of post hoc analysis of AF recurrence data).[8] Follow up ranged from 1 to 2.5 years.

The incidence of hospitalisation due to cardiovascular events or death (composite primary outcome) was statistically significantly lower in patients receiving dronedarone compared to those receiving placebo (HR 0.76, 95% CI: 0.69, 0.84, p<0.001). This outcome appeared to be mainly driven by a reduction in time to first cardiovascular hospitalisation due to a significant reduction in hospitalisation for AF (HR 0.63, 95% CI: 0.55, 0.72, p<0.001). There was no statistically significant difference in all-cause mortality between patients receiving dronedarone and those receiving placebo (HR: 0.84, 95% CI: 0.66, 1.08; p=0.18). However, cardiovascular mortality was statistically significantly lower in patients receiving dronedarone compared to those receiving placebo (0.71, 95% CI: 0.51, 0.98; p=0.03). Furthermore, whilst there was no statistically significant difference in non-arrhythmic cardiac mortality and non-cardiac vascular mortality between the two groups of patients, cardiac arrhythmia mortality was statistically significantly lower in patients receiving dronedarone compared to those receiving placebo (HR: 0.55, 95% CI: 0.34,0.88).[7].

Several post-hoc analyses were performed using data from ATHENA. Therefore, the findings from these post-hoc analyses should be interpreted with some caution.

[redacted] This finding has been used in the submission to support the suggestion dronedarone may be used for patients with multiple CV risk

[i] Correction inserted after manufacturer identified a factual error

factors (corresponding to a CHADS2 ≥4) on top of standard baseline therapy[j]. As stated earlier, it is difficult to know how valid this is since comparable analyses do not seem to have been conducted for other AADs. A number of associations between mortality and $CHADS_2$ score were performed and the statistically significant results seen here should be interpreted with caution as this may have arisen for purely statistical reasons i.e. multiple testing, for which no statistical correction was made in the analyses presented in the company submission.

A second post-hoc analysis showed that there was a statistically significant reduction in the risk of stroke in patients receiving dronedarone compared to those receiving placebo (HR: 0.66, 95% CI: 0.46, 0.96, p=0.027). However, stroke was not an a priori specified outcome but was recorded as an adverse effect. [2]

A third post hoc analysis was reported for the subgroup of patients who had entered the trial in sinus rhythm (75% of the population), in which AF recurrence measured post hoc based on scheduled ECGs, (day 7 and 14, months 1, 3 and 6 and every 6 months thereafter), hospitalisation for AF/AFL, or reported electrical cardioversion, showed a statistically significant increase in median time to first recurrence of AF/AFL in patients receiving dronedarone (737 days) compared to those receiving placebo (498 days) (HR: 0.75, 95% CI: 0.68, 0.82, p<0.001).[8] The result for this subgroup is not comparable with reports of this outcome from other trials. This is because AF recurrence in the ATHENA population was extrapolated from hospitalisation for AF/AFL, reported electrical cardioversion or AF/AFL on an scheduled ECG (day 7, 14, months 1, 3, 6 and every 6 months thereafter (clarification received from manufacturer), whereas in other trials, AF was measured measured by transtelephonic ECG whenever the patient had symptomatic AF. This means that AF recurrence in most patients in the ATHENA study, particularly those with paroxysmal AF, would have been missed unless it was symptomatic and troublesome.

DIONYSOS

The DIONYSOS trial is discussed in detail in the submission. It is an important trial because it is the only direct comparison between dronedarone and another AAD.

[j] Correction inserted after manufacturer identified a factual error

DIONYSOS was a phase III randomised double blind trial that compared the efficacy and safety of dronedarone to amiodarone in 504 adult patients with persistent AF who were indicated for electrical cardioversion and were receiving anticoagulants. Unlike earlier efficacy trials of dronedarone alone, the primary efficacy outcome of interest in DIONYSOS was a composite outcome of time to first AF recurrence or premature study drug discontinuation due to intolerance or lack of efficacy, with a follow-up period of 6 months (though some patients remained in the trial for up to 14 months. It would appear that this endpoint was selected to investigate whether the modest anti-arrhythmic activity of dronedarone would be compensated for by improved tolerability compared to the acknowledged poor tolerability of amiodarone. As reported in the submission, this was not found to be the case: the incidence of the primary efficacy was 75.1% and 58.8% in the dronedarone and the amiodarone groups respectively after 12 months of treatment (HR:1.59; 95% CI 1.28; 1.98; log-rank p-value <0.0001). The submission reports a higher rate of withdrawals due to intolerability with amiodarone but no statistical comparison was reported [redacted]. Although mortality was not an endpoint in this trial mortality data from this trial were used in the meta-analysis. As this is the only direct comparison between dronedarone and amiodarone these data are prominent in the reported results. Examination of these data by the ERG revealed the following: there were two deaths on dronedarone – [redacted], while there were five deaths on amiodarone – [redacted]. If the [redacted] deaths due to [redacted] are removed from the comparison the difference between the treatments could easily be due to chance.

DAFNE

DAFNE was not described in detail in the industry submission since it was a phase II randomised dose ranging study compared three doses of dronedarone (800mg, 1200mg and 1600mg daily) against placebo in patients with persistent AF. From the published report it appears that patients were not blinded to treatment. The primary outcome of interest was time to first AF recurrence and secondary outcomes of interest included heart rate at AF recurrence, incidence of adverse events, and premature study drug discontinuation. In the six month follow up period, time to first recurrence of AF was statistically significantly greater in patients receiving 800mg of

dronedarone compared to those receiving placebo (median of 60 days versus 5.3 days, relative risk reduction 55%, 95% CI: 28, 72%, p=0.001). No significant effect was seen at higher doses, i.e. higher doses were possibly less effective that the 800 mg dose (data not reported in the published paper). Ventricular rate on recurrence of AF was found to be statistically significantly lower in patients receiving dronedarone compared to those receiving placebo (p-value across all doses =0.0001), indicating some rate control mode of action for dronedarone. The main adverse events associated with dronedarone were gastrointestinal, including diarrhoea, nausea and vomiting. [9]Overall this trial demonstrated the basic anti-arrhythmic and ventricular rate limiting efficacy of dronedarone.

ERATO

The ERATO trial, which was conducted in patients with permanent AF, is not discussed in detail in the submission although data from this trial are used in some analyses. ERATO was a phase III randomised double blind placebo controlled trial that evaluated the effect of dronedarone 400mg twice daily and standard therapy versus placebo and standard therapy in 174 adult patients with symptomatic permanent AF in which cardioversion was not a treatment option. This trial is in a population that is outside the licensed indication of dronedarone. Its purpose is therefore to demonstrate that dronedarone can have beneficial rate control effects where control of AF is considered impossible. The primary outcome of interest was change in mean ventricular rate as measured by a 24 hour Holter recording on day 14 compared to day 0. An *a priori* subgroup analysis was performed that stratified the results of the primary outcome according to concomitant use of other rate-lowering drugs. Patients receiving dronedarone had a statistically significant reduction in mean 24 hour ventricular rate of 11.7 beats per minute compared to those receiving placebo (p<0.0001). During maximal exercise there was a statistically significant reduction in mean ventricular rate of 24.5 beats per minute compared to those patients who received placebo. The effects of dronedarone on ventricular rate were additive to those of concomitant rate controlling drugs. The incidence of overall treatment emergent adverse events was slightly higher in the dronedarone arm compared to the placebo arm, however serious adverse events and premature discontinuations due to adverse events were reported to be similar in both groups.[10]

4.1.4 Details of any relevant studies that were not included in the submission

The ERG's independent search of the literature did not retrieve any additional dronedarone studies meeting the review inclusion criteria. Given the extent of the review with a large number of comparator trials it was not possible to confirm that all comparator trials have been identified. The systematic review undertaken appears to have been very thorough and so it is unlikely that any additional trials have been missed. However, the presentation of the review in the manufacturer's submission lacks clarity. By studying the submission and the supporting systematic review and checking each RCT that appears in the former but not the latter, the ERG identified that excluded trials are those that as well as being restricted to the priority drugs or a non-active control, exclude those in which the same\priority drug is in both treatment arms (see figure 1).

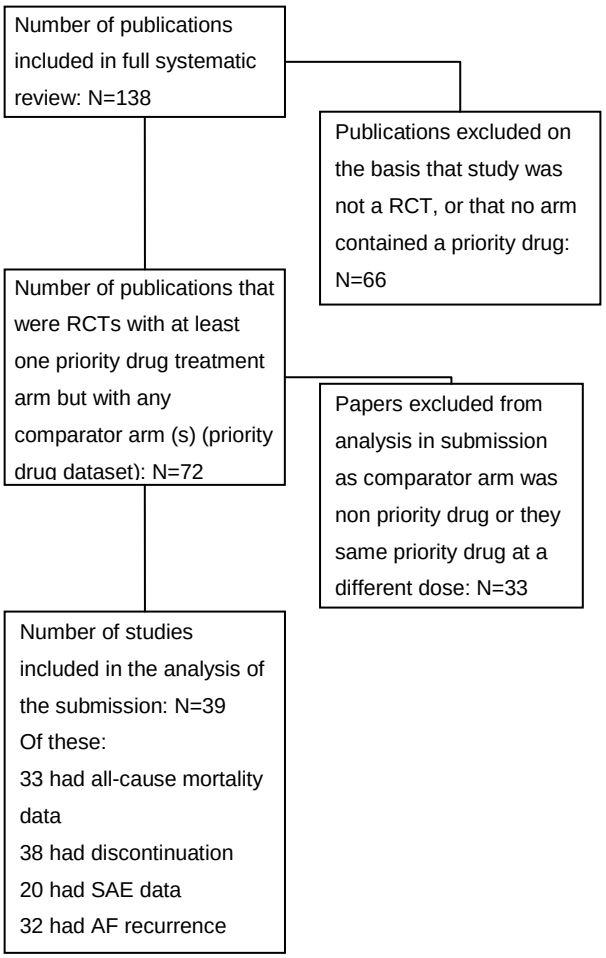


Figure 3: Study selection process as interpreted by the ERG from documentation supplied by the manufacturer

4.1.5 Description and critique of manufacturers approach to validity assessment

The systematic review protocol stated that the quality of the RCTs was to be assessed according to adequacy of randomisation, allocation concealment, blinding procedures, follow up and the use of intention to treat analysis. However, neither the manufacturer's submission nor the supporting systematic review provides any details of the validity assessment of the included RCTs, or utilises study validity in their analyses. Whilst data from RCTs can represent the best available evidence, the absence of a validity assessment means that it is not possible to determine the reliability of the data used in the analyses. In particular, in this submission, the lack of information on the comparator trials makes a lack of a quality assessment particularly important. This is because the quality of the trials should be considered when deciding whether they can be combined in a meta-analysis or MTC.

4.1.6 Description and critique of manufacturer's outcome selection

With the exception of health-related quality of life, the manufacturer's submission addresses each of the outcomes specified in the final scope issued by NICE, namely all-cause mortality, AF recurrence, stroke, cardiac events and adverse events of treatment. In addition the systematic review of clinical effectiveness and the manufacturer's submission addresses the outcome of treatment discontinuations which encompasses discontinuation not only due to adverse events but also due to lack of efficacy.

All-cause mortality

All cause mortality was defined as the reported number of deaths from any cause within the timeframe of the study and was evaluated for dronedarone and for comparator interventions. All-cause mortality was recorded as either a primary or secondary outcome, or as an adverse event in the included RCTs. In the manufacturer's submission, analysis was limited to all-cause mortality at or near 12 months. This is a reasonable criterion given the rarity of this endpoint. However, it is impossible to tell from the data provided how strictly the 12 month time point was adhered to in individual trials and in the analyses. Survival analysis of this type of

outcome would be preferable since it allows fuller consideration of the difference between treatments over the entire follow-up period (as opposed to at a single time point) and can account for any censoring. Furthermore, having specified in their methods that such trials would be included, studies in which there were zero events of mortality in both study arms were excluded from the MTC analysis. However, the meta-analyses and indirect comparisons present the results both excluding these trials and including them using the continuity correction. [k]The ERG explored the impact of this by repeating the analyses to include all relevant trials.

AF recurrence

AF recurrence was defined in the manufacturer's submission as the reported number of patients failing to maintain sinus rhythm at any point within the study timeframe or the reported number of patients with a recurrence of AF within the timeframe of the study. This outcome was evaluated for both dronedarone and for its comparators. In some analyses in the manufacturer's submission analysis was limited to AF recurrence at 12 months. AF recurrence was not an outcome measure in all dronedarone studies; AF recurrence data from the large and important ATHENA trial could not be compared to AF recurrence data of other trials due to the .way in which it was measured as discussed in section 4.1.3. In the ADONIS/EURIDIS trial(s) time to AF recurrence was the primary outcome measure.. Time to AF recurrence was not compared between drugs in the manufacturer's submission. Furthermore, although AF recurrence is an important endpoint for an AAD, if in clinical practice a patient's symptoms can be managed so that although AF occurs it is not bothersome (i.e. through rate control) AF recurrence may not be as important.

[k] Correction inserted after manufacturer identified a factual error

Stroke

Stroke was not a pre-specified outcome measure in any of the dronedarone trials. It was recorded as part of AE reporting procedures and, in the case of ATHENA, as reasons for mortality or hospitalisation. Therefore the stroke data may be subject to inter-trial variability and may not be reliable.

Cardiac events

Although cardiac events are specified in the decision problem and reported in the overview of dronedarone efficacy, they are not included in the comparative review of clinical effectiveness. Cardiac events for dronedarone were only evaluated in the ATHENA trial in which first hospitalisation due to cardiovascular events was part of the composite primary outcome (first hospitalisation due to cardiovascular events or death). The type of cardiovascular events were categorised as being due to AF, CHF, Acute Coronary Syndrome, syncope, ventricular arrhythmia or non-fatal cardiac arrest.

Adverse events of treatments

The manufacturer determined the safety of dronedarone through the evaluation of all adverse events reported in ATHENA, EURIDIS, ADONIS, ERATO and DAFNE, [11], serious adverse events (SAEs) reported in ATHENA, DAFNE and ERATO (see Section 4.6, Appendix 9 of the manufacturer's submission for further details) and treatment discontinuation due to adverse events reported in ATHENA, DAFNE, ERATO and EURIDIS/ADONIS (see Section 4.6, Appendix 7 of manufacturer's submission for further details).

With regard to the evaluation of serious adverse events, data from the EURIDIS/ADONIS trials were omitted from the analysis. The manufacturer did not provide a reason for this omission. At the request of the ERG, the manufacturer provided full and complete details of all treatment emergent adverse events for the ADONIS/EURIDIS trials. However, the information provided did not include a summary of SAEs.

It would have been useful to evaluate common non-serious adverse events (AEs). This type of AE may not be of importance from a clinical perspective but can be from the patient's perspective. This is of particular relevance considering that diarrhoea and nausea or vomiting is identified in the submission as main AEs associated with dronedarone (see Section 6.7 for further details).

Treatment discontinuations

Discontinuation events were defined as the reported number of treatment discontinuations of any cause within the timeframe of the study. The submission also presented data relating to treatment discontinuations that resulted from an adverse event. These are both important and very relevant outcomes. However comparisons between dronedarone and other trials within the submission may not be reliable due to potentially different criteria of discontinuations. For example, in some studies discontinuation due to lack of efficacy or an AE was measured, whilst in other studies incidence of AE occurrence was used as a proxy for treatment discontinuation. From the details provided in the submission and the systematic review, it is difficult to judge the reliability of the data given that the level of detail reported is poor for the comparator trials.

4.1.7 Describe and critique the statistical approach used

RCTs

The methods of analysis for the dronedarone trials are generally good. In most cases appropriate ITT analysis was used. A number of post-hoc analyses of additional outcomes and subgroups from the ATHENA trial were reported.

The methods of analysis used in the comparator trials cannot be commented upon as no details were supplied in the submission or the supporting systematic review. As the body of evidence contained a range of studies it can be anticipated that some trials e.g. SAFE-T trial (amiodarone vs sotalol vs non-active comparator), are well conducted, adequately powered RCT with appropriate analyses, whereas the analysis of much smaller trials may be less robust.

Direct and Indirect comparisons

The methods of the direct and indirect statistical (meta) analysis are not reported in the manufacturer's submission but are detailed in the supporting systematic review. However, it should be noted that the methods were not previously specified in the systematic review protocol.

As stated in the supporting systematic review, the following methodology was applied to the meta-analysis:

- [redacted]

[redacted]

MTC

No details of the methods used to perform the MTC analysis were included in the submission. The full report supplied on request by Sanofi-Aventis contained only limited details of the methods. Some further limited information was supplied by the company upon further request regarding issues of model convergence and a new methods section, which better explained the difference between the methods used and Bayesian methods utilising WinBUGs methods. However, as MTC analysis is usually conducted using Bayesian methods and implemented in WinBugs, the ERG ran an additional check on the analyses presented in the submission by rerunning one analysis using WinBUGs.

The ERG was unable to fully appraise the MTC because a copy of the full SAS program used by the manufacturer including the whole dataset used was not supplied. However, the manufacturer did supply, upon request, details of the SAS code and dataset for the outcome of all cause mortality.

When interpreting the results from any MTC it is crucial to understand the network of trials that were included. Although the data included in the final MTC analyses were supplied to the ERG, and the ERG were able to construct network diagrams for each outcome, not all trials comprising the data set for each outcome were included in the MTC. The methods document states that

[redacted]

The implication of this restriction was a substantial reduction in the number of studies entering the MTC compared to the direct and indirect analysis. For example, data

from 33 studies were available for consideration in the direct and indirect analysis for the outcome of all cause mortality, while only seven of these studies met the inclusion criteria for the MTC analysis. Furthermore, the filter used in the MTC analysis was not consistent across all outcomes. For the outcome of stroke, the restriction criteria was relaxed to consider studies where at least 50 subjects were randomised to either group and had at least one event. The ERG sought clarification with regard to the inconsistencies noted in the inclusion/exclusion criteria applied in the MTC analysis. The manufacturer's response to the ERG initial points for clarification provided further explanation, outlining that the MTC analysis was [redacted] While the ERG understands that the reason for exclusion of trials was due to methodological problems with convergence, the exact cause of not achieving convergence remains unclear. More importantly, it is unclear to what extent the exclusion of these trials impact on the relative effectiveness of the different comparators. Generally, a MTC analysis is used to increase the network of evidence available so that the relative effectiveness of the comparators can be simultaneously estimated by borrowing strength from the wider range of trials, but in the manufacturer's submission the network of evidence has been reduced considerably relative to that available for the direct and indirect analysis.

Details of the complete data set (before any studies were lost due to convergence problems) together with the *a priori* inclusion criteria were not provided to the ERG. From the limited information contained in the submission and the raw data tables, the inclusion criteria for the MTC differed from the standard meta-analysis. In addition to the exclusion of some trials from the MTC, some additional trials were included in some of the MTC outcome data sets. For some outcomes where there were not so many zero event trials in the data set, the number of trials included in the SAS MTC in the submission is greater. Therefore, there is uncertainty regarding the *a priori* inclusion criteria for these analyses and methods used to generate the results. The trials included for AF recurrence, discontinuations, discontinuations due to AEs and SAEs all included some that compared flecanide with propefonone. Given that these two drugs were treated as a single agent it is unclear how these trials were included

in the analysis. The ERG asked for clarification on the number of trials included in the meta-analyses and MTCs. However, inconsistencies were noted in the numbers provided and they did not appear to correspond to the numbers in the networks based on the data tables provided.

In general, the ERG considers that the methods used for the MTC analysis is likely to be reliable but the results from the synthesis should be treated with caution due to a lack of transparency regarding how trials were selected for inclusion in the analysis and omitted due to non-convergence of the model.

In order to check the results of the MTC, the ERG re-ran one of the analyses in WinBugs. The ERG selected all cause mortality as this is a key parameter in the economic model.

ERG MTC for all cause mortality

Re-running the model in Winbugs

A MTC using fixed effects and random effects was run in WinBUGs. The results were compared to the manufacturer's submission using the seven studies included in the manufacturer's MTC. The analysis used a logistic regression model that included correction factors for trials with zero events. The data set included in the analysis is shown in Table 2. These models are based on WinBUGS code published on the Bristol University MTC analysis webpage (https://www.bris.ac.uk/cobm/docs/RE%203-arm.odc).

Table 2: Dataset used by ERG to rerun MTC in Winbugs

Study #	Study label	Treatment	Number of events	Number of patients in study arm
1	A-COMET-II, 2006	Placebo	0	224
1	A-COMET-II, 2006	Sotalol	4	223
2	AFFIRM substudy, 2003	Sotalol	24	125
2	AFFIRM substudy, 2003	Amiodarone	15	131
71	ATHENA, 2009	Dronedarone	116	2301
71	ATHENA, 2009	Placebo	139	2327
72	DIONYSOS, 2009	Dronedarone	2	249
72	DIONYSOS, 2009	Amiodarone	5	255
21	EURIDIS/ADONIS, 2007	Placebo	3	409
21	EURIDIS/ADONIS, 2007	Dronedarone	8	828
55	SAFE-T, 2003	Placebo	3	137
55	SAFE-T, 2003	Amiodarone	13	267
55	SAFE-T, 2003	Sotalol	15	261
57	SOPAT, 2004	Placebo	0	251
57	SOPAT, 2004	Sotalol	2	264

The fixed effect had a slightly lower deviance information criterion (DIC) than the random effects model (77.25 vs 78.42). The DIC is used to compare models. A lower DIC represents a more efficient model. It combines a measure for how well the model fits the data with a measure for the complexity of the model in terms of the effective number of parameters. Therefore the results of the fixed effect model were used. The random effects results for comparison are presented in Appendix 2. The model fit was good as the residual deviance was less than the number of arms in each analysis. The results of this analysis were very similar to those generated by the SAS MTC (see Table 5, Section 4.2.2.1) confirming that the method used by the manufacturer does not appear to result in potential bias.

