Measuring health status using wearable devices for patients undergoing radical cystectomy

By

Pramit Khetrapal

A thesis submitted for the degree of

Doctor of Philosophy (Ph.D.)

University College London

Division of Surgery & Interventional Science

2021

Declaration

"I, Pramit Khetrapal, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Where assistance and materials have been obtained, I have acknowledged as appropriate"

…… ……

Date Signature

Thesis abstract

Wearable devices (WDs) are an untapped resource for measuring patient health status during the peri-operative period. The overarching aim of this thesis is to explore the potential for WDs to be used in the clinical setting for patients undergoing radical cystectomy (RC) for bladder cancer. The lack of consensus regarding the optimal approach for RC presents an opportunity to design an RCT comparing open (ORC) and robotic (RARC) RC, in which a wearable device sub-study can be embedded. While the intracorporeal Robotic vs Open Cystectomy (iROC) trial will address the comparison between ORC and RARC, my thesis focuses on exploring the clinical utility of WDs.

I present the results of a systematic review of RCTs comparing ORC and RARC. Meta-analysis shows no significant difference in peri-operative and oncological outcomes between ORC and RARC. Additionally, I systematically review healthcare studies using WDs and highlight the findings, device choices and device metrics used. Step-count is the most frequently collected WD metric, and chronic health conditions are the focus of majority of studies. Findings from these systematic reviews guided the design of the iROC trial protocol.

I present the pre-planned interim analysis of the iROC trial, and explore associations between WD data and pre-operative health measures including cardiopulmonary exercise testing (CPET). Step-count correlates with the CPET variables ($p < 0.01$) routinely used to risk-stratify patients undergoing RC, and is the only predictor of major complications following RC in a logistic regression model.

Finally, I evaluate recovery of baseline step-count at three months post-operatively as a predictor of overall survival. Applying a threshold of 50% recovery at 3 months, step-count predicts one-year survival to a sensitivity and specificity of 100% and 93% respectively.

My findings highlight the potential of WDs in peri-operative care, and my post-doctoral work will progress this work further.

Impact statement

Wearable devices (WDs) have become increasingly popular due to their appeal to fitness enthusiasts, but also offer a new opportunity in healthcare. Despite this, the uptake of WDs into routine care has been limited. In this doctoral thesis, I explored the utility of WDs in measuring health status from patients undergoing radical cystectomy, drawing on data from a multi-centre phase III RCT: intracorporeal Robotic vs Open Cystectomy (iROC) trial. This is the first RCT to explore WD data comparisons in the setting of radical cystectomy.

I have shown that it is feasible to embed WDs into clinical trials, and data can be reliably collected with minimal input from patients. WD step-count data do not correlate with patient-reported quality-of-life scores, suggesting that WDs offer new information regarding health status. Furthermore, WD data correlates with known risk-stratification tools such as cardiopulmonary exercise testing (CPET). CPET is not universally adopted, and the potential for WDs to provide similar information is interesting: WD data was an independent predictor of major complications following radical cystectomy. My work also shows that WD data also can be predictive of 1-year survival (sensitivity and specificity of 100% and 93% respectively). Taken together, the studies included highlight the potential application for WDs to measure patient health status in preoperative assessment, post-operative measurement of recovery, predicting overall and cancer-related survival. Each of these applications warrant independent studies to investigate the true value of WDs in statistically powered studies. These devices offer the ability to collect health data beyond step-count with longer battery lives and this data can be uploaded remotely to healthcare providers via 4G and Wi-Fi networks.

To harness the technological developments in the field, I have started a new prospective observational study to monitor patients who are discharged from hospital following major surgery with reported re-admission rates of >15% in 30 days. Preliminary results show that patients have signs of deterioration >48 hours before they present to A&E departments. The

work has been well received, as I have received three competitive research grants to develop this work further. Furthermore, I have published the preliminary data in a peer reviewed publication and presented it four times at international meetings.

I have no doubt that wearable devices will improve our collective understanding of surgical recovery, and have countless other healthcare applications. My post-doctoral work in the coming years will aim to test the application of wearable devices with the ambition to improve surgical care delivery for patients.

Acknowledgements

I would like to express my thanks to my supervisors Prof John Kelly and Dr Andrew Feber for their guidance and support in the last five years. I am grateful for the opportunity to have worked under their supervision to work on various research projects in bladder cancer, but also for their guidance to develop my own research ideas during my tenure in the Kelly-Feber lab. Their input, feedback and enthusiasm has been instrumental in developing a systematic academic mindset which will serve as a foundation of my academic and clinical career.

The work presented in this doctoral thesis has had input from other members of the clinical and academic team. A big thanks to all the members of the iROC trial management group and, particularly Ms Chris Brew-Graves for teaching me about the nuances of trial planning, ethics and associated processes. Mr Wei Shen Tan's input has had a strong influence in my work, and his help and guidance were instrumental in completing the systematic review and meta-analysis comparing open and robotic approaches to radical cystectomy. I am thankful to Ms Katherine Dennis and Mr Mohammed Abozaid for their help with collecting data for the preliminary experiments. I am also appreciative of the input of Professor James Catto and Prof John Kelly for providing oversight in designing and analysing the data for the iROC trial, as well as to all the sites that recruited patients into the trial. Additionally, a big thanks to Dr Patricia de Winter for being my sounding board, and providing feedback on various projects and presentations that I had to prepare.

I am grateful to the support of the Urology Foundation and the St Peter's Trust for awarding me research grants which allowed me to undertake this and other projects during my PhD.

Last but not least, I would like to thank my family for their support. I am extremely appreciative of the support, critique, proof-reading ability and academic mind of Dr Joanna Wong. I am very grateful to my parents for their unconditional love, constant encouragement and unwavering support which has enabled me to pursue my passions.

Achievements during PhD

I. List of awards during PhD

2019	NIHR Academic Clinical Fellowship in Urology (2019-2022)
2018	Best Academic Papers at BAUS 2018
2017	Best Poster Winner at the American Urological Association 2017
2016	Best Session Abstract at the 14th Urological Association of Asia Congress 2016
2017-2019	UCL Studentship for postgraduate work

II. List of funding and grants awarded during PhD

2018	The Urology Foundation Research Scholarship Award 2018 (£24,500)
2018	St Peter's Trust 2018 Grant Award 2018 (£30,125)
2018	UCLH BRC Patient and Public Involvement Starter Grant (£500)
2018	Roberts Brown Travel Award (£300)
2017	The Urology Foundation Research Scholarship Award 2017 (£42,321)

III. List of national/International presentations during PhD

2019 Expert-Guided Poster at European Association of Urology Congress 2019: *The iROC trial: An RCT comparing intracorporeal robot-assisted vs open radical cystectomy for bladder cancer*

2018 Best of BAUS session podium presentation: *The utility of plasma cell-free DNA mutations in detecting metastatic recurrence in patients after radical cystectomy for bladder cancer.*

2018 Poster presentation at American Urological Association meeting 2018: *The use of fitness tracker in monitoring functional activity of patients undergoing radical cystectomy for bladder cancer: A feasibility report and analysis as part of the multi-centre randomised iROC trial.*

2018 Extended presentation at European Association of Urology Congress 2018: *Using plasma cell-free DNA mutations to monitor patients for micro-metastatic bladder cancer after radical cystectomy.*

2017 Poster presentation at American Urological Association 2017: *Molecular tracking of bladder cancer using mutations detected in plasma cell-free DNA through radical cystectomy and chemotherapy.*

2017 Poster presentation at BAUS Annual Scientific meeting 2017: *Molecular characterisation of bladder cancer mutations using plasma cell-free DNA through chemotherapy and radical cystectomy.*

2016 Poster presentation at 14th Urological Association of Asia Congress 2016: *Blood-based liquid biopsy: Analysis of open and intracorporeal robotic cystectomy show recurrence patterns and oncological equivalence*

2016 Poster presentation at 14th Urological Association of Asia Congress 2016: *CIS with urothelial carcinoma carries worse prognosis than primary CIS: A cohort study*

2016 Poster presentation at 14th Urological Association of Asia Congress 2016: *Robotic Assisted Radical Cystectomy (RARC) with Intracorporeal Urinary Diversion: A single-centre experience reporting 90 day complications*

2016 Podium Presentation at 3rd International Symposium on Advanced Robotic Techniques in Prostate, Bladder and Renal surgery. *Does the robot matter? iROC: a multi-center randomized controlled trial comparing intracorporeal robotic cystectomy vs open cystectomy.*

2016 Poster presentation at BAUS Annual Scientific Meeting: *UroMark - A highly multiplex biomarker for the detection of bladder cancer*

2016 Poster presentation at 14th Urological Association of Asia Congress 2016: *Blood-based liquid biopsy: A comparison between plasma detection of cell-free DNA and circulating tumour cell detection in the detection of bladder cancer*

IV. List of publications during PhD

2019 **Khetrapal P**, Catto JWF, Kelly JD, iROC Trial Management Group. Robot-assisted versus open cystectomy in the RAZOR trial. Lancet. 2019 Feb 16;393(10172):644-645.

2019 Thompson JE, Sridhar AN, Tan WS, … **Khetrapal P,** … *et al*. Pathological Findings and Magnetic Resonance Imaging Concordance at Salvage Radical Prostatectomy for Local Recurrence following Partial Ablation Using High Intensity Focused Ultrasound. J Urol 2019;201:1134–43. doi:10.1097/JU.0000000000000135.

2018 **Khetrapal P**, Kelly JD, Catto JWF, Vasdev N. Does the robot have a role in radical cystectomy? BJU Int. doi: 10.1111/bju.14579

2018 Tan WS, Sarpong R, **Khetrapal P**, *et al.* Can renal and bladder ultrasound replace CT urogram in patients investigated for microscopic hematuria? *J Urol* 2018; published online April 24. DOI:10.1016/J.JURO.2018.04.065.

2018 Tan WS, Sarpong R, **Khetrapal P**, et al. Does urinary cytology have a role in haematuria investigations? BJU Int. doi: 10.1111/bju.14459

2018 Tan WS, Tan WP, Tan M-Y, **Khetrapal P**, et al. Novel urinary biomarkers for the detection of bladder cancer: A systematic review. Cancer Treat Rev 69:39–52 . doi: 10.1016/j.ctrv.2018.05.012

2018 Wong YNS, Joshi K, **Khetrapal P**, et al. Urine-derived lymphocytes as a non-invasive measure of the bladder tumor immune microenvironment. J Exp Med jem.20181003 . doi: 10.1084/jem.20181003

2018 Catto JWF, **Khetrapal P**, Ambler G, et al (2018) Robot-assisted radical cystectomy with intracorporeal urinary diversion versus open radical cystectomy (iROC): protocol for a randomised controlled trial with internal feasibility study. BMJ Open 8:e020500 . doi: 10.1136/bmjopen-2017-020500. **(Corresponding author)**

2018 Tan WS, Feber A, Sarpong R, **Khetrapal P,** … *et al.* Who Should Be Investigated for Haematuria? Results of a Contemporary Prospective Observational Study of 3556 Patients [Figure presented]. *Eur Urol* 2018; **74**: 10–4.

2018 Catto JWF*, **Khetrapal P***, Ambler G, *et al.* Multidomain Quantitative Recovery Following Radical Cystectomy for Patients Within the Robot-assisted Radical Cystectomy with Intracorporeal Urinary Diversion Versus Open Radical Cystectomy Randomised Controlled Trial: The First 30 Patients. *Eur Urol* 2018; : 8–10. **(joint first authors)**

2018 **Khetrapal P**, Lee MWL, Tan WS, *et al.* The role of circulating tumour cells and nucleic acids in blood for the detection of bladder cancer: A systematic review. *Cancer Treat Rev* 2018; **66**: 56–63.

