CALIBER – A phase II randomised feasibility trial of chemoablation with mitomycin versus surgical management in low risk non-muscle invasive bladder cancer

Journal:	*BJU International*
Manuscript ID	BJU-2019-1685.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Mostafid, Hugh; Royal Surrey County Hospital, Urology Porta, Nuria; The Institute of Cancer Research, Clinical Trials and Statistics Unit Cresswell, Joanne; South Tees Hospitals NHS Foundation Trust, Urology Griffiths, Thomas; University of Leicester, Division of Urology Kelly, John; University College London Penegar, Steven; The Institute of Cancer Research, Clinical Trials and Statistics Unit Davenport, Kim; Gloucestershire Hospitals NHS Foundation Trust McGrath, John; Royal Devon and Exeter NHS Foundation Trust Campain, Nicholas; Royal Devon and Exeter NHS Foundation Trust Cooke, Peter; The Royal Wolverhampton Hospitals NHS Trust Masood, Shikohe; Medway NHS Foundation Trust Knowles, Maggie; University of Leeds Feber, Andrew; University College London, UCL Cancer Institute Knight, Allen; Action Bladder Cancer UK Catto, James; The University of Sheffield Lewis, Rebecca; The Institute of Cancer Research, Clinical Trials & Statistics Unit Hall, Emma; The Institute of Cancer Research, Clinical Trials & Statistics Unit
Keywords:	non-muscle invasive bladder cancer, chemoablation, surgery, mitomycin C, randomised trial
Abstract:	Objectives To evaluate the activity of intravesical mitomycin C (MMC) to ablate recurrent low risk non-muscle invasive bladder cancer (NMIBC) and assess whether it may enable patients to avoid surgical intervention for treatment of recurrence. Patients and methods CALIBER is a phase II feasibility study. Participants were randomised (2:1) to treatment with four once-weekly MMC 40mg intravesical instillations (chemoablation arm) or surgical management. The surgical group was included to assess feasibility of randomisation. The primary endpoint was complete response to intravesical MMC in the chemoablation arm at three months, reported with exact 95% confidence

	intervals. Secondary endpoints included time to subsequent recurrence, summarised by Kaplan-Meier methods. Results Between February 2015 and August 2017 82 patients with visual diagnosis of recurrent low risk NMIBC were enrolled from 24 UK hospitals (54 chemoablation, 28 surgical management). Median follow-up was 24 months. Complete response at three months was 37.0% (20/54; 95%CI: 24.3-51.3) with chemoablation and 80.8% (21/26; 95%CI 60.6-93.4) with surgical management. Amongst patients with complete response at three months, a similar proportion were recurrence-free by 12 months in both groups (84%). Amongst those with residual disease at three months, the 12-month recurrence-free proportion was lower in the surgical management group (40.0%) than in the chemoablation group (84%). Recruitment stopped early as chemoablation did not meet the pre-specified threshold of 45% complete responses at three months. Conclusion Intravesical chemoablation in low risk NMIBC is feasible and safe, but did not demonstrate sufficient response in this trial. Following chemoablation there may be a reduction in recurrence rate, even in non-responders, that is greater than with surgery alone. Further research is required to investigate the role and optimal schedule of neo-adjuvant intravesical chemotherapy prior to surgery for NMIBC.

Title: CALIBER – A phase II randomised feasibility trial of chemoablation with mitomycin versus surgical management in low risk non-muscle invasive bladder cancer

Authors: A. Hugh Mostafid[1], Nuria Porta[2], Joanne Cresswell[3], TRL Griffiths[4], John D Kelly[5], Steven R Penegar[2], Kim Davenport[6], John S McGrath[7], Nicholas Campain[7], Peter Cooke[8], Shikohe Masood[9], Margaret A Knowles[10], Andrew Feber[5], Allen Knight[11], James WF Catto[12], Rebecca Lewis[2], Emma Hall[2].

Authors' affiliations:

1. Royal Surrey County Hospital NHS Foundation Trust, Guildford, UK
2. The Institute of Cancer Research, London, UK
3. South Tees Hospitals NHS Foundation Trust, Middlesbrough, UK
4. University Hospitals of Leicester NHS Trust, Leicester, UK
5. University College London, London, UK
6. Gloucestershire Hospitals NHS Foundation Trust, Cheltenham, UK
7. Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
8. The Royal Wolverhampton Hospitals NHS Trust, Wolverhampton, UK
9. Medway NHS Foundation Trust, Gillingham, UK
10. University of Leeds, Leeds, UK
11. Action Bladder Cancer UK, Gloucestershire, UK
12. University of Sheffield, Sheffield, UK

Corresponding author: Mr Hugh Mostafid, The Stokes Centre for Urology, Royal Surrey County Hospital, Egerton Road, Guildford, Surrey, GU2 7XX. Tel: +(44) 01483 571122, Fax: +(44) 01256 651115 email: hugh.mostafid@nhs.net

Abstract

Objectives

To evaluate the activity of intravesical mitomycin C (MMC) to ablate recurrent low risk non-muscle invasive bladder cancer (NMIBC) and assess whether it may enable patients to avoid surgical intervention for treatment of recurrence.

Patients and methods

CALIBER is a phase II feasibility study. Participants were randomised (2:1) to treatment with four once-weekly MMC 40mg intravesical instillations (chemoablation arm) or surgical management. The surgical group was included to assess feasibility of randomisation. The primary endpoint was complete response to intravesical MMC in the chemoablation arm at three months, reported with exact 95% confidence intervals. Secondary endpoints included time to subsequent recurrence, summarised by Kaplan-Meier methods.

Results

Between February 2015 and August 2017 82 patients with visual diagnosis of recurrent low risk NMIBC were enrolled from 24 UK hospitals (54 chemoablation, 28 surgical management). Median follow-up was 24 months. Complete response at three months was 37.0% (20/54; 95%CI: 24.3-51.3) with chemoablation and 80.8% (21/26; 95%CI 60.6-93.4) with surgical management. Amongst patients with complete response at three months, a similar proportion were recurrence-free by 12 months in both groups (84%). Amongst those with residual disease at three months, the 12-month recurrence-free proportion was lower in the surgical management group (40.0%) than in the chemoablation group (84%). Recruitment stopped early as chemoablation did not meet the pre-specified threshold of 45% complete responses at three months.

Conclusion

Intravesical chemoablation in low risk NMIBC is feasible and safe, but did not demonstrate sufficient response in this trial. Following chemoablation there may be a reduction in recurrence rate, even in non-responders, that is greater than with surgery alone. Further research is required to investigate the role and optimal schedule of neo-adjuvant intravesical chemotherapy prior to surgery for NMIBC.

Keywords: non-muscle invasive bladder cancer; chemoablation; surgery; mitomycin C; randomised trial;

Introduction

Bladder cancer is the ninth most common cancer world-wide (1), and most frequently presents as non-muscle invasive bladder cancer (NMIBC). Approximately 50% of BC patients have low risk NMIBC (2), with a 0.8-6% risk of progression to MIBC or bladder cancer death within five years and a relatively high rate of local recurrence 46-62% (2-4). Half of recurrences occur within the first year of follow-up (5). The discomfort and inconvenience of managing NMIBC recurrence combined with cost, are the key issues for patients and healthcare providers managing low risk NMIBC (6-8).

Guidelines recommend annual cystoscopy for five years for low risk NMIBC (2). Treatments for local recurrence include transurethral resection and cystodiathermy under general anaesthesia, laser ablation under local anaesthetic and watchful waiting(9, 10). This variety reflects the indolent nature of low risk NMIBC and lack of high quality evidence about optimal management.

