Summary

We would like to thank NICE for the opportunity to respond to the dossier prepared by Southampton Health Technology Assessment Centre (SHTAC).

We would like to raise two major areas of concern we have regarding the approach taken by SHTAC to the clinical evidence and the cost-effectiveness modelling. We believe that these issues may make it very difficult for the appraisal committee to make specific recommendations based on the analyses presented by SHTAC.

Firstly, SHTAC separates the patient population in to three subgroups specified in the scope.

1. Population 1 - patients at increased risk of sudden cardiac death as a result of ventricular arrhythmias despite OPT
2. Population 2 - patients with heart failure as a result of LVSD and cardiac dyssynchrony despite receiving OPT
3. Population 3 – patients with both conditions described above.

The criteria by which patients and thereby clinical trial populations are attributed to each of the groups have not been objectively defined by SHTAC. The ABHI group, including two UK clinical specialists, believe that there are no accepted *a priori* clinical criteria which allow most of the trials in this review to be categorised as a whole to the groups defined in the scope. Indeed, the cross-over between the target populations for defibrillation and cardiac resynchronisation therapy was a key driver of the merging of the ICD and CRT appraisals by NICE for this review. If the most cost-effective technology were to be selected based on the SHTAC analysis in each population, it is therefore unclear what specific guidance could be issued.

The ABHI group therefore worked from the original individual patient data in the pivotal clinical trials back to the scope's categories to determine which patients benefit on clinical and cost effectiveness criteria from the treatments being assessed, defibrillation and cardiac resynchronisation.

Using the IPD we have been able to identify patient characteristics that define these populations i.e. Population 1 (patients with an increased risk of sudden cardiac death but without dyssynchrony who have been shown to benefit from ICD therapy); those in Population 2 (patients with dyssychrony but without significantly elevated risk of sudden cardiac death who have been shown to benefit from CRT-P therapy) and those in Population 3 (patients with increased risk of sudden cardiac death and with dyssynchrony who have been shown to benefit from CRT-D therapy).

Based on the ABHI analysis, group definitions and treatment recommendations are as follows (all patients with LVEF≤35%):

- Patients with NYHA I/II and QRS<150ms without left bundle branch block (LBBB) and those with NYHA III and QRS<120ms are at elevated risk of sudden cardiac death. ICD is cost-effective in this population.

- Patients with NYHA class I/II QRS≥150ms without LBBB, NYHA I/II QRS≥120ms with LBBB, and NYHA III QRS≥120ms are at elevated risk of both sudden cardiac and heart failure death. CRT-D is cost-effective in this population.
- Patients in NYHA class IV with QRS ≥120ms are at elevated risk of heart failure death. CRT-P is cost-effective in this population.

Table 1 summarises these recommendations. Summary and detailed cost-effectiveness results to support these conclusions are reproduced in Appendix 1 of this document.

Table 1: Treatment recommendations resulting from ABHI IPD-based cost-effectiveness model (all patients with severe left ventricular systolic impairment, LVEF≤35%)

LBBB	NYHA	QRS (ms)	Treatment recommendation
No	I/II	<150	ICD
		≥150	CRT-D
	III	<120	ICD
		≥120	CRT-D
	IV	≥120	CRT-P
Yes	I-III	≥120	CRT-D
	IV	≥120	CRT-P

This issue is discussed in detail in Part 1 of this response.

Secondly, in the SHTAC analysis the comparators comprise “treatment pathways” in which patients start on a given therapy (OPT, ICD, CRT-P or CRT-D) and may subsequently cross-over to a different device. We do not believe this comparison of treatment pathways can inform decision making in this appraisal. If a device is not recommended, patients will obviously not be able to switch to it in clinical practise. This issue is discussed in detail in Part 2 of this response. Although SHTAC have provided a revised analysis (“Additional scenario analysis, Southampton Health Technology Assessments Centre (SHTAC), March 2013”) this only partially addresses this concern, as documented in Part 2.

The ABHI analysis included individual patient data from over 12,000 patients, representing 95% of the relevant randomised controlled trial evidence. The ABHI analysis responds to the difficult decision of how to select appropriate devices in a heterogeneous patient population. This has been achieved by identifying measurable patient characteristics that could be used to define when patients would benefit most from a device delivering defibrillation or resynchronisation treatment, or both. Part 3 of our response discusses the critique of our analysis by SHTAC which we do not think recognised the full value of this analysis of such a large volume of individual patient data.

In addition, Appendix 2 documents substantive technical errors in the SHTAC analysis. This appendix focuses largely on Population 2 and 3 where the concerns are greatest. Population 1 represents a simpler population both in terms of comparators and the selection of relevant clinical data. Furthermore, the revised analysis from SHTAC for Population 1 fully addresses the concern regarding cross-over documented above. Both the ABHI group and SHTAC analyses suggest ICERs under £30,000/QALY in this group (ICD vs. medical therapy).

The ABHI group would like to emphasise that even if the technical flaws in the SHTAC model are addressed (both the removal of cross-over and the large number of flaws noted in Appendix 2); the

group believes that there are significant limitations to using a model based on summary trial data to inform the current appraisal. We therefore suggest that the committee use the SHTAC model in conjunction with the individual patient data based ABHI model to inform decision making in this appraisal. And that furthermore, careful attention is paid by SHTAC and the committee to the drivers of any differences between the model results. This can only be done comprehensively once the fundamental flaws in the SHTAC analyses are addressed.

Finally, the ICD implant rates quoted in the Assessment Group report originate from TA11 published in 2000 and the subsequent update in TA95 (2006). An implant number of 100 per million has inappropriately been used as a 'national target implant rate' since 2006. This figure is from a NICE costing template[1] designed to estimate budget impact, rather than any published NICE guidance. This ceiling has led to rationing of devices through commissioning policies, contributing to the low penetration rates for device therapy in the UK. We are concerned that the continued reference to the 100 per million implant rate as 'a target' (SHTAC report p37) will reinforce this error with commissioning bodies and providers alike. We request that future Guidance issued by NICE in this therapy area should acknowledge that any reference to implants rates as used in costing templates or the appraisal process are in the context of budget impact estimation and therefore should not be used as justification by commissioning bodies or providers for rationing cost effective technologies as has been previously evident. The criteria for implant should be defined by NICE guidance rather than target implant rates.