2017 **Khetrapal P**, Tan WS, Lamb B, *et al.* The Role of Robotics in the Invasive Management of Bladder Cancer. *Curr Urol Rep* 2017; **18**: 57.

2017 Tan WS, Feber A, Dong L, **Khetrapal P,** … *et al.* DETECT I & DETECT II: a study protocol for a prospective multicentre observational study to validate the UroMark assay for the detection of bladder cancer from urinary cells. *BMC Cancer* 2017; **17**: 767.

2017 Ertemi H, **Khetrapal P**, Pavithran NMNM, Mumtaz F. Optimising renal cancer patients for nephron-sparing surgery: a review of pre-operative

considerations and peri-operative techniques for partial nephrectomy. *Urologia* 2017; **84**: 20–7.

2017 Tan WS, Lamb BW, **Khetrapal P**, *et al.* Blood Transfusion Requirement and Not Preoperative Anemia Are Associated with Perioperative Complications Following Intracorporeal Robot-Assisted Radical Cystectomy. *J Endourol* 2017; **31**: end.2016.0730.

2017 **Khetrapal P**, Tan WS, Kelly JD. Factors Affecting the Cost of Radical Cystectomy in the USA: Some Centres Are More Equal than Others. *Eur Urol* 2017; : 17–8.

2016 Tan WS, **Khetrapal P**, Tan WP, Rodney S, Chau M, Kelly JD. Robotic Assisted Radical Cystectomy with Extracorporeal Urinary Diversion Does Not Show a Benefit over Open Radical Cystectomy: A Systematic Review and Meta-Analysis of Randomised Controlled Trials. *PLoS One* 2016; **11**: e0166221.

2016 **Khetrapal P**, Shen TW, Lamb B, *et al.* Port-site metastases following robotic radical cystectomy: A systematic review and management options. *Clin Genitourin Cancer* 2016: 1–5.

Table of Contents

Abbreviations

A&E	Accident & Emergency department
ADLs	Activities of Daily Living
ASA	American Society of Anaesthesiologists physical status classification
ASC	Average step-count
AUA	American Urology Association
AUC	Area under the curve
BCG	Bacillus Calmette–Guérin
BLM30	European Organisation for Research and Treatment of Cancer QLQ-BLM30
C30	European Organisation for Research and Treatment of Cancer QLQ-C30
CD	Clavien Dindo
CSS	Cancer-specific Survival
CT	Computed Tomography
DFS	Disease-free Survival
DREAMPath	Domiciliary Return After Medicalisation Pathway
ECOG	Eastern Cooperative Oncology Group Performance Status
eCRF	electronic Case Report Form
EORTC	European Association of Research and Treatment of Cancer
eRARC	extracorporeal Robot-assisted radical cystectomy
GP	General Practitioner
HR	Hazard Ratio
HRQoL	Health-related Quality of Life

iRARC	intracorporeal Robot-assisted radical cystectomy
IRCC	International Robotic Cystectomy Consortium
iROC	intracorporeal Robotic vs Open Cystectomy
MDT	Multi-disciplinary team
MIBC	Muscle invasive bladder cancer
MiSC	Minimum step-count
MR	Misfit Ray
MSC	Maximum step-count
NAC	Neoadjuvant Chemotherapy
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMIBC	Non-muscle invasive bladder cancer
NPV	Negative Predictive Value
ORC	Open radical cystectomy
OS	Overall Survival
PPV	Positive Predictive Value
PROM	Patient-reported outcome measure
QoL	Quality of Life
RARC	Robot-assisted radical cystectomy
RC	Radical cystectomy
RCT	randomised controlled trial
ROC	Receiver operating characteristic
SCC	Squamous Cell Carcinoma

TURBT	Transurethral resection of bladder tumour
UCC	Urothelial Cell Carcinoma
UK	United Kingdom
USA	United States of America
WHODAS	World Health Organisation Disability Assessment Scale

List of figures

List of tables

Chapter 1 Introduction

1.1 Bladder Cancer

1.1.1 Cost burden and epidemiology

In 2014, the cost of bladder cancer care in the USA was estimated to be $US4.25 billion and had risen over successive years despite the static incidence of the disease[1]. Similarly, the annual cost of bladder cancer care in the European Union was estimated to be 4.9 billion euros[2]. Bladder cancer has the highest lifetime treatment cost per patient among all types of cancers[3].

Bladder cancer is the ninth most common in the world, with over 430,000 new cases diagnosed in 2012, and 165,000 bladder cancer deaths[4]. Urothelial cell carcinoma (UCC) is the most common type of cancer of the urinary bladder, and is responsible for >8,000 new cases of cancer and >4,000 deaths per year in England and Wales. Most of these cases are non-muscle invasive (NMIBC) at presentation. About 25% of bladder cancers present as muscle invasive bladder cancer (MIBC) which is defined as cancer invading into the detrusor muscle of the bladder wall. The 5-year overall survival for patients bladder cancer is 53.7%, but survival rates are worse for patients presenting with higher stage disease (down to 35.3% and 27.4% for males and females presenting with stage 4 disease)[5].

The most established risk factor for bladder cancer is tobacco smoking, with a population attributable risk of 0.50 and 0.52 for men and women respectively[6]. According to a recent meta-analysis by Cumberbatch *et al.*[7], the relative risk for current smokers is 3.47 (95% CI 3.07-3.91), and for ex-smokers is 2.04 (95% CI 1.85-2.25). Another significant risk factor is occupational exposure, accounting for 5.3% of bladder cancers (7.1% in males and 1.9% in females)[8].

While bladder cancer is nearly four times more common in men than women[9], women are more likely to be diagnosed with muscle-invasive bladder cancer (MIBC) at initial diagnosis (85%

in women and 51% in men)[10]. Consequently, the female gender is associated with a significant negative impact on cancer-specific survival (CSS) among bladder cancer patients[11].

1.1.2 Diagnosis, staging and grading

1.1.2.1 *Diagnosing bladder cancer*

Painless haematuria is the most common presenting symptom for bladder cancer. Other common symptoms include dysuria, polyuria and urgency. Patients with recurrent urinary tract infections are at increased risk of bladder cancer[12,13]. Patients with more advanced tumours can present with pelvic pain and urinary tract obstruction. Upon presentation, patients are referred to secondary care centres for urgent urological assessment as part of the 2-week suspected cancer pathway referral process.

The mainstay of diagnosis is flexible cystoscopy to assess the bladder and imaging to assess the upper urinary tract. The choice of imaging, either computerised topography urogram or renal ultrasound[14] varies according to whether haematuria is visible or non-visible, as well as the age and gender of the patient. In the UK, flexible cystoscopy is performed in patients over 45 years of age with visible haematuria, and in patients over 60 years of age with non-visible haematuria according to national guidelines published by NICE in 2015[15]. If a lesion suspicious of bladder cancer is noted on flexible cystoscopy, patients undergo a transurethral resection and the specimen obtained is then sent for pathological evaluation.

1.1.2.2 *Staging and grading*

The TNM classification's seventh edition[16] (Supplementary Table 10-1) is used for the staging of bladder cancer. This system is used to stage bladder cancer based on pre-operative imaging, as well as post-operative histology, denoted by a prefix of 'c' and 'p' respectively (e.g. cT1a or pT1a). Two versions of the WHO grading system[17] are used to grade bladder tumours, and both are outlined in Supplementary Table 10-2.

The 1973 WHO system grades tumours by level of de-differentiation, with well differentiated, moderately differentiated and poorly differentiated tumours being classified as grade 1, 2 and 3 respectively. The 2004 WHO system provides detailed architectural and cytological criteria to stratify tumours by their malignant potential: low, low-grade and high-grade.

Although both versions are used clinically, there remains some controversy about which system has better prognostic value[18]; The Royal College of Pathologists recommends using both systems but acknowledges that the 1973 classification remains in more widespread use[19].

1.1.3 Treatment pathway

Treatment offered depends on bladder cancer staging. Non-muscle invasive bladder cancer (NMIBC) is defined as ≤T1 disease and first-line treatment is transurethral resection of the bladder tumour (TURBT). Adjuvant therapy is offered on the basis of risk stratification[20]. Generally, a single instillation of intravesical chemotherapy is beneficial at the time of TURBT. Adjuvant therapy of up to six instillations is recommended for patients with intermediate risk NMIBC and adjuvant Bacillus Calmette-Guerin (BCG) for high-risk disease[20].

For patients who progress to MIBC, there are two main modalities for treatment: RC and radical radiotherapy. The decision between the two radical therapies is determined by patient preference and fitness for major surgery. In two large studies, Ghoneim *et al*. reported a 5-year overall survival (OS) of 55.5% following RC without chemotherapy[21], whilst a prior publication by Stein *et al*. reported a 5-year OS of 66% in a cohort in which 5% of patients were eligible for and received neoadjuvant chemotherapy (NAC)[22]. Prior to RC, eligible patients are offered platinum-based NAC. This recommendation is based on a systematic review and meta-analysis that showed a 5% OS benefit for NAC after 5 years of follow-up[23]. More recently, novel immunotherapy agents have shown activity in late stage disease and are currently being tested in trials in the first line setting[24], but more evidence is needed before their adoption into routine practice.

Radical radiotherapy is offered as an alternative treatment option to surgery. A meta-analysis[25] of RCTs (Bloom, 1982[26]; Sell, 1991[27]; Miller, 1977[28]) comparing surgery and radiotherapy for MIBC reported 5-year OS of 36% and 20% respectively. However, this meta-analysis draws on historical data, combined various treatment protocols and radiation dose, and draws conclusions from a relatively small population with a total of 439 patients. There is no large contemporary RCT comparing outcomes between RC and radical radiotherapy, so both options are offered to patients eligible for radical therapy.

1.2 Radical Cystectomy

RC is the surgical treatment for MIBC and recurrent NMIBC. The procedure can be broadly divided into two parts. The first part is an extirpative component, which consists of the removal of the urinary bladder, surrounding organs and lymph nodes. In females, this includes anterior pelvic exenteration (uterus, fallopian tubes and a component of the anterior vaginal wall) and the urethra in females choosing an ileal conduit. In males, this includes the prostate and seminal vesicles. RC is completed with removal of lymph nodes to the level of the common iliac vessels (standard template) or to the level of aorta bifurcation (extended template). The second part is the urinary diversion which can be classified as continent diversion (orthotopic neobladder reconstruction or a Mitrofanoff procedure) and incontinent diversion (ileal conduit or cutaneous ureterostomy). In the UK, 80.6% of urinary diversions are performed as an ileal conduit[29] and a further 6.9% as orthotopic reconstructions. This is however only an average and varies by institution. In 2017, approximately 27% of RCs were performed with continent diversion at our centre (University College London Hospitals NHS Foundation Trust).

Traditionally, open radical cystectomy (ORC) has been the approach for both cystectomy and urinary diversion performed via a lower midline incision. The first laparoscopic cystectomy was described for pyocystis and not bladder cancer, and lymph nodes were not removed (described as simple cystectomy). Sánchez de Badajoz *et al*. [30] described a laparoscopic radical

cystectomy (LRC) procedure for a patient with MIBC in which urinary diversion was performed as an extracorporeal or open approach. With the introduction of the da Vinci system in 2001, a robotic surgical system, there was considerable interest to replicate the principles of ORC and LRC using the robotic platform. The robotic approach has since gained popularity in the England, with 20.6% of cases performed robotically and 67.8% of cases being performed via open surgery[31].