As stated earlier, there is some lack of clarity regarding how this small data set was arrived at. The data set included in submission stated that for the outcome of all cause mortality the MTC analysis trials were restricted to those comparing target pharmaceutical therapy either with an untreated control condition or an alternative target pharmaceutical, with at least 100 subjects per randomised group and at least 1 event in either group (n=7). Clarification requested by the ERG established that the inclusion criteria were not specified *a priori* but were dictated by a lack of convergence of the MTC model. It is still unclear what that full data set comprised.

The ERG were concerned at the exclusion of all zero event trials, particularly as no studies of either Class 1c agents, flecainide or propafenone, were included in the

final data set. In order to investigate the effect of including all relevant trials in the MTC analysis the ERG broadened the inclusion criteria to include all studies with 12 months duration or slightly more to be consistent with the studies included in the direct and indirect analyses. There were several studies that investigated class 1c drugs that met this criterion. The data set now comprised 19 trials (see appendix 2). However, the lack of events in any of the class 1c trial arms had the effect of preventing convergence in the WinBUGS model. Therefore, a continuity correction was required in order to perform this analysis. The continuity correction used was the Laplace correction, specified in the manufacturer's systematic review for use in the standard meta-analyses (adding 1 to the numerator for every arm for every trial with at least one arm with zero events and adding 2 to every denominator).

The results of the ERG MTC WinBugs analysis together with the SAS MTC results for all cause mortality are presented in Table 5, Section 4.2.2.1

4.1.8 Summary statement

The submission included all relevant dronedarone trials and, as far as it is possible to establish, all comparator trials. However, it is unclear whether all data were used where appropriate, nor whether all were analysed appropriate. This will be explored further in section.4.2.2.3

4.2 Summary of submitted evidence

4.2.1 Summary of results

The dronedarone RCTs demonstrate the beneficial effect of dronedarone on AF recurrence and ventricular rate during recurrence. However, when compared with amiodarone, dronedarone was shown to be less effective at preventing AF recurrence. Dronedarone is well tolerated but this has not been shown to outweigh its limited efficacy.

The manufacturer also demonstrates that dronedarone results in a statistically significant reduction in the composite endpoint of cardiovascular hospitalisation and

all-cause mortality and has an adverse event profile that is not significantly different to baseline therapy. The main adverse events identified with dronedarone were diarrhoea, nausea or vomiting, serum creatinine increase, rash, bradycardia and QT prolongation.

The effects of dronedarone on mortality have been explored in the ATHENA trial, which was a very large RCT. In the trial population dronedarone demonstrated beneficial effects on cardiovascular mortality and all-cause mortality (although the latter was not statistically significant). A sub group analysis of the same trial demonstrates a beneficial effect in reducing the incidence of stroke. However, the generalisability of these findings to all patients with persistent or paroxysmal AF should be made with caution, as the study population was selected to be of moderate to high risk of mortality.

All adverse events (i.e. serious and non-serious combined) of dronedarone versus non-active control were evaluated in DAFNE, EURIDIS, ADONIS, ERATO and ATHENA. The results of this analysis were in an FDA briefing document, which the company referenced in their submission. [11] No further details of how the analysis were performed were provided. In comparison to placebo, dronedarone was associated with a statistically significantly greater risk of the following common adverse events: rate and rhythm disorders (RR, 1.89, 95% CI: 1.39, 2.59), rashes, eruptions and exanthems (RR: 1.77, 95%CI: 1.23, 2.54), nausea and vomiting symptoms (RR: 1.61, 95%CI: 1.28, 2.03), diarrhoea (excl infective) (RR: 1.55, 95%CI 1.29, 1.86). As there are no details of how this analysis was performed the reliability of the above results is uncertain. Given the importance of data synthesis in informing the model, a more detailed summary of the different syntheses are discussed in section 4.2.2.

4.2.2 Critique of submitted evidence syntheses

The following sections discuss the evidence as presented in the submission of the efficacy of dronedarone when compared with relevant comparator AADs.

Results are presented for AF recurrence, all cause mortality, treatment discontinuation (any cause), treatment discontinuation (AEs), and stroke. Results are presented for all three types of analyses (where available): direct, indirect and MTC.

Supporting information was provided in appendices. When the results tables were checked by the ERG against the appendices discrepancies between how results were reported between and even within tables were identified. In the appendices a number of analyses were run: results were pooled using three different methods: Peto model, Peto model with continuity correction (thereby allowing trials with zero events in each arm to be incorporated in the analysis), and random effects model. Depending upon the outcome and the data available sometimes 12 month follow up data were pooled separately from all time points.

The ERG have identified which trials are included in each comparison synthesis and have attempted, from the limited information available in the submission and the supporting systematic review, to identify some of the potential problems and limitations of these analyses. To discuss each of these by individual analysis would be laborious. Furthermore, without checking the details of each individual trial included in each analysis we cannot be certain of the impact of these potential limitations. Therefore only a short summary is presented here and in the following sections.

Across all drugs the inclusion and exclusion criteria applied to trials are unclear and appear to be inconsistently applied. For example, it is unclear how strictly and how consistently the 12 month end point for the outcomes all cause mortality and AF recurrence is applied.

The populations included in the analyses are also questionable and inconsistent. Some analyses include patients with permanent AF; others include only paroxysmal patients. In particular the ATHENA trial, with its moderate to high risk of stroke elderly population, may well not be exchangeable with the other trials.

In addition to these questions of clinical heterogeneity, statistical heterogeneity was not reported in the submission for any meta-analysis. Examination of the Forest plots provided in the appendices and supporting systematic review indicated that some meta-analysis were subject to statistical heterogeneity. This should have been investigated further and commented upon in the submission.

The indirect comparisons using the Bucher method utilised the results from the standard meta-analyses and so any deficiencies in the latter fed through into the former.

The inclusion of trials in the MTC did not correspond with the trials in the standard meta-analyses. Much of this disparity might be due to trials being removed from the MTC to achieve convergence. However, the inclusion of trials ineligible for the standard meta-analyses indicates that the original inclusion criteria for the MTC was different from that for the standard review for each outcome measure.

The question of the exchangeability of the trials that were included in the MTC analyses was not investigated or discussed.

4.2.2.1. Comparison of Dronedarone with Comparators

As specified previously the relevant comparators considered in this submission were amiodarone, sotolol, class 1c agents and non-active control. The comparisons are discussed by outcome below.

AF recurrence

A summary of the results for all treatments versus a non-active and active control is presented in Table 3.

Table 3: **Comments on Table 6.9 of submission: Meta-analysis summary of comparison between treatments: Odds ratio (OR) for AF recurrence**

	Direct analysis		**MTC**
	Peto OR (95% CI) (in submission)*	**Peto OR (95% CI) at 12 months[I] (in appendix 5)***	**OR (95% CI) (in submission)*** **
Non-active control			
Dronedarone v control	[redacted]	[redacted]	[redacted]
Amiodarone v control	[redacted]	[redacted]	[redacted]
Class 1c v control**	[redacted]	[redacted]	[redacted]
Sotalol v control	[redacted]	[redacted]	[redacted]
	Direct Analysis	**Indirect analysis**	**MTC**
	Peto OR (95% CI)	**Peto OR (95% CI)**	**OR (95% CI)**
Head to head			
amiodarone v dronedarone	[redacted]	[redacted]	[redacted]
Class 1c v dronedarone**	[redacted]	[redacted]	[redacted]
sotalol v dronedarone	[redacted]	[redacted]	[redacted]

*OR lower than 1 describes a lower rate of AF recurrence for the comparator drug
**Class 1c agents include flecainide and propafenone combined.
*** SAS methods
NA = not available

[I] A continuity correction was not required for this outcome as at least one arm in each study had an event of AF recurrence

As discussed in Section 4.1.7 the robustness of the results of the direct, indirect and MTC comparisons are difficult to evaluate.

Particular points to note with regard to the synthesis for AF recurrence are:

- The results presented in Table 3 confirm that the results presented in the submission were all based on 12 month data.
- ATHENA data were correctly not included in any pooled analysis
- Inclusion and analysis of AF recurrence as survival data would have been more appropriate.
- Analysis of time to occurrence of AF would have been useful in comparing the active treatments.
- Some trials had inappropriate populations, including patients with permanent AF.
- Meta-analysis of sotalol vs non-active control was subject to statistical heterogeneity, which was not reported or explored.
- Details of the number of trials included in the MTC supplied to the ERG on request do not concur with the network constructed from the raw data file supplied.
- The MTC analysis omitted trials that had been included in the standard meta-analysis, but included other trials with a non-active control that were not included in the standard meta-analysis.

Despite these limitations the results of the various analyses for AF recurrence (Table 3) indicate the same overall finding: the results from the direct comparison between active and non-active control show that, whilst all drugs statistically significantly reduce AF recurrence, dronedarone appears to have the smallest effect size albeit with the smallest confidence intervals[m]. This agrees with the submission's conclusion that all active comparators (amiodarone, class 1c agents and sotalol) are potentially superior to dronedarone at reducing AF recurrences.

[m] Correction inserted after manufacturer identified a factual error

All-cause mortality

A summary of the results for all treatments versus a non-active and active control is presented in Table 4.

Particular points to note with regard to the synthesis on all-cause mortality are:

- The ATHENA trial dominates the meta-analysis of dronedarone trials (96.1% of the weight)
- There is significant clinical heterogeneity between ATHENA and EURIDIS/ADONIS (and probably other drug trials in the MTC).
- Use of continuity corrected data was inconsistently applied in the submission, which overestimates the risk of all-cause mortality for amiodarone, class 1c agents and sotalol.
- There were reservations regarding the methods of the MTC used in the submission. The MTC performed by the ERG, which included all the trials with 12 month data, including zero event trials, produced results closer to the indirect Bucher method than the MTC of the company submission, which did not include all the trials. These are presented in Table 5.

Both the indirect comparison and MTC recalculated by the ERG using continuity corrected data showed that there was no statistically significant difference in all-cause mortality between dronedarone, amiodarone and class 1c agents. Sotolol had the highest risk of all cause mortality compared with non-active control and was associated with a statistically significant higher risk of all-cause mortality than dronedarone. This contrasts with the findings reported in the submission which concluded that dronedarone was associated with a statistically significant lower risk of all-cause mortality for both sotalol and amiodarone, and that there was insufficient evidence to include class 1c agents in the MTC analysis.

All-cause mortality

Table 4: Comments on Table 6.10 of submission: Summary of comparison between treatments: Odds ratio for all-cause mortality

	Direct comparison			Mixed Treatment Comparison	
Non-active control	**Peto OR (95% CI) (in submission)***	**Peto OR at 12 months without continuity correction (in appendix)***	**Peto OR at 12 months with continuity correction (in appendix)***	**OR (95% CI) (in submission)***	**P value (in submission)**
Dronedarone v control	0.85 (0.66, 1.09)	0.85 (0.66, 1.09) 2 studies[n]	NA	0.85 (0.67,1.09)	0.165
Amiodarone v control	2.02 (0.70, 5.80)	2.02 (0.70, 5.80) 1 study [o]	1.41 (0.60,3.29) 5 studies[p]	2.73 (1.00, 7.41)	0.049
Class 1c v control**	0.68 (0.20, 2.31)	0.12 (0.00, 6.11) 1 study[q]	0.68 (0.20, 2.31) 6 studies[r] 1.03 (0.25, 4.15) 4 studies[s]	NA	
Sotalol v control	2.72 (1.16, 6.38)	2.72 (1.16, 6.38) 3 studies [t]	1.89 (0.93, 3.84) 8 studies[u]	4.52 (1.59, 11.70)	0.013
	Direct comparison	**Indirect comparison**		**Mixed Treatment Comparison**	
Head to head	**Peto OR (95% CI) (in submission)**	**Peto OR (95% CI) (in submission)**	**Peto OR (95% CI) calculated by ERG using direct Peto OR at 12 months with continuity correction**	**OR (95% CI) (in submission)**	**OR (95% CI Calculated by ERG**
Amiodarone v dronedarone	2.32 (0.52, 10.32)[1]	2.38 (0.80, 7.07)	1.66 (0.68, 4.03)	3.19 (1.16, 8.76)	1.55 (0.79, 3.13)
Class 1c v dronedarone**	NA	0.8 (0.23, 2.49)	1.21 (0.29, 5.05)[v]	NC	1.23 (0.42,3.49)
Sotalol v dronedarone	NA	3.20 (1.32, 7.78)	2.22 (1.05, 4.72)	5.05 (1.84, 13.87)	2.22 (1.15, 4.35)

NA = not available, *Odds ratio greater than 1 indicates a higher risk of mortality for the comparator , **Class 1c includes flecainide and propafenone combined.
NC= not calculated

[n] ATHENA & EURIDIS/ADONIS
[o] SAFE-T
[p] Boos, 2008 (time point is at 16 months), Channer 2004, Galperin 2001, Kochiadakis 2000, SAFE-T 2005
[q] Dogan 2004
[r] Bellandi 2001, Carunchio 1995, Dogan 2004 (15 months), Kochiadakis 2004b, Pritchett, 2003 (10 months), Van Gelder 1989
[s] Calculated by ERG. Excluded Dogan 2004 and Pritchett 2003
[t] Fetsch, 2004, SAFE-T 2005, SOPAT 2004
[u] Bellandi 2001, Benditt 1999, Carunchio 1995, Fetsch 2004, Kochiadakis 2000, Kochiadakis 2004b, SAFE-T 2005, SOPAT 2004
[v] Calculated by ERG using continuity corrected OR excluding Dogan 2004 and Pritchett 2003.

Table 5: Results of MTC for all-cause mortality rerun by ERG

	Submission MTC using SAS code		**ERG MTC using Winbugs but same data set as SAS MTC***		**ERG MTC Winbugs + all trials using CC*****	
Comparison	**OR**	**95% Cr I**	**OR**	**95% Cr I**	**OR**	**95% Cr I**
Sotalol vs NAC	4.52	(1.59, 11.70)	4.57	(1.82,14.54)	1.87	(1.01,3.57)
Amiodarone vs NAC	2.73	(1.00, 7.41)	2.87	(1.14,9.14)	1.30	(0.68,2.56)
Dronedarone vs NAC	0.85	(0.67, 1.10)	0.85	(0.67,1.09)	0.84	(0.66,1.07)
Class 1c vs NAC					1.03	(0.36,2.86)
Amiodarone vs sotalol			0.63	(0.37,1.04)	0.70	(0.43,1.13)
Dronedarone vs sotalol	0.20	(0.07, 0.54)	0.19	(0.06,0.48)	0.45	(0.23,0.87)
Class 1c vs sotalol					0.55	(0.19,1.50)
Dronedarone vs amiodarone	0.31	(0.11, 0.86)	0.30	(0.09,0.76)	0.65	(0.32,1.26)
Class 1c vs amiodarone					0.79	(0.26,2.26)
Class 1c vs dronedarone					1.23	(0.42,3.49)

* The MTC was rerun using the same data as used in NF report (Raw data supplied by Sanofi-Aventis)
** The MTC was rerun excluding trial A-COMET-II, 2006 because the study duration was less than 12 months
*** The MTC was rerun including all trials that reported 12 month data (from Submission Appendix 6). Zero event trials were included using the Laplace continuity correction Note: It was attempted to run the model including all trials without using the continuity correction, but the model could not converge.

Stroke

The presentation of the analysis relating to stroke was less detailed than any of the other outcomes. Only the results from the MTC were presented in the submission and its appendices. The results of the MTC analysis are presented here in Table 6.

Table 6: MTC results for stroke

	MTC (SAS) OR (95% CI)
Dronedarone vs non-active control	0.69 (0.57,0.84)
Amiodarone vs non-active control	0.89 (0.48,1.65)
Sotalol vs non-active control	0.80 (0.39,1.63)
Dronedarone vs Amiodarone	0.78 (0.41, 1.45)
Dronedarone vs Sotalol	0.87 (0.42, 1.78)

Particular points to note with regard to the synthesis of stroke are:

- The data included in the MTC (Appendix 8 of the submission) are derived from trials in which stroke was not a specified efficacy variable and therefore the data are unlikely to have been collected consistently across the trials and the trials underpowered for this rare outcome

- The number of strokes in the ATHENA data set did not match those found in published sources. Due to the low event rate, small differences could have a large effect on the results of the analysis. The ERG were unable to confirm the accuracy of the data from other trials.

- The most reliable data are derived for dronedarone from ATHENA. Post hoc analyses demonstrated a small but statistically significant beneficial effect for the entire ATHENA population (HR: 0.66, 95% CI: 0.46, 0.96, p=0.027) and a greater effect in the subgroup of patients with $CHADS_2$ score ≥ 2. [2]

- The data from the dronedarone versus placebo studies was heterogeneous in terms of study population, with the ATHENA study only including elderly patients at moderate to high risk of stroke.

The manufacturer's conclusion that dronedarone is associated with a statistically significant reduction in stroke compared to control is uncertain based on the MTC results. The relatively conservative hazard ratio for stroke derived from ATHENA (whole population rate) is likely to be more reliable estimate of the effect of dronedarone. However the rate of stroke from the ATHENA trial may not be generalisable to the wider AF population for whom dronedarone is indicated.

Cardiac events

No statistical analysis was performed for cardiac events in the company submission. Evidence for first hospitalisation due to CV events comes only from the ATHENA study. There was a statistically significant difference in the percentage of patients first hospitalised due to CV events (29.3% (dronedarone) versus 36.9% (placebo); HR: 0.74, 95%CI: 0.67, 0.82, p>0.001). The same trial provides data on cardiovascular mortality and arrhythmic mortality.

No comparable data were available for other AADs and therefore the relative benefit of dronedarone and other AADs in terms of this outcome cannot be determined.

Treatment discontinuations (any cause)

A summary of the results for all treatments versus a non-active and active control is presented in Table 7

Particular points to note with regard to the synthesis of treatment discontinuations for any cause are:

- Data from all three doses (800mg, 1200mg and 1600mg) of DAFNE were used.
- There was significant statistical heterogeneity in the dronedarone meta-analysis.
- The dronedarone meta-analysis was dominated by ATHENA.
- Some included trials in some analyses were of patients with permanent AF.

- In the analysis for sotalol, the SOPAT trial was included, which used the number of SAEs as a proxy for treatment discontinuations.[12] This could have either over or underestimated treatment discontinuations.

Withdrawals, including a combination of lack of efficacy and adverse events, [w]are very important in determining the efficacy of dronedarone compared to other drugs. This is because dronedarone is less effective than amiodarone, sotalol or class 1 c agents at reducing AF recurrence. However this might be counterbalanced by a better tolerability profile.

The meta-analysis presented in the submission and in Table 7 for each drug compared with non-active control suggests that sotalol has a beneficial effect, dronedarone has no effect, amiodarone and class 1c agents have a negative effect compared with no active treatment, however, only the class 1c effect is statistically significant.
Across all the treatments the trials included in the MTC did not correspond with those in the standard meta-analysis. In part this was due to not achieving convergence in the model but also a number of non-active control trials were missing.

The MTC presented for each drug compared with non-active control suggests that sotalol has a beneficial effect, and that dronedarone, amiodarone, and class 1c agents have no effect, with none of the results reaching statistical significance. Estimates for amiodarone vs. dronedarone are inconsistent across the different synthesis: from the direct analysis, treatment discontinuation is statistically significantly lower for amiodarone than dronedarone, whilst both the indirect analysis and MTC show that there is no statistically significant difference.

The meta-analysis and MTC results generate different conclusions, but the confidence (credibility) intervals all cross 1, indicating uncertainty around all the estimates. Overall the results indicate that there is no clear evidence that dronedarone is associated with a lower rate of discontinuations than other AADs. The manufacturer's conclusion that the analysis indicates a trend towards an increase in

[w] Correction inserted after manufacturer identified a factual error

discontinuation for active treatment compared with control is conservative given the evidence presented.

Table 7: Comments on Table 6.11 of submission : Meta-analysis summary of comparison between treatments: Odds ratio for treatment discontinuations

	Direct analysis			Mixed Treatment Comparison	
Non-active control	**Peto OR (95% CI) (in submission)***	**Peto OR without continuity correction (in appendix)***	**Peto OR with a continuity correction (in appendix)***	**OR (95% CI) (in submission)*****	**P value (in submission)**
Dronedarone v control	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Amiodarone v control	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Class 1c v control**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sotalol v control	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

	Direct Analysis	Indirect Analysis	Mixed Treatment Comparison
	Peto OR (95% CI)	**Peto OR (95% CI)**	**OR (95% CI)**
Head to head			
Amiodarone v dronedarone (1)	[redacted]	[redacted]	[redacted]
Sotalol v dronedarone	[redacted]	[redacted]	[redacted]
Class 1c v dronedarone**	[redacted]	[redacted]	[redacted]

NA= not available
*Odds ratio smaller than 1 indicates a benefit (lower discontinuation) for the comparator
**Class 1c includes flecainide and propafenone combined.
*** SAS method

Treatment discontinuations due to AEs

A summary of the results for all treatments versus a non-active and active control is presented in Table 8.

Particular points to note with regard to the synthesis of treatment discontinuations due to AEs are:

- The studies included in the meta-analysis for dronedarone were clinically heterogeneous: the DAFNE study had data relating to discontiuations at 800mg, 1200mg and 1600mg dronedarone, the ATHENA study has a moderate to high risk elderly population and dominates the analysis, and the ERATO study population has permanent AF.