[1] http://www.nice.org.uk/nicemedia/live/11566/33173/33173.xls

Part 1 – Subgroup definition

Throughout their systematic review and economic evaluation SHTAC have viewed these as three distinct and mutually exclusive populations (highlighted by the fact that separate cost-effectiveness results are generated for each of the three groups).

Clinical trials appear to have been allocated to each of these subgroups on an *ad-hoc* basis rather than by using formally defined criteria. SHTAC state "LVSD was defined as reduced LVEF using the cut-off provided by the publications (an arbitrary cut-off was not imposed by this review). Similarly, cardiac dyssynchrony was as defined by the publications; usually a prolonged QRS interval. Trials clearly stating that participants had reduced LVEF, cardiac dyssynchrony and an indication for an ICD were considered as having both conditions." (Assessment Group report p43). We are concerned that the lack of clear definitions for Populations 1-3 mean that it will be very difficult to create clinically meaningful guidance based on the results of the SHTAC analysis.

Although SHTAC state "there is considerable overlap between the groups, such that people with HF due to LVSD are at risk of SCD from ventricular arrhythmia" (Assessment Group report p24), the extent of this overlap does not seem to have been fully understood and has not been reflected in the SHTAC modelling. This can be shown by looking at causes of death in SCD-HeFT (considered by SHTAC a "Population 1" trial) and Companion and CARE-HF (considered by SHTAC to be "Population 2" trials).

- In SCD-HeFT, 207 patients experienced sudden cardiac death, but an almost equivalent number (205) experienced heart failure deaths[2].
- In CARE-HF, 102 heart failure deaths and 86 sudden cardiac deaths occurred[3] and in Companion 139 heart failure and 83 sudden cardiac deaths occurred[4].

It is troubling that Companion, a trial which randomised patients to CRT-D, is included in the population of patients identified to be not at risk of sudden cardiac death (Population 2) and that CRT-D is included as a comparator in this population. Such allocation is fundamentally flawed and distorts the nature of the study populations and clinical conclusions.

SCD-HeFT, COMPANION and CARE-HF enrolled during overlapping periods (Sept 1997-July 2001; Jan 2000-Nov 2002 and Jan 2001-March 2003 respectively). The concurrent nature of the pivotal trials resulted in highly overlapping patient populations (e.g. all trials included significant numbers of patients with QRS≥120ms and NYHA III).

If the trials had not been run concurrently, it would have had important implications for their design, SCD-HeFT would have excluded patients with wide QRS due to the clear benefit of resynchronisation therapy in these patients and CARE-HF would have had to offer CRT-D to patients in NYHA III due to the clear benefit of defibrillation therapy observed in the other trials.

These changes were seen in later trials, for example a protocol amendment was applied to RAFT in 2006 to stop randomisation of NYHA III patients to ICD rather than CRT-D. This was driven by the clear benefit of resynchronisation therapy for NYHA III patients observed in CARE-HF.

---

[2] Packer 2009 http://www.ncbi.nlm.nih.gov/pubmed/19917887.
[3] Cleland 2006 http://www.ncbi.nlm.nih.gov/pubmed/16782715.
[4] Carson 2005 http://www.ncbi.nlm.nih.gov/pubmed/16360067.

This overlapping nature of the patient populations drove the ABHI group decision to develop an individual patient data analysis. This approach allowed the ABHI analysis to look both within and across the available trial evidence, and to identify where the devices could deliver most value.

The artificial separation of the evidence in to three groups means that none of the SHTAC analyses incorporate all relevant available evidence. There are many instances where studies of direct relevance to more than one population (e.g. RAFT, MADIT-CRT; CARE-HF; COMPANION, SCD-HeFT) being used in one population but not the other.

Rather than trying to artificially allocate individual trials to individual subgroups, the ABHI analysis included a synthesis of trials to identify the characteristics that define the subgroups described in the scope. The utility of this approach is recognised by one of the clinical specialists consulted by SHTAC who stated that "what the manufacturers seem to be doing is providing a more rational basis for resource allocation" (Subgroups in the ABHI submission: clinical opinion. Southampton Health Technology Assessments Centre (SHTAC), March 2013 p1). This limited and artificial allocation of trials to only one population in the scope not only weakens the analysis, but is most likely to lead to distorted non-evidence based conclusions that do not reflect the totality of the data from the randomised trials.

A final point is that SHTAC use individual trials to estimate the prognosis in each of the populations and meta-analyses of multiple trials to estimate population-specific treatment effects. This ignores within-group heterogeneity in key prognostic factors (e.g. NYHA) and treatment effect modifiers (e.g. QRS, LBBB) which could be (and according to our analysis are) important predictors of the cost-effectiveness of the different devices.

Part 2 – Inclusion of device cross-over in treatment pathways

SHTAC have included 'device crossover' in the design of the economic model. This includes patient crossover from no device (OPT) to a device and the possibility of patients switching from one device to another.

As the key decision facing the committee is whether a given device should be made available or not, the analysis should compare costs and outcomes in 'worlds' where ICDs and/or CRTs are available with those where they not. The comparators as defined in the scope are entirely aligned with this principle and explicitly exclude the option of individuals switching to device based therapy from OPT. By considering cost-effectiveness related to "treatment pathways" (Assessment Group report p298) the SHTAC analyses are incompatible with the current appraisal.

In the previous CRT appraisal (TA120), patients on OPT were allowed to switch to an ICD device. This was within the scope of that appraisal since ICD was not a comparator technology and hence was viewed as a component of standard care. This is a different situation to the current appraisal where ICD is a comparator technology and hence citing NICE's acceptance of the modelling within TA120 does not justify the current approach.

The Appraisal Committee should be aware that crossover is not a rare event in the SHTAC model. The crossover rates in each of the three patient groups are shown in Table 2 below.