Several small studies demonstrated promising results for intravesical chemotherapy alone (chemoablation) as an alternative to surgical management for NMIBC. The optimal schedule and its effectiveness in achieving a complete response in low risk NMIBC are unclear. Reviews of chemoablation (including over 1,200 patients of varying risk and different chemotherapy regimens) suggest the complete response rate is around 50%, with the therapeutic effect sustained for at least two years (11, 12). These data suggest chemoablation may be a viable treatment for low risk NMIBC.

To inform trial design one hundred patients undergoing surveillance for low risk NMIBC were surveyed. They had concerns with inpatient surgical management of recurrence under general anaesthetic, and stated a preference for a non-surgical outpatient option. A focus group of NMIBC patients was then held to discuss potential trial designs, at which, based on

available data(11, 12), chemoablation was confirmed an attractive alternative to surgical management for recurrent low risk NMIBC and suitable success criteria for a phase II trial were agreed.

CALIBER was therefore developed to investigate intravesical chemoablation as an alternative to surgical management for recurrent low risk NMIBC, incorporating patient reported outcomes to assess participants' acceptability of treatments.

Patients and Methods

Trial design, management and governance

CALIBER (NCT02070120) is a phase II multicentre feasibility study. A two-stage randomised design was used to establish chemoablation response rate whilst obtaining prospective surgical management data and assessing feasibility of randomisation between treatments for any subsequent comparative trial. Recruitment was planned to continue seamlessly between stages one and two.

The trial was approved by the Medicines and Healthcare Products Regulatory Authority and South Central – Hampshire-B Research Ethics Committee (ref: 14/SC/1223, approved 29th August 2014), sponsored by The Institute of Cancer Research (ICR) and conducted according to the principles of good clinical practice. The Clinical Trials and Statistics Unit at the Institute of Cancer Research (ICR-CTSU) co-ordinated the study, data collection, and conducted statistical analysis. The trial management group was overseen by independent data monitoring and trial steering committees.

Patients

Eligible patients had previously diagnosed, histologically confirmed, low risk NMIBC with visual diagnosis of recurrence. Patients were over 16, with an European Organisation for

Research and Treatment of Cancer (EORTC) risk of recurrence score ≤6 (2) (this criterion was revised in December 2016 from ≤5, due to inadvertent exclusion of patients for whom chemoablation may be an appropriate treatment option), with no history of high grade/≥T1 or non-urothelial transitional cell carcinoma. Participants with prior treatment of the recurrence or contraindication to trial treatment were excluded. All participants provided written informed consent.

Treatment allocation and study procedures

Participants were recruited at UK NHS hospitals and allocated by the ICR-CTSU to either chemoablation or surgical management in a 2:1 ratio. Treatment allocation was by minimisation with a random element, with balancing factors of treating site and recurrence history (first or further recurrence). Treatment allocation was not blinded.

Chemoablation participants received four once weekly intravesical instillations of 40mg mitomycin-C (MMC) as outpatients, in accordance with local policy. No dose reductions were permitted. Participants assigned to surgical management had the local standard technique for treatment of recurrence; a single instillation of 40mg MMC within 24 hours post-operatively was permitted.

A cystoscopy was conducted three months after treatment completion to visually assess response and biopsy the tumour bed. Subsequent cystoscopic follow-up was at six (if disease at three months) and 12 months after treatment, and annually thereafter.

Outcomes

The primary endpoint was complete response to chemoablation at three months post-treatment, defined as an absence of any bladder tumour both by visual assessment and biopsy.

Secondary endpoints included time from end of treatment to subsequent recurrence, subsequent transurethral resection of bladder tumour (TURBT)/biopsy rates after the three month disease assessment, safety and patient reported health related quality of life (HRQOL) outcomes.

Adverse events were assessed at end of treatment and three months, using National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) version 4.0. HRQOL was assessed with the EORTC's general quality of life questionnaire (QLQ-C30) (13) and NMIBC specific module (QLQ-NMIBC24) (14). The primary objective of the HRQOL study was to assess differences between groups in the QLQ-C30's global quality of life scale. Questionnaires were completed by participants at baseline, three, six and twelve months.

Statistical considerations

CALIBER was designed to rule out a complete response rate of less than 45% in the chemoablation group. Using a Simon's two-stage optimal design (15), complete response in at least 26/51 chemoablation patients was required in stage one. Prior to stage one analysis, the design was adapted to reduce stage two sample size and remove the randomisation (see supplementary material). In the revised design, with 85% power and α=0.10, complete response in at least 31/60 chemoablation patients was required at the end of stage 2. The total target recruitment was 89 patients, 63 chemoablation (accounting for 5% noncompliance) and 26 surgical management patients (stage one control group).

Efficacy outcomes were analysed on the evaluable population, i.e. participants with three-month assessment data who received their allocated treatment. Sensitivity analyses on the per-protocol and eligible populations were performed (supplementary table 1). Safety analyses were conducted according to treatment received.

Complete response rate was calculated based on (i) no disease on visual assessment at three month cystoscopy and (ii) where three month biopsy performed, no disease on histopathology assessment. Patients with visual disease, or positive histology when visually clear, were classified as not responding. Both definitions were considered for the stage one stop/go decision. Complete response rates were presented with exact binomial 95% confidence intervals (95% CI). The trial was not powered for the direct comparison of complete response rate between treatment groups and no formal statistical comparisons of the primary endpoint were planned.

Time to first subsequent recurrence after response status assessment at three months was summarised using Kaplan-Meier methods, and treatment groups compared by the stratified log-rank test, adjusting by response status at 3 months. The four groups defined by the combination of treatment and response status at 3 months were compared by the log-rank test. Frequency of subsequent NMIBC recurrence/TURBT was summarised by treatment; worst CTCAE grade adverse event was summarised by timepoint and treatment received. Treatment comparisons used Chi-squared or Fisher tests as appropriate. Statistical comparisons for the secondary endpoints were considered exploratory.

Standard algorithms were used to derive scores from and handle missing HRQOL data(16). Change from baseline was calculated and summarised descriptively at each subsequent timepoint with means and 99% CI. A larger confidence level was chosen for HRQOL endpoints to account for multiplicity across sub-scales and timepoints.

Analyses were based on a data snapshot taken on 10 October 2018, triggered once all patients had at least 12 months of follow-up (or earlier if loss to follow-up), and performed using STATA version 15.0(17).

Results

Participants

Eighty-two participants were enrolled (54 chemoablation, 28 surgical management) from 24 UK sites between February 2015 and August 2017 (Figure 1). Fifty-six percent (82/145) of eligible patients reported on sites' screening logs consented to participation. CALIBER ceased recruitment in August 2017, after the Independent Data Monitoring Committee concluded the trial should stop for futility based on stage one complete response rates.

Baseline features were evenly matched across treatment groups (Table 1). Fifty-three chemoablation participants (98%) received all four planned instillations, with one participant receiving three. Twenty-seven surgical management participants received surgery, of whom 16 (57%) received diathermy (Table 2).

Response rates

The stage one stop/go decision was based on the first 51 evaluable chemoablation participants: 18 complete responses were reported by visual and histopathology assessment (where available) with 23 complete responses reported by visual assessment alone. The criterion to proceed to stage two was not met by either definition of complete response.

Complete response rate in the chemoablation group overall was 37% (20/54; 95% CI 24-51) by visual and histopathology assessment and 48% (26/54; 95% CI 34-62) by visual assessment alone. Complete response rate was 81% (21/26; 95% CI 61-93) in the surgical management group by visual and histopathology assessment.

Figure 2 shows concordance between visual and histopathology assessment. In the chemoablation group, 28/54 (52%) participants had visible disease at three months (no complete response), with 23/28 confirmed histologically. Of 26/54 (48%) patients with no

visible disease, 6/26 had disease confirmed on biopsy. In the surgical management group, 3/26 (12%) patients had visible disease at three months, all confirmed histologically; 2/23 patients with no visible tumour had residual disease confirmed on biopsy. Three month histology was unavailable for nine chemoablation and eleven surgical management participants. Table 3 summarises disease found at three months.

