Full title: A Proof of Concept Study to Investigate the Feasibility of Targeted Release of Doxorubicin from Lyso-thermosensitive Liposomal (LTSL) Doxorubicin (ThermoDox®) Using Focused Ultrasound in Patients with Primary or Secondary Liver Tumours

Statistical Analysis Plan

Version 3.0 – 22 Jul 2016

Based on version 3.0 of protocol (06 Apr 2016)

	Name	Title/Role	Signature	Date
Author	Paul Lyon	Clinical Fellow		
Author	Lang'O Odondi	Medical Statistician		
Reviewer	Lucy Griffiths	Trial Coordinator, OCTO		
Reviewer	Meena Patil	Data Manager, OCTO		
Reviewer	Robert Carlisle	Associate Professor of Biomedical Engineering		
Approver	Lang'O Odondi	Medical Statistician		
Approver	Mark Middleton	Professor, Department of Oncology, Chief Investigator		

Oxford Clinical Trials Research Unit (OCTRU) and Centre for Statistics in Medicine (CSM)

Sponsored by the University of Oxford

CONTENTS

1 INTRODUCTION

This document details the proposed presentation and analysis for the main publications reporting results from the NIHR and Celsion jointly funded Single Centre Non-Randomised Phase 1 Clinical Trial of targeted chemotherapy using focused ultrasound, in the treatment of liver tumours (TARDOX). The results reported in these publications should follow the strategy set out here. Subsequent analyses of a more exploratory nature will not be bound by this strategy, though they are expected to follow the broad principles laid down here. The principles are not intended to curtail exploratory analysis (for example, to decide cut-points for categorisation of continuous variables), nor to prohibit accepted practices (for example, data transformation prior to analysis), but they are intended to establish the rules that will be followed, as closely as possible, when analysing and reporting the trial.

The analysis strategy will be available on request when the principal papers are submitted for publication in a journal. Suggestions for subsequent analyses by journal editors or referees, will be considered carefully, and carried out as far as possible in line with the principles of this analysis strategy; if reported, the source of the suggestion will be acknowledged.

Any deviations from the statistical analysis plan will be described and justified in the final report of the trial. The analysis should be carried out by an identified, appropriately qualified and experienced statistician, who should ensure the integrity of the data during their processing. Examples of such procedures include quality control and evaluation procedures.

1.1 Key personnel

Chief Investigator
Professor Mark Middleton
Department of Oncology, University of Oxford
Oxford Cancer and Haematology Centre
Churchill Hospital
Oxford OX3 7LE
Mark.Middleton@oncology.ox.ac.uk

Senior Scientific Researcher
Professor Robert Carlisle
Associate Professor of Biomedical Engineering
Institute of Biomedical Engineering
Old Road Campus Research Building
University of Oxford
Oxford OX3 7DQ
Robert.Carlisle@eng.ox.ac.uk

Clinical Researcher & Co-Investigator
Mr. Paul Lyon
NIHR Clinical Research Fellow (Nuffield Dept. Surgical Sciences)
Institute of Biomedical Engineering
Old Road Campus Research Building, University of Oxford
Oxford OX3 7DQ
Paul.Lyon@nhs.net

Trial Coordinator
Lucy Griffiths
Oncology Clinical Trials Office (OCTO)
Department of Oncology
University of Oxford
Old Road Campus Research Building
Roosevelt Drive
Oxford OX3 7DQ
OCTO-TARDOX@oncology.ox.ac.uk

Data Manager

Marian Taylor
Oncology Clinical Trials Office (OCTO)
Department of Oncology
University of Oxford
Old Road Campus Research Building
Roosevelt Drive
Oxford OX3 7DQ
Marian.Taylor@oncology.ox.ac.uk

Statistician

Lang'O Odondi
Oncology Clinical Trials Office (OCTO)
Department of Oncology & Centre for Statistics in Medicine
Botnar Research Centre
Windmill Road
Oxford OX3 7LD
Lango.Odondi@oncology.ox.ac.uk

DSMC Members: (if appropriate)

There is no Data and Safety Monitoring Committee (DSMC) for this study and an Independent Trial Steering Committee (ITSC) will be in place to monitor safety.

Key TMG Members:

Prof. Middleton (Chair)	Mark.Middleton@oncology.ox.ac.uk
Mr. Paul Lyon	Paul.Lyon@nhs.net
Mrs Lucy Griffiths	OCTO-TARDOX@oncology.ox.ac.uk
Prof. Constantin Coussios	Constantin.Coussios@eng.ox.ac.uk
Lang'O Odondi	Lango.Odondi@oncology.ox.ac.uk

Other/Optional TMG Members

Prof. Fergus Gleeson	FGleeson@mac.com
Prof. Feng Wu	Feng.Wu@nds.ox.ac.uk
Prof. Robert Carlisle	Robert.Carlisle@eng.ox.ac.uk
Miss Chrissie Butcher	OCTO-TARDOX@oncology.ox.ac.uk
Dr. Michael Gray	Michael.Gray@eng.ox.ac.uk
Dr. Victoria Woodcock	Victoria.Woodcock@oncology.ox.ac.uk

Additionally, other Oxford University Hospital Trust or Oxford University staff may be invited to a TMG meeting, where their expertise is required.

2 BACKGROUND INFORMATION

The development of distant spread (metastases) is a leading cause of death in patients with advanced solid tumours. The liver is one of the most common sites for metastatic disease. Studies have shown that the treatment of distant metastases improves survival. Systemic chemotherapy is the current treatment of choice for patients with inoperable liver metastasis, however long-term survival is rare. The majority of primary liver cancers (hepatocellular carcinoma (HCC) and cholangiocarcioma) are unresectable and survival remains dismal. Post-hoc analysis of the Phase III HEAT study is indicative that ThermoDox® treatment improves survival in patients with HCC, where Radiofrequency Ablation (RFA) was optimal.

2.1 Study Hypothesis

This proof of concept study proposes targeted delivery of broad-spectrum cytotoxic agent (doxorubicin), via a specially formulated LTSL (ThermoDox®) activated by mild hyperthermia, as both a feasible and effective solution to the drug-delivery problem for the same systemic dose. It is proposed that this method of targeted drug delivery is clinically applicable, and may achieve increased tumour uptake and local dose for the equivalent dose of doxorubicin used in systemic chemotherapy, which has a well-established and safe toxicity profile.

Key Study Hypothesis: whilst ThermoDox® is circulating systemically at therapeutic levels, doxorubicin concentration in the target tumour will be significantly increased following its exposure to safe levels of hyperthermia, as delivered by a clinically-approved focused ultrasound (FUS) device.

Further research hypotheses which relate to this key study hypothesis are listed in section 3.1.

2.2 Key Definitions

As defined in the protocol, the following definitions are assumed in the following statistical plan:

"Intervention" refers to attempted targeted release of ThermoDox® by mild hyperthermia using a focused ultrasound.

The "desired range of mild hyperthermia" is used in the context of ThermoDox® release in this study and is defined as achieved bulk tissue temperature in the range of 41-47^{o}C.

"Optimal FUS exposure parameters" are defined as those combinations of settings on the FUS device (power, duty cycle and transducer motion plan) which result in an adequate temperature rise to successfully release doxorubicin from the LTSLs in the target tumour, without causing direct thermal tissue damage.

An "evaluable participant" is defined as a participant who has received FUS following delivery of ThermoDox® in Part I or II, and for whom it has been possible to evaluate the outcome of the intervention based on biopsy or plasma samples, or radiologically.*

Please note, with regards to the primary and secondary endpoints, the definition for 'evaluable participant' does not distinguish those participants having 'optimal FUS' for drug delivery, as defined in Table 5. This means participants having both optimal and sub-optimal FUS will be included in analysis of these endpoints. However with regards to the tertiary endpoints, as stated in the protocol, these will only be evaluated for those participants having optimal FUS.

An "evaluable sample" is defined as a sample for which sufficient material was available for analysis and the analytical technique has provided quantifiable and interpretable data.

*Please note that a participant does not require the full complement of evaluable samples and follow-up

outcome data in order to be an evaluable participant for inclusion in the relevant subset of endpoints for analysis. Thus, more specifically, a participant is evaluable with respect to outcome measure(s), for example:

- In the case where a participant undergoes the intervention including plasma and biopsy samples, but is then lost to radiological follow-up for whatever reason, that participant will still be deemed evaluable with respect to the endpoints involving biopsy and plasma analysis.

- In the case where a participant undergoes the intervention including plasma samples, and radiological follow-up, but whose biopsies non-evaluable due to being misplaced, incomplete, inadequate or failed analysis, that participant will still be deemed evaluable with respect to the endpoints involving plasma analysis and radiological follow-up. An exception is allowed in Part I; if only the Released biopsy is evaluable, the primary endpoint may still be assessed with respect to the second criterion (total drug concentration ~~below~~ above minimum threshold)

- In the case of tertiary (exploratory) endpoints, analytical methods such as MR spectroscopy remain under evaluation, and consequently they may not be performed in each case, but this is not at the cost of other radiological endpoints, which will still be evaluable.