- In the meta-analysis of sotolal vs control, the largest studies in the analysis used the number of serious adverse events as a proxy for the number of treatment discontinuations due to AEs. The inclusion of these results may over or underestimate the summary OR.

- Across all the drugs, three trials were included in the MTC which were not included in the standard meta-analysis.

- As for other outcomes some trials were omitted from the MTC.

The active drug versus non-active control direct comparison for all drugs indicate a significantly higher rate of withdrawals due to AEs for active treatment than non-active control. However the rate for amiodarone is particularly high. Both the indirect analysis and MTC results indicate that there is no significant difference in discontinuations between dronedarone vs sotalol and dronedarone vs class 1c agents due to AEs.

Direct and indirect evidence show that there is a statistically significant increase of discontinuing treatment due to AEs for amiodarone compared to dronedarone. The head to head MTC also suggests that discontinuation due to AEs is more likely for amiodarone compared to dronedarone.

The results in Table 8 demonstrate the inconsistencies in the reporting of the results in the submission. .The evidence shows that patients receiving active treatment are statistically significantly more likely to discontinue due to AEs than those receiving non-active control. The evidence does support the submission's conclusion that patients receiving amiodarone are more likely to discontinue due to AEs, but there is uncertainty around this. The results show disparity between the results from the direct and MTC comparisons and between results compared with non-active control and active control.

Table 8: Treatment discontinuations due to AEs

	Direct Comparison			Mixed Treatment Comparison	
	Peto OR (95% CI) (in submission)*	**Peto OR without continuity correction (in appendix)***	**Peto OR with continuity correction (in appendix)***	**OR (95% CI) (in submission)***	**P value (in submission)**
Non active control					
Dronedarone v control	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Amiodarone v control	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Class 1c v control**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sotalol v control	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Direct analysis	**Indirect analysis**		**Mixed treatment comparison (Submission)***	
	Peto OR (95% CI)	**Peto OR (95% CI)**		**OR (95% CI)**	
Head to head					
Amiodarone v dronedarone (1)	[redacted]	[redacted]		[redacted]	
Sotalol v dronedarone	[redacted]	[redacted]		[redacted]	
Class 1c v dronedarone**	[redacted]	[redacted]		[redacted]	

NA= not available
*OR smaller than 1 indicates a benefit (lower discontinuation) for the comparator.
**Class 1c includes flecainide and propafenone combined.
***SAS method

Serious Adverse Events (SAEs)

A summary of the results for all treatments versus a non-active and active control is presented in Table 9.

- The main concern with regard to the synthesis of SAEs was the omission of EURIDIS/ADONIS serious adverse event data from the analysis.
- The dronedarone meta-analysis was subject to statistically significant heterogeneity
- The dronedarone meta-analysis was subject to statistically significant heterogeneity: dose, (all three doses from DAFNE); population (permanent AF from ERATO) and moderate to high risk elderly patients (ATHENA).
- Other permanent AF patient studies were also included.
- Across all the drugs, one non-active control trial included in the MTC was not included in the standard meta-analysis.

The results of the direct, indirect, and MTC analyses support the submission's conclusion that dronedarone has the lowest odds of SAE for dronedarone compared to other drugs. However, the omission of data from the EURIDIS/ADONIS trial, which was a large good quality RCT with a population that reflects that of the licensed indication, means that there is uncertainty in this conclusion.

Again the large differences in the results generated by the different analyses indicate discrepancies between the data included in these analyses.

Serious Adverse Events (SAEs)

Table 9: Comments on Table 6.14 of submission: Meta-analysis summary of comparison between treatments: Odds ratio for SAEs

	Direct Comparison		Mixed Treatment Comparison	
	Peto OR (95% CI) (in submission)*	**Peto OR (95% CI) [x] (in appendix)***	**OR (95% CI) (in submission)***	**P value (in submission)**
Dronedarone v control	0.96 (0.84, 1.11)	0.96 (0.84, 1.11)[y] 3 studies	0.98 (0.74, 1.30)	0.886
Amiodarone v control	8.10 (2.36, 27.81)	8.10 (2.36, 27.81)[z] 1 study	1.71 (0.97, 3.01)	0.001
Class 1c v control**	2.77 (1.78, 4.30)	2.77 (1.78, 4.30)[aa] 6 studies	0.96 (0.53, 1.73)	0.77
Sotalol v control	1.38 (1.06, 1.81)	1.38 (1.06, 1.81)[bb] 6 studies	0.995 (0.63, 1.58)	0.011
	Direct analysis	**Indirect analysis**	**Mixed treatment comparison*** **	
	Peto OR (95% CI)	**Peto OR (95% CI)**	**OR (95% CI)**	
Head to head				
Amiodarone v dronedarone	1.45 (0.89, 2.35)[1]	8.44 (2.44, 29.17)	1.86 (0.76, 4.55)	
Sotalol v dronedarone	NA	1.44 (1.07, 1.93)	1.35 (0.72, 2.51)	
Class 1c v dronedarone**	NA	2.89 (1.82 4.58)	3.06 (0.82, 11.46)	

1 one trial only (DIONYSUS)

NA=not available

*OR higher than 1 describes a higher rate of SAE for the comparator

**Class 1c includes flecainide and propafenone combined

*** SAS method

[x] A continuity correction was not required for this outcome as at least one arm in each study had a SAE

[y] ATHENA, DAFNE, ERATO,

[z] Kochiadakis 2000

[aa] Cobbe 1995, Connolly 1989, Dan Nar FMSG 1991, ERAFT 2002, Stroobandt 1997, Van Gelder 1989

[bb] A-COMET-II 2006, Benditt 1999, Fetsch 2004, Kochiadakis 2000, Singh 1991, SOPAT 2004

4.2.3 Summary

The following features of the systematic review and of the analysis may increase the risk of bias of the treatment effect of dronedarone, both in relation to relevant outcomes and to comparators of interest:

Systematic review methodology:

- Unspecified study inclusion/ exclusion criteria for direct and indirect analyses used in the submission
- Lack of study quality assessment

Analysis:

- Different inclusion/exclusion criteria to identify studies for the direct head to head analysis, the indirect approach, and the MTC.
- Lack of exploration of clinical and statistical heterogeneity (and exchangeability regarding indirect and MTC analyses)
- Limited data sets included in the MTC
- Inclusion of studies of patients with permanent AF
- Inconsistent use of continuity corrected data
- Inconsistent use of time points

Despite these limitations, evidence from EURIDIS/ADONIS, which were good quality RCTs, show that dronedarone is statistically significantly more effective than placebo for reducing AF recurrence and and in reducing ventricular rate during recurrence of AF/AFL. However, consistent evidence from direct, indirect analyses and MTC shows that dronedarone is not as effective as amiodarone, sotalol or class 1c agents in reducing AF recurrence.

Data from the large ATHENA study has shown that in comparison to placebo, dronedarone results in a statistically significant reduction in the primary composite endpoint of time to first CV hospitalisation or death from any cause in patients with AF (although the primary endpoint was mainly driven by a reduction in time to first cardiovascular hospitalisation due to a significant reduction in hospitalisation for AF). However, due to ATHENA's moderate to high-risk elderly AF population, the generalisability of this evidence to inform the management of a lower risk and younger AF population remains uncertain.

There is evidence that dronedarone statistically significantly reduces all-cause mortality in patients to patients with multiple CV risk factors, corresponding to a $CHADS_2$ score ≥ 4.

However, $CHADS_2$ is a clinical prediction score for estimating the risk of stroke in patients with AF and its validity as a prediction score for CVD risk more generally and for all-cause mortality remains uncertain. Furthermore, the all-cause mortality effect for this subgroup was derived from a post-hoc analysis. In regard to all-cause mortality for the general AF population, the estimate of effect of dronedarone on all-cause mortality is uncertain, with confidence (credibility) intervals of the direct comparison and MTC of dronedarone versus non-active control all crossing 1.

The evidence for the effect of dronedarone on the risk of stroke is highly uncertain. Only a small number of studies reported this outcome, which was not a pre-specified in any of these studies. No significant difference was reported between dronedarone and amiodarone and sotalol based on the results from the MTC. However, there was reliable evidence from the ATHENA study that in moderate to high risk elderly patients, dronedarone is associated with a statistically significiant reduction in the risk of stroke.

It is uncertain whether dronedarone is associated with a lower rate of discontinuations than other AADs as the confidence (credibility) intervals of the direct comparison and MTC all cross one.

Patients receiving active treatment are statistically significantly more likely to discontinue due to AEs than those receiving non-active control. However, the disparity between results from the direct, indirect and MTC suggest that there is unreliability as to whether treatment discontinuation due to AE is statistically significantly greater for amiodarone compared to dronedarone.

The results of direct analysis, indirect analysis and the head to head MTC support the submission's conclusion that that dronedarone has the lowest odds of SAE for dronedarone compared to other drugs. However, the omission of data from the EURIDIS/ADONIS trial means that the result is unreliable.

In summary, dronedarone is indicated in adult clinically stable patients with a history of, or current non-permanent atrial fibrillation (AF) to prevent recurrence of AF or to lower ventricular rate. The anti-arrhythmic efficacy of dronedarone has been demonstrated but found to be moderate compared to other AADs. The rate controlling effects of dronedarone have not been compared with other AADs nor have they been compared with other rate controlling drugs. The wider benefits of dronedarone, such as reduction is hospitalisation due to cardiovascular causes, and reduction in risk of stroke, which are not part of its indication,

have been demonstrated in moderate to high risk elderly population. These benefits may not be seen in the wider AF population for whom dronedarone is indicated.

5 Economic Evaluation

5.1 Overview of manufacturer's economic evaluation

This section provides a structured critique of the economic evaluation reported in the manufacturer's submission to NICE.

The manufacturer's initial economic submission to NICE included (references in brackets refer to the manufacturer's submission):

1. A description of the systematic search strategy used to identify existing cost effectiveness studies for dronedarone and for other relevant comparators with full details in a separate appendix (p63 – 65, Appendix 10)
2. A report of the economic evaluation conducted by the manufacturer describing the technology, comparators, patient population including subgroups, model structure, inputs and assumptions and finally the base-case results and sensitivity analysis (p65 – 115, Figure 7.1 – 7.5, Tables 7.2 – 7.24).
3. An electronic copy of a Discrete Event Simulation (DES) model developed in SIMUL8 with an Excel front end allowing changes to input values.
4. A detailed series of appendices including full details of the search strategy, decision tree diagrams and full list of variables used in the model, details of the assumptions and statistical approaches employed for the risk equations and other major inputs and detailed results (Appendix 10 - 22).

Following a number of points of clarification raised by the ERG, a number of addenda were submitted by the manufacturer. These included:

1. A report summarising the existing cost-effectiveness evidence for the comparators.
2. A report on the quality of life data derived from the Euro Heart Survey.
3. Amended input and results tables.
4. A revised electronic copy of the SIMUL8 model.

This section focuses on the economic evidence submitted by the manufacturer. The submission is subject to a critical review on the basis of the manufacturer's report and by direct examination of the electronic version of the economic model. The critical appraisal is conducted with the aid of a checklist to assess the quality of economic evaluations and a narrative review to highlight key assumptions and possible limitations. These areas are then used to formulate the points for clarification raised by the ERG to the manufacturer. Section 6 presents a description of the additional analyses requested from the manufacturer and a critique of their re-submitted results, alongside additional work undertaken by the ERG to address any remaining uncertainties.

A summary of the manufacturer's approach and signpost to the relevant sections in the manufacturer's submission are reported in Table 10.

Table 10: Summary of the manufacturer's economic evaluation (and signpost to manufacturer's submission)

	Assumption	**Source / Justification**	**Signpost***
Model	Cost-utility analysis using Discrete Event Simulation (DES).	To capture the treatment complexities and the different morbidities associated with the AF patient.	Section 7.2.6, Figure 7.4, p. 77 - 78
States and events	The model included 4 states including death and 7 events relevant to the disease course.	Based on the disease course and the treatment pathways according to the UK clinical guidelines and clinical opinion.	Section 7.2.6, Figure 7.4, p.77 *Appendix 11*
Comparators	Dronedarone was compared in two positions: (i) vs. standard care alone in high risk patients, and (ii) vs. other AADs (sotalol, class 1c agents and amiodarone).	The comparator treatments were modelled in line with current UK clinical guidelines.	Section 7.2.3, p. 70-71, Fig. 7.2a and 7.2b
Sub groups	5 main subgroups were defined in line with current UK clinical guidelines for treating AF. Sensitivity analysis also considered additional sub groups based on age, $CHADS_2$ score and gender.	In accordance with NICE clinical guideline of AF management and validated by UK clinical experts.	Section 7.2.2, p. 67 - 69
Natural History	Based on DES model with 4 main states and 7 events. Movements between states and the rate of events were derived from ATHENA and other external sources.	Event risk equations from ATHENA used for AF recurrences, CHF, ACS, stroke and treatment discontinuation. Published literature and simulation approaches used to estimate all-cause mortality risk and risk of mortality following CHF and stroke.	Section 7.2.7 – 7.2.8, Table 7.2 – 7.10 *Appendices 14-15*
Treatment effectiveness	Treatment effectiveness data for dronedarone and the other AAD comparators were derived from the MTC and assumptions.	Systematic review and MTC used to estimate the relative treatment effects. Separate assumptions applied to class 1c for all-cause	Section 7.2.7.2, Table 7.5, p.92 *Appendices 4-*

		mortality and stroke due to lack of data in MTC.	*9*
Health related quality of life	Utility data for the main health states and events were derived from AFTER cohort of Euro Health Survey. The disutility of adverse events was calculated separately using a time-trade off approach (TTO).	EQ-5D data from a regression analysis of 3045 AF patients from the AFTER cohort. Adverse event disutility derived from TTO study of 127 members of the general public.	Section 7.2.8, page 95 – 99 *Appendices 16-17*
Adverse events	A total of 10 adverse events and their utility decrements included in the model. Separated into short-term (28-days) and long-term adverse events.	AEs data derived from a pooled data of 5 clinical trials for Dronedarone, a single trial for amiodarone and the SPC for sotalol and class1c.	Section 7.2.7.4, Table 7.6, p.93 – 94 *Appendix 17*
Resource utilisation and costs	Acquisition costs for drugs, treatment initialisation and monitoring, event and adverse events were included.	UK reference cost, published literature and clinical expert opinion. The acquisition cost of dronedarone is not yet finalised.	Section 7.2.9, p 100 – 106
Discount rates	A 3.5% discount rate was employed for both costs and health benefits.	According to the NICE guideline.	Section 7.2.10, p.106
Sensitivity analysis	Detailed univariate and probabilistic sensitivity analysis (PSA) undertaken	Main uncertain parameters are defined and all the variables included in the PSA are listed. Scatter plot and CEAC provided.	Section 7.3.2, Table 7.23, p.111 – 114 *Appendices 19-20*

* location in the manufacturer's submission

The search strategy was described in the manufacturer's submission to identify published cost-effectiveness studies for dronedarone and for the comparators (amiodarone, sotalol and class 1c). The manufacturer searched a variety of electronic databases including Medline, Medline (R) In Process, Embase, Health Economic Evaluation Database and the NHS Economic Evaluation Database (NHS EED). Detailed search strategies for Medline, Embase and Cochrane database are presented separately in Appendix 10 of the manufacturer's submission.

A total of 15 studies were considered to be eligible for the systematic review. However, all 15 studies evaluated the cost-effectiveness of the comparator treatment and no cost-effectiveness studies of dronedarone were identified. The manufacturer did not consider these studies to be directly relevant to the decision problem and no further details of these were reported in the main submission. An additional report on the comparator cost-effectiveness studies was made available on request to the ERG. This report was submitted in response to the ERG points for clarification and included a summary and critical appraisal of the methods and results of the identified economic evaluations and commentary on their relevance to decision making in England and Wales.

The manufacturer's submission evaluates the cost-effectiveness of dronedarone in accordance with the anticipated licensed indication i.e. '400mg bd for stable adult patient with a recent history of, or current non permanent atrial fibrillation (AF).' In the absence of a final license at the time of the initial submission, the manufacturer noted that 'this population is likely to exclude patient with NYHA CHF class IV and also class III CHF with a recent haemodynamic instability' (p66 manufacturer's submission)). A brief overview of the key assumptions used in the analysis, alongside a narrative description of the main approach used, is reported below. This is followed by a more detailed critique of the economic evaluation and its assumptions.

The manufacturer evaluated the cost-effectiveness of dronedarone in two separate positions:

(i) For patients with multiple CV risk factors (corresponding to a $CHADS_2 \geq 4$) on top of standard baseline therapy (including anti-coagulation and beta blockers in line with current UK clinical guidelines and referred to within these guidelines as 1st line treatment)

(ii) For patients when it is considered appropriate to introduce an AAD, as a first line alternative to current AADs. The manufacturer notes that this position is referred to within current UK guidelines as 2nd line treatment.

To establish the cost-effectiveness in these two separate clinical positions, dronedarone was compared with a range of alternative therapeutic options (including standard care, class 1c, sotalol and amiodarone). The choice of comparator was dependent upon both the clinical position, the clinical AF type and baseline risk factors in line with clinical guidelines and are outlined schematically in the manufacturer's submission in Figures 7.2a and 7.2b (pages 71-72 of the manufacturer's submission).

A model based cost utility analysis was carried out using DES. The model includes four main health states (normal sinus rhythm, permanent atrial fibrillation with uncontrolled symptoms, permanent atrial fibrillation with controlled symptoms and death). Movements between these main health states were driven by events. A total of seven events were included (AF recurrence, ACS, stroke, CHF, treatment discontinuation for any cause, AF symptoms change for permanent patient and death) to capture the range of possible clinical pathways for the patient population.

Patients were stratified in the model according to their clinical AF type and the baseline risk factors in line with UK guidelines. 5 main patient groups were considered:

1. Paroxysmal AF patients with no structural heart disease
2. Paroxysmal AF patients with coronary heart disease
3. Paroxysmal AF patients with LV dysfunction
4. Persistent AF patients with no structural heart disease
5. Persistent AF patients with structural heart disease

In addition to these main groups of patients, a range of additional subgroup analyses were also performed in the economic evaluation including: age, $CHADS_2$ score and gender. The choice of subgroups was considered by the manufacturer to be the most clinically relevant groups to be analysed in accordance with the approach employed in current UK clinical guidelines.

A brief overview of the key assumptions used in the cost-effectiveness is reported below. This is followed by a more detailed critique of the economic evaluation and its assumptions.

- The placebo treatment arm of ATHENA is assumed to be representative of standard therapy in the UK and hence is used as the basis for informing the majority of baseline event rates applied in the model. However, given the low mortality rate observed in ATHENA, baseline all-cause mortality is estimated using external sources based on UK life table data adjusted by $CHADS_2$ score.

- The results from the MTC are used to inform the relative effectiveness estimates applied to each AAD for the reduction of AF recurrence and for 3 other outcomes (all-cause mortality, stroke and treatment discontinuation).

- Due to lack of data in the MTC for class 1c for mortality and stroke, separate assumptions are employed. For all-cause mortality, no difference is assumed between dronedarone and class 1c agents. However, for stroke it is assumed that class 1c agents have no effect compared to standard care alone for stroke.

- The relative treatment effect estimates of AADs are assumed to be constant across all relevant sub-groups, with the exception of patients with multiple CV risk factors (corresponding to a $CHADS_2 \geq 4$). For this particular group the subgroup relative effectiveness estimate for all-cause mortality is employed. Differences between subgroups are therefore modelled primarily by applying different *baseline* event rates.

- Treatment discontinuation is assumed to be independent of the events considered in the model.

- Adverse events are also modelled independently to the main health states and events considered in the model and their impact are modelled by applying cost and utility decrements.

A comprehensive list of assumptions is also provided by the manufacturer in pages 78-85 of their submission.

5.1.1 Natural history

DES is used to model the natural history for the separate patient groups considered. DES is a form of patient level simulation. That is, rather than following an entire cohort through a model by assigning proportions to different states, DES models the pathway of an individual patient. In DES models individual patients are assigned atrributes (e.g. age, sex, duration of disease, prior events etc). These values are defined at the start of the simulation and may be updated during the simulation itself. Within DES an event is defined as anything that can happen during the simulation and these differ from the conventional transitions applied in a Markov model because the event need not imply a change in the patient's health state. The rates at which these events occur in the model is dependent on the patient attributes and hence can change over time as appropriate. Time is modelled more flexibly using DES compared with Markov models which employ fixed cycle lengths Within DES models patients can experience an event at any discrete period after a previous event. Following the occurrence of an event, a DES model considers the type and timing of the next event compared with a Markov process which considers what events are occurring at a regular interval.

The use of DES is considered by the manufacturer to provide a flexible modelling approach to capture the complex clinical outcomes of the disease, the associated co-morbidities and variable treatment pathways. The events associated with atrial fibrillation alter the long term risks of subsequent events and hence it is important to keep track of the history of patients in the model. The manufacturer argues that the DES confers significant advantages over Markov models in terms of appropriately reflecting both the disease and the impact of patient history.