Recurrences subsequent to the three-month disease assessment

With a median follow-up at time of data snapshot of 24 months (IQR 15-29), 27 participants had NMIBC recurrences after their three month disease assessment. In the chemoablation group, 16 (30%) patients had at least one NMIBC recurrence, with two (4%) experiencing more than one. Eleven surgical management patients (39%) had at least one subsequent NMIBC recurrence, with four (14%) experiencing more than one. Five chemoablation patients (9%) and six surgical management patients (21%) had a TURBT. No statistically significant differences were found between the groups.

One patient had a second primary cancer diagnosed before their NMIBC recurrence and was censored from analysis of time to first post-three month recurrence. No significant difference was observed between treatment groups in recurrence rates over time (Figure 3A). When explored by disease status and treatment at three months (Figure 3B), surgical management patients with disease at three months did significantly worse (p=0.01). The proportion free of subsequent recurrence at 12 months was similar across other groups.

Progression rate and overall survival

No participants experienced disease stage progression, although five patients had grade progression to carcinoma in situ and/or G3Ta at three months (Table 3). Two participants

(one in each group) died during follow-up, both from cardiac events not considered disease-related, both had complete response at three months.

Safety and tolerability

Post-treatment adverse event data were available for 81 participants. No serious adverse events or grade 3-4 adverse events were reported. Grade two adverse events were reported for 14/81 participants (17%), and for 29/81 patients (36%) a worst grade of one was reported. No differences between groups were found (see supplementary Tables 2-4). In the surgical management group, 7/28 patients (25%) experienced complications prior to discharge from surgery, mostly haematuria (six patients; 21%).

HRQOL

Seventy-eight participants consented to the optional HRQOL sub-study (51 chemoablation, 27 surgical management). The two treatment groups exhibited similar HRQOL throughout follow-up, both in global quality of life and other key subscales of interest (Figure 4; supplementary Figures 1 to 2).

Discussion

We demonstrated feasibility of randomisation between surgical and medical management of low risk NMIBC. Chemoablation with four MMC instillations was well tolerated. The pre-defined criterion for progression to stage two was not met and the trial closed early, but a sustained reduction in recurrence rate was suggested. HRQOL was not substantially impacted by either treatment.

To our knowledge this is the first study to measure the effect of chemoablation using histological rather than visual criteria. Complete response rates in both groups were lower than expected when compared to previous studies reporting visual complete response only (11, 12). Based on our findings, visual complete response should be used with caution as a primary endpoint in NMIBC trials, although its pragmatic use in a clinical setting is probably acceptable.

At 12 months, recurrence rate was similar between patients with complete response at three months in both groups (16%). Rates were also similar in patients who 'failed' chemoablation and were 'salvaged' by surgical management at three months. On the other hand, patients who 'failed' surgical management without prior intravesical chemoablation had a 12-month recurrence rate of 60%, although caution is needed due to the small size of the groups.

The use of four instillations of MMC was chosen pragmatically to fit into the UK national 31 day target for cancer surgery and avoid delaying surgery if there was no response. A more intensive or extended regimen may result in improved response rates and any further research should consider this. Our results suggest that four MMC instillations may have some chemo-protective effect against low risk NMIBC recurrence. There remains a group of frail patients who tolerate surgery poorly, for whom a near 50% chance of complete ablation of visible tumours may be beneficial in terms of safety and improved quality of life.

A particular challenge for trials in low risk NMIBC is that the diagnosis can only be confirmed after tissue examination from TURBT. Therefore we could only recruit patients with a previous low risk NMIBC diagnosis who had a recurrence. In order to ensure consistency in definition of low risk NMIBC across multiple hospitals we used the EORTC risk score tables(18) rather than the European Association of Urology's NMIBC guideline risk categories(2). These constraints had important consequences; although 50% of newly

diagnosed NMIBC patients are low risk, over two thirds never have any subsequent recurrence(2), whilst those that do (those eligible for this study) are re-classified as intermediate risk patients, both according to EAU guidelines and the EORTC risk tables. The results should therefore be interpreted in this context.

The trial has a number of weaknesses. It was not powered for direct comparison of response rate between randomised groups, limiting ability to definitively identify differences between treatments. The study population likely reflects a group of patients with intermediate, rather than low risk NMIBC, limiting ability to extrapolate results to newly diagnosed low risk NMIBC. To assess potential comparators for phase III, the control arm permitted different surgical options including biopsy with diathermy, potentially underestimating the benefits of an expertly conducted TURBT. Only three surgical group participants (11%) received a post-operative MMC instillation - had all surgical management participants received this, the observed surgical CR may have been higher and subsequent recurrence rate reduced. Finally, there was relatively poor compliance with the biopsy at three months so visual assessment of response was not verified by histology for every participant.

Alternative strategies for managing low risk NMIBC include active surveillance (9) and office fulguration. Whilst active surveillance appears safe, our patient focus group indicated this was not a popular strategy. Office fulguration is popular in some countries since it avoids general anaesthesia and is therefore cost-effective but it is not popular amongst patients in the UK and is often painful particularly for elderly patients. Moreover, in the surgical arm of CALIBER 57% of patients had fulguration (rather than TURBT) and nearly 20% had residual disease at three months which calls into question the effectiveness of using this strategy alone.

Ultimately, all three strategies have an important role to play in reducing the burden of treatment on frail patients undergoing low risk NMIBC surveillance. One could consider chemoablation in frail patients presenting with multifocal or very large papillary tumours prior to TURBT in the expectation that some will have their tumour burden reduced at surgery. Our results indicate that a neo-adjuvant course of intra-vesical chemotherapy, given over a short period, is well tolerated and may provide additional therapeutic benefit over surgical management alone.

In conclusion, low risk NMIBC management with chemoablation as an alternative to TURBT is feasible and safe, but our study did not reach the pre-specified level of complete response. Nevertheless, following chemoablation there appears to be a sustained reduction in recurrence rate that is greater than with surgical management alone. Further research is required to investigate the role and optimal schedule of neo-adjuvant therapy prior to TURBT.

Acknowledgements

HM is the CALIBER trial Chief Investigator and EH is the methodological lead. Both led study design and acquired funding for the trial.

HM, NP, JC, TRLG, JK, SP, KD, JM, NC, MK, AF, AK, JWFC, RL and EH are members of the CALIBER Trial Management Group which contributed to study design, was responsible for oversight throughout the trial and contributed to data interpretation and manuscript preparation.

HM, JC, KD, JM, PC, SM and JWFC were involved in recruitment and treatment of participants and contributed to data collection and manuscript preparation.

EH oversaw statistical analyses and was responsible for central management of the trial at ICR-CTSU, with RL's support.

SP conducted central study management at ICR-CTSU and contributed to data acquisition, interpretation and manuscript writing.

NP conducted statistical analyses at ICR-CTSU and contributed to data interpretation and manuscript writing.

All authors reviewed and approved the manuscript.

Mr. Mostafid reports grants from the Department of Health during the conduct of the study and personal fees from Astrazeneca, Olympus, Cepheid, and Medac outside the submitted work.

Prof Hall reports grants from the Department of Health during the conduct of the study; grants and non-financial support from Merck Sharp & Dohme, Astra Zeneca, Janssen-Cilag,

Bayer, Kyowa Hakko UK, Alliance Pharma (previously Cambridge Laboratories), Aventis Pharma Limited (Sanofi), Cancer Research UK and Accuray Inc. outside the submitted work.