The robotic approach aims to emulate the keyhole approach of laparoscopic surgery and the dexterity of open surgery. The approach to using the robotic platform can be either extracorporeal urinary diversion (eRARC) or intracorporeal urinary diversion (iRARC). Extracorporeal diversion involves an extra 5 to 7 cm skin incision[32] (muscle splitting incision in the right iliac fossa for an ileal conduit and lower midline incision for an orthotopic neobladder) to access the small intestine and ureters constructing the diversion similar to open surgery following completion of the extirpative component. The intracorporeal approach results in the diversion constructed using minimal access techniques without the need for the mini-laparotomy described in eRARC. Proponents of iRARC highlight potential benefits including less bowel manipulation, lack of open incision and retraction with potential for reduced incisional pain, decreased bowel exposure and reduced fluid imbalances[33].

1.2.1 Morbidity following ORC and RARC

Parts of this section have been adapted from an article published in the British Journal of Urology International describing the results of the RAZOR trial and the ORC vs RARC debate[34].

In a study reporting peri-operative outcomes for 1142 patients undergoing ORC, a total of 1637 complications were reported in 735 (64%) patients within 90 days following surgery. Furthermore, 493 (43%) and 428 (37%) patients experienced complications during the index admission and following discharge respectively[35]. In total, 153 (13%) patients experienced major complications requiring surgical intervention (±general anaesthesia) and intensive care

admission, had organ dysfunction, or suffered complications resulting in death, as defined by the Clavien-Dindo classification (described in section 1.4.2.3). Of note, this patient group was relatively older and co-morbid, with a median age of 68 (IQR: 60-75) years and an ASA grade of ≥3 in 43% of patients. These findings are largely consistent with other contemporary reports of outcomes following ORC, as well as the recently reported RAZOR RCT[36].

The interest in RARC and its growth in the last decade stem from the expectation that morbidity related to major surgery would be reduced using a minimal access approach. A systematic review by Novara *et al*. comparing ORC and RARC reported 30-day complication rates[37]. Interestingly, the complications associated with RARC (46%) were not different to the ORC (52%). Robotic surgery has gained a foothold as a standard approach for RC, though not at the pace noted in radical prostatectomies (RP). Between 2014 and 2015, a total of 3,742 RCs were performed in the UK. The majority of these were open operations, whilst only 25% were robot assisted[38]. This data contrasts starkly with that in RP, for which most are robot assisted (79.4% of the 7,673 in 2016). Given that most pelvic surgeons have access to robotic facilities, as shown by the RP trends, the relatively slower uptake of robotic surgery for RC is surprising. Moreover, the fact that RC is a more morbid operation than RP and most patients with bladder cancer are less fit than the average man with prostate cancer, mean that reductions in morbidity in this cohort will be especially rewarding.

The ongoing debate regarding the optimal approach for performing RC has merit on both sides. ORC is less expensive, whereas RARC offers reduced blood loss and a potentially quicker recovery. Four small randomised studies have shown similar complication rates and peri-operative morbidity[39–42]. One RCT has highlighted that there may be oncological differences in terms of recurrence patterns – with ORC having increased distant metastases and RARC having increased locoregional recurrence[43]. A single large phase III RCT was designed to address oncological outcomes has refuted this[44]. The RAZOR trial randomised 350 patients to

ORC or RARC[36]. The two-year overall survival was similar in ORC and extracorporeal RARC (eRARC). This data supports the previously non-randomised evidence in terms of oncological equivalence[45,46]. Additionally, the authors reported that RARC had longer operating times, lower blood loss and lower transfusion rates. No difference in complications was identified, but patients undergoing RARC had a shorter length of stay (6 days vs 7 days, $p<0.05$).

RARC performed in RAZOR involved an extracorporeal urinary diversion, which means that there was a conversion to open surgery for each case. While it would stand to reason that a completely minimally invasive approach would provide reduced morbidity and quicker recovery, particularly with bowel-related complications such as post-operative ileus, there is ongoing debate on whether this effect is notable in currently reported data. Intracorporeal urinary diversion, in which the urinary diversion is constructed within the body (i.e. without conversion to open surgery), is considered to have theoretical advantages over ORC and eRARC. For instance, there is emerging evidence that patients with poor performance status such as reduced cardiopulmonary function recover with fewer complications following iRARC [47]. Data from the International Robotic Cystectomy Consortium suggests an advantage for iRARC over extracorporeal RARC, with lower post-operative complication rates at 90 days. Taken together, the observational studies point to the benefits of iRARC, justifying the need for high level evidence to determine the benefit of iRARC[48], of which there is currently none.

Since the exenterative part of both ORC and RARC are the same, there is no expectation that RARC will improve on the oncological outcomes of ORC. As shown in the RAZOR trial[36], surrogates for quality of surgery and oncological outcomes such as surgical margins, lymph node yield and cancer specific survival should be similar to ORC. While the RAZOR trial has addressed oncological equivalence of RARC and ORC, this does not sufficiently justify the cost of the comparatively expensive robotic platform.

ORC appears to have a direct cost advantage of RARC, both in the UK[49] and the USA[50]. As no large randomised data is available, this conclusion is based on evidence from data collected in prospective observational studies. Ideally, a large RCT which accounts for confounders such as learning curve[51], case mix[52] and institutional factors such as volume-outcome relationship[53] and failure to rescue[54] is required to address this comparison. Of note, the largest portion of the cost for RARC is attributed to robotic equipment and consumables, and to a lesser extent the additional increase in operative time[50].

The goals of curative surgery are to increase quality of life and improve survival. However, radical cystectomy, general anaesthesia and the extended hospital stay cause physiological and metabolic disturbances, tissue trauma and an increased post-operative risk of infection. Psychological and health-related quality of life measures only return to baseline values after 12 months[55]. This data must be interpreted with caution, as “baseline” in this case refers to patients’ quality of life after the diagnosis of operable bladder cancer.

I will discuss some of the commonly used metrics used to quantify different aspects of patient health status in Section 1.3.

1.2.2 Cancer-related outcomes

The five and ten-year disease-free survival (DFS) following ORC for MIBC is 55.5% and 50.03% respectively[21]. DFS and OS are related to pathological stage, with higher stage cancers having lower survival rates at both 5 and 10-year time-points. Figure 1-1 (a) and (b) show DFS following RC over 10 years, as described by Ghoneim *et al*[21]. Tumour stage, histological grade and lymph node status were found to be the only independent variables which affect survival, with the majority of oncological failure (distant or local recurrence) occurring within the first 3 years (>90%). Of note, all 1,054 patients in this analysis did not receive neoadjuvant chemotherapy.

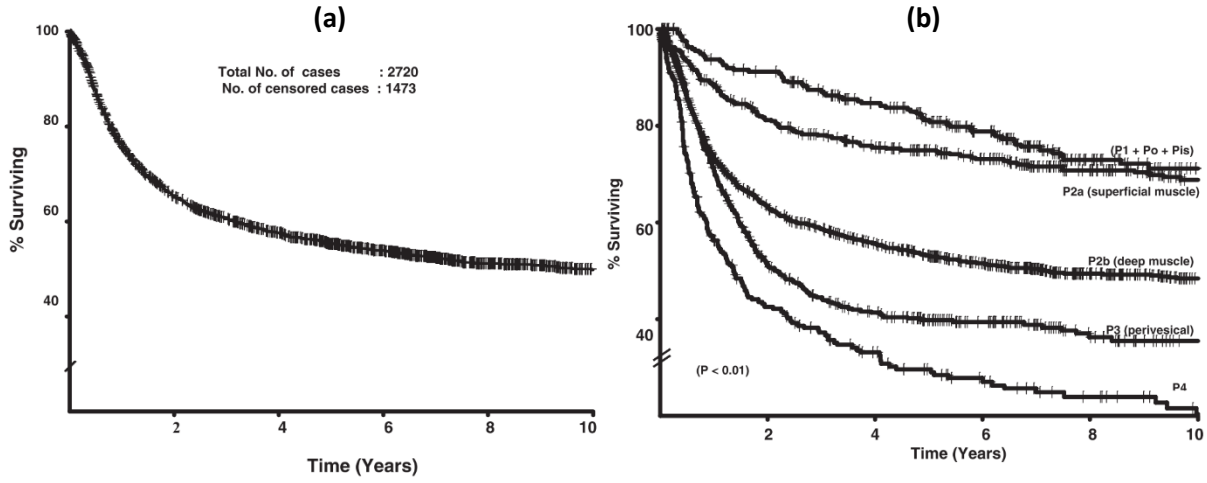


Figure 1-1: Kaplan-Meier estimates of (a) disease-free survival and (b) overall survival by tumour pathological stage.

A systematic review and meta-analysis by the Advanced Bladder Cancer Meta-Analysis Collaboration reported a 5% OS difference at 5 years. In their interpretation of the findings, the authors stated that this 5% difference provides the best estimate of effect in all stage sub-groups: T1-T2 (55% to 60%), T3 (40% to 45%) and T4 (25% to 30%). A limitation of their meta-analysis is that the individual studies used different drugs, dosages and regimens, but it serves to highlight the importance of neoadjuvant chemotherapy in the treatment of bladder cancer.

The International Robotic Cystectomy Consortium published five-year survival outcomes of 702 patients who underwent RARC in 11 institutions, making this the largest published RARC cohort[56]. The five-year DFS, cancer-specific survival (CSS) and OS are 67%, 75% and 50%

respectively. While this data suggests superior cancer control with RARC when compared with the data published by Ghoneim *et al.*[21], it is important to note that this is observational data. Hu *et al*.[57] compared outcomes between patients undergoing RARC (n = 439) and ORC (n = 7308) and reported no difference in OS and CSS. This data is consistent with the results of the RAZOR trial, which reported the non-inferiority of RARC when compared with ORC in measuring progression-free survival (PFS). The 2-year PFS was reported to be 72.3% and 71.6% for RARC and ORC respectively.

1.2.3 Institutional factors affecting outcomes

The three domains of the Donabedian model (structure, process and outcome) are of particular interest when comparing outcomes across different centres, because difference in domains in 'structure' and 'process' affect comparability. For example, there are variations in 'structure' domains such as availability of post-operative destinations (monitored bed vs ward), and 'process' domains such as the use of enhanced recovery, anaesthetic regimes and treatment protocols.

In the last decade, there has been an increase in the number of RCs performed in the UK despite no increase in the incidence in bladder cancer[29]. This increase is particularly notable in older patients, which suggests that more co-morbid patients are now being offered RC as a treatment option, as these patients would have been offered only radical radiotherapy in the past. Whilst RC is a morbid procedure typically performed in elderly patients with pre-existing cardiopulmonary diseases, recent refinements in peri-operative pathway ('process') have dramatically improved the outcomes from this operation[58]. These include the use of enhanced recovery after surgery (ERAS) programmes, centralisation of services into high volume hospitals and minimal invasive approaches (discussed in section 1.2.1).

1.2.3.1 *Role of enhanced recovery*

The ERAS programme is a peri-operative programme that aims to deliver an optimal pathway (including the pre-operative, intra-operative and post-operative periods) that is focused on optimal recovery and quicker hospital discharge for patients. It aggregates incremental gains from different optimisations made to the peri-operative period. The ERAS programme has shown to be effective in both urological pelvic cancer and colorectal surgery, resulting in reduced length of stay with lower re-admission rates, increased numbers of patients being treated and better staffing environment[59]. Dutton *et al*. reported their experience in implementing ERAS for ORC, and showed a reduction in length of stay from 14 to 9.2 days, with minor and major complication rates of 43.5% and 6.6% respectively[60]v. Julian *et al*., reported a similar experience in Southampton for ORC which noted a significant decrease in length of stay from 14 to 7 days[61]. A meta-analysis of 4048 patients in 22 studies by Williams *et al*. reported that ERAS was associated with reduced morbidity, quicker bowel recovery and shorter LOS without affecting mortality. The British Association of Urological Surgeons (BAUS) has highlighted the importance of ERAS for RC, outlining pre-operative and referral elements as well as peri-operative and post-operative recovery guidance [62], which is similar to the European Association of Urology consensus[63]. BAUS adoption of ERAS has helped standardise perioperative recovery across UK centres. It represents an important standardisation of 'process' in the UK to collect, analyse and compare outcomes.