Table 2: Crossover rates arising from SHTAC models (all values rounded to the nearest integer, all values obtained from "DetailedResults" sheets of models)

Population	Initial treatment	Switch to ICD	Switch to CRT-P	Switch to CRT-D
1	OPT	45%	N/A	N/A
	ICD	100%	N/A	N/A
2	OPT	<1%	1%	<1%
	CRT-P	6%	<1%[b]	16%
	CRT-D	16%	0%	0%
3	OPT	16%	<1%	81%
	ICD	1%	0%	5%
	CRT-P	17%[d]	0%	86%
	CRT-D	20%	0%	1%[b]

On Monday 11th March 2013 NICE supplied a revised analysis, the description of this provided is provided below:

"Scenario analysis was performed to estimate the impact of an alternative comparator arm with patients being managed with OPT only (i.e. no upgrades to a device) on the cost-effectiveness results for the three populations."

Based on a comparison of the "Base case" and "OPT only scenario" results provided and the description provided by SHTAC of the modifications made to the model, it is clear that only cross-over from OPT has been removed from the model. The cross-over between all other devices documented in Table 2 does not appear to have been removed (though there are slight differences in the total costs for CRT-P and CRT-D in Population 2). This therefore represents only a partial removal of the extensive switching documented in Table 2.

We believe that the modelling of treatment sequences is the key driver in a number of the counter intuitive issues raised later in this response (see Appendix 2).

Part 3 - Clarification of points arising from the SHTAC review of the ABHI model

On a number of occasions SHTAC state that the ABHI analysis is not in line with the final scope issued in September 2011. This criticism is based on the way we have chosen to start from individual patient data from the published clinical trials to develop a model of clinical and cost effectiveness and the way we have used this model to define which patients would fall in to Populations 1, 2 and 3. As documented in Part 1 of this response, we believe that the evidence base cannot be segregated in the way employed in the SHTAC analysis to form a basis for specific guidance development.

The ABHI analysis addresses all patients with LVEF≤35%[5] but without ICD secondary prevention indications and, within these patients, seeks to identify clinical risk factors which predict better/ worse response to each therapy. This includes clinical factors that identify patients at a higher risk of either sudden cardiac or heart failure mortality and also clinical risk factors that predict the magnitude of therapeutic effect on each of these causes of death. The ABHI results are presented for patient groups defined in terms of commonly measured clinical variables (NYHA class, aetiology (ischemic/non-ischemic), QRS duration and left bundle branch block (LBBB) conduction abnormality). The results of the analysis are presented with each device (and a no device option) included as individual comparators.

We believe this approach clearly addresses the decision problem as defined in the scope and are unsure why the ERG conclude that this approach is "challenging in terms of developing guidance" (Assessment Group report p256). On the contrary, the ABHI analysis provides clear definition of which patients comprise populations 1, 2 and 3 in terms of NYHA class, aetiology, QRS and LBBB and the cost-effectiveness of each device within these groups. These clinical features are routinely assessed in usual cardiological practice in the NHS. Despite examination of 48 subgroups defined by these characteristics in the ABHI analysis, the recommendations are relatively clear and simple. These recommendations are presented as Table 3.

Table 3: Treatment recommendations resulting from ABHI IPD-based cost-effectiveness model (replicates Table 1, all patients with severe left ventricular systolic impairment, LVEF≤35%)

LBBB	NYHA	QRS (ms)	Treatment recommendation
No	I/II	<150	ICD
		≥150	CRT-D
	III	<120	ICD
		≥120	CRT-D
	IV	≥120	CRT-P
Yes	I-III	≥120	CRT-D
	IV	≥120	CRT-P

SHTAC do acknowledge that the approach of using individual patient data to develop a network meta-analysis meta-regression is justified. SHTAC also identifies no errors in the ABHI analysis and acknowledges that the results of the analysis make intuitive sense and that the conclusions seem valid.

[5] The systematic review included studies including patients with LVEF≤40%. However, as the database included only a very small number of patients with LVEF>35% (see ABHI submission p112) we consider the analysis to be representative of a group with LVEF≤35%.

However, throughout their review of the ABHI submission SHTAC have been unfairly critical. Principally, they have incorrectly grouped together the results from the IPD network meta-analysis as well as the IPD based analyses of hospitalisations and HRQoL into one entity 'the NMA' and used concerns about one aspect (the HRQoL analysis) to cast doubt on all aspects – in particular the estimates of treatment effect modifiers.

SHTAC emphasise the uncertainty in the treatment effect estimates from the adjusted network meta-analysis. However their critique seems to ignore that this analysis represents the systematic analysis of individual patient data from over 12,000 patients (95% of the randomised controlled trial evidence base) and the trade-off between improving precision of the treatment effects and retaining statistical significance within specific subgroups.

The ABHI group believes that the network meta-analysis of the overall trial evidence (unadjusted for subgroup effects) provides clear evidence that each device is highly effective in reducing all cause-mortality. The results of this analysis are shown in Table 4 below. We would also argue that our analysis adjusting for covariables serves to refine the point estimates to better reflect risk reduction in specific subgroups.

Table 4: Results of unadjusted network meta-analysis (taken from ABHI submission p123)

Comparison	Hazard ratio, all-cause mortality (mean, 95% CI)
CRT-P vs. OPT	[redacted]
ICD vs. OPT	[redacted]
CRT-D vs. OPT	[redacted]

The Assessment Group correctly cites the fact that Figure 19 in the ABHI submission indicates that 14/52 treatment effect estimates were statistically significant (Assessment Group report p250). However, following submission of this report we have identified an error in the coding for this plot. Although the code for the analysis that entered the model (including variance parameters) was reviewed line-by-line by a second analyst, this was not the case for the associated plot code. Effectively uncertainty is overestimated in these plots due to incomplete accounting for the covariance between the model parameters. Appendix 3 of this document contains the original and revised results. These indicate that when the covariance is appropriately accounted for the treatment effect estimates are statistically significant in 28/52 subgroups.