Prof. Catto reports personal fees from Astra Zeneca, personal fees from Janssen, personal fees from Roche, personal fees from Ferring, personal fees from MSD, personal fees from Bristol-Myers Squibb, during the conduct of the study.

No other authors have any competing interests to declare.

Grateful thanks are due to all the patients who have participated in this study; all involved staff at the participating centres; and trials unit staff at ICR-CTSU. We also acknowledge on-going contributions of colleagues in the CALIBER Trial Management Group, Trial Steering Committee and Independent Data Monitoring Committee.

This report presents independent research funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0712-28112). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

The funder had no involvement in study design, data collection, analysis or manuscript preparation or publication decisions.

Programme grants from Cancer Research UK also supported the work of the ICR-CTSU (C1491/A15955; C1491/A25351).

Trial recruitment and on-going patient follow-up is supported within centres by the NIHR funded National Cancer Research Network. The Clinical Trials and Statistics Unit at the Institute of Cancer Research (ICR-CTSU) is supported by Cancer Research UK core grant (C1491/A15955). We acknowledge NHS funding to the NIHR Biomedical Research Centre at The Royal Marsden and the ICR.

Grateful thanks to the contributing Principal Investigators (no. of patients recruited): Ms Kim Davenport (13) / Prof. James Catto (9) / Ms Joanne Cresswell (7) / Mr Peter Cooke (6) / Mr Shikohe Masood (5) / Mr John McGrath (5) / Mr Hugh Mostafid (5) / Prof Sanjeev Madaan (4) / Mr Stephen Andrews (3) / Mr Robert Brierley (3) / Mr Jonathan Goddard (3) / Mr Sunjay Jain (3) / Mr Kevin Turner (3) / Mr Barnaby Chappell (2) / Mr Jaswant Mom (2) / Miss Hazel Warburton (2) / Mr Nasr Arsanious (1) / Mr John Calleary (1) / Mr Gerald Collins (1) / Mr Mark Johnson (1) / Mr Martin Nuttall (1) / Mr Ahmed Qteishat (1) / Mr Subramanian Kanaga- Sundaram (1).

Figure legends

Figure 1: CONSORT diagram

Eighty patients were included in the primary and efficacy endpoints' analysis: two surgical management patients without a three month assessment were excluded (one withdrew from trial treatment after randomisation, one was lost to follow-up before three months). All patients for whom there were completed post-treatment and/or three month adverse events forms were included in the safety analyses (N=81). Nine patients (three surgical management, six chemoablation) were found ineligible after randomisation but were included in all analyses in accordance with the CALIBER Statistical Analysis Plan.

Figure 2: Response at three month assessment – visual vs histological confirmation

Figure 3: Kaplan-Meier estimate of proportion free of subsequent recurrence after three month disease assessment, by allocated treatment (A) and by allocated treatment and disease status (B).

Patients who had a second primary cancer or died because of reasons other than bladder cancer without a prior recurrence were censored at date of second primary or date of death. Stratified log-rank test and stratified Cox model to explore the differences between treatment groups were used as appropriate to account for disease response status at three months (Figure A). When treatment and disease status were combined to form four groups, these were compared by log-rank test (not stratified). Proportional hazards were tested using Schoenfield residuals.

Figure 4 – HRQOL: Change from baseline in QLQ-C30 global health scale

High score at any timepoint represents high quality of life. Positive change from baseline (calculated score at timepoint – score at baseline) represents improvement in quality of life. Questionnaire return rates were 91% at baseline, 72% at three months post end of treatment, 92% at six months, and 85% at 12 months.

Tables

Table 1. Baseline characteristics of CALIBER participants

	Surgical management (N=28)		Chemoablation (N=54)		Total (N=82)	
	N	%	N	%	N	%
Gender						
Male	23	82.1%	40	74.1%	63	76.8%
Female	5	17.9%	14	25.9%	19	23.2%
Age (years)						
Mean (SD)	69.3 (11.5)		73.4 (7.6)		72.0 (9.2)	
Median (Q1-Q3)	70.7 (61.1-77.1)		72.5 (68.8-78.3)		72.4 (66.8-77.9)	
Number of tumours at trial entry						
1	21	75.0%	47	87.0%	68	82.9%
2-7	7	25.0%	7	13.0%	14	17.1%
Max tumour diameter at trial entry						
<3cm	27	96.4%	54	100.0%	81	98.8%
≥3cm	1	3.6%	0	0.0%	1	1.2%
Recurrence rate at trial entry						
≤ 1 year	27	96.4%	49	90.7%	76	92.7%
> 1 year[1]	1	3.6%	5	9.3%	6	7.3%
Number of previous occurrences of NMIBC[2]						
1	15	53.6%	30	55.6%	45	54.9%
2	8	28.6%	12	22.2%	20	24.4%
3	4	14.3%	4	7.4%	8	9.8%
4	0	0.0%	3	5.6%	3	3.7%
≥5	1	3.6%	55	99.3%	66	77.3%
Prior MMC (single instillation)						
Yes	19	67.9%	33	61.1%	52	63.4%
No	8	28.6%	18	33.3%	26	31.7%
Unknown	1	3.6%	3	5.6%	4	4.9%
Grade at original diagnosis						
G1	15	53.6%	22	40.7%	37	45.1%
G2	13	46.4%	32	59.3%	45	54.9%
Risk score at trial entry						
2	10	35.7%	21	38.9%	31	37.8%
3	10	35.7%	24	44.4%	34	41.5%
5	5	17.9%	3	5.6%	8	9.8%

6	2	7.1%	3	5.6%	5	6.1%
8(1)	1	3.6%	3	5.6%	4	4.9%

NMIBC:Non muscle-invasive bladder cancer; MMC: mytomicin-C; SD: standard deviation; Q1: first quartile, 25% percentile; Q3: 3rd quartile, 75% percentile.

(1) Patients found ineligible after randomisation, due to incorrect calculation of the risk score at site.

(2) Including diagnosis; overall (since diagnosis)

360

Table 2. Surgical management group: Details of surgical technique and histology at trial entry

	Surgical management (N=28)	
	N	**%**
Type of surgery		
Diathermy	16	57.1%
TURBT	12	42.9%
Single post-operative MMC instillation given		
Yes	3	10.7%
Stage		
Benign	*3*	*10.7%*
Ta	*18*	*64.3%*
Grade		
Benign	*3*	*10.7%*
G1	*6*	*21.4%*
G2	*11*	*39.3%*
GX	*1*	*3.6%*

TURBT: transurethral resection of bladder tumour;

MMC: mytomicin-C.