2.3 Unreleased and Released Nomenclature

At the time of the intervention, biopsy and plasma samples will be taken at defined time periods as per the treatment timeline (see figures 4 and 5 of the protocol for Part I and II respectively). The naming convention of the biopsy and plasma samples, 'Pre-treatment', 'Unreleased' and 'Released', relates to the *expected* representation in terms of targeted ThermoDox® release and may not necessarily translate to the true biological picture. In particular, 'Unreleased' samples are taken after systemic exposure to ThermoDox®, but before exposure to the final dose of FUS, whereas 'Released' samples are taken after the final dose of FUS. In other words, 'Unreleased' samples are obtained at the time point at which it is anticipated that the target tumour is exposed to therapeutic levels of circulating encapsulated doxorubicin, whereas 'Released' samples are obtained at the time point where it is hypothesised that encapsulated drug is released as bioavailable (free) intra-tumoural doxorubicin following FUS.

We note that due to the selected analytical technique (HPLC) used to determine doxorubicin concentration in biopsy tissue obtained for this study, it is not possible to distinguish between liposomal (encapsulated) and free doxorubicin levels; only the total doxorubicin concentration is evaluable and is directly applicable to the primary endpoint. The same holds true for plasma analysis, applicable to a tertiary endpoint, which is evaluated by the same analytical technique.

2.4 Primary, Secondary and Tertiary Objectives & Endpoints

The objectives and endpoints for the study are detailed in the following tables. Only evaluable samples will be included in endpoint analysis. Each endpoint has been designed to be independent of the others, which is why a given participant may only be evaluable for a subset of endpoints. All endpoints will be evaluated at the end of Part II.

The following qualifiers (*,†, ‡ and §) apply to **Table 1**, **Table 2** and **Table 3**:

* In Part II the Unreleased biopsy is not performed and therefore the average value for all evaluable tumours receiving intervention in Part I is used as a comparison for the two-fold increase from Unreleased to Released biopsy.

† Duration refers to the completion of sample processing but not necessarily the sample analysis. The sample may be stored in a fixed or frozen state and analysed at some time period beyond the specified duration, for example to facilitate batch processing.

‡ Analytical method will only be performed where tissue mass allows. It is expected at most one technique can be performed in addition to the primary endpoint for a given sample. Where biopsy mass falls short, the primary endpoint will be prioritised.

§ Must be shown in at least 50% of evaluable participants.

Primary Objective	**Primary Endpoint**	**How and when data is captured**
To determine whether targeted release of doxorubicin from ThermoDox® ('drug') using mild hyperthermia generated non-invasively by focused ultrasound (FUS) is feasible in cancer patients	A demonstrable two-fold increase in*, or value exceeding 10µg/g of, the concentration of intra-tumoural doxorubicin at the treated tumour site following FUS-induced mild hyperthermia § Quantification of intra-tumoural drug release will be achieved by direct analysis of the intra-tumoural biopsy samples using standardised curve(s).	Analytical chemistry performed on a recorded mass of biopsy sample processed within 12 hours of intervention†

Table 1: Primary Objective and Endpoint for the TARDOX Study

Secondary Objectives	**Secondary Endpoints**	**How and when data is captured**
Part I only: To determine optimal FUS exposure parameters for a range of participant Body Mass Indices (BMIs) and tumour locations within the liver	Achievement of the desired range of mild hyperthermia in the target tissue as monitored by the implanted thermistor during FUS exposure	Real-time thermometry monitoring during intervention
To assess the safety of FUS-induced mild hyperthermia for drug delivery	1. Persistence of cell viability stain or percentage of cell or tissue necrosis under 30% following FUS exposure, as assessed by cytological or histological methods[1] ‡,§ 2. Adverse events deemed related to FUS up to 30 days post-intervention	1. Cytological or histological analysis of biopsy samples processed within 96hr of intervention† 2. Adverse event recording for 30 days post-intervention
To assess the local and systemic cytotoxic effects of ThermoDox® in this setting	1. Significant bone marrow suppression, deranged liver function and liver toxicity 2. Adverse events deemed related to ThermoDox® up to 30 days post-intervention	1. Grade 3 and 4 laboratory results from blood tests at Day 1 and Day 15 post intervention 2. Adverse event recording for 30 days post-intervention

Table 2: Secondary Objectives and Endpoints for Parts I & II of the TARDOX Study

[1] In some necrotic tumours, base line necrosis may be high (e.g. >30%) even prior to treatment. Consequently an increase in necrosis by no more than 25% over that of a suitable control can be used as an alternative endpoint.

Regarding participants for which doxorubicin has been released by optimal FUS exposure parameters[2]:

Tertiary (Exploratory) Objectives	**Tertiary (Exploratory) Endpoints**	**How and when data is captured**
1. To establish the validity of other minimally-invasive and non-invasive methods of evaluating intratumoural uptake of doxorubicin as an alternative to that used in primary endpoint	Combined analysis of relevant end points to establish most effective alternative analytical method(s) for quantifying drug release	See below
a. Fluorescence microscopy of biopsy	Released biopsy samples exhibiting regions of nuclear or background fluorescence levels at least twice that of the background levels of a suitable control, as assessed by fluorescence microscopy ‡,§	Microscopy analysis of biopsy samples processed within 96hr of intervention†
b. MR-Spectroscopy	MRI/MR-Spectroscopy indicating[3] targeted delivery of doxorubicin to the tumour by demonstrating a statistically significant increase in the tumour: background ratio of signal intensity following intervention compared to baseline	MRI/MR-Spectroscopy signal within 24 hours of intervention
c. Plasma pharmacokinetics	A statistically significant decrease in total doxorubicin plasma levels over time-matched clearance levels, following FUS-induced release	Analytical chemistry of plasma samples processed within 12hr of intervention†
2. To determine whether the dose of doxorubicin released has a therapeutically significant effect on the target tumour	Radiological evidence of tumour response over up to 60 days in the target tumour alone, as assessed by principles of CHOI and RECIST response evaluation using MRI and CT [1, 2] and SUV_{max} using PET-CT [3] §	Scans to be performed within 60 days of intervention.

Table 3: Tertiary (Exploratory) Objectives and Endpoints for Parts I & II of the TARDOX Study

[2] "Optimal FUS exposure parameters" are defined in section 2.2.
[3] The spectroscopy signal analysed may not be doxorubicin itself; metabolites or breakdown products of ThermoDox® may be utilised.

2.5 Study Design

The TARDOX trial is designed as a Phase 1 prospective non-randomised safety cohort study, with all patients recruited from a single UK site (Oxford). The study has an open label design with all participants receiving ThermoDox® and FUS. The study is split into two parts. Part I is required to identify optimal FUS exposure parameters using real time thermometry data from an implanted thermistor. Part II does not require thermistor implantation, and is designed to more accurately reflect how the therapy would be implemented in clinical practice. All evaluable participants from both Part I and Part II will be included in the endpoint analysis.

Part I

For each participant, a single liver tumour will receive the intervention and follow-up imaging. The ability to achieve targeted release of doxorubicin from ThermoDox® at the tumour site will be determined using three tumour biopsies with corresponding peripheral blood samples, taken during intervention, and/or using a radiological method (MR-Spectroscopy).

Part II

For each participant, a single liver tumour will receive the intervention and follow-up imaging. Where available, a second 'control' tumour, exposed to ThermoDox® alone, will also be identified for follow-up imaging. The optimal FUS exposure parameters determined in Part I will be selected for Part II interventions, accounting for anatomical differences. The ability to achieve targeted release of doxorubicin from ThermoDox® at the tumour site will be determined using the same methods as for Part I, with the exception of requiring only one tumour biopsy.

In the first 60 days following intervention, both Part I and II participants will be followed up with a maximum of two CTs, two PET-CTs and three MRI/MR-Spectroscopy scans. Detailed criteria for evaluation of efficacy and safety are described in subsequent sections.

Date of start of recruitment:	04Jul2014
Date of expected end of recruitment:	31Mar2017
Date expected end follow-up:	30Apr2017
Date expected analysis:	01Sep2017
Target number of subjects:	Up to a maximum of 28 evaluable participants will be recruited to the study overall.
Participating Centres:	One

2.6 Eligibility

Inclusion Criteria

Patients will only be eligible for inclusion in this study if all of the following criteria apply:

1) Pathologically confirmed advanced solid tumour with liver metastasis suitable for intervention (as assessed by ultrasound or other radiological methods). In addition confirmed primary liver tumours (hepatocellular carcinoma or cholangiocarcinoma) can be included.

2) Will have progressed or remained stable on conventional chemotherapy.

3) Male or Female, Age ≥ 18 years.

4) Have life expectancy of ≥ 3 months.

5) Left Ventricular Ejection Fraction (LVEF) ≥ 50% on echocardiogram.

6) Have not received radiotherapy to the target area within the preceding 12 months.

7) A World Health Organisation (WHO) performance status of ≤ 1 (Appendix 1).

8) Able and willing to give written informed consent, indicating that they are aware of the investigational nature of this study and potential risks, and able to comply with the protocol for the duration of the study, including scheduled follow-up visits and examinations.

Exclusion Criteria
Patients will be ineligible for recruitment to the study if any of the following conditions hold:

1) Have surgery or other procedure requiring general anaesthesia planned to be undertaken during the period of the study.

2) Have serious illnesses including, but not limited to, congestive heart failure (NYHA class III or IV functional classification); life threatening cardiac arrhythmia; or myocardial infarction or cerebral vascular accident within the last 6 months.

3) Have on going significant infection (chest, urine, blood, intra-abdominal).

4) Have uncontrolled diabetes.

5) Have received a life-time dose of doxorubicin > 400mg/m^2 or a life-time dose of epirubicin > 800mg/m^2 or any dose or both.

6) Pregnant or breast-feeding. In women of childbearing potential, a negative pregnancy test (serum) is required within 30 days prior to study intervention.