In addition, we feel that the following aspects of the methodology were not clearly reflected in the SHTAC critique:

1) The use of conservative assumptions regarding treatment effect durations. All-cause mortality treatment effects were not assumed to be permanent; instead outside the follow-up period of the available trials (7.5 years) we assume that the hazard ratio converges to 1.0 over a 20 year period. Evidence that this assumption may be conservative is presented in Section 3.9 of the ABHI submission. Any HRQoL benefit associated with each device is also assumed to diminish over time with any benefit set to diminish after five years. Both of these assumptions are in contrast to those used in the academic models created during TA95 and TA120 which both assumed constant treatment effects. Our sensitivity analyses showed that both of these assumptions in the ABHI submission do indeed bias against treatments that provide clinical benefit (ABHI submission p194-199).

2) SHTAC also state that we did not provide cost and QALY breakdowns for all interventions in the 48 subgroups. This is incorrect on two counts. Firstly, we provided cost-effectiveness planes for all of the groups in an appendix to our dossier from which the information could have been derived. Secondly, we also provided SHTAC with a fully working model containing a separate worksheet from which they could directly examine the relevant information should they have wanted to.
3) SHTAC state that we did not included adverse events into our model. This statement is incorrect. While not explicitly modelled as health states (as in the SHTAC model), device related complications are included in the calculation of all implant costs (original and upgrade) as detailed in our submission dossier (ABHI submission p167).
4) The utility values used in the model are criticised by SHTAC on the basis of some counterintuitive device effects which were not used in the cost-effectiveness model. The critique provided does not acknowledge the strength of the utility analysis. The analysis incorporates all available EQ-5D data from the trials of interest, was corroborated by analyses of a large database of Minnesota Living with Heart Failure quality of life data and is therefore the most reliable data available for estimating quality of life adjustments in device eligible patients.
5) Finally, there was a concern raised by SHTAC that no evidence was provided to demonstrate whether or not the model had been validated. The model underwent two rounds of external cell-by-cell validation undertaken by an experienced health economist. We are happy to provide the validation reports to SHTAC or the committee if requested.

Appendix 1: Results of ABHI analysis

Table A1: Summary of cost-effectiveness results for all patients with LVEF ≤35%) (Replicated from Assessment report Table 91)

Heart failure severity	QRS duration	Results summary
NYHA class I/II	QRS duration<120ms	The ICERs for ICD vs. OPT are below £25,200 per QALY gained.
	QRS duration 120-149ms	ICD is a cost-effective treatment option[a] (ICER < £17,000 / QALY) patients with no LBBB. For CRT-D all ICERs are below £25,000 per QALY gained in LBBB patients (£20,608 to £24,343)
	QRS duration≥150ms	CRT-D is cost effective treatment with an ICER of less than £28,000 per QALY for all options.
NYHA class III	QRS duration<120ms	ICD vs. OPT generates ICERs below £30,000 per QALY
	QRS duration 120-149ms	CRT-P is cost-effective[a]. CRT-D generates ICERs between £23,900 and £27,400 per QALY gained relative to CRT-P.
	QRS duration≥150ms	CRT-P is cost-effective vs. OPT (ICER < £20,000 per QALY). Compared with CRT-P, CRT-D generates ICERs below £30,000 per QALY gained. ICD is either dominated or extended dominated.
NYHA class IV	QRS duration<120ms	No comparative analysis was possible in this patient group, as no patients were identified for this combination.
	QRS duration≥120ms	For CRT-P compared with OPT, all ICERs are close to or below £20,000 per QALY gained. For the comparison of CRT-D to CRTP, all ICERs are above £30,000 per QALY gained.

[a] According to willingness to pay threshold of £20,000 - £30,000 per QALY gained.

These conclusions can be summarised as follows:

Table A2: Treatment recommendations resulting from ABHI IPD-based cost-effectiveness model. All patients have severe left ventricular systolic impairment

LBBB	NYHA	QRS (ms)	Treatment recommendation
No	I/II	<150	ICD
		≥150	CRT-D
	III	<120	ICD
		≥120	CRT-D
	IV	≥120	CRT-P
Yes	I-III	≥120	CRT-D
	IV	≥120	CRT-P

Table A3: Deterministic base case results for ABHI analysis (patients without LBBB), replicated from ABHI submission

				C-E Sequence				ICERs			
NYHA Class	Etiology	QRS Duration	N	1st	2nd	3rd	4th	1st	2nd	3rd	4th
I	Non-Ischemic	<120ms	66	OPT	ICD	N/A	N/A	Referent	£24,304	N/A	N/A
I	Non-Ischemic	>=120, <150 ms	11	OPT	CRTD	ICD	N/A	Referent	Dominated	£16,619	N/A
I	Non-Ischemic	>=150ms	8	OPT	ICD	CRTD	N/A	Referent	£18,074	£1,080,057	N/A
I	Ischemic	<120ms	272	OPT	ICD	N/A	N/A	Referent	£24,016	N/A	N/A
I	Ischemic	>=120, <150 ms	216	OPT	CRTD	ICD	N/A	Referent	Dominated	£16,234	N/A
I	Ischemic	>=150ms	106	OPT	ICD	CRTD	N/A	Referent	Ext Dominated	£21,086	N/A
II	Non-Ischemic	<120ms	710	OPT	ICD	N/A	N/A	Referent	£25,110	N/A	N/A
II	Non-Ischemic	>=120, <150 ms	232	OPT	CRTD	ICD	N/A	Referent	Dominated	£17,016	N/A
II	Non-Ischemic	>=150ms	141	OPT	ICD	CRTD	N/A	Referent	£20,312	£27,175	N/A
II	Ischemic	<120ms	788	OPT	ICD	N/A	N/A	Referent	£23,884	N/A	N/A
II	Ischemic	>=120, <150 ms	756	OPT	CRTD	ICD	N/A	Referent	Dominated	£16,749	N/A
II	Ischemic	>=150ms	470	OPT	ICD	CRTD	N/A	Referent	£20,697	£22,777	N/A
III	Non-Ischemic	<120ms	255	OPT	ICD	N/A	N/A	Referent	£29,402	N/A	N/A
III	Non-Ischemic	>=120, <150 ms	150	OPT	CRTP	ICD	CRTD	Referent	Ext Dominated	£19,760	£27,336
III	Non-Ischemic	>=150ms	109	OPT	ICD	CRTP	CRTD	Referent	Dominated	£13,227	£24,350
III	Ischemic	<120ms	438	OPT	ICD	N/A	N/A	Referent	£26,923	N/A	N/A
III	Ischemic	>=120, <150 ms	426	OPT	CRTP	ICD	CRTD	Referent	£19,670	Ext Dominated	£24,796
III	Ischemic	>=150ms	192	OPT	ICD	CRTP	CRTD	Referent	Dominated	£14,392	£25,734
IV	Non-Ischemic	<120ms	5	OPT	N/A	N/A	N/A	Referent	N/A	N/A	N/A
IV	Non-Ischemic	>=120, <150 ms	12	OPT	CRTP	CRTD	N/A	Referent	£17,324	£30,624	N/A
IV	Non-Ischemic	>=150ms	9	OPT	CRTP	CRTD	N/A	Referent	£16,304	£33,901	N/A
IV	Ischemic	<120ms	42	OPT	N/A	N/A	N/A	Referent	N/A	N/A	N/A
IV	Ischemic	>=120, <150 ms	52	OPT	CRTP	CRTD	N/A	Referent	£24,366	£43,500	N/A
IV	Ischemic	>=150ms	10	OPT	CRTP	CRTD	N/A	Referent	£18,065	£37,802	N/A