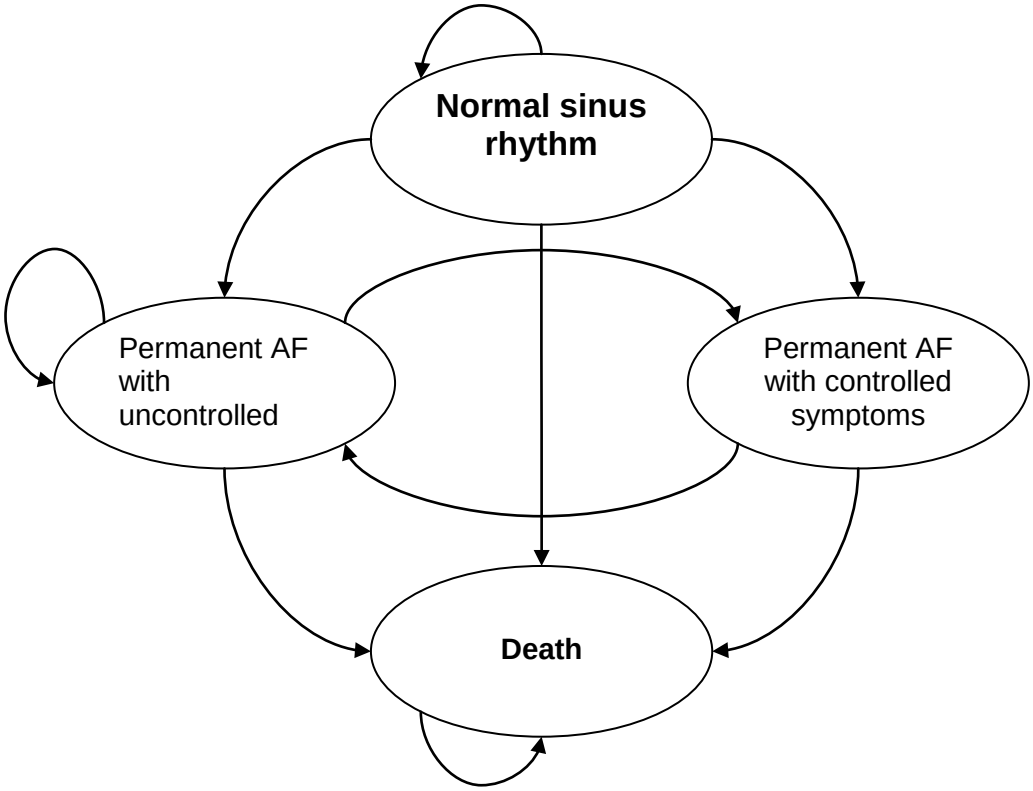
The states and events included in the structure of the model should be consistent with the clinical course and the relevant management of the particular disease. The manufacturer states that the model is constructed in accordance with the UK guidelines for AF and that this has also been validated by the UK clinicians.

A schematic of the model is reported in Figure 4. It is assumed that all patients are in sinus ryhthm before entering the model (Fig. 7.4, MS p. 78). A total of 4 health states are included:

1. Normal sinus rhythm (NSR),
2. Permanent AF with uncontrolled symptoms,
3. Permanent AF with controlled symptoms and
4. Death.

Movement of patients between these states is driven by events. There are 7 events included in the model: (i) AF recurrence, (ii) ACS, (iii) stroke, (iv) CHF, (v) treatment discontinuation of any cause, (vi) AF symptoms change for permanent patients and (vii) death. Each of the 7 events may have different clinical outcomes (except death) and different probabilities. For example, AF recurrence is an event that may have at least 3 major clinical outcomes with its subsequent treatment success or failure. A patient who experiences an AF recurrence may have a spontaneous return to NSR or need electrical or pharmacological cardioversion or may develop CV events. Within-event decision trees are used to estimate the probabilities of which health state the patient remains in until the next event and the associated cost and quality of life outcomes.

Figure 4: Schematic of the main health states in the model (MS p.78)



The DES model estimates costs and outcomes over a patient's lifetime. The model uses the placebo arm of the ATHENA trial to inform the baseline risk of different events (AF recurrences, CHF, ACS, stroke and treatment discontinuation). Survival analysis based on the time to event data for each event were used to generate risk equations which included the following covariates: age, gender and baseline SHD, CAD and $CHADS_2$ score. These covariates were used to reflect the different subgroups considered in UK clinical guidelines. Statistical goodness of fit tests were employed to determine the appropriate curve fit for these equations.

Given the low number of deaths reported to have occurred during the ATHENA trial, the manufacturer considered that the use of risk equations, from the ATHENA data, for extrapolation purposes for this outcome would over-estimate the remaining life years. Instead, the manufacturer estimated all-cause mortality using UK all cause mortality data from life tables adjusted for $CHADS_2$ score. Simulation approaches were then used to estimate a risk equation based on age and sex. Mortality following a stroke and CHF were also based on external sources given the small number of overall events in ATHENA. For both of these outcomes, published data were used to estimate the mortality rates after these events. The approaches are summarised in Table 11. Full details are reported in Appendices 14 and 15 in the manufacturer's submission.

Table 11: Sources and adjustments used for key events

Events in the model	Sources	Approach
AF Recurrence, CHF, ACS, Stroke, All cause discontinuation	ATHENA	Event risk equations derived directly from patient level data from ATHENA. Covariates include age, sex and baseline SHD, CAD and $CHADS_2$ score.
All cause mortality	UK life tables adjusted for $CHADS_2$ score	Event risk equations derived using simulation methods adjusting for age and sex.
Mortality – following stroke and CHF	Published literature	Published literature adjusting for age, sex (CHF only) and time since event.

The risk equations for the events derived from the ATHENA trial are dependent on the patient's baseline $CHADS_2$ score. The base-case analysis assumes the same $CHADS_2$ distribution as in the ATHENA trial. Sensitivity analysis was also carried out by using $CHADS_2$ distribution from non-trial sources which the manufacturer considered may be more reflective of 'real world' data. Two separate sources were used within the sensitivity analysis: (i) an international observational study of 5,600 AF patients (RECORD-AF) and (ii) a GPRD dataset of UK AF patients (n=55,412). The distribution of $CHADS_2$ score across the different studies is presented in Table 12.

Table 12: Baseline $CHADS_2$ score distributions used in the model (MS p.90)

$CHADS_2$ Score	ATHENA (base case) n=4628	% of population	[redacted]
0	3%	[redacted]	[redacted]
1	32%	[redacted]	[redacted]
2	36%	[redacted]	[redacted]
3	18%	[redacted]	[redacted]
4	8%	[redacted]	[redacted]
5	3%	[redacted]	[redacted]
6	1%	[redacted]	[redacted]

5.1.2 Treatment effectiveness within the submission

The relative effectiveness of dronedarone and the comparator AADs were estimated using the MTC results described in section 4.2.2.1. Due to lack of effectiveness evidence in the MTC for class 1c agents in mortality and stroke prevention, it is assumed in the model that it has similar effect of dronedarone on mortality and no effect on stroke. The odds ratios applied in the DES model are reported in Table 13. No treatment effect was assumed for either dronedarone or any comparator AAD for CHF and ACS.

Table 13: Odds ratios (95% CI) relative to control treatment reported in the MTC (updated values in response to the ERG's points for clarification)

Treatment	Parameter			
	All-cause mortality	AF recurrence	Treatment discontinuation	Stroke
Dronedarone	0.85** (0.67- 1.09)	[redacted]	[redacted]	0.69 (0.57 – 0.84)
Amiodarone	2.73 (1.00 – 7.41)	[redacted]	[redacted]	0.89 (0.48 – 1.65)
Sotalol	4.52 (1.59 – 11.70)	[redacted]	[redacted]	0.80 (0.39 – 1.63)
Class 1c	1.00*	[redacted]	[redacted]	1.00*

*No evidence available so assumed to be = 1.00

**Odds ratio differs by subgroup: OR = 0.53 for the comparison of dronedarone versus standard therapy as first line treatment; OR = 0.85 for the comparison of dronedarone versus sotalol or

amiodarone as first line anti-arrhythmics; OR = 1.0 for the comparison of dronedarone versus class 1c as first line anti-arrhythmic.

A total of 10 individual adverse events were considered and their short term and life time effects on costs and utility decrement is incorporated in the model. These adverse events were modelled independently from the main states and events considered. The probability of each adverse event was derived from a pooled analysis of 5 RCTs for dronedarone and a single RCT for amiodarone (DIONYSOS). Data for sotalol and class 1c were derived directly from their respective Summary of Product Characteristics (SPCs). The adverse event rates applied in the model are reported in Table 7.6 (p94) of the manufacturer's submission.

5.1.3 Health related quality of life

The manufacturer reported that quality of life had not been collected within any of the dronedarone clinical trials. The utility estimates applied in the model were therefore derived from external data sources. Separate sources and approaches were employed to estimate utility values applied to the main health states or events and to adverse events.

Utility values for the health states and main events were estimated using EQ-5D data from the AFTER cohort. The AFTER cohort is a part of Euro Heart Survey and a separate report was made available to the ERG on request. A total of 3045 patients with AF had complete data on survival status and EQ-5D. Regression approaches were used to derive the coefficients reported in Table 14 (also reported in Table 7.7, MS p.96). The utility decrement for AF symptoms was applied for the duration of the event. The utility decrements for stroke and CHF were applied for the remainder of a patient's lifetime and the decrement for ACS was applied for a year.

Table 14: Health state and event utility weights use in the model.

Factor	Coefficient	SD
Constant	[redacted]	[redacted]
Demographic characteristics		
Age (years)	[redacted]	[redacted]
Male sex		
Health states/events		
AF symptoms	[redacted]	[redacted]
Stroke*		
CHF*		
ACS**		

* Applied for the life-time of the patient post event
** Applied for 1 year after the event

A separate literature review was undertaken to identify utility decrements for the adverse events of AF related treatment options (MS Appendix 16). In the absence of any suitable published evidence, a separate study was undertaken by the manufacturer involving 127 members of the general public using time-trade off (TTO) approaches. A report on this study is included in Appendix 17 of the manufacturer's submission. The utility decrements for treatment related adverse events and the overall weighted utility loss for each AAD are reported on pages 97-99 of the manufacturer's submission. A summary of the weighted utility loss by treatment is provided in Table 15. Adverse events related to pulmonary and thyroid dysfunctions were considered as long term adverse events and causing continuous utility loss while all other adverse events were assumed to incur a 28 day utility decrement.

Table 15: Weighted utility decrement for adverse events

	Weighted utility decrement	
Treatment	**Permanent utility loss**	**One-off utility loss**
Dronedarone	0.0012	0.0014
Amiodarone	0.0052	0.0014
Sotalol	0.0021	0.0018
Class 1c	0.0017	0.0021

5.1.4 Resources and costs

Resource use and costs were assigned to the following broad areas: (i) drug acquisition, initiation and monitoring costs, (ii) health states and events costs and (iii) adverse event costs.

Drug acquisition, initiation and monitoring costs

The acquisition costs for amiodarone, sotalol and class 1c agents were derived from the British National Formulary (vol. 57) using recommended doses from the SPC for each intervention. The daily costs calculated according to the recommended dose of each intervention are reported in Table 16. In the absence of a finalised price for dronedarone, a daily cost of £ 2.30 was assumed in the base case analysis. A range of £2.20 - £2.50 for the daily cost of dronedarone is considered in the sensitivity analysis.

Table 16: Drug doses and costs used in the economic model (MS p.100)

Treatment	Dose (mg)	Pack cost	Tablets per pack	Tablet size (mg)	Daily cost

Dronedarone	400 b.i.d	£66.00	60 (30 day)	400	£2.30
Amiodarone	200	£1.42	28	200	£0.05
Sotalol	320	£2.21	28	160	£0.16
Class 1c*	n/a	n/a	n/a	n/a	£0.25

+ This is the assumed base case cost. The final cost will be confirmed before launch.
* Although different doses and pack sizes are used for Flecainide and Propafenone, the daily cost work out the same.

The manufacturer's submission assumes that dronedarone can be initiated by a specialist during an outpatient visit. Patients also require an additional serum creatinine test after 7 days that is assumed to be undertaken within a routine GP visit. The initiation cost of dronedarone is estimated as £213 (Reference cost: Consultant Led First Attendance Outpatient Face to Face code 320 and additional cost of GP visit and creatinine test). In contrast, the manufacturer assumes that all other AADS (sotalol, amiodarone and class 1c) require an inpatient hospitalisation for initiation at a cost of £249 (Reference cost: Observation Wards, Code: VEB07I).

The manufacturer's submission reports that there are no requirements for monitoring with dronedarone after the initial initiation follow-up and hence the manufacturer assumes zero monitoring costs for dronedarone. Additional monitoring costs were applied every 6 months for amiodarone, sotalol and class 1c. For amiodarone it was assumed that patients need to visit their GP every 6 months for a thyroid function test (TFT), liver function test (LFT), digoxin level test and an electrolyte test. For sotalol and class 1c it was assumed that patients need to visit their GP for an electrolyte test and an ECG every 6 months. The monitoring costs applied every 6 months were £51.91 for amiodarone and £75.71 for sotalol and class1c. Costs of all the tests were taken from reference costs (UK reference costs : DAP841 [TDAPS] – Biochemistry, UK reference costs : DA01 ECG [12 Lead], UK reference costs : DAP840 [TDAPS] and the costs of a GP visit were derived from the literature.

Health states and events costs

A range of published literature sources and reference costs were used to estimate the costs for different events included in the model. These are described in detail in pages 102-103 of the manufacturer's submission. The majority of events were assumed to incur a one-off cost (e.g. ablation, ACS, AF, cardioversion) whereas stroke and CHF were assumed to incur both a one-off cost (reflecting the initial acute treatment) and an ongoing daily cost.

Adverse events costs

The costs of adverse events were based on the event rate for each AAD and the proportion of these events which were considered serious enough to require hospitalisation. Expert

clinical opinion was used to estimate the proportion of events requiring hospitalisation and the remainder were assumed to require an outpatient consultant visit. NHS reference costs were applied to these estimates to derive treatment specific adverse event costs. These are summarised in Table 17 and are applied as a one-off cost at the time of treatment initiation. The adverse events with life-time effects (pulmonary, hyperthyroidism and hypothyroidism) were assumed to require an additional GP visit every 6 months.

Table 17: Adverse event costs for each AAD (updated values in response to the ERG's points for clarification)

AAD	Cost
Dronedarone	£106.28
Amiodarone	£178.92
Sotalol	£135.77
Class 1c	£205.09

5.1.5 Discounting

The economic evaluation by the manufacturer used 3.5% discount rate for both cost and health benefits.

5.1.6 Sensitivity analysis

Both univariate and probabilistic sensitivity analyses (PSA) were undertaken by the manufacturer. A full list of parameters (and their associated distributions) included in the PSA are reported in Appendix 19 of the manufacturer's submission. The main sources of parameter uncertainty considered in the PSA were: (i) the time to event equations, (ii) treatment effect estimates, (iii) the increase risk of all-cause mortality due to $CHADS_2$, (iv) post-stoke mortality, (v) utilities assigned to the main events and (vi) costs (AF hospitalisation, ablation and initiation costs).

A detailed series of univariate sensitivity analyses were also undertaken to explore a range of alternative scenarios. These are summarised in Table 18 and are reported in full in Appendix 20 of the manufacturer's submission.

Table 18: Summary of the scenarios and assumptions included in the sensitivity analysis.

Scenarios considered	Assumptions / Analysis
Alternative acquisition price of dronedarone	Varying the acquisition price between £2.20 and £2.50 per day.

Subgroup analysis by $CHADS_2$ score (0 – 6)	Separate cohorts of patients were considered with $CHADS_2$ score 0 – 6.
Alternative sources for the distribution of $CHADS_2$ score	$CHADS_2$ distribution from 2 different sources (RECORD AF, GPRD) were considered.
Best fit curve based on AIC criteria	Best fit and 2nd best fit curve compared.
Time horizon	The impact of alternative time horizons was explored (1 year to life time).
Varying mortality benefit	Separate scenarios were considered to explore the maximum and minimum benefit of dronedarone. For mortality benefits the lower limit of 95% CI for comparators and upper limit for dronedarone were used to model the minimum benefit and vice versa for the maximum benefit.
Varying starting age of the patient	The impact of assuming a starting age at 65 years (compared to 72 years in the base-case analysis) was considered.
Gender	Separate male and female cohorts were analysed.

Finally, tornado diagrams were used to assess the robustness of the base-case ICER results to particular groups of parameters. These analyses applied the lower and upper end of the confidence intervals or varied the input by +/- 20% where no range existed. The groups of parameters considered were: mortality treatment effect, stroke treatment effect, treatment discontinuation, adverse event rate (+/-20%), cost excluding dronedarone acquisition price (+/-20%), cost of dronedarone acquisition and the utilities (+/-20%).

5.1.7 Model Validation

The manufacturer reported that the analysis and assumptions employed had been validated with both UK and international clinicians. In addition, the manufacturer reported that the outputs from the model had been compared against the results from the ATHENA study to demonstrate the face validity of the results.

5.2 Critique of the manufacturer's economic evaluation

The ERG has considered the methods applied in the manufacturer's economic evaluation in the context of a detailed checklist reported in Appendix 3 which is used for quality assessing decision analytic models.[13] Table 19 compares the manufacturer's submission to that of the NICE reference case.[14]

Table 19: NICE reference case checklist

Attribute	Reference Case	Included in submission	Comment on whether de-novo evaluation meets requirements of NICE reference case
Comparator(s)	Alternative therapies including those routinely used in NHS	Yes	
Perspective - costs	NHS and PSS	Yes	NHS and PSS costs have been taken into account.
Perspective - benefits	All health effects on individuals	Yes	QALY benefits to treated individuals were considered.
Time horizon	Sufficient to capture differences in costs and outcomes	Yes	The economic model has a lifetime time horizon. Alternative time horizons are also explored.
Synthesis of evidence	Systematic review	Yes	Systematic review and MTC used. However, due to problems with convergence a restricted set of studies were considered in the MTC. Also, in the absence of data for particular outcomes, assumptions were employed.
Outcome measure	QALYs	Yes	
Health states for QALY measurement	Described using a standardised and validated instrument	Yes	The main health states and events were derived using EQ-5D data. Time-trade off (TTO) was used to estimate the utility decrement of adverse events.
Benefit valuation	Time Trade Off or Standard Gamble	Yes	TTO
Source of preference data	Sample of public	Yes	Societal tariffs from EQ-5D and sample of general public for the TTO study.
Discount rate	Health benefits and costs	Yes	Benefits and costs have both been discounted at 3.5%.
Equity	No special weighting	Yes	No special weighting was undertaken.
Sensitivity analysis	Probabilistic sensitivity analysis	Yes	Probabilistic sensitivity analysis was undertaken but the parameters are varied lower and upper bound of the confidence interval and a (+/-) 20% range where no interval is available. Results are presented graphically using cost-effectiveness acceptability curves (CEACs). Mortality benefit sensitivity analysis is presented separately.

A critical review of the methods used in the manufacturer's economic evaluation has been undertaken. The checklists have been used to identify key issues together with close scrutiny of the electronic model. However, it should be noted that a comprehensive validation by the ERG of the Simul8 model has not been possible due to the complexities of the coding and delays with the receipt of an updated model following the initial points for clarification. The Simul8 coding received comprised over 60 different subroutines of code, which made a detailed line by line validation impossible within the time constraints of an STA. However, the ERG has undertaken a series of detailed checks of the Excel front end and the Simul8 code to ensure that the general methods outlined were followed. The ERG was assisted in the validation process by an external researcher experienced in the use of SIMUL8 (Dr Matt Stevenson from ScHARR). In addition, further validation was undertaken to ensure that the Simul8 coding was correctly reading the input data from the Excel front end.

5.2.1 Comparators

The manufacturer presents separate cost-effectiveness results for the following main comparisons:

- Dronedarone plus standard therapy (i.e. beta-blockers and anticoagulants) versus standard therapy alone as first line treatment in patients with multiple CV risk factors corresponding to $CHADS_2 \geq 4$.
- Dronedarone versus sotalol as first line anti-arrhymic in patients with paroxysmal or persistent AF, with or without CAD, and who do not have LVD.
- Dronedarone versus class 1c as first line anti-arrhymic in patients with paroxysmal or persistent AF with no SHD.
- Dronedarone versus amiodarone as first line anti-arrhymic in patients with paroxysmal AF and LVD or persistent AF with SHD.

The comparisons made by the manufacturer raise a number of important issues. The first of these relates to the use of dronedarone as a first line treatment in patients with multiple CV risk factors (corresponding to $CHADS_2 \geq 4$) on top of standard therapy. For this analysis the comparator is standard baseline therapy alone. While a comparison of dronedarone versus placebo, as an adjunctive treatment to standard therapy, reflects the randomisation in ATHENA, the ERG has concerns about how the results of ATHENA have been interpreted by the manufacturer in their

submission. In particular, the manufacturer's assertion that dronedarone is an appropriate as first line therapy in high risk patients is based on a post hoc analysis of a subgroup of patients of ATHENA categorised by their $CHADS_2$ score. [redacted] While the results show mortality benefits for the higher risk group, ATHENA was not set up to directly collect, examine or address any questions relating to $CHADS_2$ subgroups. Furthermore, the results of ATHENA do not indicate whether any of the effects are dronedarone-specific nor do they indicate whether another anti-arrhythmic agent could have achieved the same or greater benefits. The ERG is unclear how the pharmacological mechanism of dronedarone acts differently from that of any other AAD for dronedarone to be considered as first line treatment against no other relevant alternatives.

The ERG also considers that the use of dronedarone as a first line treatment in the high risk sub-population may not be in accordance with its licence. The agreed indication in the latest draft of the SPC for dronedarone (September 2009) states:

"MULTAQ is indicated in adult clinically stable patients with a history of, or current non-permanent atrial fibrillation (AF) to prevent recurrence of AF or to lower ventricular rate."