7) Female participants of child bearing potential and male participants whose partner is of child bearing potential who are not willing to practice an acceptable form of contraception (i.e. oral contraceptive, diaphragm, cervical cap, condom, surgical sterility) during the study and for 6 months thereafter. Women whose partner has or men who have undergone a vasectomy must use a second form of birth control.

8) Have any known allergic reactions to any of the drugs or liposomal components or intravenous imaging agents to be used in this study.

9) Have portal or hepatic vein tumour invasion/thrombosis.

10) Inadequate haematological and biochemical function (see section 4.1.2 of protocol)

11) Have contraindications to receiving doxorubicin including prior sensitivity (rash, dyspnoea, wheezing, urticarial or other symptoms) attributed to anthracyclines or other liposomal drugs.

12) Use of chemotherapy or of an investigational drug within 30 days or 5 half-lives, whichever is longer, preceding the intervention.

13) Have medically significant active infection.

14) Have Child-Pugh Class C liver disease or Class A-B with encephalopathy and/or refractory ascites (Appendix 3).

15) Documented HIV positive.

16) Documented diagnosis of haemochromatosis.

17) Documented history of contrast-induced nephropathy.

18) Have any of the following contraindications for liver biopsy:

a) Suspected liver haemangioma or other vascular tumour

b) Tense ascites

c) Known cystic liver disease[4]

d) Extra-hepatic biliary obstruction[4]

19) Other medical or psychiatric conditions or laboratory abnormalities that the investigator considers would make the patient a poor trial candidate.

2.7 Treatment Interventions

There is a single intervention day for Parts I and II, during which the components of the intervention are performed as outlined below.

Part I

Target tumour is exposed to FUS alone, and then a combination of FUS and ThermoDox®, under General Anaesthetic. Temperature readings are recorded during FUS exposure. Three liver tumour biopsies and corresponding peripheral blood samples are taken for analysis.

Part II

Target tumour is exposed to a combination of FUS and ThermoDox®, under General Anaesthetic. Two biopsies at a single timepoint are taken of the target tumour receiving intervention, along with three peripheral blood samples for analysis.

2.8 Sample Size

Up to a maximum of 28 evaluable participants[5] will be recruited to the study across two separate Parts (Part 1 and Part 2). A minimum of 5 participants are required in Part 1 before Part 2 can be opened. Non-evaluable participants with respect to the primary endpoint may be replaced if still within the recruitment period. Primary, secondary and tertiary endpoint will be applied to all evaluable endpoints for all evaluable participants, unless specified otherwise. There will be an interim analysis (see section 7). Sample size will not be expected to change during the trial.

[4] Relative contraindications only and may be non-exclusive at discretion of the study team

[5] An "Evaluable participant" is defined in section 2.2 (Key Definitions).

The relatively small sample size was determined based on the available resources and structure of trial funding, with consideration to this being a Phase I proof of concept study, and the anticipated limited cohort of patients suitable for enrolment. The sample size aims to recruit enough patients within a reasonable time frame to allow sufficient statistical power of study outcome measures to help inform design (Phase II-III) studies with higher statistical power in the future.

2.9 Strategies for achieving adequate recruitment

Recruitment prediction is for 28 patients over 24 months, on average 14 patients annually from the initially selected combined patient cohort; liver metastasis from a colorectal, breast or lung primary, and additional patient cohort; primary liver cancers and all other liver metastasis. It may be that the trial closes before the full quota of 28 evaluable patients has been achieved.

Colorectal carcinoma accounts for approximately 70% of liver metastases, therefore it was originally anticipated that the majority of potential participants were likely to be from this patient group. Approximately 100 patients with newly diagnosed colorectal liver metastases are referred to the Oxford Cancer Centre for treatment every year. It is anticipated that of these patients, approximately 50% (50 patients) will be offered chemotherapy, 30% will undergo surgery or other treatment such as ablation therapy, and the remaining 20% will be considered for entry into this study. From this 20%, it was estimated that just over half would offer consent and meet the eligibility criteria, equating to recruitment of approximately 12 patients per year. It was predicted that an additional 4 eligible and consenting patients per year would be recruited from the other presenting groups (breast or lung primary).

Although not originally considered for inclusion, primary liver cancers are the third highest cause of cancer-related death globally [4] with around 3000 new cases per year in the UK. HCC, also known as malignant hepatoma, is the most commonly occurring primary liver cancer accounting for 90%, and usually results from chronic liver conditions such as hepatitis B and C and/or cirrhosis. Cholangiocarcinoma (CCA), cancer of the biliary tract, is the largest contributor to the remaining primary liver cancers and is the most common cause of death from primary liver cancer in the UK, where it has an annual mortality rate of 1500 [5]. Although the combined incidence of HCC and CCA is less than the much larger cohort of colorectal liver metastases, due to lack of alternative local therapies, it is anticipated this study will offer a unique treatment option and thus a higher referral and recruitment rate. Furthermore at this stage, for this Phase I study which focuses on *drug delivery*, it is prudent to open up the inclusion criteria to liver metastases from any advanced solid tumour. In addition to improving recruitment rates, observing tumour response in a wide range of tumour types will better inform design of future Phase II/III studies, which will shift focus to *response* and overall survival. Following the protocol amendment (V2.0) to additionally include liver metastasis from any confirmed advanced solid tumour and primary liver tumours, it is anticipated that recruitment will be recoverable to meet the original predictions defined above.

Approaches to promote better recruitment include widening patient population to any liver tumour type, stream-lining patient screening to allow consent for ultrasound screening in advance of the main study consent, opening of patient identification centre (PIC) sites and extended recruitment period until March 2017. A protocol amendment has been effected to allow Part I to remain open in parallel to Part II, i.e. allow Part II to open sooner hence more flexibility on the location of the target tumour and extending selection cohort without excluding patients who can only be treated as Part I participants. In addition, Part II is likely to be a considerably shorter intervention and therefore may be more acceptable to referring clinicians and patients alike in addition to giving more flexibility in available anaesthetic booking times.

2.10 Randomisation

No randomisation is required in this Phase I proof of concept study.

3 HYPOTHESES AND OUTCOME MEASURES

3.1 Clinical Research Hypotheses

The study proposes the following clinical research hypotheses and the primary, secondary and tertiary outcome measures, or endpoints, are designed to test them:

1. For the target tumour, following exposure to optimal (sub-ablative) levels of FUS for mild hyperthermia[6], there will be persistence of cell viability, or absence of tissue or cellular damage, which would otherwise be expected with ablative levels of hyperthermia

2. For the target tumour, whilst there are therapeutic levels of circulating ThermoDox®, exposure to optimal levels of FUS for mild hyperthermia will result in increased total intra-tumoural concentration of doxorubicin

3. Following optimal FUS treatment to the target tumour, there is release of *bio-available* doxorubicin from ThermoDox®, which is then able to bind to cellular DNA.[7]

4. For the target tumour, when compared with any available non-target (control) tumour in the liver (receiving ThermoDox® alone), there will be improved radiological response following combined treatment (receiving ThermoDox® and FUS)

3.2 Definition of Primary, Secondary and Tertiary Outcomes

Success of this study requires evidence of successful targeted drug release, in other words a quantifiable increase of doxorubicin concentration in the tumour following exposure to optimal levels of FUS for mild hyperthermia. A number of alternative analytical methods will be employed in order to obtain this evidence, each with its own statistical requirement for success, as shown in Table 4. The primary endpoint will employ a robust and validated analytical technique (HPLC), but at the same time requires an invasive biopsy. As tumour biopsy is already required for the primary endpoint, it is prudent to obtain as much information from this biopsy as possible. Consequently, in Part I, if enough material is available, each biopsy will be divided for microscopy for secondary and/or tertiary endpoints, using two approximate halves. In Part II, a single 'Released' biopsy will be taken for HPLC immediately followed by another for analysis of the secondary and tertiary endpoints by microscopy, and so more tissue will be available. For the secondary biopsy-related endpoint, microscopy will be used to quantify levels of tissue necrosis, following optimal levels of FUS exposure, to explore hypothesis (1). For the tertiary biopsy-related endpoint, fluorescence microscopy will be used to quantify intra-tumoural drug concentration (hypothesis (2)) and its distribution, in particular the presence of nuclear doxorubicin (hypothesis (3)).

If this study leads on to Phase II-III trials, it might be that other non-invasive methods of estimating intra-tumoural drug quantification can be employed. For this reason such non-invasive exploratory objectives have also been included in this trial, namely MR-Spectroscopy and Plasma Pharmacokinetics (PKs).

---

6 "Mild hyperthemia" is defined in section 2.2, Key Definitions.

7 When ThermoDox® is in its unreleased form it is non-bioavailable and thus cannot bind to cellular DNA.

4 PATIENT GROUPS FOR ANALYSIS

An "evaluable participant" is defined in section 2.2, Key Definitions. All evaluable participants in both Part I and Part II will be included in the final analyses. All patients enrolled in the study, will be accounted for and included in the analyses. The number of patients who were not evaluable, who died or withdrew before treatment began will be recorded.

Variables will be analysed to determine whether the criteria for the study conduct are met. This will include a description of patients who did not meet all the eligibility criteria, an assessment of protocol violations and other data that impact on the general conduct of the study.

Baseline characteristics will be summarised for all recruited and treated patients as detailed in section 6.2. Patients who died or withdrew before treatment started or do not complete the required safety observations will be described and evaluated separately.

Treatment related toxicity will be tabulated by type and grade of toxicity as detailed in Section 8 (Table 9 and Table 10). There will also be a table of trial treatment and observed adverse events. All patients will be evaluable for toxicity from the time of their intervention using NCI CTCAE v 4.0 scoring.