Table A4: Deterministic base case results for ABHI analysis (patients with LBBB), replicated from ABHI submission

				C-E Sequence				ICERs			
NYHA Class	Etiology	QRS Duration	N	1st	2nd	3rd	4th	1st	2nd	3rd	4th
I	Non-Ischemic	<120ms	0	OPT	N/A	N/A	N/A	Referent	N/A	N/A	N/A
I	Non-Ischemic	>=120, <150 ms	21	OPT	ICD	CRTD	N/A	Referent	Ext Dominated	£21,021	N/A
I	Non-Ischemic	>=150ms	33	OPT	ICD	CRTD	N/A	Referent	Ext Dominated	£18,118	N/A
I	Ischemic	<120ms	0	OPT	N/A	N/A	N/A	Referent	N/A	N/A	N/A
I	Ischemic	>=120, <150 ms	76	OPT	ICD	CRTD	N/A	Referent	£19,989	£24,343	N/A
I	Ischemic	>=150ms	165	OPT	ICD	CRTD	N/A	Referent	Ext Dominated	£17,335	N/A
II	Non-Ischemic	<120ms	0	OPT	N/A	N/A	N/A	Referent	N/A	N/A	N/A
II	Non-Ischemic	>=120, <150 ms	385	OPT	ICD	CRTD	N/A	Referent	Ext Dominated	£20,608	N/A
II	Non-Ischemic	>=150ms	1,308	OPT	ICD	CRTD	N/A	Referent	Ext Dominated	£17,794	N/A
II	Ischemic	<120ms	0	OPT	N/A	N/A	N/A	Referent	N/A	N/A	N/A
II	Ischemic	>=120, <150 ms	477	OPT	ICD	CRTD	N/A	Referent	£20,640	£21,277	N/A
II	Ischemic	>=150ms	982	OPT	ICD	CRTD	N/A	Referent	Ext Dominated	£17,479	N/A
III	Non-Ischemic	<120ms	0	OPT	N/A	N/A	N/A	Referent	N/A	N/A	N/A
III	Non-Ischemic	>=120, <150 ms	189	OPT	ICD	CRTP	CRTD	Referent	Dominated	£12,550	£23,831
III	Non-Ischemic	>=150ms	775	OPT	ICD	CRTP	CRTD	Referent	Dominated	£9,798	£27,592
III	Ischemic	<120ms	0	OPT	N/A	N/A	N/A	Referent	N/A	N/A	N/A
III	Ischemic	>=120, <150 ms	355	OPT	ICD	CRTP	CRTD	Referent	Dominated	£15,449	£25,540
III	Ischemic	>=150ms	773	OPT	ICD	CRTP	CRTD	Referent	Dominated	£11,408	£29,912
IV	Non-Ischemic	<120ms	0	OPT	N/A	N/A	N/A	Referent	N/A	N/A	N/A
IV	Non-Ischemic	>=120, <150 ms	22	OPT	CRTP	CRTD	N/A	Referent	£14,715	£31,920	N/A
IV	Non-Ischemic	>=150ms	81	OPT	CRTP	CRTD	N/A	Referent	£12,076	£35,660	N/A
IV	Ischemic	<120ms	0	OPT	N/A	N/A	N/A	Referent	N/A	N/A	N/A
IV	Ischemic	>=120, <150 ms	38	OPT	CRTP	CRTD	N/A	Referent	£22,340	£41,695	N/A
IV	Ischemic	>=150ms	97	OPT	CRTP	CRTD	N/A	Referent	£17,722	£46,445	N/A

Appendix 2: Technical critique of Assessment Group model

We were given the opportunity to review the three Excel models developed by SHTAC. Each model is complex and has a relatively large number of health states. It was therefore not possible to review all aspects of the models within the time period available.

Two fundamental concerns regarding the definition of the decision problem have been described in detail in Part 1 and Part 2 of this document. The following additional methodological concerns and technical errors were also identified:

Derivation of treatment effects

The Assessment Group have effectively conducted a set of indirect comparisons to inform the cost-effectiveness analysis.