Table 3. Three month assessment: Details of surgical technique and histology

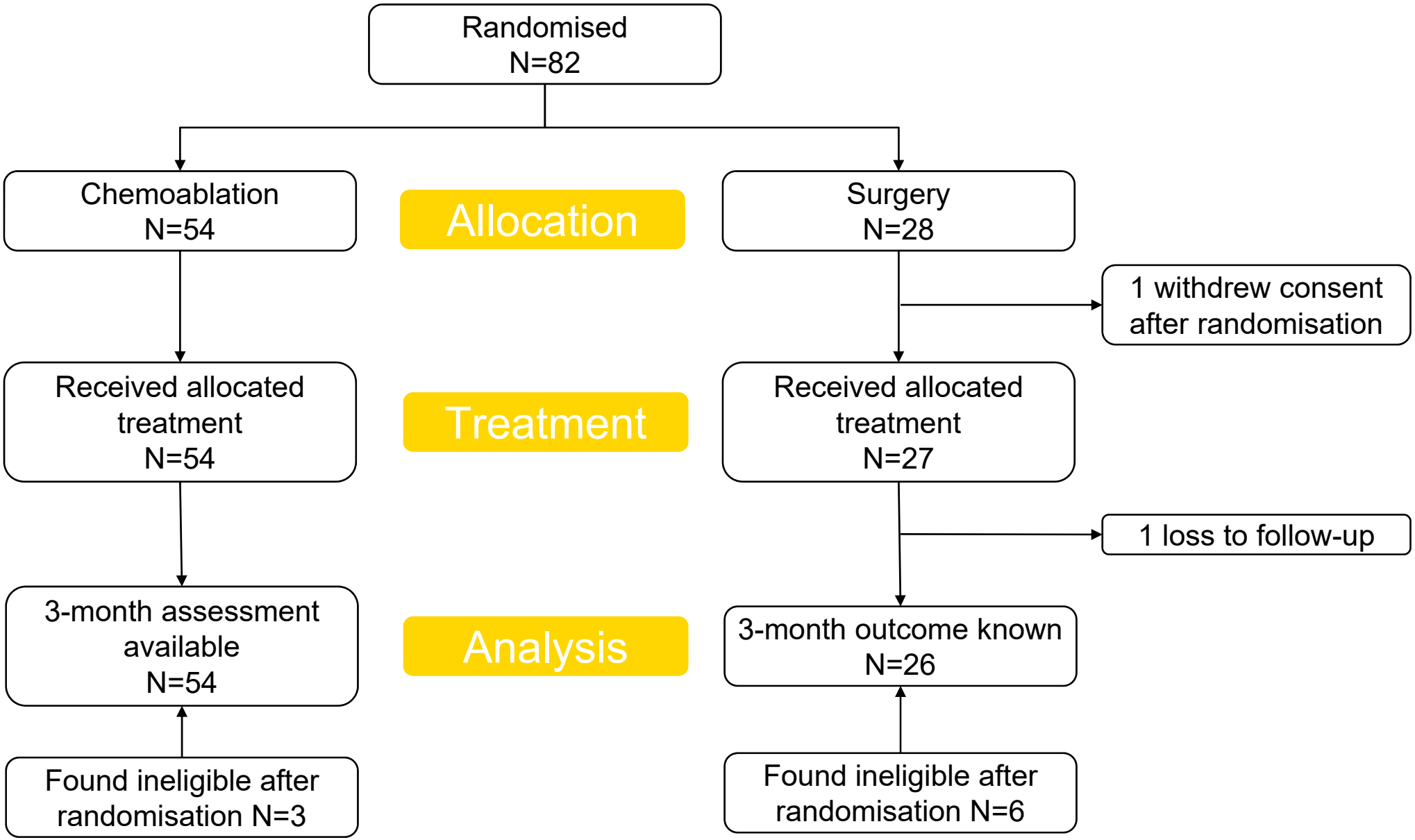
	Surgical management		Chemoablation		Total	
	N	%	N	%	N	%
Patients with disease present at three months (visual and histologically, where available)	**5**	**100%**	**34**	**100%**	**39**	**100%**
Treatment for residual disease						
Diathermy	1	20%	11	32.4%	12	30.8%
TURBT	3	60%	19	55.9%	22	56.4%
Biopsy alone	1	20%	3	8.8%	4	10.3%
Cystoscopy alone	0	0	1	2.9%	1	2.6%
Single post-operative MMC instillation given						
Yes	0	0	2	5.9%	2	5.1%
Number of tumours						
1	5	100%	20	58.8%	25	64.1%
2-7	0	0	12	35.3%	12	30.8%
Unknown	0	0	2	5.9%	2	5.1%
Max tumour diameter						
<3cm	4	80%	29	85.3%	33	84.6%
≥3cm	1	20%	2	5.9%	3	7.7%
Unknown	0	0	3	8.8%	3	7.7%
Stage						
Benign	0	0	3	8.8%	3	7.7%
Ta	5	100%	27	79.4%	32	82.1%
Ta+CIS	0	0	1	2.9%	1	2.6%
CIS	0	0	1	2.9%	1	2.6%
Unknown	0	0	2	5.9%	2	5.1%
Grade						
Benign	0	0	3	8.8%	3	7.7%
G1	0	0	10	29.4%	10	25.6%
G2	4	80%	13	38.2%	17	43.6%
G3	1	20%	3	8.8%	4	10.3%
GX	0	0	1	2.9%	1	2.6%
Unknown	0	0	4	11.8%	4	10.3%
Disease location						
Same as trial entry	5	100%	32	94.1%	37	94.9%
Different location	0	0	2	5.9%	2	5.1%

TURBT: transurethral resection of bladder tumour; MMC: mytomicin-C.

366

References

1. Antoni S, Ferlay J, Soerjomataram I, et al. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. European urology. 2017;71(1):96-108.
2. Babjuk M, Burger M, Comperat EM, et al. European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and Carcinoma In Situ) - 2019 Update. European urology. 2019.
3. Linton KD, Rosario DJ, Thomas F, et al. Disease specific mortality in patients with low risk bladder cancer and the impact of cystoscopic surveillance. The Journal of urology. 2013;189(3):828-33.
4. Boustead GB, Fowler S, Swamy R, et al. Stage, grade and pathological characteristics of bladder cancer in the UK: British Association of Urological Surgeons (BAUS) urological tumour registry. BJU international. 2014;113(6):924-30.
5. Noon AP, Albertsen PC, Thomas F, Rosario DJ, Catto JW. Competing mortality in patients diagnosed with bladder cancer: evidence of undertreatment in the elderly and female patients. Br J Cancer. 2013;108(7):1534-40.
6. Burke DM, Shackley DC, O'Reilly PH. The community-based morbidity of flexible cystoscopy. BJU international. 2002;89(4):347-9.
7. Mason SJ, Downing A, Wright P, et al. Health-related quality of life after treatment for bladder cancer in England. Br J Cancer. 2018;118(11):1518-28.
8. Svatek RS, Hollenbeck BK, Holmang S, et al. The economics of bladder cancer: costs and considerations of caring for this disease. European urology. 2014;66(2):253-62.
9. Soloway MS, Bruck DS, Kim SS. Expectant management of small, recurrent, noninvasive papillary bladder tumors. The Journal of urology. 2003;170(2 Pt 1):438-41.
10. Wong KA, Zisengwe G, Athanasiou T, O'Brien T, Thomas K. Outpatient laser ablation of non-muscle-invasive bladder cancer: is it safe, tolerable and cost-effective? BJU international. 2013;112(5):561-7.
11. Gofrit ON, Zorn KC, Shikanov S, Steinberg GD. Marker lesion experiments in bladder cancer--what have we learned? The Journal of urology. 2010;183(5):1678-84.
12. Bono AV, Hall RR, Denis L, Lovisolo JA, Sylvester R. Chemoresection in Ta-T1 bladder cancer. Members of the EORTC Genito-Urinary Group. European urology. 1996;29(4):385-90.
13. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute. 1993;85(5):365-76.
14. Blazeby JM, Hall E, Aaronson NK, et al. Validation and reliability testing of the EORTC QLQ-NMIBC24 questionnaire module to assess patient-reported outcomes in non-muscle-invasive bladder cancer. European urology. 2014;66(6):1148-56.
15. Simon R. Optimal two-stage designs for phase II clinical trials. Controlled clinical trials. 1989;10(1):1-10.
16. Fayers P, NKAaronson, Bjordal K, et al. EORTC QLQ-C30 Scoring Manual. 3rd ed2001.
17. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
18. Sylvester RJ, van der Meijden AP, Oosterlinck W, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. European urology. 2006;49(3):466-5; discussion 75-7.