5 QUALITY CONTROL AND DATA VALIDATION

The following measures will be taken to ensure quality control and data validation:

- Trained radiologists and/or study clinicians will perform assessment of radiological outcome measures.
- In terms of data quality, the biopsy (for primary endpoint) and plasma pharmacokinetic analyses will be performed in Good Clinical laboratories Practice (GCP) by GCP-trained staff using validated techniques.
- Trained NHS, University of Oxford laboratory staff and/or study team members will perform microscopic analyses. A qualified histopathologist will perform reporting of haematoxylin and eosin (H&E) slides.
- In the case of HPLC analysis pertaining to the biopsy analysis for primary endpoint[8] and plasma analysis for tertiary endpoint:
    - An internal standard will be used to normalise the HPLC curves, according to GCP principles.
    - Quality controls of high, medium and low concentration will be used before and after the patient sample chromatography run, and checked to ensure all QCs are in specification to ensure assay validity (see GCP GI/042/1 SOP).
    - A single individual will perform the analytical technique, validation checks and post-processing data analysis.
    - A second individual will independently check the results including validation, entry of correct tumour weights and correct transcription of the data results.
    - A summary of the concentrations of the QC and patient samples are printed using Empower

[8] The analytical technique (HPLC) can be performed only once per biopsy sample due to limited tissue availability and thus best efforts are made to ensure that samples are successfully processed on the first (and only) attempt.

software. These will be signed and dated by the analyst and checked, signed and dated by an independent checker. This document will be stored in the trial file; a copy will be stored with the chromatograms. For reporting purposes, patient data are exported into an Excel spreadsheet and checked against the original. These will be signed and dated by the analyst and an independent checker prior to submission of the data.

  - Although HPLC results are not generally open to interpretation, in the case of any disagreements, a third party will perform an additional independent assessment on the primary endpoint data used in the final analysis.

  - The departmental leads for the Bioanalysis Core (Dr Lisa Folkes or Qualified Deputy) and the Churchill GCP Laboratories (Dr Kevin Myers) will sign off the resulting reports.

- Where tissue quantities allow, microscopy analysis will be performed with averaged multiple observations ($n \geq 3$) to eliminate spurious results.

The TARDOX Trial Office may instigate audits as and when required. Advanced notice will be issued. Site staff are required to co-operate with the requirements of the audit and allow auditors Direct Access to all source information.

A review of the final statistical report will be formed by a separate independent individual and documented in the Statistical report to ensure principles of this Statistical Analysis Plan have been followed.

6 DATA ANALYSIS

6.1 Outcomes & Endpoint Analysis

Part I and Part II data will be combined for overall analysis of all endpoints at the end of the study, unless otherwise specified. Table 4 provides details of the statistical analysis of the primary outcome measure for this Phase I study (i.e. a two-fold increase in intra-tumoural doxorubicin concentration in at least 50% of evaluable participants).

Specifically, Table 4 enumerates the methods of intra-tumoural drug quantification under evaluation; a cross (†) implies the observation will be performed at the following time points during the study intervention:

- Pre-treatment (Part I only)
- Unreleased (Part I only)
- Released (Parts I and II)

For more information on the Pre-treatment, Unreleased and Released nomenclature, please refer to section 2.2, Key Definitions. Timelines incorporating the time points are shown schematically in Figure 4 and Figure 5 of the protocol.

Table 5 covers the secondary endpoints, which concern safety aspects of the study and achieving optimal FUS hyperthermia, relating to hypothesis (1).

Table 6 covers the remaining tertiary endpoint, concerning tumour response, relating to hypothesis (4).

A double cross (‡) applies to Table 4 and Table 6 and implies that only patients receiving optimal FUS for drug release (as defined in Table 5) should be included in the statistical analysis for these endpoints

Evidencing the primary endpoint criteria determines overall success pertaining to the key study hypothesis. As summarised in Table 4, the concentration of intra-tumoural doxorubicin in a section of the Released biopsy will be compared to a section of the Unreleased biopsy (the control). A two-fold increase in total doxorubicin concentration from the Unreleased control following FUS, or a final (Released) concentration above a minimum threshold level would determine success for a given treatment. The primary endpoint requires independent success either criterion in at least 50% of evaluable patients for overall success of the study. In particular the first criterion aims to establish if there is increased *accumulation* of doxorubicin in the tumour following FUS exposure as per hypothesis (2). Currently there is no validated method to determine if there is increased intra-tumoural *release* as per hypothesis (3), as the HPLC assay is only able to quantify *total* doxorubicin. However release will be explored as a tertiary endpoint by fluorescence microscopy where possible[9]. The tertiary endpoints look at other methods of evidencing targeting doxorubicin delivery and are detailed in Table 4 and Table 6.

The endpoints concerned with quantifying drug, Plasma Pharmacokinetics, MR-Spectroscopy and Radiological Response Evaluation all require statistical significance testing. A two-tailed Fisher Exact Probability test of 90% power with a 5% level of significance will be used to determine if there is a statistically significant increase following treatment relative to the control. GraphPad Prism software or equivalent will be used for this purpose (versions will be recorded in the Statistical report).

During analysis of the Part I and Part II data, a modified CONSORT flow diagram will be completed as shown in the appendices, section 10.1. The information recorded on this chart will be useful in informing future similar studies.

Type of endpoint	**Analytical Technique**	**Quantifiable measure**	**Primary measure**	**Comparator (baseline)**	**Statistical Requirement for success**
Primary	Tumour pharmacokinetics using laboratory extraction technique followed by fluorescence HPLC[10]	Concentration of total intra-tumoural doxorubicin†	Section of Released biopsy from the treated tumour site following FUS-induced mild hyperthermia	Section of Unreleased biopsy from the treated tumour site[11] (Prior to final FUS exposure)	Two-fold increase in doxorubicin concentration in at least 50% of evaluable participants
				N/A	Concentration above minimum threshold of 10µg/g in at least 50% of evaluable participants
Tertiary	Fluorescence microscopy	Average fluorescence intensity levels over a minimum area, in selected regions of interest†	If tissue mass allows, section of Released biopsy from treated tumour	If available, section of Unreleased biopsy from treated tumour or other suitable control[12]	Two-fold increase in at least 50% of evaluable participants‡
		Average nuclear fluorescence intensity	If tissue mass allows, section of	If available, section of Unreleased	Two-fold increase in nuclear fluorescence

[9] As doxorubicin is inherently fluorescent, fluorescence microscopy may be used as a technique to explore fluorescence distribution. If doxorubicin co-localises with a nuclear stain (e.g. DAPI) this shows that it is in its bio-available (released) form.

[10] Quantification using area under curve (AUC) on HPLC chromatogram, normalised by an internal standard, referenced from a standard curve to infer intra-tumoural doxorubicin concentration

[11] In Part II the Unreleased biopsy is not performed and therefore the average value for all evaluable tumours receiving intervention in Part I is used as a comparison for the two-fold increase from Unreleased to Released biopsy

[12] In Part II the Unreleased biopsy is not performed and therefore the average value for all evaluable tumours receiving intervention in Part I, or other suitable control tissue, is used as a comparison for the two-fold increase from Unreleased to Released biopsy

		levels over a minimum number of nuclei, in selected regions of interest†	Released biopsy from treated tumour	biopsy from treated tumour or other suitable control[11]	intensity in at least 50% of evaluable participants‡
Tertiary	MR-Spectroscopy	Tumour:background ratio of signal intensity[13] at the following time points: 1. Pre-treatment (on day -1) 2. Post-treatment (on day 1 or 2)	Mean post-treatment tumour:background ratio	Mean pre-treatment tumour:background ratio	Statistically significant (P<0.05) increase‡
Tertiary	Plasma pharmaco-kinetics using laboratory extraction technique followed by fluorescence HPLC [14]	Plasma concentration of total doxorubicin	Released plasma sample	Time-matched doxorubicin plasma pharmacokinetic data with clearance levels obtained from other clinical studies using ThermoDox®	Statistically significant decrease (p<0.05) over time matched clearance levels‡

Table 4: Summary of quantifiable measures employed to evidence increase in the concentration of intra-tumoral doxorubicin in the TARDOX study, presented in terms of the primary and tertiary outcome measures. Combined Part I and Part II data will be analysed for all endpoints.

Purpose	**Analytical Technique**	**Quantifiable measure**	**Primary measure & Comparator**	**Statistical Requirement for success**
Part I only: To determine optimal FUS exposure parameters for a range of participant Body Mass Indices (BMIs) and tumour locations within the liver	Analysis of real time thermometry obtained by thermistor	1. Both cumulative and continuous minutes above drug release threshold 2. Cumulative Equivalent in Minutes at 43°C (CEM43), in the region of the target tumour being heated by FUS	Real-time thermometry data obtained during FUS intervention, in the first two hours post drug infusion, using implanted thermistor No comparator is required for either measure, but standards are set in the respective statistical requirements.	1. Efficacy: Temperature is maintained above the drug release threshold (39.5°C) for at least 60 continuous seconds AND at least 300 cumulative seconds. Any FUS treatment with a thermometry meeting these criteria will be defined as having 'optimal FUS' for drug release. 2. Safety: Temperature does not exceed 25% of CEM43 thermal dose threshold for cell death [6, 7][15]

[13] The spectroscopy signal analysed may not be doxorubicin itself; metabolites or breakdown products of ThermoDox® may be utilised

[14] Quantification using area under curve (AUC) on HPLC chromatogram, normalised by an internal standard, referenced from a standard curve to infer intra-tumoural doxorubicin concentration

[15] In-house software is used to convert temperature history into cumulative equivalent minutes (CEM) at a reference temperature of 43°C (CEM43). The thermal dose model by Sapareto and Dewey defines that it takes 240min for >96% cell death at 43°C, and this time halves for each subsequent degree rise i.e. t=900s at 47°C, t=3.52s at 55°C, t=0.11s at 60°C. As the critical thermal dose for liver tissue is CEM43=340 minutes, we have defined CEM must be below 25% of this, i.e. 85 minutes, to satisfy safety.