1.2.3.2 *Effect of centralisation of surgical services*

The 2001 NHS Improving Outcomes from cancer Guidance (IOG)[64] centralised RC into cancer centres within each cancer network. This dramatically reduced the number of providers of RC and increased the volume of surgery within single teams. Recent data have shown that there has been a dramatic improvement in the perioperative outcomes from RC since the implementation of IOG. Post-operative 30-day and 90-day mortality rates have halved over the

last 10 years[65]. Furthermore, in-hospital mortality (during the same admission as cystectomy) has decreased by 60% (absolute difference -2%, $p < 0.001$).

There is increasing evidence of a volume-outcome relationship[66] – high-volume centres produce better outcomes. This phenomenon is attributed largely to surgical experience, with the hypothesis that performing a certain operation repeatedly and frequently improve surgeons' skills as they get past their 'learning curve'[67]. Leow *et al*. described outcomes for 49,792 patients undergoing RC in the USA, and found that high-volume surgeons had significantly lower complication rates than low-volume surgeons (OR 0.45, 95% CI 0.31–0.67; $p<0.001$)[53]. Furthermore, dealing with complications effectively is an important aspect of post-operative care. Termed as "failure to rescue", this metric reflects the ability of the surgical team to identify and effectively manage complications early and effectively, ensuring that deviations in the recovery process are appropriately treated. Even in surgeries where complication rates are similar across low and high-volume centres, the ability of a hospital to effectively rescue patients from complications is significantly better than lower volume centres. This effect also explains why patients are more likely to have worse outcomes if they are re-admitted to a non-index hospital following major surgery. This effect has been observed in all surgeries[54] including RC[68], and makes a strong case for centralisation of major surgery[69].

1.2.3.3 *Benefits of the robotic platform in surgical training*

Robotic surgery has gained adoption in the last fifteen years. As its adoption has grown, there has been a need for training curriculums to be developed specifically for the robotic platform. Standardised curriculum have been beneficial in delivering education[70], but have also enabled global standardisation of accreditation and certification of surgeons[71]. Robotic procedures have been dissected into smaller tasks (termed 'task deconstruction'), allowing for modular training of trainees. This allows for focused familiarisation with specific parts of the operation, enabling trainees to learn complex procedures in parts. Furthermore, the robotic platform

enhances mentoring and proctorship using dual consoles – this allows the trainee to view the operation from the perspective of the trainer, or for the trainer to supervise and take over parts of the operation as required[72]. The robotic platform also provides metrics such as economy of potion, instrument collision etc to trainees in an objective manner, which can lead to an improvement in technical performance[73]. To aggregate the benefits of these and other aspects of training, consensus statements and training courses have been designed to encourage safe adoption of the robotic platform[74]. While many of these principles are built upon the foundations laid by many decades of open surgery, they cannot be replicated in open surgery as they are enabled by new technologies built into the robotic platform.

1.3 Measuring peri-operative health status

1.3.1 Peri-operative outcomes after radical cystectomy

Radical cystectomy (RC) is a morbid procedure. Over 50% of patients experience a complication in the first 90 days, and up to 24% of patients experience a major complication[35,75]. Re-admission rates during the first 90 days are as high as 26.6%, with 19.7%, 10.8% and 3.9% having an early (first 30 days), late (31-90 days) and both (early and late) readmissions respectively[76]. Open radical cystectomy (ORC) had been the traditional approach for performing cystectomy, until interest developed in the robotic minimal-access approach in the last decade. The robotic platform purports benefits such as reduced morbidity and quicker return to normal function. Although there is evidence attesting to these benefits, the majority is from observational case series and there is limited high-level evidence from well-designed studies to support this. When comparative trials are combined for meta-analysis, the benefits of RARC are not apparent. A systematic review by Novara *et al*. in 2015[37] reported no difference in high-grade complication and mortality rates between ORC and robotic radical cystectomy (RARC), and that morbidity with RARC remained high with 30 and 90-day complication rates of 45.7% and 59% respectively. These results corroborate with those of a randomised controlled trial (RCT) reported in 2018[36] showing no difference in adverse events between RARC and ORC (67% and 69% respectively).

Despite the lack of high-quality evidence showing objective benefits of RARC over ORC, the uptake of RARC has steadily increased and may eventually become the standard surgical approach for the treatment of high risk and muscle-invasive bladder cancer.

1.3.2 Measuring the quality of surgical care

Morbidity from surgery is a major public health concern with associated health economic impact relating to delay in recovery and readmission to hospital. The number of surgical procedures undertaken globally has risen from 234.2 million cases[77] in 2004 to 312.9 million cases[78] in 2012 representing a 33.6% increase in eight years. However, the quality of surgery offered is

highly variable between centres[79,80] and key indicators of quality such as morbidity related to surgery are measured in a standardised and well-reported manner.

The Donabedian model is a tool used to assess the quality of surgery and divides the surgical pathway into three components: structure, process and outcome[81]. 'Structure' refers to the hospital environment in which surgery and peri-operative care is delivered. 'Process' accounts for the actions of the healthcare team that affect the patient pathway including pre-operative, operative, post-operative care. 'Outcome' refers to the patient's post-operative health status, including morbidity, mortality and quality of life.

All three domains are interdependent, and when optimised are instrumental in ensuring a good result from surgery. Well-trained staff and available infrastructure such as theatre and necessary equipment (structure) are essential for the surgery to be performed and create a platform where pathways (process) can be constructed to improve the quality of care. Measurement of outcome is therefore instrumental to understand whether 'structure' and 'process' are working, and whether optimisations need to be made to improve patient outcomes.

In the context of this thesis, the Donabedian model is an essential framework to consider when designing a randomised controlled trial (RCT). An RCT allows for the comparisons of 'outcomes' while standardising the 'structure' and having different 'processes' in the different study arms. Additionally, new ways of measuring outcome following surgery will be explored using wearable device data.

1.3.3 Remit and scope of doctoral work

The work undertaken in this thesis will focus on the use of wearable sensor devices in measuring health status and recovery for patients undergoing RC. Preliminary experiments and systematic review will guide the trial design of a phase III RCT to compare ORC and RARC. Subsequently, I will use the secondary outcome collected as part of the trial to explore the value of wearable device data in measuring health status and exploring correlations with other metrics as well as post-operative outcomes.

Whilst the patient group of interest are those undergoing RC, the results may be generalisable to other major index procedures. RC is similar to other major procedures in terms of morbidity. Ghaferi *et al*. reported 30-day mortality and morbidity in 84,730 patients undergoing general and vascular surgery in the US.[79] Mortality varied between 3.5% and 6.5% across centres, and rates of all complications and major complications were reported as 24.6-26.9% and 16.2-16.8% respectively. In a separate study by the same group, 30-day complication rates of 38.9-44.3% and mortality rates of 5.3-17.5% were reported for three high risk cancer operations (gastrectomy, pancreatectomy and oesophagectomy)[54]. While neither publication reported readmission rates, other contemporary reports suggest that patients undergoing major abdominal and pelvic surgery have re-admission rates of over 20% in the 90 days after surgery[82,83].

1.4 Tools for measuring health status in the peri-operative period

There are many tools that have been developed to measure health status in the peri-operative period. Some commonly used measures of baseline fitness for surgery which will be discussed in this section include performance status, cardiopulmonary exercise testing (CPET), tests of frailty, and anaemia status. On the other hand, some commonly used measures of post-operative recovery which are included in this section are length of stay, days alive and out of hospital, and complication rates.

1.4.1 Measures of pre-operative health status

1.4.1.1 *Performance status*

All patients preparing for surgery undergo a pre-surgical assessment, which includes checking performance status to estimate their ability to cope with the physiological stress of undergoing a general anaesthetic and surgery. The American Society of Anaesthesiologists (ASA) physical status classification system is the most commonly used measure of performance status, and is associated with post-operative mortality[84]. Developed in 1963, it separates patients into different risk groups based on severity of systemic diseases. While functional limitations are mentioned, the focus is on co-morbidities. The full ASA classification is summarised in Table 1-1.

ASA Classification	Definition	Examples, including, but not limited to:
I	A normal healthy patient	Healthy, non-smoking, no or minimal alcohol use
II	A patient with mild systemic disease	Mild diseases only without substantive functional limitations. Examples include (but not limited to): current smoker, social alcohol drinker, pregnancy, obesity (30 < BMI < 40), well-controlled DM/HTN, mild lung disease
III	A patient with severe systemic disease	Substantive functional limitations; One or more moderate to severe diseases. Examples include (but not limited to): poorly controlled DM or HTN, COPD, morbid obesity (BMI ≥40), active hepatitis, alcohol dependence or abuse, implanted pacemaker, moderate reduction of ejection fraction, ESRD undergoing regularly scheduled dialysis, premature infant PCA < 60 weeks, history (>3 months) of MI, CVA, TIA, or CAD/stents.

IV	A patient with severe systemic disease that is a constant threat to life	Examples include (but not limited to): recent (<3 months) MI, CVA, TIA, or CAD/stents, ongoing cardiac ischemia or severe valve dysfunction, severe reduction of ejection fraction, sepsis, DIC, ARD or ESRD not undergoing regularly scheduled dialysis
V	A moribund patient who is not expected to survive without the operation	Examples include (but not limited to): ruptured abdominal/thoracic aneurysm, massive trauma, intracranial bleed with mass effect, ischemic bowel in the face of significant cardiac pathology or multiple organ/system dysfunction
VI	A declared brain-dead patient whose organs are being removed for donor purposes	

Table 1-1: Summary of American Society of Anaesthesiologists physical status classification

The Eastern Cooperative Oncology Group (ECOG) Performance Status has been used to record performance status of cancer patients. The main reason for its conception was to standardise the recording of performance status of cancer patients in research studies. Unlike ASA grade, ECOG grading (summarised in Table 1-2) is based on ability to perform activities of daily living (ADLs) and not co-morbidities. While it has not been reported to be related to morbidity or mortality, it is often used as a measure of performance status in cancer trials.

GRADE	ECOG PERFORMANCE STATUS
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
2	Ambulatory and capable of all selfcare but unable to carry out any work activities; up and about more than 50% of waking hours
3	Capable of only limited selfcare; confined to bed or chair more than 50% of waking hours
4	Completely disabled; cannot carry on any selfcare; totally confined to bed or chair
5	Dead

Table 1-2: Summary of Eastern Cooperative Oncology Group (ECOG) Performance Status

1.4.1.2 *Cardiopulmonary exercise testing*

Approximately 30,000 cardiopulmonary exercise tests (CPET) are performed in the UK every year. 64% of urology patients and 89.5% of colorectal surgical patients undergo CPET, the most frequently tested group of patients[85]. CPET is a non-invasive method used to assess the performance of the heart and lungs at rest and during exercise. Patients are consented to walk on a treadmill or ride a cycle ergometer with the aim to maximally stress the oxygen transport

and use system. Physiological parameters are measured during peak stress. Parameters such as low anaerobic threshold (<11 mL/kg/min) and high VE/VC02 (≥33) are used as cut off points for high risk of complications following major surgery[86]. As impaired cardiopulmonary reserve is associated with post-operative morbidity and increased length of stay[86,87] CPET is therefore used routinely in some centres to risk assess patients prior to RC.