The adjusted indirect comparison in Population 2 is conducted on heart failure death and sudden cardiac death separately. Our fundamental concern with this comparison is with respect to the heart failure analysis. This comparison uses data from COMPANION and CARE-HF to inform the CRT-P vs. OPT treatment effect and from COMPANION only for the CRT-D vs. OPT treatment effect. Patients in CARE-HF were much more likely to have wide QRS ([redacted] QRS<150ms in CARE-HF; [redacted] in COMPANION) (ABHI submission p125) and LBBB (89-90% in CARE-HF, 69-73% in COMPANION). Individual trial subgroup analysis (Assessment Group report p244), previous meta-analyses[6] and the ABHI submission analysis identify QRS and LBBB as predictors of enhanced CRT efficacy. This indirect comparison is therefore likely to bias the treatment effect of CRT-P vs. CRT-D in favour of CRT-P. The treatment effect for CRT-P is also unlikely to reflect a clear patient population (as defined in terms of QRS and LBBB).

The indirect comparison in Population 3 is conducted on all-cause mortality. The CRT-D vs. ICD treatment effect is informed by eight trials (MADIT-CRT, MIRACLE-ICD II, RAFT, Pinter 2009, Contak-CD, Rhythm-ICD, MIRACLE ICD, RethinQ). The CRT-D vs. OPT treatment effect is taken directly from COMPANION. For CRT-P vs. CRT- D the Assessment Group have assumed equivalence. This appears to be illogical on the following grounds: (1) the direct evidence available from COMPANION (a group who may be *less* likely to receive additional benefit of CRT-D over CRT-P than the patients included in the CRT-D vs. ICD comparison due to Companion patients being more likely to die from heart failure causes) supports a stronger treatment effect for CRT-D than CRT-P; and (2) if ICD and CRT-P are effective it is illogical to assume that CRT-D is less effective than the two individual devices as a CRT-D device combines the treatment modalities of ICD and CRT-P.

For Population 1 pairwise analyses are presented. Patients with and without remote-MI are separated, without clear evidence of differential treatment effects from either individual trial subgroup analyses or the ABHI submission analysis (see ABHI submission Appendices p38). Neither analysis can therefore be thought to represent the full evidence base for these patients.

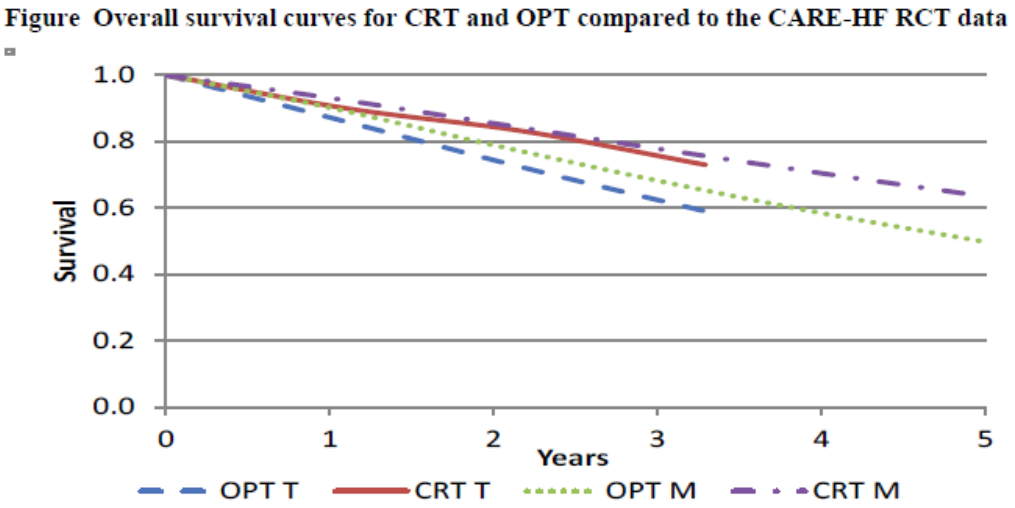
Derivation of survival curves

The SHTAC analysis applies relative risks as if they are hazard ratios in a number of instances. This is a violation of statistical principles and undermines the credibility of the analyses.

[6] Sipahi 2012 http://www.ncbi.nlm.nih.gov/pubmed/22305845

In addition, the proportion of patients alive at any given time point in the models does not reflect the data on which the models were largely parameterised. This is particularly apparent for the 'competing risk' approach used by SHTAC to model Population 2. To aid the Appraisal Committee we have reproduced the plot prepared by SHTAC comparing the output from their model with the all-cause mortality data from the CARE-HF study below. Assuming that T refers to trial and M to model, the fit to the OPT arm is extremely poor and leads to an inflation in survival. This in turn will lead to a decrease in incremental benefit for CRT-P and a corresponding inflation in the ICER. We believe this is one reason why the ICER for CRT-P vs. OPT is a lot higher than in previous evaluations (another being the presence of device crossover mentioned earlier).

Figure A1: Comparison of SHTAC model output and CARE-HF clinical data (Assessment Group report appendices, appendix 17, p231)



We have concerns about other curves and would have liked to have seen validation plots with longer follow up than five years in order to make an informed review. This is especially the case as an age specific scaling factor is applied to "all-cause mortality" data to reflect increasing risk of death in Population 1 and Population 3. Gamma values used in all survival analyses are greater than one, meaning that the rate of death is naturally increasing. Application of the scaling factor has therefore accelerated this rate of death, potentially "double counting" some mortality and biasing against treatments with a better survival profile. No validation of this approach is provided.