To assess the safety of FUS-induced mild hyperthermia for drug delivery	Standard H&E Microscopy	Percentage of cell or tissue necrosis following FUS exposure, as assessed by microscopy report	If tissue available, microscopy using the Unreleased (and/or Released) Biopsy sample in Part I and the Released Biopsy sample in Part II The Pre-treatment biopsy sample or mean of other of other suitable controls will be used as the baseline comparator[16]	Mean percentage cell or tissue necrosis less than 30%, or an increase in necrosis by no more than 25% over baseline comparator in at least 50% of evaluable participants Note: some tumours are highly necrotic and so persistence of cell viability stain is another method, which may be used.
	Fluorescence microscopy	Persistence of cell viability stain		Presence of cell viability stain anywhere in the sample confirms absence of complete ablation
	Adverse event reports	Adverse events deemed related to FUS	Quantity and severity of adverse events deemed related to FUS up to 30 days post-intervention No comparator required	No statistical test required. Safety data with regards to FUS-related adverse events to be reviewed by the Trial Steering Group on completion of both Parts I and II
To assess the local and systemic cytotoxic effects of ThermoDox® in this setting	Adverse event reports	Significant bone marrow suppression, deranged liver function and liver toxicity	Quantity of Grade 3 and 4 laboratory results from blood tests at each time point post intervention	No statistical test required. Safety data with regards to ThermoDox®-related adverse events to be analysed at the end of the study
	Adverse event reports	Adverse events deemed related to ThermoDox® up to 30 days post-intervention	Quantity and severity of adverse events deemed related to ThermoDox® up to 30 days post-intervention	

Table 5: Summary of safety-related quantifiable measures assessed in the TARDOX study, presented in terms of the secondary outcome measures

Purpose	**Quantifiable measure**	**Primary measure & Comparator**	**Statistical Requirement for success**
To determine therapeutic effect on the target tumour	Radiological evidence of tumour response in the target tumour alone, as assessed by Principles of CHOI and RECIST response evaluation using MRI and CT [1, 2] and SUV_{max} using PET-CT [3] Scan(s) to be performed within 60 days of	CHOI response evaluation [1] using follow-up CT scan demonstrating most significant response to target * For both criterion, the baseline CT scan prior to intervention (day -1) will be used as the comparator In addition, for the first criterion, control tumours exposed to drug alone will be used as a comparator to demonstrate an overall statistically significantly response rate	For each of the following paired analyses‡: • CT / CHOI • CT / RECIST • MRI / RECIST At least one of the following must hold true: 1. A statistically significant ($p<0.05$)

[16] In Part II the Pre-treatment biopsy is not performed and therefore the average value for all evaluable tumours receiving intervention in Part I, or other suitable control tissue, is used as the baseline comparator

	intervention	RECIST response evaluation[17] [2] using both follow-up CT and MRI scans demonstrating most significant response to target *,[18] For both criterion, the baseline CT and MRI scans prior to intervention (day -1) will be used as the comparator In addition, for the first criterion, control tumours exposed to drug alone will be used as a comparator to demonstrate an overall statistically significantly response rate	number of FUS-targeted tumours showing Complete or Partial Response (CR or PR) over all assessed control tumours showing Stable or Progressive Disease (SD or PD) 2. Complete Response (CR) or Partial Response (PR) in one or more modalities for 50% or more of FUS-targeted tumours overall
		PET-CT SUV_{max} [3] response evaluation using follow-up PET-CT scan demonstrating the most significant response * The baseline PET-CT scan prior to intervention (day -1) is used as the comparator and, in addition, a control tumour where available	For the following paired analyses‡: • PET-CT / PERCIST The following must hold true: An SUV_{max} response in 50% or more targeted lesions overall, where *response* is defined, using PERCIST principles [8], as a reduction in SUV_{max} of over 30% over baseline

Table 6: Summary of response-related quantifiable measures required in the TARDOX study, presented in terms of the tertiary outcome measures.

* There will be two follow up scan visits, both for MR/MR-Spectroscopy and CT/PET-CT, in the 60-day follow-up period. In this proof of concept study only a single cycle of chemotherapy is given thus any radiological response may be transient. Consequently the scan demonstrating the most significant response should be used for response evaluation.

** If more than one control tumour has been evaluated for the same PET-CT scan, the average SUV_{max} value of all evaluated control tumours will be used.

6.2 Patient Characteristics

During the final analysis a summary of patient characteristics will be generated to summarise the various attributes and evaluability of the participants recruited and treated in the study. The following table details the patient characteristics recorded for each participant, and the summarised data that will be reported.

Patient Characteristic	**Data-type**	**Summarised Data for Report**
Treated Tumour Type	Enumeration (Primary, Secondary)	Total and % Primary Total and % Secondary
Primary cancer site	Free Text	Total and % of each category e.g. Colorectal, Lung, Breast, Upper GI, Primary HCC, Primary CCA...
Age	Integer	Median
Sex	Enumeration (Male, Female)	Total Female Total Male
BMI	Decimal	Mean + standard deviation

[17] RECIST criteria allows up to a maximum of five tumour foci to be considered in the evaluation of response. In this proof of concept study, for any one participant, only a single tumour will receive the intervention (targeted drug delivery with focused ultrasound). Any other 'control tumours' in the liver will be exposed to ThermoDox® alone.
[18] Note the MRI scan performed on the day following treatment is excluded as it is not a "follow-up" scan for tumour response.

WHO Performance Status		Enumeration (0,1,2,3,4)	Total PS=0 Total PS=1
FUS targeted tumour	Anatomical location	Free-text including segment as per Couinaud Classification	Total for each segment (I-VIII)
	Estimated tumour volume	Decimal	Mean + standard deviation
	Estimated % of tumour volume treated with FUS	Decimal	Mean + standard deviation
	Depth from skin to closest border (cm)	Decimal	Mean + standard deviation
	Depth from skin to furthest border (cm)	Decimal	Mean + standard deviation
	Largest axial dimension (cm)	Decimal	Mean + standard deviation
No. selected control liver lesions		Integer	Median
Consenting patients completing intervention		Enumeration (Y, N)	Percentage completing
Follow-up 1 completed		Enumeration (Y, N)	Percentage completing
Follow-up 2 completed		Enumeration (Y, N)	Percentage completing
Evaluability Assessment	1^{o} Total Doxorubicin	Enumeration (Y, N)	Percentage Evaluable Percentage Non-evaluable, of which breakdown for each endpoint of relevant cause, for example: Not performed due to: 1. Method unavailable 2. Patient refusal 3. Lost in follow-up Attempted but failed due to: 4. Inadequate sample 5. Inadequate control 6. Method failure
	2^{o} Safety – thermometry		
	2^{o} Safety - cell viability		
	2^{o} Safety - FUS AEs		
	2^{o} Safety - drug AEs		
	2^{o} Safety – Grade 3,4 blood results		
	3^{o} Fluorescence microscopy		
	3^{o} Spectroscopy		
	3^{o} Response – CHOI by CT		
	3^{o} Response – RECIST by CT		
	3^{o} Response – RECIST by MRI		
	3^{o} Response - SUV_{max} by PET-CT		
	3^{o} Plasma pharmacokinetics		

Table 7: Patient characteristics and evaluability recorded individually for each patient and summarised in final report

6.3 Missing Data

For a given subject, a single piece of non-evaluable data does not mean the entire subject in non-evaluable, as each endpoint evaluation will include all available evaluable data (see section 2.2). However, if for a given endpoint, the total number of evaluable data points (participants) is less than 10 then that endpoint will be deemed non-evaluable. Details on replacement can be found in section 2.8 (Sample Size).

6.4 Pre-specified Subgroup Analysis

Note that in large tumour volumes (>50cc), it is often not possible to cover the entire tumour volume in a single FUS session. Consequently any radiological response may be confined to a smaller volume of a larger tumour volume. Radiological endpoints do not take this into consideration and consider the tumour in its entirety. It is likely that, following close of study and initial endpoint analysis, an additional post-hoc analysis will be performed to look at radiological response in the subset of smaller tumours receiving optimal FUS.

No other subgroup analyses are planned for this open label Phase I study.

6.5 Sensitivity Analysis

Not applicable: No sensitivity analyses will be performed for this open label Phase I study.

6.6 Blinded analysis

Not applicable in this open label Phase I study.

6.7 Meta-analyses

No meta-analyses will be performed in this open label Phase I study.