A systematic review by Moran *et al*. consolidated results from 37 manuscripts reporting the role of CPET for risk-assessment prior to intra-abdominal surgery[88]. Thresholds used for risk stratification varied between studies and type of surgery, endpoints (e.g. prediction of 90-day to 3-year survival, length of hospital admission or post-operative ITU admission) and which metrics were used for risk stratification (anaerobic threshold, peak VO_2 or VE/VCO_2). Moran and colleagues concluded that despite these variations, CPET is a useful risk stratification tool that can predict post-operative outcomes. For radical cystectomy, the anaerobic threshold (AT) and VE/VCO_2 are commonly used[86].

However, a recent publication by Lamb *et al.*[47] observed that patients with low cardiopulmonary fitness did not have significantly longer length of stay or major complications after RARC in their cohort. This study did not have a control group of patients undergoing ORC, but the authors hypothesise that the robotic approach may be of particular benefit to patients with a low cardiopulmonary reserve. Similarly, a systematic review assessed the value of CPET testing in predicting early outcomes after major cancer surgery, and reported that CPET testing had mostly poor to average discriminatory accuracy to predict post-operative morbidity in non-lung cancers. In lung cancer, VO_{2peak}≤15ml/kg/min was associated with an increased risk of respiratory complications and death[89].

1.4.1.3 *Tests of frailty*

Pre-operative tests of frailty such as the timed up and go (TUG) test and 30-second chair to stand test are used in clinical practice as quick clinical assessments of frailty prior to surgery. As

discussed in Section 1.2.3.2, the number of older patients undergoing RC have increased in the last decade, and it is important to assess their physiological reserve to undergo the stress of prolonged anaesthetic and major surgery.

In a systematic review by Lin *et al.*, 23 studies using 21 different frailty instruments were identified. The Fried Criteria was a popular instrument for measuring frailty, with seven studies using it or incorporating it into their frailty assessment. The Fried index[90] consists of five criteria – unintentional weight loss, exhaustion, low physical activity, slowness and weakness. Despite the large variations in frailty measurement instruments used, high frailty scores were found to be associated with increased mortality at 30 days, 90 days and at one-year follow-up, as well as post-operative complications and length of hospital stay.

1.4.1.4 *Anaemia status*

Pre-operative anaemia and blood transfusion has previously been associated with higher morbidity and mortality following major surgery[91–93]. In the current literature, the only publication exploring this association is by our research group. Tan *et al.* reported outcomes on 166 patients[94] undergoing iRARC as part of a single-centre study. Pre-operative anaemia was identified in 72 patients (43%), but no association between pre-operative anaemia and post-operative complications. However, post-operative blood transfusion was found to be identified to associated with all 30-day complications, 90-day complications and 90-day major complications.

Additionally, a systematic review and meta-analysis by Xia *et al*. identified 17 studies evaluating the effect of pre-operative anaemia on outcomes following RC[95]. Data of 4,525 patients from 9 studies was meta-analysed. While definitions of anaemia varied across different studies (10.5-13.5 g/dL for male, 10.5-13.4 g/dL for female), anaemia was associated with an increased all-cause and cancer-specific mortality, as well as disease recurrence.

1.4.2 Measures of post-operative recovery

1.4.2.1 *Length of stay*

Length of stay (LOS) in hospital is often used as an easily measurable indicator of efficiency. Given that hospital stays are expensive, any changes to structure, process or outcome that can reduce the LOS can have a big impact on the cost-effectiveness of a treatment or procedure. Patients undergoing RARC and ORC in UK high volume centres in 2014-2015 had a median LOS of 8 and 11 days respectively[31]. However, LOS while easily measurable is a crude marker which only includes the index admission but not any subsequent re-admissions and complications in the peri-operative period. As such, it is an insufficient indicator of quality of care.

1.4.2.2 *Days alive and out of hospital (DAOH)*

Days alive and out of hospital (DAOH) is a relatively new metric in measuring surgical recovery. Instead of measuring only the length of the index admission for surgery, it extends into the peri-operative period and includes re-admissions to hospital. The mathematical calculation for DAOH is as follows:

DAOH = study duration – index admission length of stay – readmission days

1.4.2.3 *Complication rates – the Clavien-Dindo classification*

The Clavien-Dindo (CD) classification of surgical complications was proposed by Clavien *et al*. to standardise the way complications are reported following surgery, in an effort to ease data interpretation of surgical outcome[96]. Prior to its conception and acceptance, complications were described as minor, moderate, major or severe with varying definitions for each category across different surgery types and centres. Complications are collected as per the CD classification at 30 days and 90 days post operatively, to arbitrarily divide early post-operative complications (0-30 days) from late post-operative complications (31-90 days). Additionally, the measurement of a total CD score at 90 days is indicative of all complications in the peri-operative period. The CD classification is summarised in Table 1-3.

Grades	Definition
Grade 0	No deviation from the normal postoperative course
Grade I	Any deviation from the normal postoperative course without the need for pharmacological treatment or surgical, endoscopic and radiological interventions Allowed therapeutic regimens are: drugs as antiemetics, antipyretics, analgesics, diuretics and electrolytes and physiotherapy. This grade also includes wound infections opened at the bedside.
Grade II	Requiring pharmacological treatment with drugs other than such allowed for grade I complications. Blood transfusions and total parenteral nutrition are also included.
Grade III	Requiring surgical, endoscopic or radiological intervention
- IIIa	Intervention not under general anesthesia
- IIIb	Intervention under general anesthesia
Grade IV	Life-threatening complication (including central nervous system complications) requiring IC/ICU-management
- IVa	Single organ dysfunction (including dialysis)
- IVb	Multi-organ dysfunction
Grade V	Death of a patient

Table 1-3: The Clavien-Dindo Classification of Surgical Complications

1.4.3 Peri-operative Quality of Life (QoL) assessments

Self-ratings of quality of life are one of the most commonly used measures of health status. Unlike the CD system of classifying complications, QoL assessments measure the impact of major surgery and recovery process from the perspective of the patient. Validated QoL questionnaires can be given to patients pre-operatively and post-operatively to measure any changes in QoL having undergone surgery.

Yang *et al.* performed a systematic review to determine if differences exist in health-related quality of life (HRQOL) outcomes among different types of urinary diversions after RC[97]. A total of 32 studies used various bladder cancer-specific and generic questionnaires to compare HRQOL of patients undergoing RC. The Short Form (36) Health Survey (SF-36) and European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) were the most commonly utilised generic QoL instruments, with ten studies using each of them in their design. EORTC Quality of Life Questionnaire for Muscle Invasive Bladder Cancer (QLQ-BLM30) and Functional Assessment of Cancer Therapy-Bladder (FACT-Bl) which are bladder cancer-specific questionnaires were also used by nine and two studies respectively. The commonest comparisons were made between continent diversion and ileal conduit, with twenty studies comparing the two groups. They concluded that there is no overall difference in overall QoL in patients undergoing continent diversion and ileal conduit during RC, but ileal conduit patients have a greater improvement in physical health while continent diversion patients had superior emotional function and body image. However, none of the studies reviewed offered a comparison of QoL between ORC and RARC.

A further systematic review by Lauridsen *et al*. reported HRQOL outcomes as measured in three RCTs comparing ORC and RARC[98]. Between the three studies, 4 different HRQOL instruments (FACT-Bl, FACT-G, FACTV-CI and EORTC QLQ-C30) were collected at different time schedules in all three studies. Overall, only one study identified any difference: Messer *at al*. found a 2.5 point

lower physical well-being score on the FACT-VCI questionnaire in the ORC, but this difference was not considered clinically relevant by the authors[99]. Additionally, the study by Messer *et al*. reported results from only 20 patients in each arm (40 in total), so it can be argued that their findings are based on a relatively small sample size. A statistically powered comparison of HRQOL between ORC and RARC is required to conclude if a meaningful difference exists.

1.5 Use of wearable devices in healthcare

The work undertaken by me in this thesis will be to explore novel application of new technology to measure recovery. Specifically, wearable devices have the potential to collect data about patient health non-invasively and through passive data collection. Wearable devices come with variable features in variable forms. Devices like the Apple Watch (Apple Inc., California) offer continuous heart rate tracking, step tracking, ECG and fall detection with a battery life of approximately one day. On the other hand, The Misfit Shine (Misfit Inc, California) offers step-counting and sleep duration measurement, but offers a battery life of up to six months. The features offered by these devices can allow for large amounts of data to be collected about a patient's health status. Once a baseline trend has been established, post-operative data can be compared to assess return to normal activity.

The market for wearable technology that can monitor activity and fitness is projected to grow to $48.2 Billion by 2023[100]. This is largely owed to their success as a social phenomenon, with companies such as Fitbit Inc. allowing users to track their physical activity continuously, and to compare their activity levels with their friends and family. Modern wearable devices are capable of monitoring physical activity, often include an interface to input lifestyle information such as diet. Raw data can be collected on multiple physiological outputs, from movement to heart rate, these raw data can be used to calculate step-count, distance travelled, measure sleep duration, energy expenditure etc. Modern wearable devices often pair with smartphones, which can analyse and translate this data and upload it to the cloud for individuals to view, record and monitor their data.

This surge in popularity has also translated to wearables being utilised in healthcare. For example, insurance companies provide their clients with smartwatches capable of activity tracking, and offer incentives on the basis of targets[101]. This also allows clinicians the opportunity to remotely monitor patients based on objective quantifiable data, even after

patients have left their direct care as hospital inpatients. These devices have immense potential in the peri-operative setting, where patient mobility and health trends can be monitored after discharge from hospital – particularly while patients remain high risk for complications and resulting re-admissions.

There has been growing research interest in this field, but this has mostly focused on chronic health conditions. Jakicic *et al.* conducted an RCT recruiting 471 participants to compare the results of a weight loss program with and without a fitness tracker over 2 years[102]. Interestingly, they found that patients with the fitness tracker lost less weight over 2-year period when compared with the control group. In a different RCT, Finkelstein *et al.* reported that in 800 patients, adding a fitness tracker did not lead to an increase in physical activity when compared with lifestyle advice (control) but providing monetary incentives did[103]. There is some evidence about the use of such devices in weight-loss programs[104], but they have not been used in the setting of peri-operative recovery such as prehabilitation or rehabilitation programs.

Post-operative mobility is an integral part of the enhanced recovery programme discussed in section 1.2.3.1, with benefits in minimising risks of basal atelectasis, reducing risk of hospital acquired pneumonia, reducing venous thromboembolisms, etc[105]. However, post-operative mobility is difficult to quantify, particularly after patients have been discharged from hospital. Wearable devices provide an easy avenue to collect objectively measured mobility data from patients in hospital as well as after discharge, for goal-setting as part of a rehabilitation programme or even remote monitoring to identify mobility and health trends associated with complications.

1.6 Conclusions

After a decade of using the robotic platform in RC, there is no consensus on whether it offers sufficient benefit over the traditional ORC. Furthermore, local context is important in evaluating any treatment, as the costs of pre-operative, operative and post-operative care are varied in different models of healthcare[106]. Two trials in the UK have aimed to compare ORC and RARC, but both closed before meeting their recruitment or feasibility targets. However, since those trials, adoption of RARC increased, and there is an opportunity to open a new clinical trial to explore a comparison between the two approaches.