Randomised
N=82
Chemoablation
N=54
Surgery
N=28
Allocation
1 withdrew consent after randomisation
Received allocated treatment
N=54
Received allocated treatment
N=27
Treatment
1 loss to follow-up
3-month assessment available
N=54
3-month outcome known
N=26
Analysis
Found ineligible after randomisation N=3
Found ineligible after randomisation N=6

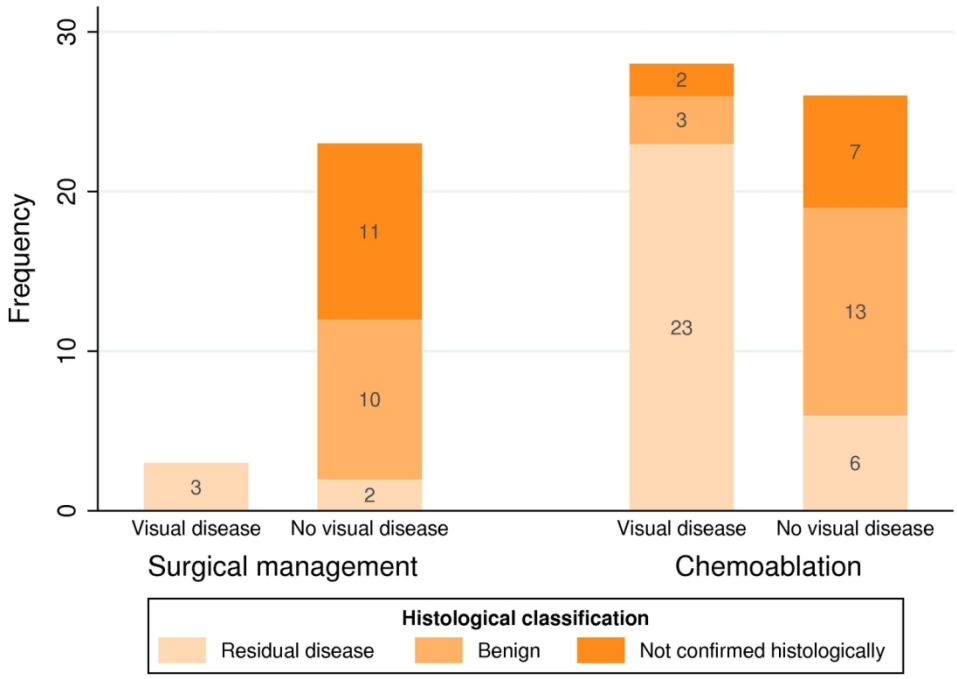


Figure 2

139x101mm (300 x 300 DPI)

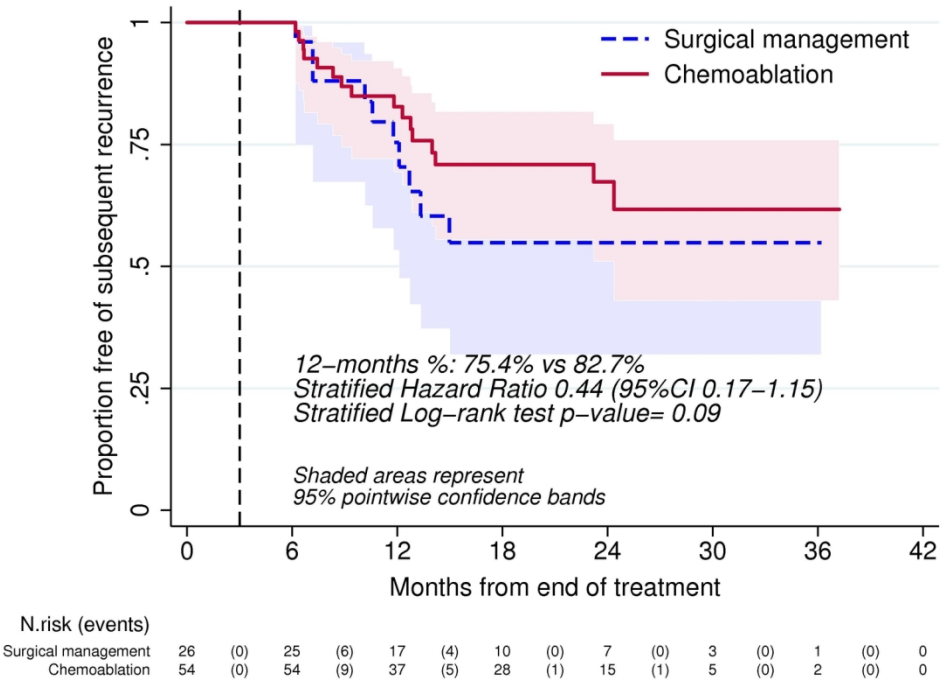


Figure 3A

139x101mm (300 x 300 DPI)

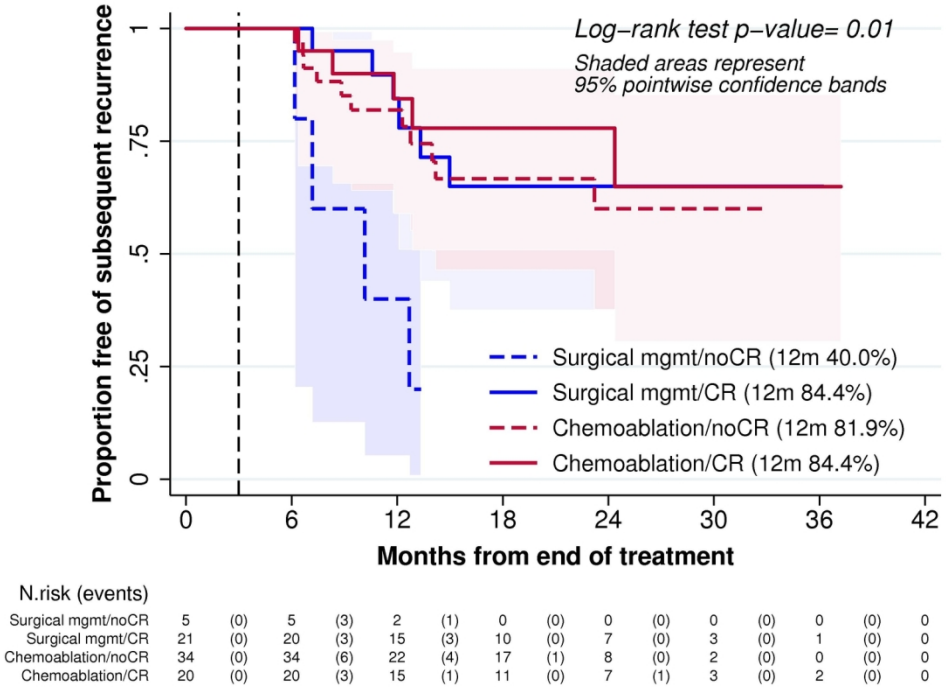


Figure 3B

139x101mm (300 x 300 DPI)