6.8 Outcomes Assessment Schedule

Table 8 summarises the data that will be captured for the purposes of outcome assessment in the format of a schedule and consists of both CRF and non-CRF data (see Data Management and Sharing Plan for more information).Specifically, the non-CRF will be captured in the electronic Trial Master File (eTMF) only, and they include anatomical, FUS exposure parameters, and real-time thermometry data (see section 6.9). Blood results will also be captured in OpenClinica. Standard procedures will be followed with respect to retention and archiving of the eTMF. Where an item is missing (e.g. limited biopsy tissue), or non-evaluable due to technical difficulties it will be entered as "Non-Evaluable". Trial statistician will bring all the data together both for analysis and for archiving at the end of the trial. Both 'raw' data and processed data from the various imaging scans will be retained and archived for a minimum of 5 years as per University of Oxford's policy. Digital image files will not be stored in the eTMF to avoid unnecessary duplication. In case of an image-related query that arises following image analysis and data capture, it is possible for the radiology department to revisit any given scan and amend a report, due to the automatic archive facility on the Oxford University Hospitals NHS Trust servers.

Data Item to be Captured	**Data type description**	**Data type storage**	**Relevant Endpoint**	**Day(s) when data captured** (tolerance) **Day 1=intervention**			
				1	**1-2**	**15** (±3)	**<60** (flexible)
Biopsy for drug level	Doxorubicin conc.***	Decimal Quantity	Primary	X			
Real-time thermometry for duration of FUS treatments (Part I only) (Duplicated for each FUS exposure attempt)	Duration of FUS exposure attempt	Decimal Quantity	Secondary	X			
	Baseline, mean, median, peak and trough temperatures[19]	Decimal Quantity		X			
	Cumulative drug-release time (>39.5°C)	Minutes : Seconds					
	Max. continuous drug-release time (>39.5°C)	Minutes : Seconds					
	CEM43 time	Minutes : Seconds					
Haematology tests	Blood results[20]	Decimal Quantities	Secondary		X	X	X
Biochemistry tests	Blood results[21]	Decimal Quantities	Secondary		X	X	X
Adverse Event Review	Adverse events details inc. severity	Tabular	Secondary	X	X	X	X
Plasma for drug level	Doxorubicin conc.**	Decimal Quantity	Tertiary	X			

[19] Note that this data will be extracted from the 'raw' thermometry trace data using in-house analysis software. In particular this data will be presented in the interim analysis (see section 5) and used to determine if optimal FUS parameters have been established.

[20] Full Blood Count (FBC) to include Haemoglobin (Hb), white blood cells (WBC) with differential count (neutrophils and lymphocytes) and platelets

[21] Biochemical profile to include U&Es, (Na+, K+, Ur, Cr), phosphate, calcium, LFTs (total protein, ALP, Albumin, ALT or AST, Bili, LDH)

MR Spectroscopy	Tumour:bkgd Ratio	Decimal Quantity	Tertiary		X		
CT Liver	CHOI Outcomes (target & controls)	Enumeration (CR, PR, SD, PD)	Tertiary			X*	X*
CT Liver	RECIST 1.1 Outcomes (target & controls)	Enumeration (CR, PR, SD, PD)	Tertiary			X*	X*
MRI	RECIST 1.1 Outcomes (target & controls)	Enumeration (CR, PR, SD, PD)	Tertiary			X*	X*
Full Body PET-CT	SUV_{max} value (target & controls)	Decimal Quantity	Tertiary	X		X*	X*
Biopsy for fluorescence	Avg. Bkgd Fluorescence	Decimal Quantity	Tertiary	X			
	Avg. Nuclear Fluoresc.	Decimal Quantity	Tertiary	X			
Biopsy for cell viability	% Necrosis	Percentage	Tertiary	X			
	Persistence of cell viability stain	Enumeration (Y/N)	Tertiary	X			

Table 8: schedule of possible data capture points for purposes of outcome assessment. Note that additional data captured that is not relevant to outcome assessment has been excluded.

***** In Part II, the number of follow-up MRI/MR-Spectroscopy, Liver pCT and FDG PET-CT scans may be reduced, if the data from Part I informs that a more efficient (in terms of reducing patients & site requirements) schedule is possible

****** For plasma pharmacokinetics, three blood samples in both Phases I and II (Pre-Treatment, Unreleased drug, Released drug) will be assayed.

******* For the biopsy tissue for HPLC, section(s) of three biopsies in Part I (Pre-Treatment, Unreleased drug, Released drug) and a two serial biopsies in Part II (both at the Released drug time-point) will be assayed.

6.9 Data Management Responsibility

The OCTO data management team will design TARDOX CRF and identify which data items will be captured in the subjects' clinical care and electronic records, therefore can be captured in the CRFs ('CRF data'), and those which will be generated by collaborating groups without being returned to the clinical care record ('non-CRF data').

The non-CRF data includes:

- Pre-treatment, Unreleased and Released Plasma and Biopsy PK analytical chemistry results:
    - Obtained by HPLC in the Bioanalysis Core and GCP Laboratories based in the Old Road Campus Research Building
    - Batch analysis of the samples will produce chromatograms generated electronically by the HPLC equipment, which can then be used to calculate doxorubicin concentrations by hand
- FUS delivery parameters:
    - Captured from reports written by members of the study team at the time of treatment and the treatment report generated by the HIFU device
- Thermistor data output:
    - The thermistor captures data on the tumour temperature over time and produces an electronic dataset which will be analysed in the Institute of Biomedical Engineering
- If available, Pre-treatment, Unreleased and Released Microscopy study data output:
    - Samples will be analysed by fluorescence microscopy in the GCP Laboratories based in the Old Road Campus Research Building
    - Microscopic analysis of the samples will produce images that may be used to evaluate cell viability post-FUS and/or presence of nuclear doxorubicin

- If available, MR-spectroscopy data output:
    - Generated by the Churchill Radiology Department, Oxford University Hospitals NHS Foundation Trust

The collaborators undertaking these analyses are responsible for producing datasets using validated methods where possible. The unique trial patient ID will be used to identify these data sets. The raw data output will be stored securely and backed up regularly. The final listings produced from these data, for use in statistical analysis, will be retained by the collaborating investigators, and stored securely and backed up regularly.

A report on the methodology used to derive the listings will be produced for methods that are not standard practice. This report will be filed in the eTMF, along with information indicating the location and archiving arrangement of the raw data (in the event that this is not the eTMF).

All other data items will be captured on the trial-specific CRFs.

The data management of CRF data will be compliant with GCP and OCTRU data management SOPs, CRF data will be managed via a web-based, bespoke trial database using OpenClinica. OpenClinica is a dedicated and validated clinical trials database designed for electronic data capture. The OCTO data management team are responsible for producing as complete and clean data set as possible. The CRF data management plan is incorporated in the central monitoring plan.

The trial statistician will receive a copy of the final non-CRF listings and also the CRF data extractions. The statistician will be responsible for linking the various datasets. Data transfer will comply with OCTRU SOP GEN-031. The listings will be stored unamended in the eTMF. The statistical analysis processes will be undertaken in accordance with OCTRU SOPs.

7 DATA SAFETY MONITORING COMMITTEE AND INTERIM ANALYSES

There is no Data Safety Monitoring Committee (DSMC) and an Independent Trial Steering Committee (ITSC) will be in place to monitor safety (please refer to section 18 of the study protocol).

The investigator team (including clinical fellow(s) and lead FUS clinician) will independently monitor FUS performance for at least 5 patients in Part I. When the investigator team are confident in achieving optimal levels of hyperthermia for a range of BMIs and anatomical tumour sites, a meeting with the Trial Management Group (TMG) will be called. This decision is expected between the 5th and 14th recruited patients, with target recruitment being 1.2 patients per month. The TMG will use results from interim analysis review to determine if and when Part II can open in parallel with Part I. In particular only the secondary objective will be considered, which concerns optimal FUS exposure parameters. This is because in Part II there will be no real time thermometry during FUS exposure as a thermistor is not placed. Consequently, at the time point at which the investigator team feel enough experience in determining optimal FUS parameters has been obtained in Part I, at TMG interim analysis will be called to review the following data for each treated patient:

- Demographic data:
    - Sex
    - Age
- Anatomical data:
    - Body Mass Index (BMI)
    - Anatomical position of liver lesion treated (including segment)
    - Estimated percentage volume of liver lesion treated

  - Depth from skin to closest border in cm
  - Depth from skin to further border in cm
  - Largest tumour dimension in cm
- FUS exposure parameters[22] for each FUS exposure, including no less than the following:
  - Unique patient identifier and date of treatment
  - Overall exposure time (s)
  - Frequency (MHz)
  - Duty cycle (%)
  - Average Power (Watts)
  - Overall Energy (KJ)
- Real-time thermometry data for each FUS exposure attempt, including no less than:
  - Unique patient identifier and date of treatment
  - Baseline body temperature[23]
  - Date and time of the start of the real time thermometry monitoring
  - Duration of FUS exposure attempt (MM:SS)
  - Peak & trough temperatures attained (^{o}C)
  - Mean and median (to one decimal place) (bulk) temperature[24]
  - Cumulative and maximum continuous durations at which the drug-release threshold was met (MM:SS)
  - CEM43 time (MM:SS)
- Where possible (in some cases limited tissue biopsies may mean this data is not available), cell viability data obtained for the purposes of the secondary objective
- Up-to-date adverse event data relating to FUS exposure

The summary data listed will be provided by the investigator team, in tabular format, for example as shown in Appendices 4 and 5. Start time and end times of the real-time thermometry recording for each FUS exposure attempt (or treatment plan) will be recorded by hand and transcribed into the real-time thermometry proforma (Appendix 4). Real-time raw thermometry data will be acquired using a clinically approved Temperature Probes that is connected to a personal computer (PC) running bespoke data acquisition software (running on the LabView platform) developed by the Institute of Biomedical Engineering (IBME). An in-house MatLab script, developed by the IBME, will be used to process thermometry data in order to generate the required data summary for each FUS exposure attempt (see Appendix 3). The JC-200 therapeutic device software will automatically generate the FUS exposure parameters in the final treatment report for each FUS exposure attempt, which will be transcribed to the FUS exposure proforma (Appendix 5).