As the licence does not indicate the use of dronedarone for preventing mortality or other CVD events, the ERG considers that the manufacturer's consideration of dronedarone in this position may not be within its licensed indication. Existing NICE clinical guidance for other AADs clearly place their use where symptomatic suppression of AF is not achieved with standard therapy including beta blockers. Consequently, the ERG does not consider that the approach to modelling dronedarone as a first line treatment in high risk patients reflects how existing AADs are currently used in the NHS.

A second issue relates to the choice of comparators in specific subgroup populations. The manufacturer evaluates the incremental cost-effectiveness of dronedarone via a series of pairwise comparisons. While this approach may be appropriate in the populations where there is only one relevant comparator due to treatment contraindications, the approach does not directly address the full decision problem where there are two or more relevant comparators, since this would require a

simultaneous assessment of all relevant treatment options. For example, in the existing NICE guidance, the treatment pathway for patients with paroxysmal or persistent AF with no structural heart disease who have failed to achieve suppression of AF with standard beta blockers includes first line AAD therapy consisting of sotalol or class 1c agents. Therefore the correct comparators for dronedarone in this population are sotalol and class 1c agents. The manufacturer considers these comparators but compares each agent (sotalol or class 1c) to dronedarone in a pairwise comparison. The correct approach requires the mean lifetime costs and QALYs of all the relevant strategies to be compared simultaneously and their cost-effectiveness assessed, estimating ICERs as appropriate using standard decision rules (i.e. establishing whether particular treatments are ruled out on the grounds of dominance or extended dominance). [15] Consequently, the ICERs, as they are presented by the manufacturer, do not necessarily reflect the correct estimate of the ICER for dronedarone. This can be observed in the manufacturer's results where the estimate of mean lifetime costs and QALYs for dronedarone differs in value in the individual pairwise comparisons when presented for the same population.

Finally, it is worth noting that the manufacturer does not explicitly consider the cost-effectiveness of alternative sequences of treatment. That is, the manufacturer only considers the cost-effectiveness of dronedarone at specific points within the treatment pathway, thus precluding a full assessment of the cost-effectiveness of using dronedarone at different points of the pathway. For example, for paroxysmal patients with CAD, the manufacturer compares the incremental cost-effectiveness of sotalol and dronedarone as alternative 1st line AADs. Patients who withdraw from these treatments are then assumed to receive amiodarone as a 2nd line AAD. However, no assessment is made of the potential cost-effectiveness of alternative management approaches which could involve using dronedarone as an alternative 2nd line AAD to amiodarone. This approach would require additional strategies to be included in the model (e.g. a strategy of sotalol followed by dronedarone or amiodarone) and their cost-effectiveness to be assessed against the strategies currently included. As a result, it is not possible to establish based on the current results whether dronedarone might be more cost-effective when used at later points in the treatment pathway.

5.2.2 Treatment effectiveness

The ERG has previously discussed (Section 4.2.2) a number of issues and concerns related to the synthesis of evidence presented by the manufacturer. The ERG has identified several specific issues which relate to the subsequent use of this data in the economic model. The key issues identified by the ERG include:

- The inclusion/exclusion criteria applied by the manufacturer in selecting studies for the synthesis;
- The appropriateness of using the treatment effects estimated from the MTC in the economic model;
- The assumption that absence of evidence implies absence of effect.

Each of the areas is considered in more detail below, outlining the key assumptions and the potential uncertainties surrounding them.

The inclusion/exclusion criteria applied by the manufacturer in selecting studies for the synthesis

The ERG is concerned about the different inclusion/exclusion criteria employed to identify studies for the 3 types of synthesis: direct, indirect, and MTC, leading to each type of analysis comprising a different set of studies for each outcome analysed. The ERG is particularly concerned about the reduced number of studies which entered the MTC analysis, given that the results of the MTC are subsequently employed in the economic model. For the direct and indirect approaches, a broad range of studies incorporating all available evidence identified as part of a wider systematic search (see Section 4.1.3) were considered. In contrast, a filter was applied to the studies which entered the MTC analysis due to methodological issues with not achieving convergence. This filter restricted trials entering the analysis to those which compared target pharmaceutical therapy either with an untreated control group or an alternative target pharmaceutical with at least 100 subjects per randomised group and at least one event in either group. The consequence of this restriction was a substantial reduction in the number of studies entering the MTC compared to the direct and indirect analysis. For example, data from 33 studies were available for consideration in the direct and indirect analysis for the outcome of all cause mortality, while only 7 of these studies met the inclusion criteria for the MTC analysis. Furthermore, the filter used in the MTC analysis was not consistent across all outcomes. For the outcome of stroke, the restriction criteria was relaxed to consider

studies where at least 50 subjects were randomised to either group and had at least one event.

While the ERG understands that the reason for exclusion of trials was due to methodological problems with convergence, the exact cause of not achieving convergence remains unclear. More importantly, it is unclear to what extent the exclusion of these trials impact on the relative effectiveness of the different comparators. Generally, a MTC analysis is used to increase the network of evidence available so that the relative effectiveness of the comparators can be simultaneously estimated by borrowing strength from the wider range of trials, but in the manufacturer's submission the network of evidence has been reduced considerably relative to that available for the direct and indirect analysis.

The merit of using the treatment effects estimated from the MTC in the economic model

The economic model utilises the treatment effects estimated as part of the MTC analysis. While generally MTC approaches have the advantage that the network of evidence can be synthesised simultaneously to provide estimates of the comparative effectiveness of all included treatments using an evidence base of trials that individually do not compare all treatment options, the exclusion of much of the available evidence from the manufacturer's MTC calls into question the validity of subsequently applying these estimates in the economic model. A second key advantage of a MTC analysis is the fact that it allows a formal assessment of the consistency of the evidence with estimates that can be established by assuming a direct and indirect evidence fit. [16] Despite having provided estimates for the different comparators from a direct and indirect analysis where possible, the manufacturer's submission does not provide any assessment or discussion on how consistent the MTC results are with the wider set of studies. This is an important omission given that each type of analysis comprised a different set of studies for each outcome analysed. Furthermore, the ERG noted some inconsistencies in the direction of effect between the results reported for the direct and indirect analysis and those reported for the MTC. For example, for the outcome of treatment discontinuations due to any cause, amiodarone has a positive effect relative to dronedarone in the direct analysis [redacted] and MTC [redacted] while it has a negative effect in the indirect analysis [redacted] The ERG strongly feels that some form of validation between the 3 types of synthesis should have been undertaken in order to assess

the validity of using the treatment effects estimated from the MTC analysis in the economic model.

The assumption that absence of evidence implies absence of effect

Where there is an absence of evidence, the manufacturer's submission appears to assume an absence of treatment effect. However, a lack of evidence does not necessarily imply an absence of a treatment effect. No studies of class 1c agents, either flecainide or propafenone, met the inclusion criteria of the MTC for the outcomes of all cause mortality and stroke. Consequently, the economic model assumed that there was no treatment effect associated with class 1c agents relative to standard therapy. Although the assumption of no treatment effect is relative to standard care, this assumption has direct implications on any comparison made against dronedarone, since the treatment effect for dronedarone applied in the model is also relative to standard care. The ERG does not consider this an adequate assumption, particularly for those outcomes where there was evidence available but the evidence did not meet the restrictive inclusion criteria of the MTC. For the outcome of all cause mortality, the direct evidence suggests a positive effect of class 1c agents relative to standard therapy (OR = 0.68), while the indirect analysis suggests a positive effect of class 1c agents relative to dronedarone (OR = 0.80). The assumption that class 1c agents have no mortality benefit suggests a potential positive bias in favour of dronedarone.

Following the ERG's request to the manufacturer to provide additional justification for this assumption in the initial points for clarification, the manufacturer updated the model so that there was no mortality benefit for dronedarone compared to class 1c agents but maintained the treatment effect for dronedarone for the outcome of stroke. It is clear that the exclusion of the available data on class 1c agents may introduce potential bias into the manufacturer's cost-effectiveness results for this comparison.

5.2.3 Baseline event rates

A central component of the manufacturer's submission is the assumption that the population of ATHENA is reflective of typical AF patients in the UK NHS. Consequently, the key source of data for baseline event rates employed in the economic model is the control arm of ATHENA. The generalisability of ATHENA to the NHS is an important consideration and has been discussed in detail in Section 4.1.3. Clearly the population recruited into the ATHENA trial is a moderate- to high-

risk AF population and hence the subsequent cost-effectiveness results are not necessarily generalisable to a more general AF population.

Baseline events rates for AF recurrence, CHF, ACS, stroke and treatment discontinuation were derived from ATHENA by estimating event-risk prediction equations and extrapolating these equations to a lifetime risk of events. While the ERG recognises that the risk equations have been established from the same source, the ERG considers that there is limited contextual information provided to establish whether there maybe viable alternative risk equations, which may estimate different relationships for some of the key risk factors considered or which may be more generalisable to a broader AF population. Given the importance of the risk equations as a source of contemporary event rates relevant to the UK, the ERG considers that additional information and a more detailed critique of the equations would have helped to confirm the relevance of the equations used. In particular, the ERG considers it important to establish that the event rates predicted from the extrapolation of ATHENA accurately reflect the rates expected in the UK.

Survival analysis was used to extrapolate the time to event data of ATHENA for each of the modelled events. The manufacturer examined various curve fits (exponential, weibull, Gompertz, log-normal, gamma, and log-logistic) for stroke, CHF and treatment discontinuation, and the most appropriate fit was chosen on the basis of Akaike's information criterion (AIC) and Bayesian information criterion (BIC). Given that both measures of goodness of fit predicted the same distribution for the extrapolation (see Appendix 14 of the manufacturer's submission), with the exception of treatment discontinuation, the ERG were satisfied with the choice of curves used to generate these time to event rates. In addition, the manufacturer undertook a sensitivity analysis examining the second best curve fits. However, the ERG felt that there was a lack of clarity regarding the exploration of alternative curve fits for the outcomes of AF recurrence and ACS. Therefore additional clarification was sought by the ERG concerning the justification for assuming an exponential curve fit for these two outcomes without examining any alternative fits. Only limited additional information was reported by the manufacturer stating that they used the simplest form of parameterisation for these outcomes because they are based on multiple events occurring and that the "theoretical underpinnings of the use of more advanced functional forms for the hazard were questionable" (p12 Manufacturer's response to clarification questions, September 2009).

The risk equations are dependent on a patient's baseline $CHADS_2$ score. The base case model assumes the same $CHADS_2$ distribution as observed in ATHENA. The ERG noted discrepancies between the baseline values of ATHENA reported in Table 7.3 of the manufacturer's submission and the corresponding values used in the economic model. Following the ERG's initial points for clarification, the model was corrected to use the values reported for ATHENA in Table 7.3.

As noted in earlier sections, the control arm of ATHENA was not used to estimate baseline mortality rates. The manufacturer justified this on the basis that a relatively low number of deaths occurred during the trial and a Weibull regression fit to predict survival after the end of the trial would overestimate the number of remaining life years. While the ERG accepts that the trial may not have been sufficiently long to observe enough events and extrapolate over a lifetime, it was felt that there was a general lack of clarity around the subsequent approach used for estimating baseline mortality rates. A mortality risk equation was generated by a simulation exercise which utilised UK all cause mortality lifetables, adjusting for an increased risk based on $CHADS_2$ score. The ERG considers that there are a number of uncertainties surrounding the risk equation employed in the manufacturer's model and that a more transparent and critical approach would have provided greater reassurance in relation to the underlying mortality rates applied in the model. In particular, the ERG would like to have seen some validation of the estimated mortality rates with the rates observed in ATHENA for the length of follow-up of the trial. The relationship between all cause mortality and $CHADS_2$ score is also subject to uncertainty. Furthermore, the ERG had initial concerns that the mortality rates estimated by the manufacturer may include mortality from stroke which could potentially constitute double counting as stroke represents a separate event in the model. However, following the ERG's initial points for clarification, a supplementary Excel based simulation model for estimating time to mortality was submitted. The ERG was satisfied that the mortality effect associated with stroke had been removed from the all cause mortality data.

5.2.4 Adverse events

Despite having performed a detailed synthesis around adverse events and serious adverse events (see Section 4.2.2.1) to inform the relative effectiveness of the different comparators in reducing adverse event rates, the manufacturer did not utilise this information in the economic model. Although the rationale for this has been explained by the manufacturer (i.e. individual adverse events data were not

available from the meta-analysis and MTC analysis), adequate justification for the different approaches and sources subsequently used to populate the economic model for adverse events were not provided. Instead, the manufacturer simply estimat[cc]ed absolute event rates for each of the comparators from a variety of different sources. Pooled data from 5 trials was used to inform the absolute event rates for dronedarone, while a single source of data was used to inform its comparators. The ERG is concerned about the lack of consistency in the methods used. For example, the head to head RCT of amiodarone versus dronedarone (DIONYSOS) was used as a single source to inform the absolute event rates for amiodarone, while it was used as part of the pooled data to inform the rates for dronedarone. The absolute adverse event rates employed in the model form an integral part of the evidence base as the events have associated quality of life decrements and resource implications.

5.2.5 Resource utilisation and costs

In general, the ERG considers that the manufacturer's approach to resource utilisation and costing is appropriate although some of the assumptions employed in relation to treatment initiation and monitoring costs were not considered to be sufficiently justified. In particular, it is unclear why dronedarone can be initiated in an outpatient setting while the other AADs (amiodarone, sotalol and class 1c) all require hospitalisation for initiation. The assumption applied to the other AADs is referenced to the Sheffield formulary and there is no discussion of whether this is representative of NHS practice. Consequently, the initiation costs for dronedarone are slightly lower than for its comparators (£213 vs £247). The ERG also notes that there are no requirements for monitoring with dronedarone outside of initiation follow up. This contrasts with the other treatments where there are monitoring costs every 6 months on treatment. Treatment related adverse event costs were based on the absolute incidence rate of adverse events, as discussed above. The rates of hospitalisation for each of these events were estimated by UK clinicians. No details are given in the submission on the number of clinicians contacted or the approach used to arrive at a consensus.

[cc] Correction inserted after manufacturer identified a factual error

5.2.6 Health-related quality of life (HRQoL)

In the absence of quality of life data collected in any of the dronedarone trials, the use of external data sources to derive utility weights in order to estimate QALYs is appropriate. However, the utility weights applied to the different health states were derived from a single external source, the AFTER cohort from the Euro Heart Survey. The rationale for selecting this particular source is not justified and a systematic search for HRQoL data for the main health states and events does not appear to have been performed. Although the data from the AFTER cohort meets the reference case requirement since EQ-5D data is available, there is no attempt by the manufacturer to compare these values with other estimates that may be reported in existing literature. However, it should be recognised that the AFTER cohort is based on a large sample of patients (n=3045) and the use of a single source ensures consistency in the values assigned to the different states and events.

No details were provided in the main submission on the regression approach used to estimate the health state and event utilities. In response to the ERG's points for clarification the manufacturer submitted a supplementary report on the AFTER study. However, it should be noted that the regression model reported in the AFTER study does not appear to the same as the model used to estimate the utility weights applied in the manufacturer's economic model. The regression model reported in the main submission includes age, gender and 4 states/events (AF symptoms, stroke, CHF and ACS). Consequently this model does not allow for any difference according to the type of AF (paroxysmal and persistent) or any existing comorbidities (e.g. SHD, CAD etc). Hence, the manufacturer assumes that there is no difference in the underlying HRQoL between different subgroups and the differences in QALYs are thus driven entirely by the incidence of events. This approach contrasts with the regression model that is reported in the supplementary report of the AFTER study where additional covariates were also included for the type of AF. This would allow for an adjustment for the type of AF as well as incorporating the impact on HRQoL due to different event rates. No explanation is provided for why a different regression model is applied in this submission or the implications that this may have for the subsequent QALY estimates.

The ERG also has some concerns that the utility values applied in the economic model potentially imply a higher estimate of quality of life than that expected from the general UK population. For example, the utility value estimated for a 70 year old AF

male without any symptoms is 0.918 (see p97 Table 7.8 of the manufacturer's submission). For comparative purposes, the UK population norm, based on EQ-5D scores, for a 70 year old male is 0.78, decreasing to 0.75 for aged 75 years and over.[17] These values are lower than the estimates quoted in the manufacturer's submission for an AF patient with and without AF symptoms. Consequently, the manufacturer's overall estimates of QALYs associated with the different treatments are likely to be overly optimistic. This also has important implications for the cost-effectiveness results since this assumption is likely to work in favour of dronedarone given the mortality effect estimates from the MTC and the additional life-years that are subsequently attributed to dronedarone. A more appropriate approach would have been to use the regression model to estimate the utility decrements associated with particular health states and events and apply these to the UK population norms for EQ-5D. This approach would ensure that the utility values applied in the model do not exceed those of the general population.

5.2.7 Sensitivity analysis

The manufacturer presents a detailed set of univariate and probabilistic sensitivity analyses in Appendix 20 of their submission. In general these appear to be relatively comprehensive and several of these scenarios consider a number of the issues identified by the ERG in their critique of the submission.

The results of the univariate sensitivity analyses demonstrate that the base-case results appear relatively robust to the majority of the inputs considered. However, while the majority of the analyses considered had only a minor impact on the ICER estimates, it is unclear what impact these may have in combination. The robustness of the cost-effectiveness results could have been reinforced by using multi-way sensitivity analyses.

The manufacturer also presented summary results of a probabilistic sensitivity analysis (PSA), indicating the probability that dronedarone is cost-effective against its comparator at thresholds of £20,000 and £30,000 per additional QALY. Although the results of the PSA are essential for reflecting uncertainty across the range of inputs, the manufacturer did not provide detailed information on the distributions used to represent parameter uncertainty in the model. Appendix 19 of the manufacturer's submission indicates that a log normal distribution was assumed for all parameters except costs, which utilised a beta distribution. The manufacturer does not give any

justification for the choice of distribution or the parameters values used to represent each distribution.

5.2.8 Results included in manufacturer's submission

The results of the model are presented in the manufacturer's submission from p107 to 114. However, following a number of issues identified by the ERG in their initial points for clarification and subsequent correspondence, the model was updated and a new set of results submitted (Addendum: Updated Results 7th October 2009, supplementary document, October 2009).

For comparative purposes both the original base case cost-effectiveness results and those based on the revised model are reported. Table 20 summarises the incremental cost-effectiveness of dronedarone versus each of the alternative comparators in the populations considered as part of the base case analysis presented in the initial submission. The revised results based on the updated model are presented in Table 21. Only the revised results are subsequently discussed.

Table 20: Summary of incremental cost per QALY results for each of the basecase population (original submission)

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£ 4,070	£ 4,365	£ 3,699	£ 3,424	£ 3,254
Dronedarone vs. sotalol	£ 1,797	£ 1,888	NA	£ 1,927	NA
Dronedarone vs. class 1c	£ 20,143	NA	NA	£ 18,239	NA
Dronedarone vs. amiodarone	NA	NA	£ 2,112	NA	£ 2,570

SHD, structural heart disease; CAD, coronary artery disease; LVD, left ventricular dysfunction; NA, not applicable

Table 21: Summary of incremental cost per QALY results for each of the base case population (updated model)

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£7,885	£8,142	£7,865	£7,007	£7,163
Dronedarone vs. sotalol	£1,980	£2,246	NA	£2,082	NA
Dronedarone vs.	£21,026	NA	NA	£21,770	NA

class 1c					
Dronedarone vs. amiodarone	NA	NA	£2,724	NA	£3,185

Across the base case populations, the ICER of dronedarone in the updated model varied between £7,007 and £8,142 per QALY when compared to standard therapy as first line treatment, £1,980 to £2,246 per QALY when compared to sotalol as first line AAD, £21,026 to £21,770 per QALY when compared to class 1c as first line AAD, and £2,724 to £3,185 per QALY when compared to amiodarone as first line AAD. The only individual pairwise ICERs for dronedarone which exceeded £20,000 per QALY are for class 1c.

The results of the probabilistic sensitivity analyses are summarised in Table 22. Importantly, it should be noted that these results are only reported for the original version of model and due to time constraints were not updated by the manufacturer for the latest version of the model. The results indicate that dronedarone has a high probability of being cost-effective at a threshold of £20,000 per QALY relative to most of the comparators considered. The exception is for the comparison against class 1c, where the probability of dronedarone being cost-effective was between 50%-52% in the different AF types. Since the revised ICER results are less favourable for dronedarone in the revised model, the probability that dronedarone is cost-effective will be lower than the results presented here. However, the ERG does not consider that these probabilities would be markedly different to those presented here.

Table 22: PSA results - probability dronedarone is cost-effective at a threshold of £20,000 (£30,000) per QALY

	Paroxysmal AF			Persistent AF	
	No SHD	**CAD**	**LVD**	**No SHD**	**SHD**
Dronedarone vs. standard therapy	72% (84%)	74% (86%)	74% (85%)	74% (84%)	74% (85%)
Dronedarone vs. sotalol	96% (98%)	95% (98%)	NA	94% (98%)	NA
Dronedarone vs. class 1c	50% (82%)	NA	NA	52% (84%)	NA
Dronedarone vs. amiodarone	NA	NA	94% (97%)	NA	94% (97%)

The results of the deterministic univariate sensitivity analyses were presented in Appendix 20 of the manufacturer's submission. Again, these analyses were not updated to correspond to the latest version of the electronic model submitted. The

ICER estimates across the majority of the analyses in Appendix 20 were broadly consistent with the base case results. The largest impact on the ICER of dronedarone occurred under the following scenarios:

- Time horizon of the model was reduced from lifetime to 1 year.
- Lower 95% confidence interval of mortality benefit was used for comparators versus upper 95% confidence interval of mortality for dronedarone.