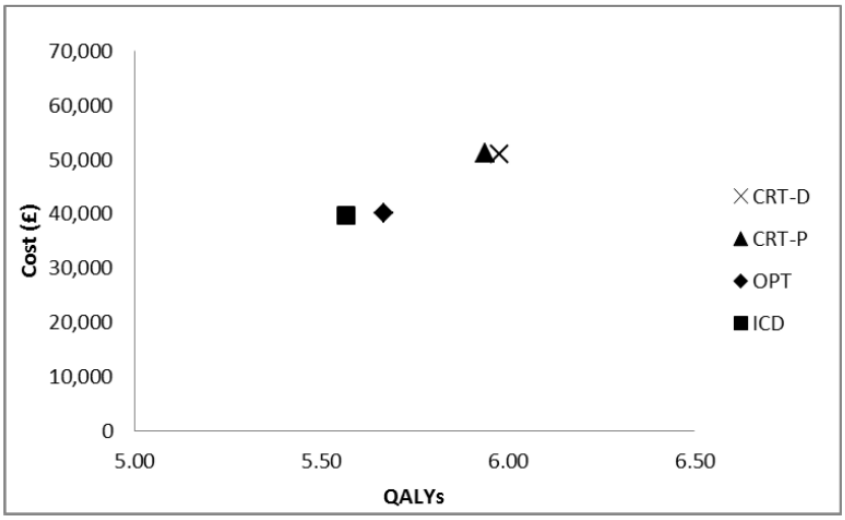
Use of RAFT mortality data in Population 3

Device-specific mortality in Population 3 is modelled by applying treatment effects for each treatment (vs. CRT-D) to the CRT-D mortality rates in RAFT. Although RAFT is a large trial with long follow-up, it cannot alone be viewed as representative of Population 3. Patients in RAFT were predominantly NYHA II (80%), whereas the patient population considered for CRT-D includes both patients with more severe heart failure (NYHA III/IV) and those with NYHA class I.

Lack of face validity of results for Population 3

One of the key tests of any cost-effectiveness model is whether or not the results are clinically plausible. To demonstrate why we believe the results for SHTACs Population 3 fail this simple test, we have reproduced the cost-effectiveness plane from the main report below.

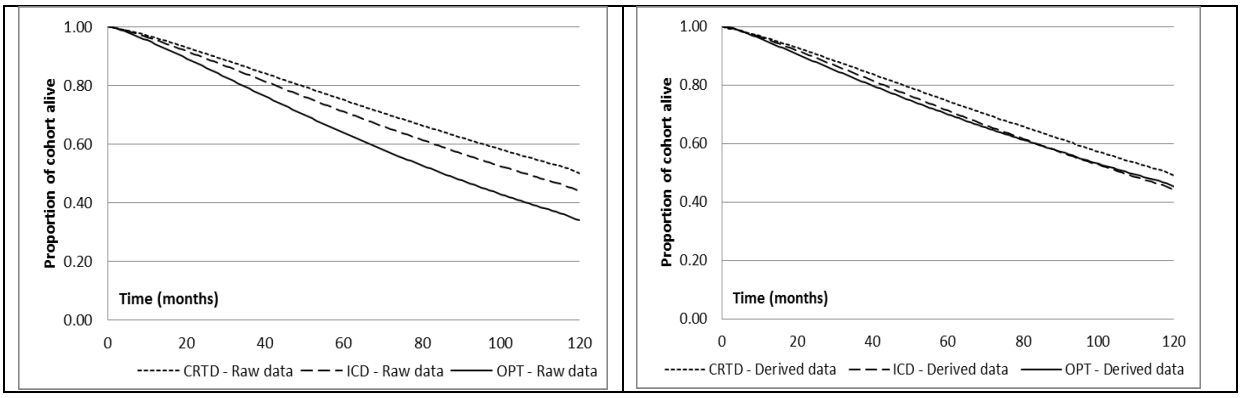
Figure A2: SHTAC base case results for Population 3 (Assessment Group report, p358)



By definition, this patient group are at an increased risk of sudden cardiac death but yet the use of ICD therapy (which includes OPT) results in the generation of fewer lifetime QALYs than were patients treated with OPT alone. This result is clinically implausible and undermines the validity of the results for this group.

We have not been able to undertake a full review of the model but we believe that this anomaly may have occurred due to the very high levels of device crossover. To demonstrate this point, consider the two plots below. The one to the left documents the treatment specific survival (as noted on the 'mortality' worksheet) and the one to the right the proportion of patients alive as derived from information on the relevant Markov traces. The extra benefit for the OPT arm is likely to be driven by the very high rate of cross-over to CRT-D.

Figure A3: Comparison of SHTAC mortality inputs (left panel) and outputs (right panel) in Population 3



Lack of face validity of results for Population 3 compared to Population 2

The fact that CRT-D is associated with poorer cost-effectiveness in Population 3 compared to Population 2 lacks face validity. CRT-D is the treatment arm providing both defibrillation and resynchronisation – the modalities from which a population with an increased risk of sudden death and heart failure despite OPT (Population 3 as defined in the scope) would be expected to benefit

most. This is likely to be again driven by the high level of cross-over in the OPT arm for Population 3 compared to Population 2 (see Table 2).

No quality of life benefit of CRT-P in Population 3

SHTAC state that no benefit is assumed based on the fact that "For Population 3, robust evidence of the effect of devices on heart failure progression was not found". However, the failure to model a quality of life benefit with CRT-devices in this population appears to contradict SHTAC's own systematic review findings. SHTAC state that "An improvement in QoL score was seen with CRT-D when the trials were pooled (MD -6.9, 95% CI -10.4 to -3.4, p=0.0001)" (Assessment Group report p224).

Transplant mortality

20% of transplant patients are still alive after 34 years (i.e. at age 100 in Population 3)

Utility data

NYHA class specific utility data were taken from a study which recruited only patients who had survived an acute MI (not an average device population). The relevance of this study to the current appraisal is debatable.

Systematic inflation of costs in the device arm

Focussing on the model used to generate results for SHTAC population 3, the higher cost of CRT-P when compared to CRT-D also appears implausible. This has occurred for two reasons: i) the very high rates of crossover to CRT-D and ii) incorrect modelling of device related infections and lead displacement.

<u>Reason 1: very high rates of crossover to CRT-D</u>

To explore the impact of device crossover, we have set the probability of hospitalisation due to non-fatal arrhythmia to zero. The impact on treatment specific costs is presented in Table A5. Device specific lifetime costs are lowered by approximately 20-30% for ICD and CRT-D and by approximately 70% for CRT-P. This is driven by the cross-over associated with non-fatal arrhythmia in the model.