Cell viability will be evaluated either using microscopy assessments or flow cytometric analysis using specialist software (ImageJ/FIJI for microscopy or equivalent) at the University of Oxford. All raw data will ultimately be stored on the Statistical eTMF to allow for future reproducibility of results, along with the analysed data to be presented.

With regards to optimising FUS exposure parameters, it is difficult to perform a blanket statistical test for this rather disparate dataset, as there are many parameters that can be adjusted when applying FUS and 'optimal FUS' is somewhat subjective. When reviewing thermometry traces for drug delivery, the key

[22] If it was required to modify the FUS parameters in order to achieve the optimal range of hyperthermia, for example by increasing changes the in duty cycle, then each set of FUS parameters will be presented.
[23] This is the temperature immediately prior to first FUS exposure, as shown on the thermometry trace
[24] Please note that large volumes of real time thermometry data may be generated for each FUS session and there may be periods of acquisition where no heating is taking place. It is the 'bulk' temperature recorded by the thermistor during FUS exposures, which is of most physiological relevance.

consideration is to ensure that the bulk tumour temperature, as measured by the thermistor, exceeds the drug delivery threshold (39.5°C) as quickly as possible after drug administration and is sustained for as long as practically possible, given the patient is under a general anaesthetic. Ideally the desired range of hyperthermia is maintained (41-47°C), such that temperature remains below the safety limits for ablation, and remains comfortably above the drug release threshold for a prolonged and ideally continuous period. The statistical tests in Table 5 go some way to capture these thermometry attributes in order to define what is 'optimal FUS' in the context of this study.

In view of this collective data set, at the point at which both the investigator team and the TMG are satisfied that sufficient FUS experience has been gained to safely achieve the desired range of mild hyperthermia for a range of anatomical tumour locations and BMIs, the trial will progress to Part II. In Part II the FUS exposure parameters will be determined without a thermistor using Part I data combined with planning calculations and/or simulations performed at the Institute of Biomedical Engineering (IBME). Once the decision has been made to open Part II, patients can be recruited to either part based on previous experience and safety considerations. The TARDOX trial office will also inform the trial site and notify Celsion. At the point at which this decision is made, any patients who have already been registered and assigned to Part I, but have not yet had intervention, will be considered for transfer to Part II, subject to participant consent.

The study will recruit up to a maximum of 28 evaluable participants across two separate Parts (Part 1 and Part 2). A minimum of 5 participants are required in Part 1 before Part 2 can be opened. Recruitment into Part II will not take place until the TMG board is in agreement for this transition.

Following TMG review, other possible outcomes are as follows:

- It is possible to open to Part II in parallel to Part I
- Ethical approval will be sought to continue recruiting up to the maximum of 28 patients to Part I, with a thermistor, without transferring to Part II at any stage
- The study should be terminated at this stage due to safety concerns resulting from Part I data

Following this TMG review the Statistical Analysis Plan will also be reviewed in light of available patient data. Although analysis will not be completed until the end of Part II, the statistical tests will examined at this point to ensure test validity in anticipation of the final combined analysis of Parts I and II.

8 SAFETY ANALYSIS

Adverse event (AE) monitoring starts at intervention (Day 1) until 30 days post intervention (day 30). The Investigator will monitor each patient for clinical and laboratory evidence of adverse events on a routine basis throughout the study. Should an Investigator become aware of any study drug related SAEs following this period these must also be reported as stated below. All reportable AEs will be followed to a satisfactory conclusion. Any reportable AEs that are unresolved at the patient's last visit in the study are followed up by the Investigator for as long as medically indicated, but without further recording in the CRF.

All AEs reported to the trial office will be processed according to internal SOPs, this includes the reporting procedure and assessment of causality and expectedness for SAEs. The TARDOX trial office may request additional information for any AE as judged necessary.

AEs are captured in secondary endpoints if deemed to be related to ThermoDox® or related to FUS.

AEs related to ThermoDox® will be reported in tabular format as detailed in Table 9. This table follows a similar structure to that shown in a previous publication regarding use of ThermoDox® in human subjects, in particular, for illustrative purposes, Table 2 of that publication is populated [9]. Furthermore, procedural-related AEs will be presented in a similar tabular format as demonstrated by Table 10.

System organ class / AE	Total AEs n (%)	AEs Grade 3 or more n (%)	Drug-related n (%)	SAEs n (%)
Metabolic/laboratory				
Blood and lymphatic system disorders				
Gastrointestinal disorders				
Constitutional symptoms				
Skin and subcutaneous tissue disorders				
Pain				
Hepatobiliary disorders				
Pulmonary/upper respiratory				
Infections				
Renal/genitourinary disorders				
Haemorrhage				
Cardiac general				
Musculoskeletal				
Lymphatics				
Neurology				

Table 9: Frequency listing of example drug-related adverse events, occurring in patients treated with ThermoDox®.

System organ class / AE		Total AEs n (%)	AEs Grade 3 or more n (%)	Drug-related n (%)	SAEs n (%)
General Procedural complication deemed not related to FUS or biopsy	Anaesthetic-related e.g. dysrhythmias				
	Post-operative ileus				
	DVT/PE				
	Other				
Procedural complication deemed related to FUS	Procedural pain				
	Discomfort at treatment site				
	Skin toxicity / blistering at treatment site				
	Oedema at treatment site				
	Damage to other structure(s)				

Procedural complication deemed related to Biopsy	Procedural pain				
	Haemorrhage				
	Vasovagal hypotension				
	Damage to other structure(s)				

Table 10: Frequency listing of example procedural adverse events deemed unrelated to FUS or biopsy, related to FUS, or related to biopsy

9 APPENDICES

9.1 Appendix 1: Glossary of Abbreviations

AE	Adverse Event
CI	Chief Investigator
CEM43	Cumulative Equivalent in Minutes at 43°C
DSMC	Data and Safety Monitoring Committee
DMP	Data Management and Sharing Plan
IBME	Institute of Biomedical Engineering
IHC	Immunohistochemistry
ITSC	Independent Trial Steering Committee
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
TMG	Trial Management Group