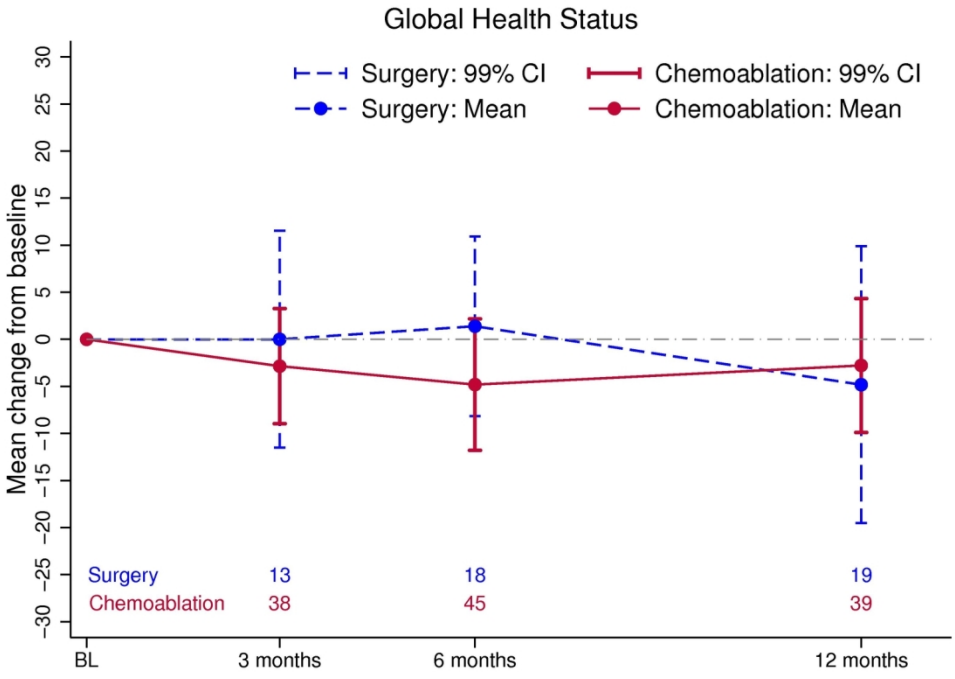


Figure 4

139x101mm (300 x 300 DPI)

Supplementary material: Revised CALIBER design following challenges in recruitment

Original sample size (from protocol v1.0 06/08/2014)

Following consultation with patient representatives, CALIBER was designed to exclude a CR rate of less than 45%. This was on the basis that if the CR rate is less than 45%, chemoablation would not be an attractive alternative to surgical management as it would delay rather than prevent surgical intervention in the majority of patients and hence would be unlikely to reduce the burden of treatment of recurrence. If the CR rate is>60% then the strategy of chemoablation would warrant further investigation. Using a Simon's 2 stage phase II optimal design (to allow early stopping for futility) with α=0.05, 90% power, p0=0.45, p1=0.60 the required sample size is 51 chemoablation patients in the first stage. If fewer than 26 CRs are seen in chemoablation patients in the first stage, then recruitment would cease (there having been no previous break in recruitment, to allow determination of CR in stage 1 patients, on completion of accrual to stage 1). At the end of the second stage, if at least 58/110 chemoablation patients have a CR then it would be concluded that chemoablation demonstrated adequate activity to warrant further investigation. An allocation ratio of 2:1 to the chemoablation group was selected to maximise information in the experimental group whilst providing contemporaneously collected information in an unbiased control group to enable informal comparisons that would support development of a phase III trial. Therefore, inflating to account for 5% noncompliance, and to include a control group, this gives a total target recruitment of 174 patients, 116 in the chemoablation group (54 in stage 1; 62 in stage 2) and 58 in the surgical management group (27 in stage 1; 32 in stage 2).

Revised sample size (implemented in protocol v6.0 20/06/2017)

Due to slower than anticipated accrual and with advice and approval of the Independent Data Monitoring and Trial Steering Committees, the 2 stage trial design was adapted. This adaption was made without knowledge of the CR rate in stage 1 i.e. prior to the decision to stop/go at the completion of stage 1. Based on good acceptance rates amongst eligible patients (approximately 56% as at February 2017 as reported on screening logs), the TSC recommended that the control group could be dropped for stage 2 – the feasibility of randomisation being proven during stage 1. To address recruitment timelines, the TSC advised that the overall power and significance levels could be relaxed, whilst maintaining the original stage 1 decision rule, to achieve a reduced overall total sample size.

Following completion of recruitment to stage 1 (51 evaluable chemoablation patients) and in the absence of any safety concerns raised by the Independent Data Monitoring Committee, recruitment to stage 2 would commence with all patients receiving chemoablation. With 51 chemoablation patients recruited in stage 1 and an additional 9 chemoablation patients recruited in stage 2, the adapted 2-stage design retains p0=0.45 and p1=0.60 and the threshold for activity at stage 1 (stop/go criteria) of at least 26 responders in 51 chemoablation patients and provides 85% power and 10% one-sided significance. If at the end of stage 2, at least 31/60 chemoablation patients had a CR then it would be concluded that chemoablation demonstrated adequate activity to warrant further investigation.

Therefore, nine additional chemoablation patients would be required in stage 2 (giving a total of 60 chemoablation patients) with an overall target sample size of 89 patients, including the control group patients at stage 1 (26 patients) and allowing a 5% drop out (unevaluable) rate in the chemoablation group.

Supplementary Table 1. Complete response rates three months after end of treatment – sensitivity analyses

Sensitivity analyses of the primary endpoint have been performed on the per protocol and the eligible populations. In addition, the following sensitivity analyses have been performed:

- Sensitivity analysis 1, surgery group: excluding evaluable patients found to be benign at baseline
- Sensitivity analysis 1, chemoablation group: evaluable patients with visual disease at three months found to be benign are considered CR in the combined visual/histological assessment.
- Sensitivity analysis 2, surgery group: Exclude from analysis two evaluable patients who received 6-course MMC following surgery and before the 3-month check.

	Surgery				Chemoablation			
	N	**CR**	**Rate**	**95% CI**	**N**	**CR**	**Rate**	**95% CI**
Per protocol population	*Exclude ineligible, 3-m visit deviations, benign at baseline*				*Exclude ineligible, 3-m visit deviations*			
Visual assessment only	19	18	94.7%	74.0-99.9	43	20	46.5%	31.1-62.3
Visual and histological assessment (where available)	19	17	89.5%	66.9-98.7	43	15	34.9%	21.0-50.9
Eligible population	*Exclude ineligible patients*				*Exclude ineligible patients*			
Visual assessment only	23	21	91.3%	72.0-98.9	48	23	47.9%	33.3-62.8
Visual and histological assessment (where available)	23	19	82.6%	61.2-95.0	48	18	37.5%	24.0-52.6
Sensitivity analysis 1	*Exclude patients benign at surgery*				*Classify benign residual disease as CR in combined assessment*			
Visual assessment only	24	21	87.5%	67.6-97.3	54	26	48.1%	34.3-62.2
Visual and histological assessment (where available)	24	19	79.2%	57.8-92.9	54	23	42.6%	29.2-56.8
Sensitivity analysis 2	*Exclude patients who received MMC before 3 months*							
Visual assessment only	24	21	87.5%	67.6-97.3				
Visual and histological assessment (where available)	24	19	79.2%	57.8-92.9				

CR: Complete Response; CI: confidence interval

Supplementary Table 2 - Worst CTCAE grade adverse event reported by visit

	CTCAE grade	Surgery		Chemoablation		Total		p-value
		N	%	N	%	N	%	
Pre-randomisation	**Total**	**28**	**100.0%**	**54**	**100.0%**	**82**	**100.0%**	0.63
	0	21	75.0%	42	77.8%	63	76.8%	
	1	6	21.4%	8	14.8%	14	17.1%	
	2	1	3.6%	4	7.4%	5	6.1%	
Post-treatment	**Total**	**22**	**100.0%**	**53**	**100.0%**	**75**	**100.0%**	0.30
	0	11	50.0%	30	56.6%	41	54.7%	
	1	6	27.3%	18	34.0%	24	32.0%	
	2	5	22.7%	5	9.4%	10	13.3%	
3 month post-treatment	**Total**	**26**	**100.0%**	**53**	**100.0%**	**79**	**100.0%**	0.74
	0	15	57.7%	35	66.0%	50	63.3%	
	1	9	34.6%	14	26.4%	23	29.1%	
	2	2	7.7%	4	7.5%	6	7.6%	

Supplementary Table 3 - Worst CTCAE grade treatment emergent adverse events

A treatment-emergent adverse event is defined as an event not present prior to the initiation of trial treatment or an event already present that worsens at end of treatment or at 3 month follow-up.