Table A5: Impact of non-fatal arrythmia on lifetime device related costs

Treatment pathway	Original cost	Cost if parameter iHA_CRT set to zero	Difference
ICD + OPT	£40,006	£28,046	£11,960
CRT-P + OPT	£51,202	£16,868	£34,334
CRT-D + OPT	£50,911	£40,369	£10,543

Of note is the fact that the device specific probabilities of a non-fatal arrhythmia requiring hospitalisation for individuals in Population 3 is very different to those in Populations 1 and 2. The values used in the model are replicated in Table A6. The large discrepancy between the values for Population 3 and for populations one and two (with the former being approximately, 400% higher than the latter) is supportive of the argument that device costs are inflated.

Table A6: Device specific four week probabilities of non-fatal arrhythmia requiring hospitalisation

	Population 1[a]	Population 2[b]	Population 3[c]
ICD + OPT	0.0075	0.007	0.032
CRT-P + OPT	N/A	0.007	0.029
CRT-D + OPT	N/A	0.007	0.029

N/A: not applicable; **a)** reproduced from SHTAC report Table 96; **b)** reproduced from SHTAC report Table 100; **c)** CRT-D value reproduced from SHTAC report table 103, ICD and CRT-P values derived from information in SHTAC report table 103.

The model includes defribillation upgrades following non-fatal arrhythmic hospitalisations (CRT-P to CRT-D; OPT to ICD or CRT-D). However, when non-fatal arrhythmic hospitalisations are set to zero (iHA_CRTD=0, reflecting a situation where nobody is at risk of the event) the number of patients being upgraded from ICD to CRT-D reduces from 48 to 8. It is unclear why the number of arrhythmic events is impacting on the number of switches to resynchronisation therapy. However, a brief review of the model suggests that this may be due to the joint modelling of arrhythmic and heart failure hospitalisations as one health state.

Reason 2: Incorrect modelling of device related infections and lead displacement

Lead displacements are predicted to occur in approximately 43% of CRT-P and CRT-D patients in Population 3 (cells D20 and E20, worksheet 'DetailedResults'). The corresponding values in population 2 for both devices are approximately 30% (cells AR24 and AS24, worksheet 'DetailedResults'). These figures are much higher than published long-term rates and appears to reflect the fact that SHTAC have applied the same event probability every month. Clinically, patients are at an increased risk of such a displacement for a short period of time after implant, after which, the risk is much lower. By comparison, the weighted average of reported rates in the studies from which the data was derived (and arguably more reflective of the actual lifetime risk) is approximately 6% (data sources as per SHTAC report p313).

Similarly, once the device is in place and the wound healed, the risk of device related infection is much lower than in the peri-operative period. Again, SHTAC have applied a common probability in all cycles a patient is alive resulting in implausibly high probabilities of infection. The event occurs in approximately 7.5% of CRT patients in Population 3 and approximately 5% of CRT patients in Population 2. The weighted average event rate in the studies used to inform the model parameter (again, arguably more reflective of lifetime risk) is approximately 1.5% (data sources as per SHTAC report p313).

Surgical failures are assumed to occur in approximately 8% of patients implanted with CRT-P or CRT-D every four week cycle. Again failure is more likely in the peri-operative period thus the approach exaggerates device failure. In addition to this, on failure of a CRT-D, the additional total cost of an ICD implant is incurred (£15,248, value taken from worksheet 'Inputs'). This is unrealistic. The vast majority of CRT-D system failures relate to pacing or sensing problems requiring the reposition or replacement of a lead, not a new defibrillation device. This clearly carries a much lower cost.

It is also illuminating to compare the predicted device specific event rates in the SHTAC model to those generated by the model developed during TA120. In this previous model, the lifetime likelihood of lead displacement for CRT-P and CRT-D was 10.3% and 11.6% respectively. The current SHTAC models therefore differ from those in this model by a factor of at least three.

Quantifying the impact of this error on lifetime costs is challenging since it is not possible to include time dependant values into the model structure for both parameters. What is clear, however, is that the SHTAC models are predicting far too many events when compared to either the clinical data or the model developed during TA120, resulting in an increase in lifetime costs for all devices.

Device longevity data

Despite being provided with device specific longevity data for approximately 40,000 implants sourced from the UK National Health service, SHTAC largely disregarded these data on two counts:

i) By assuming each device lasts for a fixed period of time, after which all will fail immediately. The ABHI model included device replacement as a time dependant function and the parameters used are listed in the SHTAC dossier (p327). These data show the rate of failure is not constant. The approach used is therefore inconsistent with what is observed in clinical practice.

ii) By performing a sensitivity analysis based on mean device specific survival times from the previous CRT appraisal. The justification for this was 'clinical advice'. We cannot understand how seven year old crude estimates of device longevity were thought to more accurately reflect current UK practice than up-to-date NHS data and would request the committee disregard these sensitivity analyses.

Appendix 3: Corrected treatment effect presentation (statistically significant results highlighted in bold)

Subgroup defining covariables				Results submitted to NICE (mean, 95% CI)			Revised results (mean, 95% CI)		
QRS	**LBBB**	**AGE**	**GEN-DER**	**CRT-D OMT**	**CRT-P OMT**	**ICD OMT**	**CRT-D OMT**	**CRT-P OMT**	**ICD OMT**
QRS<120	N	<60	F	NA	NA	[redacted]	NA	NA	[redacted]
QRS<120	N	<60	M	NA	NA	[redacted]	NA	NA	[redacted]
QRS<120	N	>=60	F	NA	NA	[redacted]	NA	NA	[redacted]
QRS<120	N	>=60	M	NA	NA	[redacted]	NA	NA	[redacted]
QRS>=120-<150	N	<60	F	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=120-<150	N	<60	M	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=120-<150	N	>=60	F	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=120-<150	N	>=60	M	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=120-<150	Y	<60	F	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=120-<150	Y	<60	M	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=120-<150	Y	>=60	F	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=120-<150	Y	>=60	M	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=150	N	<60	F	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=150	N	<60	M	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=150	N	>=60	F	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=150	N	>=60	M	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=150	Y	<60	F	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=150	Y	<60	M	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=150	Y	>=60	F	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
QRS>=150	Y	>=60	M	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]