Traditional endpoints such as length of stay, complication rates and procedure costs may not demonstrate differences in recovery between ORC and RARC. In particular, they do not capture patient experience, time taken to return to normal activity or account for hospital readmission – all of which impact the health economics of RC. Length of hospital stay is traditionally used as a metric to measure aspects of recovery and direct costs, but re-admission rates after RC are often high, making this inaccurate in assessing the actual cost of peri-operative care. A more comprehensive metric such as days alive and out of hospital (DAOH) can be used instead to include the total time patients spend in hospital[107]. DAOH accounts for re-admission rates post-surgery and can give a more realistic estimate of recovery time. It indirectly measures the number of days a patient is hospitalised for, whether that is for the primary admission recovering from surgery, or for complications requiring re-admission to hospital. By capturing the duration of all hospitalisations, DAOH provides a readily comprehensible summary of the treatment difference in two groups [107].

There are various methods to compare such shorter-term patient experience, recovery and QoL. These include operative outcomes, such as time to discharge and post-operative complications, and patient reported HRQOL such as WHODAS-2, EORTC QLQ-BLM30 and FACT-VCI. Post-operative complication rates and QoL-related outcomes can be used to measure patient's

perspectives on their own recovery, as they reflect post-operative return to normal function during the peri-operative period. These measure patients' perception of recovery, which while subjective, can be instrumental in comparing the two techniques of surgery.

With both RARC and ORC, pre-operative assessment is a key component of achieving good surgical outcomes. CPET has been used in pre-assessment clinics, but it is unclear whether patients with poor cardiopulmonary reserve and co-morbidities like hypertension are at a higher risk of post-operative complications [47,86,87]. With the recent availability of wearable devices which offer activity tracking, heart rate recording and sleep monitoring, it is becoming easier to assess a patient's pre-surgical activity levels and fitness as well as measure their return to baseline function. These devices can be used to quantify mobility across a long time-interval or within specified periods after surgery.

Taken together, these gaps in our understanding of recovery from RC represent an opportunity to explore a comparison between ORC and truly keyhole iRARC. Such a trial would need to be undertaken in large volume, tertiary centres with established enhanced recovery protocols in place, so that both treatment modalities can be compared in an optimal setting.

In this thesis, I will set out the landscape of the current use of wearable devices in patient-centred healthcare research. Additionally, I will report on the current understanding of the comparison between RARC and ORC in the context of clinical trials. With this information, I will embed a fitness-tracker based study into the protocol of a multi-centre RCT in the UK comparing the optimal intracorporeal approach for RC (iRARC) vs the gold standard ORC. Using the data collected, I will explore the use of wearable devices in measuring health status in patients during the peri-operative period.

1.7 Aims of thesis

The aim of this thesis is to explore the role of different metrics in measuring and monitoring peri-operative recovery from radical cystectomy. Specifically, I intend to:

- Undertake a systematic review and meta-analysis of RCTs comparing open and robotic cystectomy, including identifying the metrics used to measure recovery from surgery.
- Undertake a systematic review of use of fitness trackers and wearable devices in healthcare research, with the aim to apply this technology to patients undergoing radical cystectomy.
- Complete a series of experiments comparing different wearable devices in measuring activity levels.
- Conduct a prospective observational study collecting patient reported outcome measures (PROM) from patients undergoing radical cystectomy.
- Design a multi-centre RCT comparing recovery from iRARC and ORC, with a sub-study to collect activity data using wearable devices
- Compare tracker data to PROMs and clinical metrics used in the peri-operative period following radical cystectomy
- Assess the impact of recovery in mobility at the end of the peri-operative period (3 months) in 1-year outcomes

Chapter 2 Systematic Reviews

2.1 Chapter summary

In this chapter, I will discuss the two systematic reviews that were necessary to understand the current evidence for the use of wearable devices and the accepted approaches for radical cystectomy, before designing the protocol for the iROC trial.

In the first systematic review, the aim is to understand the use of wearable devices in healthcare research. It sets out to summarise of the type of fitness trackers, metrics, study designs and patient populations that have been of interest to researchers, and to summarise their findings. Prior to embarking on a fitness tracker-based sub-study in the iROC trial, it informs us of the benefits and limitations of different fitness trackers used by other healthcare researchers.

As the iROC trial aims to compare iRARC and ORC in terms of peri-operative recovery, the second systematic review and meta-analysis reports the current landscape of reported RCTs that have been performed comparing RARC and ORC. In particular, this systematic review will discuss the endpoints and results of the RCTs that have previously reported in this field. While the remit of this thesis does not include a comparison between open and robotic cystectomy, this systematic review is informative in the design of the iROC trial – which is instrumental to the study design of the fitness-tracking sub-study.

2.2 Systematic review: The role of wearable devices in healthcare

2.2.1 Introduction

The sales revenue for fitness trackers for 2016 was 16.1 billion US dollars in 2016, and is projected to grow to $73.3 billion by 2023[108]. This has largely been driven by the consumer electronic sector, but their application in healthcare and research has also increased. For example, insurance companies provide their clients with smartwatches capable of activity tracking, and offer incentives on the basis of targets[101]. By the end of 2015, only 26 studies were identifiable on PubMed with the term "fitness tracker" in the title or as a MeSH term. In the subsequent two years, an additional 143 manuscripts were published using the same search criteria. Despite this growing research interest, fitness trackers have not been adopted into clinical pathways.

Modern fitness trackers are wearable devices capable of monitoring physical activity, and data can be collected for multiple physiological outputs such as movement and heart rate. This can be used to calculate step-counts, distance travelled, sleep duration, energy expenditure etc. In addition, modern fitness trackers can interface with companion devices such as smartphones to aggregate, translate and analyse data. Mobile network technology can upload data to cloud servers for individuals to view and share. With this 'smart technology', there is an opportunity for continuous interfacing and interaction between patients and clinicians outside of the hospital and clinic environment.

There is undeniable potential for wearable devices to impact clinical pathways, but evidence of their utility in healthcare has yet to be conclusively proven, or even consolidated. In this systematic review, we present a comprehensive overview of all published reports of the use of fitness trackers in patient-centred research.

This is a modified version of a manuscript submitted for publication in the PLOS One journal.

2.2.1.1 *Aims*

In this systematic review, we present a comprehensive overview of all published reports of the use of fitness trackers in patient-centred research.

2.2.2 Methods

2.2.2.1 *Identification of relevant articles*

An initial systematic literature search was performed in April 2018, and repeated in May 2019 using the MEDLINE and Web of Science databases with the same search terms. The searches included a free-text protocol using the terms *fitness trackers* in all fields of the records for PUBMED and the Topic and Title fields of Web of Science searches. No time limits were applied to the searches. Protocols, conference proceedings, animal studies, review papers, editorials, population-based studies, case reports and book chapters and extended abstracts were included. The reference lists of systematic reviews were searched for further relevant articles. All data retrieved from selected studies were recorded in an electronic database. All articles were reviewed in accordance with the PRISMA statement. The review is registered with the PROSPERO database[109] (CRD42018098993).

2.2.2.2 *Study selection*

All studies reporting data collected from fitness trackers issued to patients (step-counts, sleep tracking, calorie-count, heart-rate, etc) were included. Studies using objective and self-reported measures were included. Exclusion criteria were (1) no empirical data collected (2) physical activity not measured or not reported (3) the study evaluated or described sensor or algorithm ('technical experiment') (4) the sensor was not mobile (5) the sensor was an implant (6) case reports.

2.2.2.3 *Data extraction and quality assessment*

All abstracts and full-text articles were reviewed independently by two authors. Any discrepancies were discussed between the two authors. Risk of bias assessment was

performed independently by two authors. Any conflicts were resolved by consulting with the senior author. The risks of bias assessment was performed for all RCTs and case-control studies using the Cochrane RoB 2[110] (Risk of Bias) and ROBINS-I[111] (Risk Of Bias In Non-randomized Studies - of Interventions) tools respectively.

2.2.2.4 *Data synthesis and analysis*

A data extraction proforma was developed to include (1) study type (RCT, case-control, cohort, cross-sectional, etc;) (2) Fitness tracker used (brand and model) (3) Patient group or inclusion criteria (4) summary of study objectives (5) Metrics collected (6) sample characteristics (number of participants, age, gender). Based on this data, a further analysis of the trackers used in this study was performed using publicly available published information regarding device specifications – company, device model, in-built display, placement (wrist, ankle, waist etc), measurements (steps, heart rate, etc), size, weight, software, battery life and local data storage duration (number of hours or days the device can store data without syncing to a companion device).

2.2.3 Results

2.2.3.1 *Summary of search results*

A total of 412 records were retrieved during the initial database search as set out in the PRISMA diagram in Figure 1. After screening titles, 61 duplicates were identified and removed. Of 351 abstracts, one non-English abstract was excluded. Full-text versions of the remaining 350 articles were screened, and 313 articles were excluded. A total of 37 articles representing 63 unique studies met the inclusion and exclusion criteria. We identified a further 23 study protocols of which 17 are randomised controlled trials (Figure 2-1) and as such are not included in the analysis.

The risks of bias in the 23 included RCTs were variable. Seven studies had had an overall low risk of bias in all five domains assessed, while an additional seven were at a high overall risk of bias. Of the 6 case-control studies, three had low risk in all seven domains assessed. Further details of the risk of bias assessment are available in Supplementary Table 10-3 to Supplementary Table 10-7.

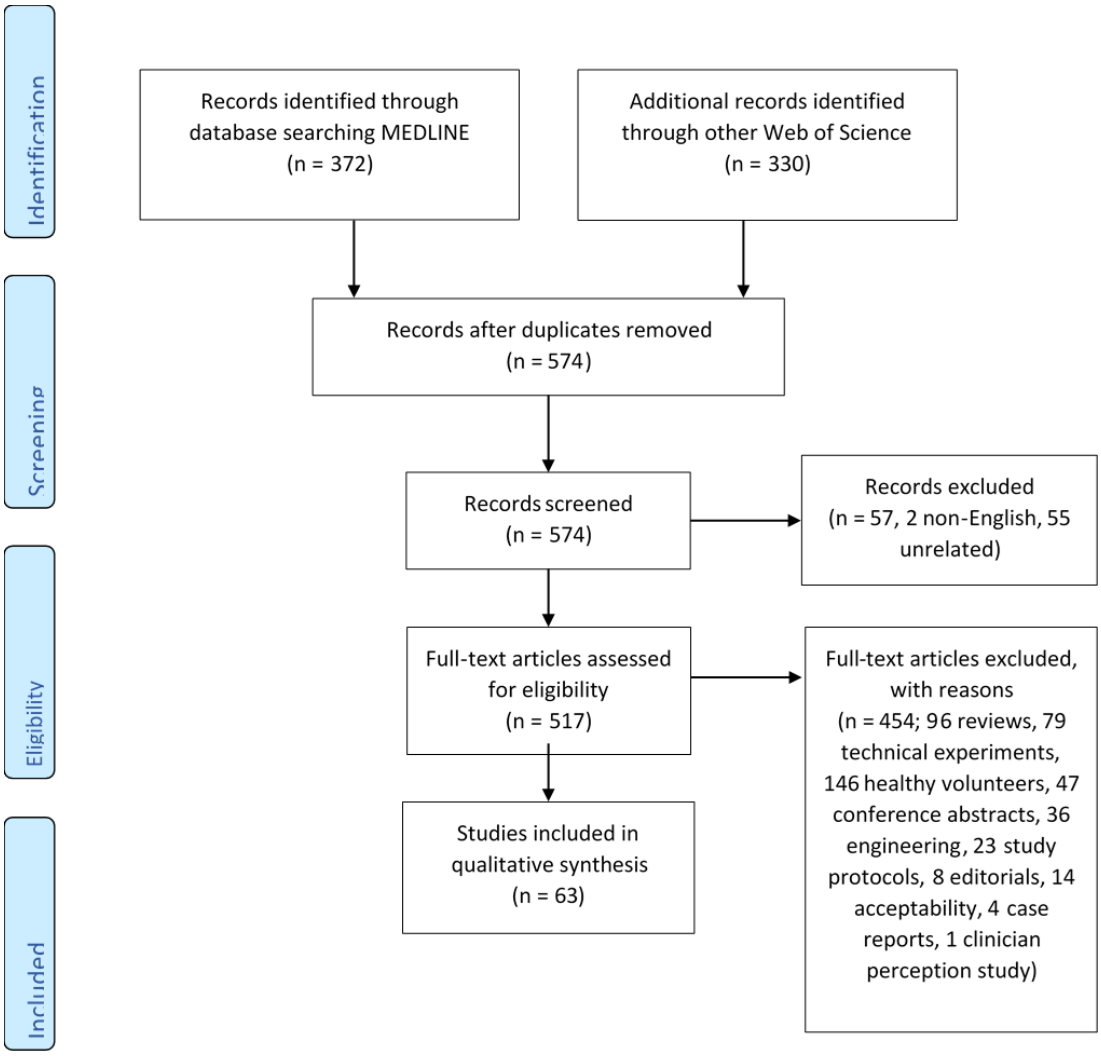


Figure 2-1: PRISMA chart outlining selection of studies using fitness trackers for health research

2.2.3.2 *Study characteristics*

Of 63 total studies, 23 randomised controlled trials, 21 prospective cohort studies, 6 prospective case-control studies, 8 cross-sectional studies, 3 qualitative studies and 2 retrospective studies were identified. Results of these studies are summarised in Table 2-1.