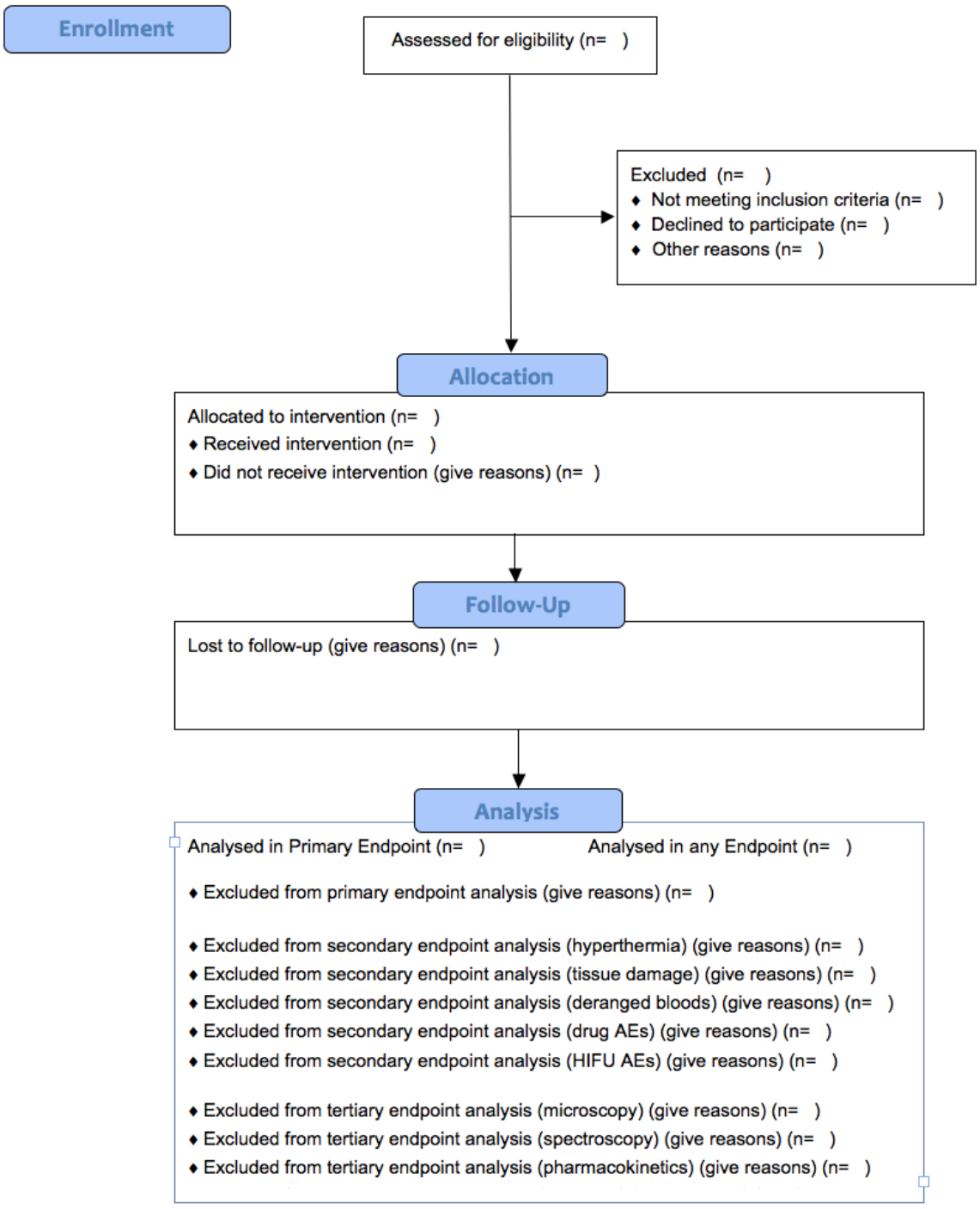
10 DOCUMENT HISTORY

Ver	Date	Who	Comments
2.1	26Jan2016	PCL	Third major revision of SAP based on Protocol version 3.0 Feb, 2016 (substantial amendment), including updates to reflect: • Staff members who have been most actively involved in the study. • Minor changes to secondary endpoints with respect to microscopy analysis. • Implementation of CEM43 thermal dose threshold model to evaluate FUS safety • Provision for parallel opening of Parts I & II. • Provision for scan follow-up period of up to 60 days • Changes to recruitment strategy.
2.1	31Mar2016	LO	Co-author and revision to conform with new SAP template
2.1	01May2016	LO	Revision to reflect Protocol V3.0
2.2	14Jun2016	LO	Harmonise comments from PL for consistency
2.3	16Jun2016	LO	A new version reflecting minor changes on opening Part II
2.3	16Jun2016	PCL	Minor changes to formatting and clarifications
2.4	16Jun2016	LO	Minor changes relocated history of changes
2.5	28Jun2016	LO	Updated designations and key personnel details

10.1 Appendix 2: Modified CONSORT Flow Diagram

CONSORT
TRANSPARENT REPORTING of TRIALS

CONSORT 2010 Flow Diagram (modified for TARDOX Study)

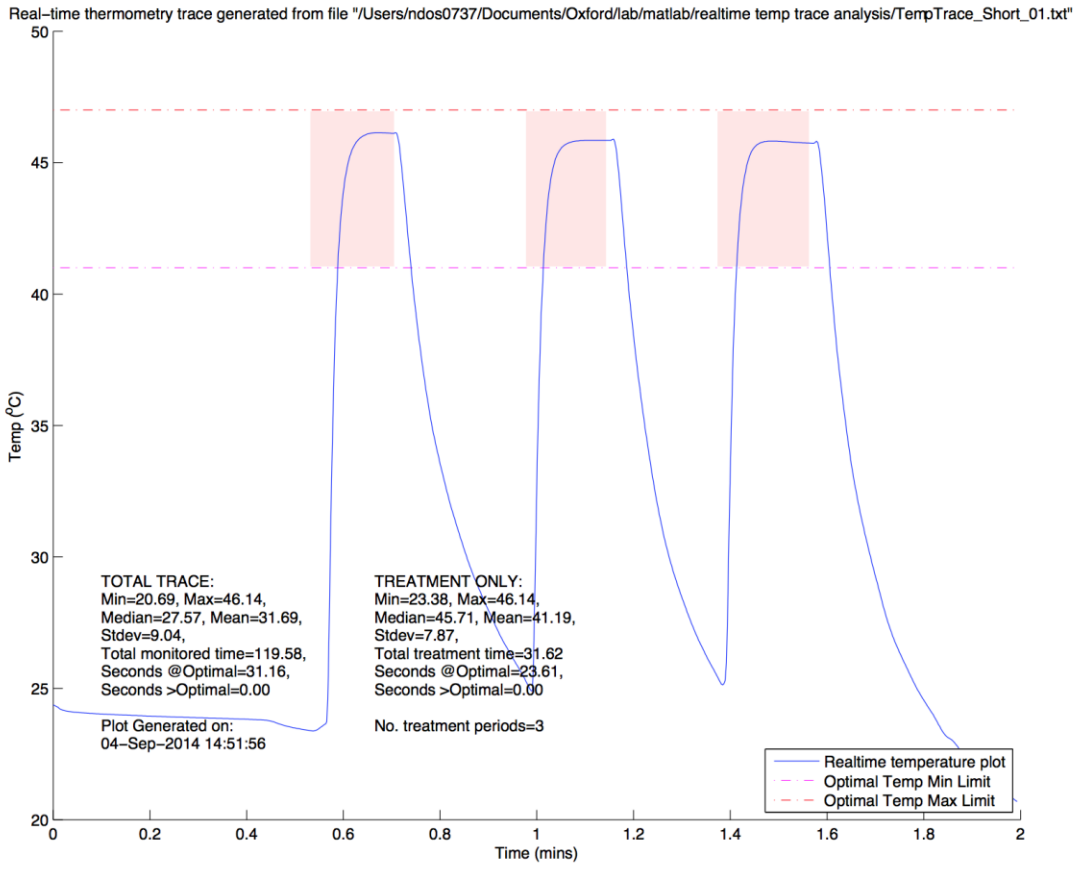


10.2 Appendix 3: Example Processed Real-Time Thermometry Trace

The example thermometry trace below has been generated using a clinically approved thermistor and processed using an in-house MatLab program developed at the IBME, which is generates the CEM43 calculations. Please note that this example real-time trace spans only two minutes, whereas for the intervention the traces are likely to span for several hours given time required for treatment planning and other procedures (biopsy). Thus, during any given patient intervention, the real-time thermometry traces are likely to extend significantly before and after FUS exposure(s), which may only represent a small portion of the total monitored period.

For this reason, the software used to capture the thermometry data allows the user to define discrete treatment periods, representing FUS exposures, in real time using a switch function on the in-house LabView graphical user interface. To demonstrate this, three treatment periods were simulated in the trace below. During these user-defined treatment periods, the thermistor tip was placed into hot water. Note that during the study intervention it is intended that a separate thermometry trace be generated and processed for each individual FUS exposure attempt, each with a single treatment period, rather than multiple periods over a single continuous trace.

The blue plot represents the real-time thermometry trace, the area between the dotted horizontal lines represent the optimal temperature range and the pink rectangles represent the user defined treatment periods. In addition, the MatLab program automatically performs statistical analyses on the raw thermometry data, for both the total trace and treatment periods separately, which is output as an overlay on the plot (bottom left). If available, the treatment only data will be preferentially captured in the real-time thermometry proforma (appendix 4), to avoid analysis of trace during prolonged periods of FUS inactivity.



10.3 Appendix 4: TARDOX Real-Time Thermometry Data Capture Proforma

Unique Subject Trial Number:			Baseline body temp (°C):	Tumour Location (liver lobe, segment):
Sex:	Age:	BMI:	ThermoDox® Infusion Start-Stop Time (HH:MM)	Water Temp. (°C):
Comments:				

FUS Exposure Attempt No.	Monitoring Start Time (HH:MM)	Monitoring Stop Time (HH:MM)	Temps. (°C) during total monitoring / treatment* period (*Circle measure used)				Durations (MM:SS), or MILLISEC (MS)			
			Peak	Trough	Mean	Median	39.5-47	40-47	41-47	>47

Optimal FUS achieved? (Y/N): Post-drug cumulative time > 39.5°C (>300s) = Post-drug max. continuous time > 39.5°C (>60s) =	CEM43 Threshold Exceeded? (Y/N): Post-drug CEM43 time =

10.4 Appendix 5: FUS Exposure Parameters Data Capture Proforma

Unique Subject Trial Number:			Depth to closest and furthest tumour borders from skin (cm):	Tumour Dimensions (mm) and Volume (cc):
Sex:	Age:	BMI:	Intercostal space (mm)	Thickness of skin & Subcutaneous Tissues (mm)
JC200 Transducer used, frequency and focal length:			Estimated percentage of tumour volume treated (%):	Prescribed FUS tumour volume post-drug (cc)
Comments:				

FUS Exposure Attempt No.	FUS Start Time (HH:MM)	FUS Stop Time (HH:MM)	Power (Watts)	Pulse Int.	Slice Thickness & Separation	Duty Cycle (%)	Mode (Dot/Linear) and Spacing /Speed	Overall Energy (KJ)	Overall time (s)

11 REFERENCES

1. Choi, H., et al., *Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria.* J Clin Oncol, 2007. **25**(13): p. 1753-9.
2. Eisenhauer, E.A., et al., *New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1).* Eur J Cancer, 2009. **45**(2): p. 228-47.
3. Kinahan, P.E. and J.W. Fletcher, *Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy.* Semin Ultrasound CT MR, 2010. **31**(6): p. 496-505.
4. Parkin, D.M., et al., *Global cancer statistics, 2002.* CA Cancer J Clin, 2005. **55**(2): p. 74-108.
5. Khan, S.A., et al., *Rising trends in cholangiocarcinoma: is the ICD classification system misleading us?* J Hepatol, 2012. **56**(4): p. 848-54.
6. Sapareto, S.A. and W.C. Dewey, *Thermal dose determination in cancer therapy.* Int J Radiat Oncol Biol Phys, 1984. **10**(6): p. 787-800.
7. Graham, S.J., et al., *Quantifying tissue damage due to focused ultrasound heating observed by MRI.* Magn Reson Med, 1999. **41**(2): p. 321-8.
8. Wahl, R.L., et al., *From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors.* J Nucl Med, 2009. **50 Suppl 1**: p. 122S-50S.
9. Poon, R.T. and N. Borys, *Lyso-thermosensitive liposomal doxorubicin: a novel approach to enhance efficacy of thermal ablation of liver cancer.* Expert Opin Pharmacother, 2009. **10**(2): p. 333-43.