	Surgery		Chemoablation		Total	
	N	%	N	%	N	%
Total CTCAE grade	**28**	**100.0%**	**53**	**100.0%**	**81**	**100.0%**
0	15	53.6%	25	47.2%	40	49.4%
1	8	28.6%	21	39.6%	29	35.8%
2	5	17.9%	7	13.2%	12	14.8%

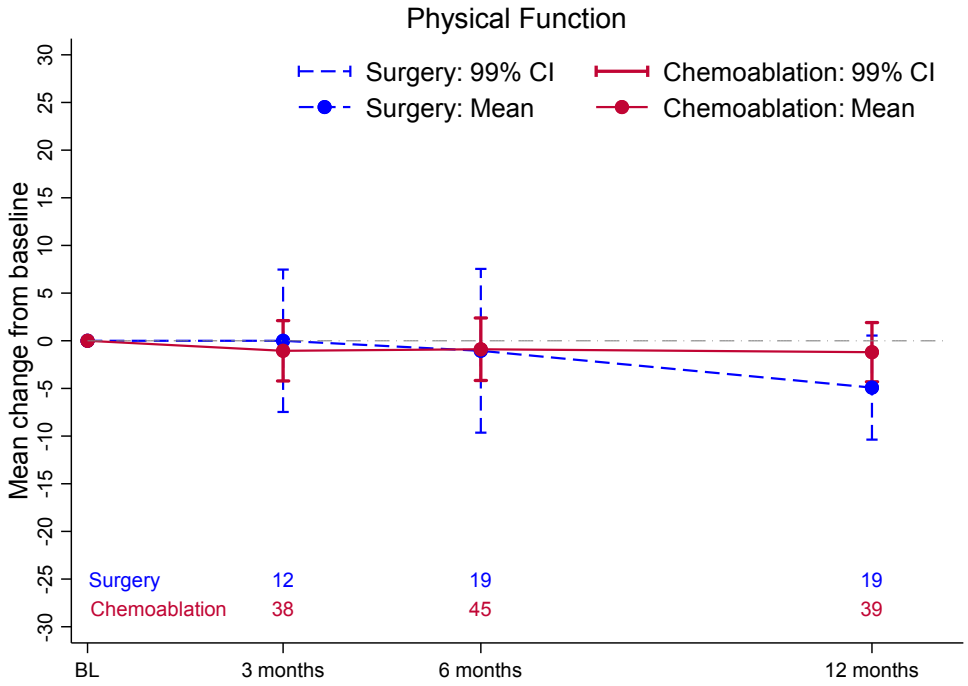
Chi-square p-value: 0.59

Supplementary Table 4 - Worst CTCAE grade treatment emergent adverse events by type of event

	CTCAE grade	Surgery		Chemoablation		p-value*
		N	%	N	%	
Total patients		**28**	**100.0%**	**53**	**100.0%**	
Anorexia	1	1	3.6%	0	0.0%	0.35
Bladder infection	2	3	10.7%	0	0.0%	0.02
Bladder spasm discomfort	1	4	14.3%	2	3.8%	0.09
Haematuria	1	3	10.7%	3	5.7%	0.34
	2	2	7.1%	1	1.9%	
Malaise	1	4	14.3%	2	3.8%	0.18
Nausea	1	1	3.6%	3	5.7%	0.65
	2	0	0.0%	1	1.9%	
Platelet count decreased	1	0	0.0%	1	1.9%	0.81
Rash	1	0	0.0%	3	5.7%	0.29
	2	0	0.0%	1	1.9%	
Urinary frequency	1	8	28.6%	9	17.0%	0.27
	2	1	3.6%	1	1.9%	
Urinary incontinence	1	1	3.6%	0	0.0%	0.12
	2	1	3.6%	0	0.0%	
Urinary obstruction	1	1	3.6%	2	3.8%	>0.99
Urinary retention	1	2	7.1%	3	5.7%	>0.99
Urinary tract pain	1	6	21.4%	4	7.5%	0.10
	2	1	3.6%	1	1.9%	
Urinary urgency	1	6	21.4%	6	11.3%	0.13
	2	2	7.1%	1	1.9%	
Other conditions reported						
Diarrhoea	1	1	3.6%	1	1.9%	
Abdominal pain	1	1	3.6%	0	0.0%	
Back pain	1	1	3.6%	0	0.0%	
Candida infection	2	0	0.0%	1	1.9%	
Cough	1	1	3.6%	0	0.0%	
Epistaxis	2	0	0.0%	1	1.9%	
Fatigue	1	0	0.0%	1	1.9%	
Feeling of body temperature change	1	0	0.0%	1	1.9%	na
Gouty arthritis	2	0	0.0%	1	1.9%	
Headache	1	0	0.0%	1	1.9%	
Labyrinthitis	1	0	0.0%	1	1.9%	
Nocturia	1	0	0.0%	1	1.9%	
Polymyalgia rheumatica	1	0	0.0%	1	1.9%	
Pruritus	1	0	0.0%	1	1.9%	

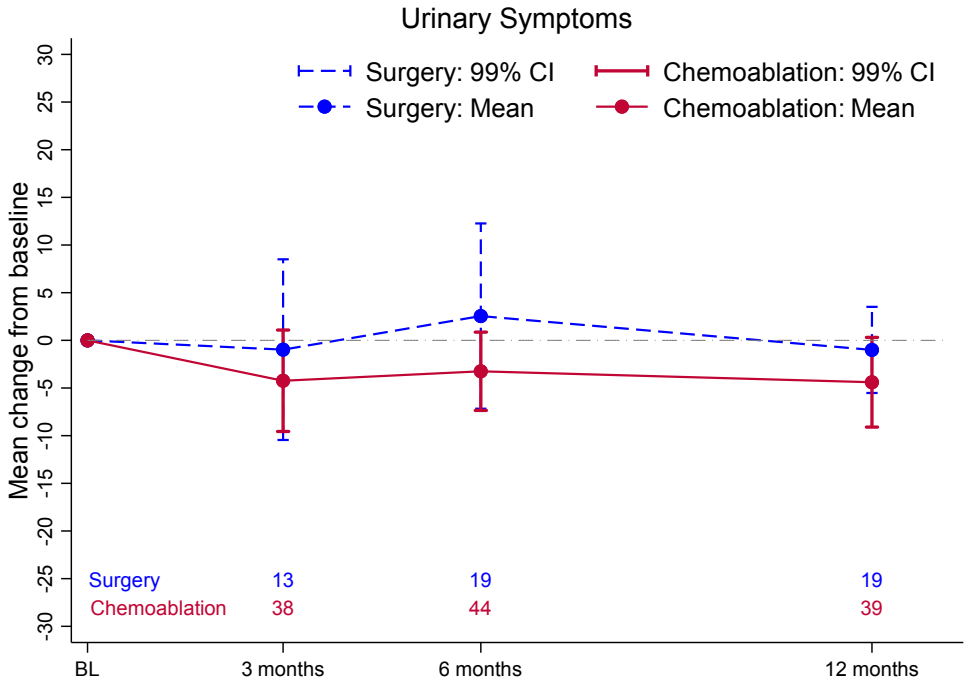
*Fisher exact test comparing Gr1+ vs Gr 0 or missing.

Supplementary Figure 1 – HRQOL Change from baseline in QLQ-C30 Physical function



High score represents high quality of life
Positive change from baseline (computed as timepoint – baseline scores) represents

Supplementary Figure 2 – HRQOL Change from baseline in QLQ-NMIBC24 Urinary symptoms



High score represents worse urinary symptoms. Positive change (computed as difference baseline – timepoint scores) represents improvement in symptoms.