5.2.9 Validity of results presented with reference to methodology used

The manufacturer's results indicate that the ICER of dronedarone is highly likely to be less than £20,000 per QALY relative to its comparators (except class 1c) based on the wide range of assumptions employed. These conclusions appear to be most sensitive to the assumptions related to mortality benefits. The validity of the findings is subject to a number of potential uncertainties, which are outlined by the ERG in the section below.

It should be noted that a complete validation by the ERG of the coding used in SIMUL8 was not possible. The complexity of the individual patient level model, combined with the inexistence of a detailed 'road map', made line by line validation of the code impossible within the time line of the STA. However, the ERG could decipher sufficient code to ensure that, in general, the approach outlined within the report was followed within the detailed patient level evaluation. In addition, the ERG was able to scrutinise in depth the Excel front end to SIMUL8, which revealed a number of potential issues which were identified within the initial points for clarification and addressed by the manufacturer in the revised model.

Further validation of the revised model, however, revealed a number of additional issues that may impact on the validity of the revised ICER results. In particular, the ERG noted an error in relation to the length of time that mortality benefits are accrued in the revised model. The base case analysis assumes that mortality benefits are incurred over a lifetime, but due to a technical error in the coding the results presented by the manufacturer more closely reflect mortality benefits incurred for two years only. The ERG also noted an inconsistency between adverse event costs inputted in the Excel front end and those values utilised in the SIMUL8 model. The potential impact of these issues is explored in more detail as part of Section 6.

5.3 Summary of uncertainties and issues

In general, the ERG considered the manufacturer's economic submission to be of high quality meeting the requirements of the NICE reference case approach. The economic model structure was considered appropriate for the decision problem and the detailed sensitivity analyses were thorough and informative in exploring the robustness of the results.

The main concerns expressed by the ERG relate to the following issues:

- Inconsistency in the reporting of the input parameters in the manufacturer's report and those applied in the economic model making validation more difficult.

- The complexity of the coding used in the SIMUL8 model. Although the ERG recognises the potential advantages of using DES, the use of SIMUL8 is not currently part of the standard software packages accepted for NICE submissions. While an experienced external researcher familiar with this package was added to the ERG team, other members of the team had to invest considerably more time than would normally be allocated to reviewing the electronic submission. Furthermore, the complexity of the coding and the use of multiple subroutines made it virtually impossible to fully validate the model. In addition, the use of individual patient sampling approaches can potentially increases the computational burden and run time of the analysis. Running the model probabilistically took approximately 3 hours which clearly limits the ability of the ERG to validate the full set of probabilistic results given the number of different subgroups and potential scenarios.

- The ERG is unclear whether the use of dronedarone as a first-line treatment alongside standard therapy is in accordance with the licensed indication. In addition, the positions for dronedarone evaluated by the manufacturer do not represent the full range of relevant strategies and potentially relevant treatment sequences have not been considered. Consequently it is possible that dronedarone might be more cost-effective when used at later points in the treatment pathway.

- The use of baseline data from the ATHENA trial to populate many of the event risks reflects the high-risk population recruited and hence the cost-

effectiveness results may not be generalisable to a more general AF population. Furthermore the high-risk population ($CHADS_2 \geq 4$) included for the comparison against standard care alone was based on a post-hoc subgroup analysis. The cost-effectiveness results for this group were based on applying subgroup estimates for both the baseline and the relative mortality effect.

- The use of the restricted set of studies included in the MTC, and the assumptions used for class 1c agents, to inform the relative effectiveness estimates applied in the model raises a number of issues that affect the potential validity as well as increasing the uncertainty in the cost-effectiveness results that are presented.

- The approach used to estimate adverse events is not based on a systematic consideration of the evidence and relevant data included in the meta-analyses and MTC do not appear to have been used.

- The assumptions concerning the initiation costs of dronedarone and the other AADs may not be reflective of current NHS practice. The costing of adverse events is based on expert opinion and is subject to additional uncertainties.

- The modelling approach used to derive utility weights for the main health states and events is not adequately justified and the different regression model applied in the manufacturer's submission and that reported in the report of the AFTER cohort is not explained. Consequently, differences in utility are driven entirely by the different health states and events predicted for the different patient subgroups and AF types as opposed to any potential differences in their underlying HRQoL.

- The utility weights applied in the model appear to be potentially optimistic compared to the UK norms for the general population. This is likely to over-estimate the QALYs gained for dronedarone due to the additional years of life gained due to the mortality benefits that are assigned. This is likely to over-state the cost-effectiveness of dronedarone for those comparisons where mortality differences are the main driver.

- A number of coding issues and errors were identified by the ERG. Although a number of these were addressed by the manufacturer in their revised model, detailed results for the univariate and probabilistic analyses based on the revised results were not presented due to the lack of time available to the manufacturer.

- Despite addressing many of the issues and errors in the revised model, several additional issues were identified particularly related to the duration of mortality benefit that is applied in the model and the costing of adverse events.

Given the importance of a number of these issues, additional information from the manufacturer was requested by the ERG in their initial points for clarification. These are considered in more detail below, alongside additional analyses undertaken by the ERG to consider the potential impact of the remaining uncertainties.

6 Additional work undertaken by the manufacturer and the ERG

6.1 Overview

As discussed in Section 5, the ERG was unable to comprehensively validate line by line the patient-level simulation model in SIMUL8. However, detailed checks of the SIMUL8 coding were performed and a number of logical checks were used to assess the internal validity of the results. In addition, the ERG focused on the Excel-based front end which enabled a thorough investigation of the input parameters to the model. Detailed checks for consistency between the values reported in the submission, the Excel front end, and those subsequently employed in SIMUL8 were made. These checks identified a number of inconsistencies and issues related to the SIMUL8 coding and input values employed. These issues were outlined in the points for clarification with the manufacturer and included:

- The treatment effects reported
- The baseline $CHADS_2$ score distributions for ATHENA
- The initiation cost for dronedarone
- The monitoring costs of the treatments
- The cost of adverse events
- The length of time that mortality benefits are incorporated
- Functionality of the second choice extrapolation curve fits

In addition, the ERG noted potential errors in the SIMUL8 coding in relation to:

- The wait time in the queue for ACS
- Discounting of the monitoring costs

The ERG requested an updated model from the manufacturer to address the inconsistencies and potential errors noted. This section provides details of the manufacturer's response and further critique of these revised results. Since receiving the manufacturer's response to the points for clarification, the ERG have also noted that the treatment effects for discontinuation of treatment due to any cause have not been incorporated into the SIMUL8 model.

T*he ERG has undertaken additional exploratory work to address several of the remaining issues and uncertainties identified during the structured critique of the manufacturer's initial submission and the revised model.*

This additional work undertaken by the ERG has three main elements:

1. *A critique of the revised cost-effectiveness results presented by the manufacturer including additional minor corrections performed by the ERG.*
2. *Exploratory work by the ERG to identify the main drivers of cost-effectiveness and key assumptions for the different comparisons.*
3. *More detailed work exploring the robustness of the cost-effectiveness to specific assumptions and additional uncertainties identified by the ERG.*

6.2 Critique of the revised cost effectiveness results presented by the manufacturer

Manufacturer revisions

In response to the points for clarification from the ERG, the manufacturer submitted a revised SIMUL8 model and Excel front end with the following amendments:

- *The errors in the wait time for ACS and discounting of monitoring costs were amended in SIMUL8.*
- *Inconsistencies in the treatment effects reported in the main submission, Appendix 12, and those used in the model were amended by submitting new tables of values for the treatment effects and updating the model to incorporate these values.*
- *Baseline CHADS2 score distributions used in the model were amended to incorporate the values from ATHENA.*
- *The initiation cost of dronedarone and the monitoring costs of the treatments were amended in the model.*
- *Problems with executing the second choice extrapolation curve fits were fixed.*
- *A random number seed was set within the model to enable the ERG to reproduce the manufacturer's results.*

The manufacturer updated their base case results following the amendments made (Addendum: Updated Results 7th October 2009, supplementary document, October 2009). The results presented in Table 21 correspond to the manufacturer's updated analysis.

ERG revisions

The ERG made a couple of further revisions to the model to address some remaining issues:

- Following a thorough validation of the model, the ERG noted an inconsistency between the adverse event costs reported in the Excel front end and those utilised in the SIMUL8 model. This inconsistency was caused by a missing link between worksheets in the Excel front end. Table 23 presents the revised costs of adverse events.

Table 23: Revised costs of adverse events for each treatment

	Manufacturer's cost used in model	ERG's amended cost (based on manufacturer's analysis)
Beta blockers	£82.64	£71.54
Dronedarone	£127.28	£106.28
Amiodarone	£181.18	£178.92
Sotalol	£277.39	£135.77
Class 1c	£0.00	£205.09
Placebo	£82.64	£71.54

- The ERG also noted inconsistencies between some of the worksheet cells in the Excel front end in relation to the length of time that mortality benefits are accrued within the model. In fact the ERG believes that the revised results presented by the manufacturer closely reflect results corresponding to 2-year mortality benefits. The manufacturer's base case analysis assumes that mortality benefits are incurred over a lifetime, but due to a technical error in the coding the results presented by the manufacturer are for two years only. Consequently, the results reported in Table 21,are biased against dronedarone if dronedarone is believed to have the most favourable treatment effect on mortality.

The base case results of the revised model are presented in Table 24 for each of the populations assuming lifetime mortality benefits. The ICERs differ from the results reported in Table 21,primarily due to different assumptions regarding length of time of mortality benefits. Given that the MTC estimates for all cause mortality show

dronedarone to be more effective than standard therapy in the high risk population (OR = 0.53), and more effective than sotalol and amiodarone, the ICERs for these comparisons are considerably reduced when lifetime mortality benefits are assumed. The comparison of dronedarone versus class 1c should be unaffected by the change in mortality benefits since both drugs are assumed to have the same effect as standard care. For this comparison, the reduction in ICER in Table 24 from that reported in Table 21 is primarily due to differences in adverse event costs between the manufacturer's model and the amended model by the ERG. The ICERs are more favourable towards dronedarone because no adverse events costs were incorporated for class 1c agents in the manufacturer's original model.

The ERG also undertook an analysis where the mortality benefits were restricted to the mean length of follow-up of ATHENA (1.8 years). Table 25 presents the results of the revised model for each of the populations assuming mortality benefits last for 1.8 years. It can be seen that the ICERs for the comparisons of dronedarone versus standard therapy, sotalol, and amiodarone are broadly similar to the manufacturer's revised results presented in Table 21.

Further exploration of the submitted evidence by the ERG incorporates the model changes mentioned above.

Table 24: ERG revised incremental cost per QALY results for each of the base case populations, assuming lifetime mortality benefits.

	Paroxysmal AF			Persistent AF	
	No SHD	**CAD**	**LVD**	**No SHD**	**SHD**
Dronedarone vs. standard therapy	£3,620	£4,014	£3,577	£3,358	£3,520
Dronedarone vs. sotalol	£1,692	£1,988	NA	£1,848	NA
Dronedarone vs. class 1c	£18,206	NA	NA	£18,955	NA
Dronedarone vs. amiodarone	NA	NA	£1,895	NA	£2,349

Table 25: ERG revised incremental cost per QALY results for each population assuming mortality benefits only last for 1.8 years.

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£7,816	£8,062	£7,748	£6,947	£7,061
Dronedarone vs. sotalol	£1,977	£2,254	NA	£2,088	NA
Dronedarone vs. class 1c	£17,829	NA	NA	£19,927	NA
Dronedarone vs. amiodarone	NA	NA	£2,674	NA	£3,084

6.3 Exploratory work by the ERG to identify the main drivers of cost-effectiveness and key assumptions for the different comparisons

Although the manufacturer undertook a detailed series of univariate and probabilistic sensitivity analyses, the ERG considered that it was difficult to establish the main drivers of the cost-effectiveness results and the impact of particular assumptions based on the evidence submitted by the manufacturer. Additional exploratory work was therefore undertaken by the ERG to further clarify the main drivers of cost-effectiveness and to identify the key assumptions within the submission. These key assumptions were then subjected to additional scrutiny and further re-analysis by the ERG (reported in Section 6.4).

The initial exploratory work by the ERG examined the individual contribution that each of the main treatment benefits assumed in the submission made to the overall cost-effectiveness estimates. This work was performed in a series of steps:

1. To begin, the ERG excluded all treatment effects (mortality, stroke, treatment-related adverse events) except those associated with a reduction in AF recurrences. Given that dronedarone is licensed primarily to reduce recurrences in AF, a sensible starting point was to assume no other benefits from treatment.
2. The ERG then included the treatment effect associated with mortality, excluding any additional benefits over and above mortality and reduction in AF recurrences. In this way, the ERG could examine how much of the cost-effectiveness was due to mortality benefits.

3. Following this, the ERG included the treatment effect associated with stroke allowing an assessment of the impact of treatment-related stroke benefits on cost-effectiveness.
4. Finally, the ERG noted that any remaining differences between the results from step 3 and the base case analysis were attributed to differences in adverse events between treatments.

The results of the ERG's analyses are outlined below.

Treatment effect on AF recurrences alone

Table 26 summarises the cost-effectiveness results for the base case populations considered by the manufacturer when all treatment effects are excluded except those associated with a reduction in AF recurrences. None of the comparisons for dronedarone across any of the populations are cost-effective since dronedarone is the least effective treatment at preventing AF recurrences (see Table 13). In most cases, dronedarone is dominated by its comparator as it is more costly and produces fewer QALYs.

Table 26: ERG's incremental cost per QALY results for each of the base case populations when the model assumes a treatment effect on AF recurrences alone.

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£7,486,908	£70,323,846	£1,355,984	£1,630,715	£2,254,522
Dronedarone vs. sotalol	£5,232,678	D	NA	D	NA
Dronedarone vs. class 1c	D	NA	NA	D	NA
Dronedarone vs. amiodarone	NA	NA	£5,694,862	NA	D

D, dominated

Treatment effect on AF recurrences and mortality

Table 27 summarises the cost-effectiveness results for the base case populations when the treatment effect associated with mortality is included, while any benefits over and above mortality and reduction in AF recurrences are excluded. The incremental cost-effectiveness is altered considerably for all comparisons except class 1c. The ICERs for dronedarone versus standard therapy, sotalol, and amiodarone closely reflect the results of the base case analysis demonstrating that

mortality is the main driver of cost-effectiveness for these comparisons. The difference in ICERs between the values reported in Table 27 and Table 26 is due to mortality benefits, while the difference between Table 27 and Table 24 is due to additional benefits beyond that explained by mortality and reduction in AF recurrences. The additional benefits have a very small impact on cost-effectiveness relative to the benefits from mortality for all comparisons except class 1c. The comparison of dronedarone versus class 1c in paroxysmal AF is unaffected by the inclusion of mortality benefits since dronedarone and class 1c are assumed to have no mortality effects relative to standard therapy for this comparison. The same assumption holds in persistent AF but the change in ICER from dominated to £370,690 in this population is most likely due to the use of amiodarone as second line anti-arrhythmic.

Table 27: ERG's incremental cost per QALY results for each of the base case populations when the model assumes a treatment effect on AF recurrences and mortality.

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£4,119	£4,566	£4,069	£3,833	£3,901
Dronedarone vs. sotalol	£1,815	£2,105	NA	£1,966	NA
Dronedarone vs. class 1c	D	NA	NA	£370,690	NA
Dronedarone vs. amiodarone	NA	NA	£2,081	NA	£2,519

D, dominated

Treatment effect on AF recurrences, mortality, and stroke

Table 28 summarises the cost-effectiveness results for the base case populations when the treatment effects associated with stroke are included, excluding any additional benefits over and above mortality, stroke and reduction in AF recurrences. The ICERs for dronedarone versus standard therapy, sotalol, and amiodarone are broadly similar to Table 27 and the base case results. The impact of stroke on these comparisons is limited because most of the effect is driven by mortality. The ICERs in Table 28 are slightly more favourable than the base case results for these comparisons because differences in adverse events have not been incorporated. The comparison of dronedarone versus class 1c is affected most by the inclusion of treatment-related stroke benefits. The ICERs are no longer dominated for this

comparison but have a high incremental cost-effectiveness ratio of around £45,000 per QALY, indicating that stroke is a main driver of the cost-effectiveness results. However, large differences remain between the ICERs reported in Table 28 and the base case results (Table 24), implying that treatment-related differences in adverse events is also a main driver of cost-effectiveness for this comparison.

Table 28: ERG's incremental cost per QALY results for each of the base case populations when the model assumes a treatment effect on AF recurrences, mortality and stroke.

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£3,574	£3,964	£3,528	£3,327	£3,486
Dronedarone vs. sotalol	£1,688	£1,983	NA	£1,847	NA
Dronedarone vs. class 1c	£46,500	NA	NA	£43,543	NA
Dronedarone vs. amiodarone	NA	NA	£1,932	NA	£2,389

Summary of the main drivers of cost-effectiveness

Based on the evidence submitted by the manufacturer, dronedarone is not cost-effective relative to its comparators when the only effect of treatment is a reduction in AF recurrences. However, if dronedarone is believed to offer additional benefits such as a reduction in mortality then the treatment can become cost-effective. The main driver of cost-effectiveness for the comparisons of dronedarone versus standard therapy as first line treatment, and sotalol or amiodarone as first line anti-arrhythmics, is mortality benefits. Stroke benefits and differences in treatment-related adverse events have only a very limited impact on cost-effectiveness for these comparisons. In contrast, the main driver of cost-effectiveness for the comparison of dronedarone versus class 1c agents is a combination of the benefits from stroke and reduction in adverse events. Neither of these benefits on their own is sufficient to reduce the ICER of dronedarone to below the £20,000 per QALY threshold but in combination the ICER falls just below this threshold.

6.4 More detailed work exploring the robustness of the cost-effectiveness to specific assumptions and additional uncertainties identified by the ERG

6.4.1 Treatment effects for individual comparators

During the critique of the manufacturer's submission in Section 5.2, the ERG identified a number of uncertainties in relation to the treatment effect estimates applied in the model. In particular, the ERG observed a number of inconsistencies between values reported from the direct and indirect analyses and those from the MTC. These inconsistencies were largely due to different inclusion/exclusion criteria applied for each synthesis. Given the importance of the evidence as a driver of cost-effectiveness, the ERG has examined a set of individual comparisons where the uncertain treatment effects have been varied.

Treatment effect of dronedarone on all cause mortality

The treatment effect of dronedarone on all cause mortality differs according to subgroup population. Based on the results of the post-hoc analysis of ATHENA where the

[redacted], the manufacturer applied a separate treatment effect on mortality for the high risk subgroup. Given the uncertainty of the post-hoc analysis, the ERG has examined a scenario where the treatment effect of dronedarone on mortality is assumed the same across subgroup populations (OR = 0.85 relative to standard care). The results are presented in Table 29 for the comparison of dronedarone versus standard therapy as first line treatment. Compared to the base case results, the ICERs have more than doubled but are still well within a £20,000 per QALY threshold.

Table 29: Incremental cost per QALY results for each base case population when the effects of dronedarone on mortality are assumed the same across subgroups

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£8,690	£9,147	£8,683	£7,589	£7,769

Treatment effect of sotalol and amiodarone on all cause mortality

From the manufacturer's MTC estimates, sotalol and amiodarone have a negative effect on mortality relative to standard therapy (OR = 4.52 and 2.73 for sotalol and amiodarone, respectively). The ERG has examined a scenario where sotalol and amiodarone are assumed to have no effect on mortality relative to standard therapy, while maintaining a treatment effect for dronedarone. Table 30 presents the results for the populations where sotalol and amiodarone are relevant comparators. The ICERs are around 4 times as large as the base case ICERs but they still remain below a £20,000 per QALY threshold. The results suggest that the cost-effectiveness of dronedarone relative to sotalol and amiodarone is driven by the differential treatment effect on mortality rather than requiring that amiodarone and sotalol are worse than standard therapy. The ERG also examined a scenario where sotalol and amiodarone were assumed to have the same effect on mortality as dronedarone (Table 31). Under this assumption, dronedarone is no longer considered to be cost-effective. This analysis reinforces the conclusion above that the mortality benefits from dronedarone are the main driver of cost-effectiveness. Other treatment-related differences between dronedarone and sotalol or amiodarone play a limited role.

Table 30: Incremental cost per QALY results for each base case population when amiodarone and sotalol are assumed to have no effect on mortality relative to standard therapy

	Paroxysmal AF			Persistent AF	
	No SHD	**CAD**	**LVD**	**No SHD**	**SHD**
Dronedarone vs. sotalol	£7,242	£7,550	NA	£7,577	NA
Dronedarone vs. amiodarone	NA	NA	£7,289	NA	£8,839

Table 31: Incremental cost per QALY results for each base case population when amiodarone and sotalol are assumed to have the same effect on mortality as dronedarone

	Paroxysmal AF			Persistent AF	
	No SHD	**CAD**	**LVD**	**No SHD**	**SHD**
Dronedarone vs. sotalol	£119,704	£102,668	NA	£92,009	NA
Dronedarone vs. amiodarone	NA	NA	£55,063	NA	£71,306

Treatment effect of class 1c on all cause mortality and stroke

The manufacturer's submission assumes an absence of treatment effect where there is an absence of evidence. No studies of class 1c agents met the inclusion criteria of the MTC for the outcomes of all cause mortality and stroke. Consequently, the economic model assumed that there was no treatment effect for class 1c relative to standard therapy for these outcomes. To explore this assumption in more detail, the ERG examined a set of scenarios where the treatment effect for class 1c was varied. In the first of these scenarios, class 1c was assumed to have the same effect on mortality as dronedarone, but different from standard therapy. This resulted in only a marginal change in the base case ICERs. In a second scenario, the ERG assumed that class 1c may have a better effect on mortality than dronedarone. The evidence from the direct analysis suggested a positive effect of class 1c agents relative to standard therapy (OR = 0.68), while the synthesis of indirect evidence suggested a positive effect of class 1c relative to dronedarone (OR = 0.80). Using the treatment effect from the direct analysis, the results showed that dronedarone was dominated by class 1c (i.e. dronedarone was more costly and produced fewer QALYs).