. Only one study published[112] by the end of 2015, and 26 studies were published in 2017, reflecting an exponential increase in research relating to fitness trackers in healthcare. The number of subjects recruited to individual studies varied largely and ranged from 7 to 2113 patients. Overall, the median study size was 45, with an interquartile range of 29-147 participants. Of note, 5(7.9%) studies were completed on a paediatric population[113–117].

No.	Author, year	Fitness tracker	Patient group/ Inclusion criteria	Metrics collected	Sample (n; age; gender)	Study objective	Summary of conclusions
Randomised controlled trials							
1	Tran, 2017[118]	Yamax SW-200	Metabolic Syndrome	Step-count	Intervention: 175; 57.6, SD 4.9; 31M, 144F Control: 162, 57.2, SD 4.9; 35M, 127F	Health promotion activities improve physical activity levels and dietary behaviours	Intervention group showed significant increases in moderate intensity activity, walking and total physical activity, and a decrease in mean sitting time
2	Takahashi, 2016[119]	Omron HJ-112	Overweight	Step-count	130; 63, SD 15; 36M, 94F	Feasibility and preliminary efficacy of activity trackers to prevent weight gain	Pedometer use and goal-setting did not improve step-count
3	Valle, 2017[120]	Withings Pulse	Breast cancer survivors	Self-monitoring of exercise behaviour	Control: 11; 52.4, SD 11.1; 11F Intervention: 13; 52.6, SD 9.4; 13F Intervention + Control: 11; 52.2, SD 6.9;13F	Effects of activity intensity awareness in reducing blood glucose	Fitness tracker group had the best weight control, followed by weighing scale alone, and then control. Both intervention groups would recommend the program
4	Miyauchi, 2016[121]	MT-KT01 Terumo	Diabetes	Step-count, time spent on physical activity	Activity monitoring: 92; 62.7, SD 9.2; 72M, 20F Pedometer: 95; 62, SD 10.6; 54M, 41F.	Feasibility of undertaking a definitive trial to investigate the effectiveness of STAK-D	The group with the activity tracker that displays information on exercise intensity showed a significantly lower HbA1C by 2 months post-intervention compared to the pedometer group
5	Gordon, 2017[122]	Fitbit Charge HR	Back Pain	Step-count, calories, sedentary minutes, lightly active minutes, fairly active minutes, very active minutes	19; 51 +/- 17; not stated 9 Fitbit, 8 Pedometer	Efficiency of activity tracker and online weight loss programme in losing weight	Significant difference in Fitbit group compared to pedometer group, but no change in body composition after 6 weeks
6	Thomas, 2017[123]	Activelink	Obese	Level of physical activity (using METS)	Control: 86; 54.9, SD 11.3; 17M, 69F WLP: 94; 55.1, SD 11.5; 17M, 69F WLP + FT: 91; 54.9, SD 11.9; 22M, 69F	Evaluate online weight loss program (WLP) +/- fitness tracker (FT)	WLP alone group outperformed WLP + FT and control groups in terms of weight loss at 3 and 12 months.
7	Li, 2017[124]	SenseWear Mini	Knee Osteoarthritis	Time spent on physical activity	Immediate: 17; 52.3, SD 9.7; 3M, 14F Delayed: 17; 58.7. SD 6.0; 3M, 14F	Effect of pedometer use and behavioural goal setting with physical therapist, either immediately or delayed	Immediate intervention group improved in moderate to vigorous physical activity time compared to the delayed intervention group
8	Jakicic, 2016[102]	BodyMedia FIT Core	Overweight	Activity intensity	Treatment: 237; Median 31.0, IQR 27.4-33.3; 69M, 168F Control: 233; Median 30.9 IQR 28.0-33.9; 67M, 166F	Effect of fitness tracker use on weight loss	Significantly less weight loss in the wearable device group
9	Williams, 2017[125]	Fitbit® Flex™	Haemodialysis	Step-count, duration of sleep	Intervention; 15; 56, SD 13; 9M, 6F Control: 14; 48, SD 15; 11M, 3F	Measure physical activity levels and sleep in feedback (intervention) and observation (control) group	Haemodialysis patients in a suburban population have lower activity levels than those in an urban population. Providing feedback did not increase activity
10	Han, 2016[126]	Fitbit® Flex™	Haemodialysis	Step-count, distance, sleep duration	Intervention: 14; 52, SD 12; 6M, 8F Control: 15; 53, SD 10; 10M, 5F	Measure physical activity levels in feedback (intervention) and observation (control) group	Providing feedback (intervention group) did not increase activity

11	Lynch, 2019[127]	Garmin Vivofit 2 (intervention arm - continuous) ActiGraph GT3X+ and ActivPAL (both arms for 1 week at start and end of study)	Post-menopausal breast cancer survivors	Step-count, MVPA, Sedentary behaviour	Intervention: 43; 61.3, SD 5.9 Control: 40, 61.9, SD 7.0	Evaluate the efficacy of a three-part intervention (goal-setting, wearable device, behavioural counselling) to increase MVPA and reduce sedentary behaviour	Significant difference favouring intervention group after 3 months: intervention arm had higher MVPA, reduced sitting time
12	Smith, 2019[128]	Fitbit Flex Fitbit One	Obese patients after total knee arthroplasty (TKA)	Resting heart rate 6MWT QoL (WOMAC)	Intervention: 24; 63.9, SD 9.7; 64.5 SD 8.2; 10M, 14F Control: 24; 64.5, SD 8.2;	To assess the impact of a 16-week home-based resistance and aerobic training on exercise tolerance and QoL	Patients in both study arms improved function and QoL. Trackers did not improve programme compliance. Patients more comfortable with technology felt it added value, and patients who did not engage with technology regularly found the additional information unhelpful
13	McNeil, 2019[129]	Polar A360 ActiGraph GT3X+	Breast cancer survivors	MVPA VO_2max	Higher intensity: 15; 58, SD 10; 15F Lower intensity: 15; 58, SD 9; 15F Control: 15; 60, SD 9; 15F	To prescribe different physical activity (PA) intensities using activity trackers to increase PA, reduce sedentary time, and improve health outcomes among breast cancer survivors.	Both intervention groups had an increase in MVPA, VO_2max at 12 weeks.
14	Falck, 2018[130]	Fitbit Flex	Knee Osteoarthritis	MVPA, METs, Sedentary behaviour	Intervention: 30; 61.73, SD 9.40; 8M, 22F Control: 31; 62.61, SD 8.54; 3M, 28F	To assess the efficacy of biweekly physical activity counselling and fitness tracker in improving cognitive function and physical activity	While this intervention increased MVPA and improved QoL among patients, no improvement in cognitive function
15	Varas, 2018[131]	OMRON Walking Style X Pocket HJ-320e	COPD	Step-count	Intervention: 21; 69.5, SD 7.4; 18M, 3F Control: 19; 64.8, SD 9.1; 13M, 6F	To assess the impact of an 8-week pulmonary rehabilitation programme on exercise capacity and steps/day at 2,3 and 12 months	Intervention arm had significantly improved exercise capacity, and QoL at all timepoints. Significant correlation between increased activity level, improved exercise capacity and QoL
16	Duscha, 2018[132]	Fitbit Charge	Cardiac rehabilitation (CR) patients	Step-count	Intervention: 16; 59.9, SD 8.1; 13M, 3F Control: 9; 66.5, SD 7.2; 6M, 3F	To evaluate if a mHealth program (telephonic coaching) can sustain increment in physical activity levels achieved during CR.	physical activity increase following CR completion is sustainable using health coaching, as measured by fitness trackers.
17	Phan, 2018[113]	Fitbit Flex	[P] Obese adolescents	Step-count MVPA	Intervention: 43; 14.7, SD 1.2; 13M, 30F Control: 45; 15, SD 1.4; 14M, 31F	To assess the impact of providing fitness trackers to caregivers in terms of satisfaction, utilization patterns and physical activity levels. All adolescents given a fitness tracker	Adolescents who used tracker had higher daily steps than those who didn't. Despite high satisfaction, tracker dropout was high (68%). Dropout was higher in patients whose carers stopped wearing tracker
18	Van der Walt, 2018[133]	Garmin Vivofit 2	Total hip arthroplasty (THA) and total knee arthroplasty (TKA)	Step-count	Intervention: 81; 67, SD 9; 45M, 36F (52THA, 29TKA) Control: 82; 66, SD 9; 36M, 46F; (43THA, 39TKA)	To determine if feedback from activity monitor improves activity levels during 6 weeks after TKA and THA.	Intervention group had significantly higher activity levels after TKA and THA over 6 weeks and 6 months.
19	Orme, 2018[134]	ActiGraph wGT3X-BT	COPD	Step-count	3 arms (education, education + feedback, control) All patients: 33; 71, SD 20; 47M, 23F	To assess the feasibility of delivering a program developed to reduce sedentary behaviour at home for COPD patients.	52% patient retention at 2 weeks, main reason for dropout was being overwhelmed following exacerbation. Feasible study, but needs modification to improve retention