In a third scenario, the ERG examined the treatment effect of class 1c on stroke. The manufacturer's submission assumed that there was no effect on stroke from class 1c, while there was an effect from dronedarone. The ERG examined the assumption that class 1c has the same effect on stroke as dronedarone. Table 32 presents the results for the base case populations. The ICERs for dronedarone are doubled from the base case analysis and are well above thresholds likely to be considered cost-effective. Therefore the cost-effectiveness of dronedarone relative to class 1c is reliant on the assumption that class 1c has no stroke benefits and has significantly more adverse events.

Table 32: Incremental cost per QALY results for each base case population when class 1c is assumed to have the same effect on stroke as dronedarone

	Paroxysmal AF			Persistent AF	
	No SHD	**CAD**	**LVD**	**No SHD**	**SHD**
Dronedarone vs. class 1c	£36,975	NA	NA	£38,584	NA

Treatment effects from the ERG's MTC on all cause mortality

All the scenarios described above in relation to the treatment effect estimates for the individual comparators were identified by the ERG on the basis of the values

employed in the model from the manufacturer's MTC synthesis. However, as discussed previously, the ERG has a number of concerns about the manufacturer's MTC. In particular, the ERG is concerned about the exclusion of studies from the MTC due to issues with convergence. To address these issues, the ERG firstly examined the robustness of the manufacturer's MTC by re-running their analyses for all cause mortality in the Bayesian software WinBUGs. The resulting estimates were broadly similar to the results of the manufacturer's MTC when the same set of studies was included (see Table 5, Section 4.2.2.1). Secondly, and more importantly, the ERG performed a separate MTC synthesis which incorporated all the studies reporting 12 month data on mortality and allowing trials with zero events to be included by using a continuity correction. A comparison of the estimates using all the studies reporting on mortality and those of the manufacturer's MTC, which used a limited number of studies, is summarised in Table 33 below. The main difference lies in the estimates of amiodarone and sotalol relative to standard care, where the odds of all cause mortality are more than halved in the ERG's MTC. An estimate of the odds ratio for class 1c relative to standard care was also achievable. Full details of the synthesis are reported in Section 4.1.7.

The ERG has examined a scenario on cost-effectiveness where the estimates from the ERG's MTC, which incorporated all studies on mortality, were used in the economic model. Table 34 presents the results for the base case populations. For the comparisons of dronedarone versus standard therapy as first line treatment, sotalol or amiodarone as first line anti-arrhytmics, the ICERs have increased relative to the base case analysis, but they are still well below conventional thresholds that are considered to be cost-effective. The comparison of dronedarone versus class 1c is most affected by the results of the new synthesis. The ICERs for this comparison in both the paroxysmal and persistent populations has dropped from around £18,000 to £5,000 per QALY. The increase in cost-effectiveness of dronedarone relative to class 1c is due to the assumption that there is a differential effect on mortality between dronedarone and class 1c agents. It was noted in Section 6.3 that the main drivers of cost-effectiveness for the comparison of dronedarone versus class 1c was a combination of stroke benefits and reduction in adverse events, but the results from the synthesis incorporating all studies on all cause mortality suggest that mortality is also a driver of cost-effectiveness for this comparison.

Table 33: Odds ratio (95% CI) for all cause mortality estimated by the manufacturer's MTC (using a limited set of studies) and the ERG's MTC (using all studies reported in Appendix 6 of the manufacturer's submission)

	Manufacturer's MTC	ERG's MTC using all studies
Dronedarone vs. standard therapy	0.85 (0.67, 1.10)**	0.84 (0.66, 1.07)
Amiodarone vs. standard therapy	2.73 (1.00, 7.41)	1.30 (0.68, 2.56)
Sotalol vs. standard therapy	4.52 (1.59, 11.70)	1.87 (1.01, 3.57)
Class 1c vs. standard therapy	Not estimable (assume 1.0)	1.03 (0.36, 2.86)

**Odds ratio differs by subgroup: OR = 0.53 for the comparison of dronedarone versus standard therapy as first line treatment; OR = 0.85 for the comparison of dronedarone versus sotalol or amiodarone as first line anti-arrhythmics; OR = 1.0 for the comparison of dronedarone versus class 1c as first line anti-arrhythmic.

Table 34: Incremental cost per QALY results for each of the base case populations when the wider set of studies on all cause mortality is incorporated in the MTC analysis

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£8,222	£8,666	£8,018	£7,214	£7,407
Dronedarone vs. sotalol	£2,322	£2,645	NA	£2,495	NA
Dronedarone vs. class 1c	£4,459	NA	NA	£4,909	NA
Dronedarone vs. amiodarone	NA	NA	£3,431	NA	£4,247

6.4.2 Uncertainty related to resource utilisation and costs

The ERG noted in its critique of the manufacturer's submission that some of the assumptions in relation to treatment initiation and monitoring costs were not considered to be sufficiently justified. The ERG examined a scenario which assumed that all treatments (as opposed to just dronedarone) could be initiated in an outpatient setting. The resulting effect on the ICERs was marginal (Table 35) and the overall conclusions on cost-effectiveness is not altered. The manufacturer assumed that there were no requirements for monitoring with dronedarone. To assess the impact of this assumption, the ERG examined a scenario where it was assumed that all treatments have the same monitoring costs. The impact on the ICERs for the comparisons of dronedarone versus standard therapy, sotalol, and

amiodarone were marginal, but the ICERs for the comparison of dronedarone versus class 1c increased from around £18,000 per QALY to £23,000 per QALY (Table 36).

Table 35: Incremental cost per QALY results for each of the base case populations assuming all treatments can be initiated in an outpatient setting

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£3,607	£4,000	£3,568	£3,354	£3,521
Dronedarone vs. sotalol	£1,711	£2,007	NA	£1,868	NA
Dronedarone vs. class 1c	£19,051	NA	NA	£19,756	NA
Dronedarone vs. amiodarone	NA	NA	£1,930	NA	£2,386

Table 36: Incremental cost per QALY results for each of the base case populations assuming all treatments have the same monitoring costs

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£4,137	£4,542	£4,106	£3,855	£4,034
Dronedarone vs. sotalol	£1,898	£2,192	NA	£2,052	NA
Dronedarone vs. class 1c	£23,540	NA	NA	£23,920	NA
Dronedarone vs. amiodarone	NA	NA	£2,095	NA	£2,544

6.4.3 Uncertainty in health-related quality of life

The ERG expressed some concern that the utility values applied in the economic model potentially imply a higher estimate of quality of life than that expected from the general UK population. In an attempt to address this issue, the ERG adjusted the constant value of the regression model used to estimate utility values to ensure that the values applied in the model do not exceed those of the general population. The adjustment was made to the regression constant such that the utility value estimated for a 70 year old AF male without any symptoms was reduced from 0.918 to 0.78, with the same adjustment applied throughout the model for all patients. The ERG recognises that this is not the most appropriate way to change the utility values but without access to the individual patient level data of the AFTER study this was the

best that the ERG could do to take account of the manufacturer's potentially overly optimistic estimates of overall QALYs. Table 37 presents the ICERs for the base case populations incorporating the adjustment in HRQoL. The implications for the cost-effectiveness results were limited. The ICERs were less favourable towards dronedarone but for all comparisons, except class 1c, they remained well below a £20,000 per QALY threshold. For the comparison of dronedarone versus class 1c, the ICER for persistent AF was pushed just above this threshold.

Table 37: Incremental cost per QALY results for each of the base case populations after adjusting the quality of life estimates to be in line with the general UK population

	Paroxysmal AF			Persistent AF	
	No SHD	CAD	LVD	No SHD	SHD
Dronedarone vs. standard therapy	£4,346	£4,825	£4,296	£4,032	£4,237
Dronedarone vs. sotalol	£2,029	£2,386	NA	£2,217	NA
Dronedarone vs. class 1c	£19,139	NA	NA	£20,181	NA
Dronedarone vs. amiodarone	NA	NA	£2,272	NA	£2,822

7 Discussion

7.1 Summary of clinical effectiveness issues

The evaluation of clinical effectiveness in the submission comprised a summary of evidence from all relevant RCTs of dronedarone together with a systematic review in which dronedarone was compared with of all relevant AADs. Within the systematic review direct, indirect and MTC syntheses were presented.

The evidence presented demonstrates that dronedarone has a beneficial effect on AF recurrence and ventricular rate during recurrence but results based on direct and indirect analyses indicate that dronedarone is a less effective AAD than direct comparators. A direct comparison with amiodarone demonstrated that whilst dronedarone is well tolerated this does not outweigh its limited efficacy in terms of AF recurrence. The large outcomes trial demonstrated that over 21 months follow-up dronedarone was associated with a significant reduction in first CV hospitalisation or death, but the reduction in all cause mortality alone was not statistically significant.

The relative effectiveness of dronedarone versus other AADs remains subject to a number of uncertainties.

- The relative efficacy of dronedarone compared to other AADs remains highly uncertain. The short-term (6 months) DIONYSOS trial is the only head-to-head RCT identified comparing dronedarone with another AAD

- Existing clinical evidence across the AADs appears most robust for AF recurrences and appears considerably more uncertain for the other major clinical endpoints. Although dronedarone was reported to have a statistically significant reduction in the odds of all-cause mortality compared to both sotalol and amiodarone based on the MTC, neither the results from the head-to-head RCT (DIONYSOS) nor the results from the indirect comparisons reported a statistically significant difference. The existing evidence for stroke is also highly uncertain and only a small number of studies have reported this outcome. This was not a pre-specified outcome in any of these studies and no significant difference was reported between dronedarone and amiodarone and sotalol based on the results from the MTC.

- The key ATHENA study included a moderate to high-risk elderly AF population, which differed from that in other trials. Furthermore, the generalisability of this evidence to inform the management of a lower risk and younger AF population remains uncertain.

- The comparison of dronedarone and the other AADs was based on a range of alternative synthesis approaches incorporating direct and indirect evidence. Given the lack of consideration of clinical and statistical heterogeneity across the different studies, the validity of pooling the individual studies in the different synthesis approaches remains uncertain. In addition there appeared to be some inconsistencies in the selection of trials across the analyses which may have introduced some bias and uncertainty.

- Although dronedarone is also licensed to lower ventricular rate, rate control was not included as an outcome measure of the scope or in the submission.

- Because of a lack of the necessary trial data the effects of dronedarone of HRQoL have not been investigated.

7.2 Summary of cost effectiveness issues

The manufacturer's submission included a discrete event simulation model which was used to estimate the cost-effectiveness of dronedarone with other licensed anti-arrhythmic drugs and standard therapy alone. The model was used to evaluate the cost-effectiveness over five main patient groups in accordance with the clinical pathways for these populations as employed in current UK clinical guidelines. The results from the manufacturer's submission demonstrated that dronedarone was highly cost-effective in each of the populations relative to using standard baseline therapy alone as first line treatment, or sotalol or amiodarone as first line anti-arrhythmics, while the results for dronedarone relative to class 1c agents showed that dronedarone was marginally cost-effective with an ICER just above £20,000 per QALY and a 50% probability of being cost-effective at this threshold. The findings were reported to be robust across a wide range of alternative assumptions. The results were most sensitive to the time horizon of the model and assumptions regarding the benefits from AADs on mortality.

A detailed critique of the manufacturer's initial submission and revised model following points for clarification was undertaken by the ERG. The economic model structure was considered appropriate for the decision problem, and the general approach employed by the manufacturer to estimate lifetime cost-effectiveness was deemed appropriate and met the requirements of the NICE reference case approach. However, the ERG identified a number of potential issues related to the submission and electronic model which were considered to compromise the validity of the model results. These included: (i) the use of dronedarone as a first line treatment may be outside its licensed indication; (ii) the treatment pathways evaluated by the manufacturer may not represent the full range of relevant strategies or sequences; (iii) the use of baseline data from the ATHENA trial may not be generalisable to the UK AF population; (iv) the use of a restricted set of studies, and the assumptions used for class 1c agents, to inform the relative effectiveness estimates applied in the model; (v) uncertainty surrounding the HRQoL data used in the model; (vi) uncertainty in relation to the costs of dronedarone.

The ERG attempted to address some of these issues by conducting separate analyses using the manufacturer's model. In particular, the ERG examined a series of scenarios to establish the main drivers of the cost-effectiveness results and to consider the impact of particular assumptions based on the evidence submitted. The additional analyses based on the manufacturer's evidence suggested that the reduction in mortality from dronedarone, inferred by the manufacturer's evidence, was the main driver of cost-effectiveness for all comparisons that had a differential mortality effect between comparators. Treatment-related stroke benefits and differences in adverse events had only a limited impact on the cost-effectiveness of dronedarone for all comparisons except class 1c. For the comparison of dronedarone versus class 1c, the marginal cost-effectiveness of dronedarone was achieved from the combined benefits of reducing stroke and preventing adverse events. Given that these conclusions were based on the relative effectiveness estimates derived from the manufacturer's synthesis, the ERG explored the robustness to a number of specific assumptions in the evidence. The ICER of dronedarone remained relatively robust throughout (< £20,000 per QALY) except for the assumptions: (i) amiodarone and sotalol have the same effect on mortality as dronedarone; and (ii) class 1c has the same effect on stroke as dronedarone. Under these situations, the ICER of dronedarone was well above £30,000 per QALY.

The ERG also ran additional analyses to explore uncertainty surrounding the assumptions applied to treatment initiation and monitoring costs and HRQoL data. Although the cost-effectiveness results remained robust, the ERG was unable to resolve the uncertainty in relation to the utility weights applied to the specific health states in the model. There also remain a number of important sources of uncertainty related to the cost-effectiveness of dronedarone which the ERG has been unable to address. These include establishing the most appropriate source of data to inform the baseline event rates applied in the model; the position for dronedarone in the pathway of treatment sequences; HRQoL benefits of dronedarone and the maintenance of benefits over the longer term.

7.3 Implications for research

Further trials or the implementation of registries would be helpful to establish the efficacy and safety of dronedarone relative to other AAD treatments that are regularly used in this indication within UK clinical practice. Additional evidence related to the

effectiveness of AADs for patients with AFL would also be valuable. Longer-term follow-up of trials, with prespecified outcome measures and analyses, are required to better establish the longer term efficacy or safety of dronedarone compared to other AADs. This is of particular importance in regard to outcomes of all-cause mortality and stroke since these appear to be the key drivers of the cost-effectiveness results. Given the lack of existing health related quality of life data, future RCTs of dronedarone should also consider using a relevant HRQOL measure.

References

1. Singh BN, Connolly SJ, Crijns HJ, Roy D, Kowey PR, Capucci A, et al. Dronedarone for maintenance of sinus rhythm in atrial fibrillation or flutter. *The New England journal of medicine [serial online]* 2007:987-99. [cited Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/900/CN-00611900/frame.html

2. Connolly SJ, Crijns HJGM, Torp-Pedersen C, van Eickels M, Gaudin C, Page RL, et al. Analysis of stroke in ATHENA: a placebo-controlled, double-blind, parallel-arm trial to assess the efficacy of dronedarone 400 mg BID for the prevention of cardiovascular hospitalization or death from any cause in patients with atrial fibrillation/atrial flutter. *Circulation* 2009;120:1174-80.

3. Køber L, Torp-Pedersen C, McMurray JJ, Gøtzsche O, Lévy S, Crijns H, et al. Increased mortality after dronedarone therapy for severe heart failure. *The New England journal of medicine [serial online]* 2008:2678-87. [cited Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/355/CN-00640355/frame.html

4. *Summary of positive opinion for Multaq*. London: Committee for Medicinal Products for Human Use, European Medicines Agency; 2009. Available from: http://www.emea.europa.eu/pdfs/human/opinion/Multaq_37683909en.pdf

5. Fang MC, Go AS, Chang Y, Borowsky L, Pomernacki NK, Singer DE, et al. Comparison of risk stratification schemes to predict thromboembolism in people with nonvalvular atrial fibrillation. *J Am Coll Cardiol* 2008;51:816-17.

6. Hohnloser S, Crijns H, van Eickels M, Gaudin C, Page R, Torp-Pedersen C, et al. Effect of dronedarone on cardiovascular events in atrial fibrillation. *Rev Port Cardiol* 2009;28:345-47.

7. Hohnloser SH, Crijns HJ, van Eickels M, Gaudin C, Page RL, Torp-Pedersen C, et al. Effect of dronedarone on cardiovascular events in atrial fibrillation. *The New England journal of medicine [serial online]* 2009:668-78. [cited Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/822/CN-00666822/frame.html

8. Page RL, Connolly SJ, Crijns HJ, van Eickels M, Gaudin C, Torp-Petersen C, et al. Abstract 4097: Rhythm- and rate-controlling effects of dronedarone in patients with atrial fibrillation: insights from the ATHENA trial. *Circulation* 2008;118:S_827.

9. Touboul P, Brugada J, Capucci A, Crijns HJ, Edvardsson N, Hohnloser SH. Dronedarone for prevention of atrial fibrillation: a dose-ranging study. *Eur Heart J [serial online]* 2003:1481-7. [cited Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/982/CN-00439982/frame.html

10. Davy JM, Herold M, Hoglund C, Timmermans A, Alings A, Radzik D, et al. Dronedarone for the control of ventricular rate in permanent atrial fibrillation: the Efficacy and safety of dRonedArone for the cOntrol of ventricular rate during atrial fibrillation (ERATO) study. *Am Heart J [serial online]* 2008:527.e1-9. [cited Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/858/CN-00650858/frame.html

11. *Multaq (Dronedarone) briefing document: Advisory Committee meeting of the Cardiovascular and Renal Drugs Division of the US Food and Drug Administration*. Bridgewater: Sanofi-Aventis Pharmaceuticals; 2009 March 18. Available from: http://www.fda.gov/ohrms/dockets/ac/09/briefing/2009-4417b1-02-Sanofi_Aventis.pdf

12. Patten M, Maas R, Bauer P, Luderitz B, Sonntag F, Dluzniewski M, et al. Suppression of paroxysmal atrial tachyarrhythmias--results of the SOPAT trial. *Eur Heart J* 2004;25:1395-404.

13. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics* 2006;24:355-71.

14. *Guide to the methods of technology appraisal*. London: National Institute for Clinical Excellence; 2008. Available from: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf

15. Johannesson M, Weinstein MC. On the decision rules of cost-effectiveness analysis. *J Health Econ* 1993;12:459-67.

16. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* 2006;101:447-59.

17. Kind P, Hardman G, Macran S. *UK population norms for EQ-5D: Centre for Health Economics discussion paper 172*. York: Centre for Health Economics, University of York; 1999.

Appendix 1: Detailed comments on search strategies

Description of manufacturer's search strategy and comment on whether the search strategy was appropriate.

The manufacturer's submission described the search strategies used to identify relevant studies of dronedarone for atrial fibrillation and atrial flutter, and full details of the search strategies used in each section were reported in the appendices or in the clarifications provided. Overall, the search strategies employed for each of the sections of the submission were appropriate.

Search strategy for section 6, clinical effectiveness

The submission gave detailed descriptions of the search strategies and largely met NICE requirements. It included the specific databases searched (MEDLINE, EMBASE and the Cochrane Library); the service providers used; the dates when searches were conducted; the date spans of the searches; the complete strategies used; the number of records identified for each search set; and the final result number. The MEDLINE In-Process database was not searched.

The search strategies were structured using a combination of subject indexing and free text search terms; thesaurus terms were exploded when relevant; truncation and wildcards were appropriately used; and search facets were combined using Boolean operators. Animal studies were excluded. The date spans ran from database inception to May 2009. Cited references from included studies and previously published reviews were also searched separately.

The description of the strategies in the manufacturer's submission stated the searches "combined both MeSH and free-text terms for 'atrial fibrillation/flutter' with the interventions 'dronedarone', and publication type 'randomised clinical trial', or studies reporting quality of life outcomes". The search strategies reported however did not contain any atrial fibrillation/flutter terms (subject indexing or free-text), and searched for dronedarone studies for any condition, and for randomised controlled trials and systematic review publication types. However, as the search strategies employed were therefore broader than as described, the searches will still have retrieved all studies of dronedarone for atrial fibrillation and atrial flutter amongst wider results.

The filters used to identify trials and systematic reviews in the searches were appropriate, but several subject indexing terms used in the EMBASE strategy were not translated from MeSH to EMTREE. These terms should however have been translated automatically to the appropriate EMTREE term by the OVID host when the searches were run. Given the small number of results for the dronedarone search terms alone in MEDLINE (118) and in EMBASE (274) it may have been appropriate to have simplified the search strategies employed by not using filters at all, and to have reviewed all of the dronedarone results.

The search strategy for section 6, clinical effectiveness, was appropriate.

Search strategy for section 6, meta-analysis and mixed treatment comparison

The submission gave detailed descriptions of the search strategies and met NICE requirements. It included the specific databases searched (MEDLINE, EMBASE and the Cochrane Library); the service providers used; the dates when searches were conducted; the date spans of the searches; the complete strategies used; and the number of records identified for each search set. The search strategies were structured using a combination of subject indexing and free text search terms; thesaurus terms were exploded when relevant; truncation and wildcards were appropriately used; and search facets were combined using Boolean operators. The date spans ran from database inception to April 2009. Conference proceedings were also searched separately to identify studies of interest from 2003 onwards.

The thesaurus terms used were appropriate to each database searched and the comparator terms searched for were comprehensive.

The filters used to identify study types in the searches were appropriate to each database searched.

The search strategy for section 6, meta-analysis and mixed treatment comparison, was appropriate.

Search strategy for section 7, cost effectiveness

Several clarifications relating to the searches were requested after the initial submission was received. Some additional details of search dates were provided, but the manufacturer largely referred us for clarification to "the report submitted on 22nd

September 2009 in response to question B1", and specifically to Section 3.1 and Table A-4. We understood this report to be the file received named 'updated priority report_050509', but on checking this report Section 3.1 was an executive summary containing no search criteria information, and no Table A-4 could be found. Therefore our assessment is largely based upon the initial submission.

The databases searched for the cost effectiveness literature included MEDLINE, MEDLINE In-Process, EMBASE, and NHS EED as required by NICE, but also DARE and HTA (both via the Cochrane Library search) which are not required. HEED was not searched, but use of this database is being reviewed at present. This is a subscription only database and the manufacturers may not have had access.

The submission and clarifications gave detailed descriptions of the search strategies and these largely met NICE requirements (no detailed description of the EMBASE search strategy was provided in the requested clarifications, but we have assumed that the EMBASE strategy listed in Appendix 1 of 'Protocol-v4 OF ABACUS systematic review.doc' was used). The submission included the specific databases searched; the service providers used; the dates when searches were conducted; the date spans of the searches; the complete strategies used; and the number of records identified for each search set. The search strategies were structured using a combination of subject indexing and free text search terms; thesaurus terms were exploded when relevant; truncation and wildcards were appropriately used; and search facets were combined using Boolean operators. The date spans ran from 1990 to December 2008. Conference abstracts were also searched separately to identify studies of interest from 2005 January 2009.

The thesaurus terms used were appropriate to each database searched.

The filters used to identify study types in the searches were appropriate to each database searched.

The search strategy for section 7, cost-effectiveness, was appropriate.

Appendix 2: Details of ERG MTC

An MTC using fixed effect and random effects models were run and compared using the seven studies included in the original MTC by the authors. This analysis uses a logistic regression model that does not include correction factors for zero events. These models are based on the WinBUGS code published on the Bristol University MTC analysis webpage (https://www.bris.ac.uk/cobm/docs/RE%203-arm.odc).

The fixed effect had a slightly lower DIC than the random effects model (77.25 vs 78.42), so these results should be used. The random effects results are presented for comparison. The model fit was good as the residual deviance was less than the number of arms in each analysis.

The results were as follows:

Fixed effect model

Comparison	OR	95% Cr I
Sotalol vs NAC	4.57	(18.2,14.54)
Amiodarone vs NAC	2.87	(1.14,9.14)
Dronedarone vs NAC	0.85	(0.67,1.09)
Amiodarone vs sotalol	0.63	(0.37,1.04)
Dronedarone vs sotalol	0.19	(0.06,0.48)
Dronedarone vs amiodarone	0.30	(0.09,0.76)

Random effects model

Comparison	OR	95% Cr I
Sotalol vs NAC	5.15	(1.75,20.88)
Amiodarone vs NAC	3.15	(1.04,12.1)
Dronedaron vs NAC	0.91	(0.47,2.27)
Amiodarone vs sotalol	0.61	(0.26,1.31)
Dronedarone vs sotalol	0.18	(0.04,0.62)
Droneadarone vs amiodarone	0.30	(0.08,1.00)

Authors' MTC excluding A-COMET-II, 2006

The MTC was rerun excluding the study with label A-COMET-II, 2006 because the study duration was less than 12 months. The results are:

Fixed effect model

Comparison	OR	95% Cr I
Sotalol vs NAC	3.72	(1.43,11.82)
Amiodarone vs NAC	2.40	(0.93,7.51)
Dronedarone vs NAC	0.85	(0.66,1.09)
Amiodarone vs sotalol	0.65	(0.39,1.07)
Dronedarone vs sotalol	0.23	(0.07,0.61)
Dronedarone vs amiodarone	0.35	(0.11,0.93)

Random effects model

Comparison	OR	95% Cr I
Sotalol vs NAC	4.00	(1.28,16.33)
Amiodarone vs NAC	2.56	(0.84,9.72)
Dronedarone vs NAC	0.90	(0.47,2.10)
Amiodarone vs sotalol	0.64	(0.29,1.38)
Dronedarone vs sotalol	0.23	(0.06,0.80)
Dronedarone vs amiodarone	0.36	(0.09,1.18)

For the analysis including all studies with inclusion criteria of 12 months study duration. Continuity correction was made adding 1 to the numerator for every arm for every trial with at least one arm with zero events and adding 2 to every denominator.

Fixed effect model results

Comparison	OR	95% Cr I
Sotalol vs NAC	1.865	(1.01,3.57)
Amiodarone vs NAC	1.302	(0.68,2.56)
Dronedaron vs NAC	0.8399	(0.66,1.07)
Class 1c vs NAC	1.033	(0.36,2.86)
Amiodarone vs sotalol	0.6997	(0.43,1.13)
Dronedarone vs sotalol	0.4501	(0.23,0.87)
Class 1c vs sotalol	0.5535	(0.19,1.50)
Droneadarone vs amiodarone	0.6455	(0.32,1.26)
Class 1c vs amiodarone	0.79	(0.26,2.26)
Class 1c vs dronedarone	1.23	(0.42,3.49)

Random effects

Comparison	OR	95% Cr I
Sotalol vs NAC	1.816	(0.92,3.72)
Amiodarone vs NAC	1.307	(0.63,2.75)
Dronedaron vs NAC	0.8437	(0.49,1.50)
Class 1c vs NAC	1.024	(0.35,2.98)
Amiodarone vs sotalol	0.7143	(0.40,1.32)
Dronedarone vs sotalol	0.4622	(0.20,1.09)
Class 1c vs sotalol	0.556	(0.19,1.64)
Droneadarone vs amiodarone	0.6482	(0.28,1.51)
Class 1c vs amiodarone	0.7822	(0.25,2.44)
Class 1c vs dronedarone	1.212	(0.37,3.91)

Model details

A burn-in of 10,000 iterations was performed followed by 50,000 iterations with the over relax function set to minimise autocorrelation.

Fixed effect model

1=Placebo
2=Sotalol
3=Amiodarone
4=Dronedarone

```
model{
sw[1] <- 0
for(i in 1:N)  {

# model
      logit(p[i])<-mu[s[i]]+ d[t[i]] - d[b[i]]

# binomial likelihood
  r[i]~dbin(p[i],n[i])

    #residual deviance

   rhat[i]<-p[i] * n[i]

dev[i]<-2 * (r[i] * (log(r[i]/rhat[i])) + (n[i] - r[i]) * (log((n[i] - r[i])/(n[i] - rhat[i]))))
  }

resdev<-sum(dev[])


# vague priors for 7 trial baselines

for(j in 1:NS){ mu[j]~dnorm(0,.0001) }

# vague priors for basic  parameters

d[1]<-0
for (k in 2:NT)  {d[k] ~ dnorm(0,.0001) }

# Pairwise ORs

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]
              }
         }
}

#initial 1
list(
d=c(NA,0,0,0),

mu=c(0,0,0,0,0,  0,0)
)

#initial 2
list(
d=c(NA,0.1,-0.5,-0.2),

mu=c(0.5,-0.5,-0.8,0,0,    -0.8,0.5)
)
```

Random effects model

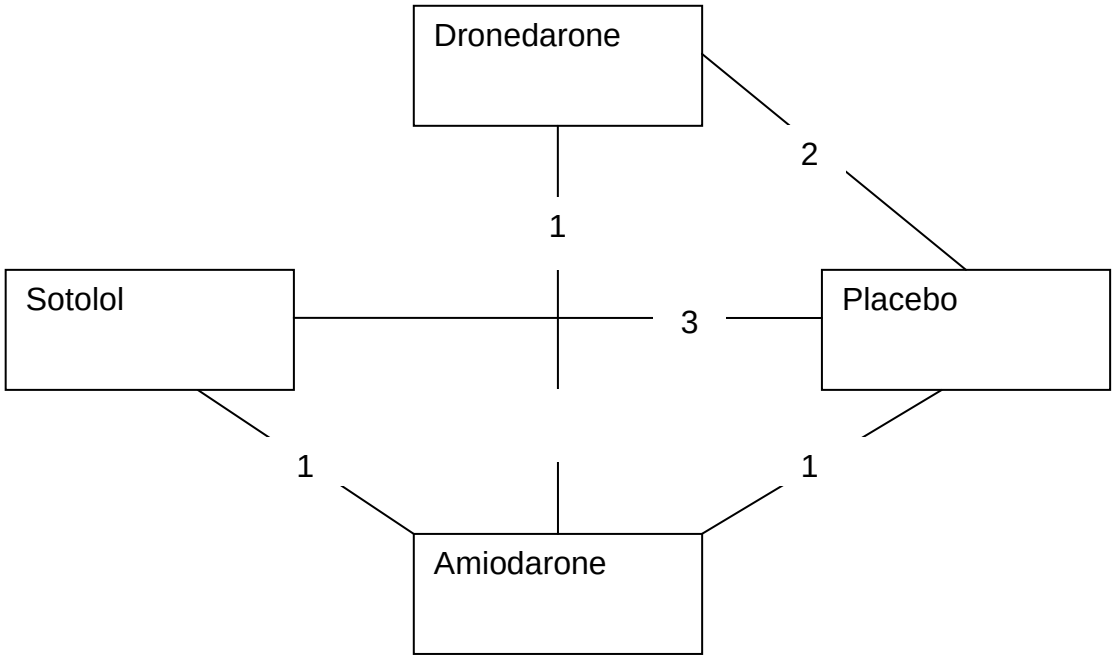
In the random effects model, correction was made for correlation within 3-arm trials. For the random effects model, the upper limit of the range of the prior distribution for the standard deviation of the random effects distribution was varied from 1 to 2 to 5 and 1 gave the best model fit.

```
model{

for(i in 1:N)  {

 #  model
      logit(p[i])<-mu[s[i]]+ delta[i]  * (1-equals(t[i],b[i]))
 #  binomial likelihood
      r[i]~dbin(p[i],n[i])
 # trial-specific LOR distributions
       delta[i] ~ dnorm(md[i],taud[i])
 # precisions of LOR distributions
       taud[i] <- tau * (1 + equals(m[i],3) /3)
# means of LOR distributions
      md[i] <- d[t[i]] - d[b[i]]  +  equals(m[i],3) * sw[i]

#residual deviance

   rhat[i]<-p[i] * n[i]

dev[i]<-2 * (r[i] * (log(r[i]/rhat[i])) + (n[i] - r[i]) * (log((n[i] - r[i])/(n[i] - rhat[i]))))
  }

resdev<-sum(dev[])

# adjustment for 3-arm trials
for (i in 2:N)  {   sw[i] <- (delta[i-1] -  d[t[i-1]] + d[b[i-1]])/2}

# vague priors for trial baselines
for(j in 1:NS){ mu[j]~dnorm(0,.0001) }

# vague priors for basic parameters
d[1]<-0
for (k in 2:NT)  {d[k] ~ dnorm(0,.0001) }

 #  vague prior for random effects standard deviation
sd~dunif(0,1)
tau<-1/pow(sd,2)


# Pairwise ORs
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]
              }
         }
}
```

Data for MTC (SAS method in Submission): All cause mortality

STUDY	Treatment	N	R
A-COMET-II, 2006	Placebo	224	0
A-COMET-II, 2006	Sotalol	223	4
AFFIRM substudy, 2003	Amiodarone	131	15
AFFIRM substudy, 2003	Sotalol	125	24
ATHENA, 2009	Dronedarone	2301	116
ATHENA, 2009	Placebo	2327	139
DIONYSOS, 2009	Dronedarone	249	2
DIONYSOS, 2009	Amiodarone	255	5
EURIDIS/ADONIS, 2007	Placebo	409	3
EURIDIS/ADONIS, 2007	Dronedarone	828	8
SAFE-T, 2003	Amiodarone	267	13
SAFE-T, 2003	Sotalol	261	15
SAFE-T, 2003	Placebo	137	3
SOPAT, 2004	Placebo	251	0
SOPAT, 2004	Sotalol	264	2

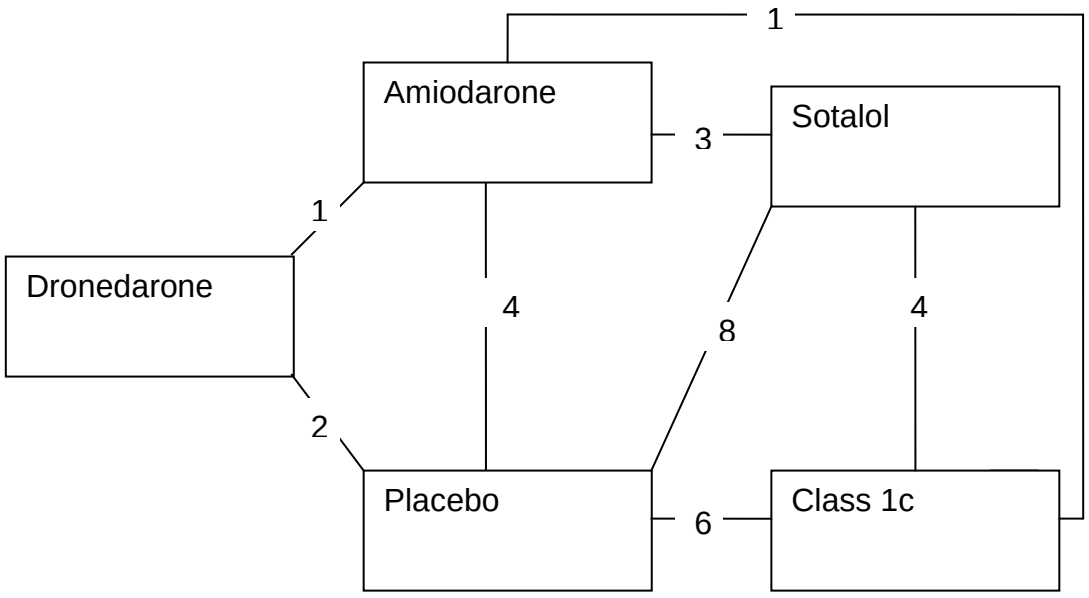
Network diagram for MTC (SAS method in Submission): All cause mortality



Data for MTC: All cause mortality (ERG re-run with all relevant trials)

STUDY	Treatment	N	R
AFFIRM substudy, 2003	Sotalol	24	125
AFFIRM substudy, 2003	Amiodarone	15	131
Bellandi, 2001	Placebo	0	92
Bellandi, 2001	Sotalol	0	106
Bellandi, 2001	Class 1c	0	102
Benditt, 1999	Placebo	0	69
Benditt, 1999	Sotalol	0	184
Boos, 2008	Placebo	0	18
Boos, 2008	Amiodarone	0	17
Carunchio, 1995	Placebo	0	26
Carunchio, 1995	Sotalol	0	20
Carunchio, 1995	Class 1c	0	20
Channer, 2004	Placebo	0	38
Channer, 2004	Amiodarone	0	61
Dogan, 2004	Placebo	1	52
Dogan, 2004	Class 1c	0	58
EURIDIS/ADONIS, 2007	Placebo	3	409
EURIDIS/ADONIS, 2007	Dronedarone	8	828
Fetsch, 2004	Placebo	0	88
Fetsch, 2004	Sotalol	6	383
Kochiadakis, 2000	Placebo	0	60
Kochiadakis, 2000	Sotalol	0	61
Kochiadakis, 2000	Amimodarone	0	65
Kochiadakis, 2004a	Amiodarone	0	72
Kochiadakis, 2004a	Class 1c	0	74
Kochiadakis, 2004b	Placebo	0	83
Kochiadakis, 2004b	Sotalol	0	85
Kochiadakis, 2004b	Class 1c	0	86
Pritchett, 2003 (RAFT)	Placebo	0	126
Pritchett, 2003 (RAFT)	Class 1c	0	397
Reimold, 1993	Sotalol	2	50
Reimold, 1993	Class 1c	0	50
SAFE-T, 2003	Placebo	3	137
SAFE-T, 2003	Sotalol	15	261
SAFE-T, 2003	Amiodarone	13	267
SOPAT, 2004	Placebo	0	251
SOPAT, 2004	Sotalol	2	264
Van Gelder, 1989	Placebo	0	36
Van Gelder, 1989	Class 1c	0	36
ATHENA, 2009	Placebo	139	2327
ATHENA, 2009	Dronedarone	116	2301
DIONYSOS, 2009	Amiodarone	5	255
DIONYSOS, 2009	Dronedarone	2	249

Network diagram for MTC: All cause mortality (ERG re-run with all relevant trials)



Appendix 3: Checklist for quality assessing the economic model

Table 1.8: Quality Assessment of Economic Model

Quality criterion	*Question(s)*	*Response (√, X, or NA)*	*Comments*
S1	Is there a clear statement of the decision problem?	√	
	Is the objective of the evaluation and model specified and consistent with the stated decision problem?	√	
	Is the primary decision-maker specified?	√	Not explicitly stated, although the report is written for NICE.
S2	Is the perspective of the model stated clearly?	√	NHS and PSS perspective were stated.
	Are the model inputs consistent with the stated perspective?	√	
	Has the scope of the model been stated and justified?	√	
	Are the outcomes of the model consistent with the perspective, scope and overall objective of the model?	√	
S3	Is the structure of the model consistent with a coherent theory of the health condition under evaluation?	√	Structure of the model is developed as per the existing UK guidelines.
	Are the sources of data used to develop the structure of the model specified?	√	
	Are the causal relationships described by the model structure justified appropriately?	√	

S4	Are the structural assumptions transparent and justified?	√		
	Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?	√		
S5	Is there a clear definition of the options under evaluation?	√		
	Have all feasible and practical options been evaluated?	√		
	Is there justification for the exclusion of feasible options?	*na*		
S6	Is the chosen model type appropriate given the decision problem and specified causal relationships within the model?	√		
S7	Is the time horizon of the model sufficient to reflect all important differences between options?	√		
	Are the time horizon of the model, the duration of treatment and the duration of treatment effect described and justified?	√	Lifetime time horizon. Duration of treatment was until death or withdrawal due to intolerance.	
S8	Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions?	√		
S9	Is the cycle length defined and justified in terms of the natural history of disease?	na		
D1	Are the data identification methods transparent and appropriate given the objectives of the model?	*√/?*	With a few limitations, most of the data sources are well described and justified.	
	Where choices have been made between data sources,	*√/?*	Most of the choices between data sources were justified.	

	are these justified appropriately?		
	Has particular attention been paid to identifying data for the important parameters in the model?	√	
	Has the quality of the data been assessed appropriately?	√	
	Where expert opinion has been used, are the methods described and justified?	√/?	A list of experts is provided in an appendix but the process of compiling their opinion or arriving at a consensus is not described in any detail.
D2	Is the data modelling methodology based on justifiable statistical and epidemiological techniques?	√	
D2a	Is the choice of baseline data described and justified?	√	
	Are transition probabilities calculated appropriately?	na	
	Has a half-cycle correction been applied to both cost and outcome?	na	
	If not, has this omission been justified?	na	
D2b	If relative treatment effects have been derived from trial data, have they been synthesised using appropriate techniques?	√	Direct, indirect, and MTC synthesis. There are concerns regarding the inclusion/exclusion of trials in the MTC.
	Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified?	√	
	Have alternative extrapolation assumptions been explored through sensitivity analysis?	√	
	Have assumptions regarding the continuing effect of treatment once treatment is complete been documented	√	

	and justified?		
	Have alternative assumptions regarding the continuing effect of treatment been explored through sensitivity analysis?	?	Treatment is continued over a lifetime or until withdrawal due to intolerance.
D2c	Are the costs incorporated into the model justified?	√	
	Has the source for all costs been described?	√	
	Have discount rates been described and justified given the target decision-maker?	√	
D2d	Are the utilities incorporated into the model appropriate?	√	
	Is the source for the utility weights referenced?	√	
	Are the methods of derivation for the utility weights justified?	√	
D3	Have all data incorporated into the model been described and referenced in sufficient detail?	√	
	Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate)?	√	
	Is the process of data incorporation transparent?	√	
	If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?	√	
	If data have been incorporated as distributions, is it clear that second order uncertainty is reflected?	√	
D4	Have the four principal types of uncertainty been addressed?	*√/?*	Structural uncertainty has not been addressed.

	If not, has the omission of particular forms of uncertainty been justified?	na	
D4a	Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?	X	No alternative model was tested.
D4b	Is there evidence that structural uncertainties have been addressed via sensitivity analysis?	X	
D4c	Has heterogeneity been dealt with by running the model separately for different subgroups?	√	
D4d	Are the methods of assessment of parameter uncertainty appropriate?	√	
	If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?	na	
C1	Is there evidence that the mathematical logic of the model has been tested thoroughly before use?	?	
C2	Are any counterintuitive results from the model explained and justified?	X/?	
	If the model has been calibrated against independent data, have any differences been explained and justified?	/	
	Have the results of the model been compared with those of previous models and any differences in results explained?	X	Not compared, although there have been no other previous CEA studies on dronedarone.