20	Katz, 2018[135]	Fitbit Zip (intervention) Jawbone Up (all groups; no display)	Rheumatoid Arthritis	Step-count	Pedometer+ targets: 34; 50.2, SD 14.1; 4M, 30F Pedometer + self-monitoring: 34; 55.9, SD 12.4; 4M, 30F Education only (control): 28; 59.1, SD 12.4; 4M, 24F	To test the effect of a pedometer-based intervention on increasing physical activity and decreasing fatigue among individuals with RA.	Both pedometer groups, with and without step targets, achieved higher step-counts than control.
21	Kooiman, 2018[136]	Fitbit Zip	Type 2 diabetes mellitus	Step-count MVPA	Intervention: 40; 56.8, SD 11.4, unknown Control: 32; 55.8, SD 11.4, unknown	To determine the efficacy of an online self-tracking program on physical activity and HbA1c	Self-tracking of physical activity improved physical activity, but no significant difference in BMI or HbA1c
22	Kanai, 2018[137]	Fitbit One	Ischaemic stroke	Step-count	Intervention: 23; 66.8, SD 10; 15M, 8F Control: 25; 62.9, SD 9.1; 13M, 12F	To evaluate the effect of accelerometer-based targets on physical activity	Intervention group had significantly higher physical activity than control at study completion (hospital discharge)
23	Mitchell, 2019[138]	GENEActiv wrist-worn accelerometers	Sedentary adults	LPA, MVPA	Intervention: 85; 51.7, SD 12.8; 23M, 62F Control: 86; 49.5, SD 12.2; 14M, 72F	To evaluate the effectiveness of a 12-week online-delivered walking intervention	Increased LPA and MVPA, and decreased sedentary time in both groups during intervention period. At six months, only LPA difference favouring intervention group. By twelve months, no difference in all metrics
Prospective cohort study							
24	Rossi, 2018[139]	Fitbit Alta	Endometrial cancer	Step-count	35; 62, SD 9	To evaluate acceptability and validity of device for endometrial cancer survivors, compare data with GLTEQ questionnaire	Devices were well accepted. Self-reported physical activity not associated with recorded steps. Tracker data suggests this cohort is insufficiently active.
25	Marthick, 2018[140]	Misfit Shine	Cancer patients	Step-count	49; 54, SD 11; 11M, 38F	To evaluate the feasibility, usability, and acceptability of an interactive Web portal developed to support patients with cancer to increase daily physical activity levels.	40/49 patients completed intervention, with higher participation with more health professional contact groups.
26	Xue, 2018[141]	SIBET CAS in-house multi-sensor device	Parkinson's Disease	Movement (triaxial acceleration and angular velocity)	29; 67.5, SD 7.0; 17M, 12F	To assess the relationship between patient-reported motor symptoms and device-measured sleep quality	Number of turns in bed correlated with patient reported sleep symptoms as well as total sleep time
27	Thijs, 2019[142]	Fitbit Charge HR	Patients undergoing coronary artery bypass surgery	Step-count Physical activity levels	Robotically assisted minimally invasive coronary artery bypass (RA-MIDCAB): 10; 68 (55, 83); 9M, 1F. Conventional off-pump coronary artery bypass (OPCAB): 12; 69 (50, 82); 10M, 2F	To compare physical activity in cardiac rehabilitation using wearable fitness trackers in patients undergoing OPCAB and RA-MIDCAB	Number of steps and physical activity level measured by the Fitbit Charge HR were trending to be higher in RA-MIDCAB patients compared to OPCAB patients, but this relationship was not statistically significant.
28	Knight, 2018[143]	Patient choice of trackers (Fitbit, Garmin) and platforms (Misfit, HealthKit, Moves, MyFitnessPal and Strava)	Psychological distress	Daily activity duration	53; 20.7, SD 3.2; 12M, 41F	To assess if early identification of warning signs from digital footprints could facilitate adaptive monitoring and care for individuals with common mental disorders.	Continuous monitoring using commercial apps and wearables is feasible. Daily activity duration was greater from wearable devices compared to smartphone. Increase in entropy of daily activity related to higher anxiety symptoms
29	Heale, 2018[114]	Misfit Flash	[P] Juvenile Idiopathic Arthritis	METs MVPA	31; 15.1 (IQR 12.8-18.6); 8M, 23F	To determine the feasibility of a wearable activity tracker intervention, and estimate variability in response to	All patients synchronised data to companion smart watch and completed study measurements. 72% of activity period logged on average

						a tracker intervention on physical activity levels	
30	Champ, 2018[144]	Misfit Shine	Breast Cancer patients receiving adjuvant radiotherapy	Step-count	10; 68 (IQR 52-79); 10F	To study the change in activity levels and sleep using a wearable device in patients undergoing radiotherapy.	Patients had a statistically significant change in steps, distance and calories, but not clinically significant (54 steps, 0.02 miles and 3 calories per day)
31	Nyrop, 2018[145]	Fitbit Zip	Breast Cancer receiving adjuvant or neoadjuvant chemotherapy	Step-count	127; 48.3, SD 9.4; 127F	To assess adherence to an exercise in women who were asked to walk 150 min/week throughout chemotherapy	79% of women had analysable data, 19% were adherent with the target of 6686 steps/day, and additional 24% were moderately adherent.
32	Van Leuteren, 2018[146]	ProMove-3D accelerometer	COPD	Physical activity Step-count	35; 65 (IQR 59-70); 23M, 12F	To investigate the relationship between dynamic hyperinflation (DH) and physical activity (PA)	No significant correlation identified between parameters describing DH and PA, but significant correlation between static hyperinflation and PA.
33	Osadnik, 2018[147]	Actigraph DynaPort SenseWear	COPD	Step-count	236; 65, SD 8; 178M, 58F	To measure the effect of pulmonary rehabilitation on baseline exercise tolerance and changes in physical activity using wearable devices and 6-minute walk distance (6MWDi)	Proportion of PA responders greater in higher 6MWDi group, and 6MWDi is the strongest predictor of PA improvement
34	Le, 2017[148]	Fitbit One	Cancer survivors	levels of physical activity	19; 24.5, SD 5.8; 5M, 14F	feasibility, patient preferences and beliefs regarding physical activity	Fitness tracking is feasible, but no changes in preferences after using fitness trackers
35	Wilson, 2017[149]	Not stated	[P] Overweight	Daily calorific expenditure, physical activity	20; 16.8, SD 1.2; 8M, 12F	Feasibility and receptivity of a community-based group fitness program	Patients were receptive to fitness trackers & goal-setting, with positive effects on weight, blood pressure etc.
36	Shen, 2017[150]	ActiGraph GT3	Heart Failure (HF)	Heart rate	40; 54.4, SD 11.7; 30M, 10F	Using heart rate and physical activity recordings to assess chronotropic response during exercise stress test	Wearable tracker could help identify HF patients with impaired chronotropic response
37	Klassen, 2017[115]	Fitbit One, Step Watch Activity Monitor	Stroke	Step-count	21; 55, SD 10; not stated	Accuracy of activity monitors during inpatient stroke PT sessions	Fitbit One placed on ankle can accurately measure steps in stroke patients
38	Kroll, 2017[151]	Fitbit Charge HR	Intensive Care Unit	Heart rate, sleep	50; 64; 26M, 24F	Feasibility of activity tracker among patients recovering from critical illness	98.8% and 69.5% sensitivity and specificity of tachycardia detection, good correlation between wearable derived sleep data and questionnaire data
39	Hooke, 2016[152]	Fitbit One	[P] Acute Lymphoblastic Leukemia (ALL)	Step-count (level of physical activity)	16; 7.69, SD 3.1, 5M, 11F	Feasibility and efficacy of activity trackers in increasing physical activity and decrease fatigue	Fitbit coached children showed a non-significant increase in daily steps during ALL treatment
40	Pérez-Alenda, 2018[153]	Fitbit Charge HR	Haemophilic arthropathy	Step-count, distance per day, duration of activity	7; Median 36.0, IQR 29.5- 44; not stated	Quantify daily physical activity in patients with haemophilic arthropathy	Feasible to quantify physical activity of arthropathic patients, patients remain physical active while on treatment
41	Abrantes, 2017[154]	Fitbit Charge, Fitbit Alta	Depressed alcohol-dependent	Step-count	20; 39.5, SD 10.6; 20F	Develop a lifestyle physical activity intervention	Fitbit was worn on 73% of days, patients reported increase in using physical activity to cope with withdrawal
42	Gardner, 2017[155]	StepWatch3	Peripheral Artery Disease	Number of strides (step-count), Time spent walking	244; 65, SD 10; 49M, 195F	Amount and pace of walking is associated with circulating antioxidant capacity	Walking > 2440 strides and faster than 31.6 strides/min for 30 minutes/day associated with greater circulating antioxidant capacity
43	Sievi, 2017[156]	SenseWear Pro™	Chronic Obstructive Pulmonary Disease	Time spent on physical activity	178; Median 64, IQR 60-69; 119M, 59F	Accelerometer vs questionnaire in measuring physical activity levels	No significant relationship between patient-reported and objectively measured activity

44	Cook, 2017[157]	Fitbit® Flex™, Actiwatch2	Major Depressive Disorder	Total sleep time, sleep onset latency, wake after sleep, sleep efficiency	21;26.5, SD 4.6; 4M, 17F	Estimate sleep in patients with major depressive disorder	In the normal setting, overestimated sleep time and efficiency
Prospective case-control							
45	Jacquemin, 2018[158]	Withings Activité Pop	Rheumatoid Arthritis and Axial Spondyloarthritis	Step-count, proportion of morning steps, duration of total activity, level of physical activity	RA: 83; 49.9, SD 12.9; 14M, 69F AS: 74; 43.3, SD 10.4; 43M, 31F Controls: 19; 45, SD 11; 8M, 11F	Compare physical activity between patients with rheumatoid arthritis (RA) and axial spondyloarthritis (AS)	Activity levels similar in both patient groups, good adherence to fitness trackers by patients
46	Van't Hul, 2016[159]	DynaPort® MoveMonitor®	Bronchial asthma	step-count, energy expenditure, daily time (minutes) spent doing physical activity	Patients: 226; 27.3, SD 15.3; 86M, 140F Controls: 201; 42.3, SD 16.3; 49M, 152F	Compare physical activity between adults with bronchial asthma and apparently healthy controls	Bronchial asthma patients have a significant lower physical activity compared to healthy controls
47	Peacock, 2017[160]	SenseWear Armband model MF-SW	Vertebroplasty	Step-count, sleep efficiency, total sleep time, levels of activity	Patients: 15; 70.1, SD 11.6; 8M, 7F Controls: 4; 70.5, SD 17.8; 0M, 4F	Determine the correlation between patient-reported outcomes, quantitative activity metrics at baseline and at 30 days	No significant correlations between reported main, disability scores and activity monitor data were identified
48	Glaviano, 2017[161]	Fitbit Charge HR	Patellofemoral pain (PFP)	Step-count, Time spent (minutes) on physical activity	Patients: 20; 22.2, SD 2.6; 5M, 15F Controls: 20; 20.8, SD 1.8; 5M, 15F	Identify activity levels in patients with and without PFP	Daily activity for patients with PFP are significantly less than controls. This relationship correlates with patient-filled questionnaire scores (subjective function)
49	Colón-Semenza, 2018[162]	Fitbit Zip	Parkinson's disease	Mean Step-count, time spent on physical activity as part of training program	Patients: 5; 63.4, SD 2.1; 3M, 2F Controls: 5; 64.6, SD 4; 3M, 2F	Feasibility, safety, and acceptability	Remote peer coaching is feasible, safe and acceptable, 4/5 patients had increased daily step counts after coaching
50	Kuenze, 2019[163]	ActiGraph wGT3X-BT	ACL reconstruction (ACLR)	MVPA	Male Healthy: 22; 20.4, SD 1.7 Male ALCR: 25; 20.8, SD 2.6 Female healthy: 33; 20.6, SD 1.8 Female ALCR: 34; 20.1, SD 2.1	To investigate the effects of sex as a modifier of MVPA following ACLR	No significant difference in odds (X^2 = 2.33, OR = 2.13, CI_{95} = 0.80–5.69) of meeting national physical activity guidelines for males, but significantly worse for females (X^2 = 4.18, OR = 2.54, CI_{95} = 1.03–6.27) undergoing ACLR vs controls
Cross-sectional studies							
51	Ezeugwu, 2017[164]	activPAL3 Micro	Stroke	Sleep duration, levels of physical activity (step-count)	30; 63.8, SD 12.3; 17M, 13F	Sleep duration, sedentary behaviour, physical activity and QOL after 1 month of rehabilitation	Stroke patients sleep longer, are more sedentary & engage in minimal walking
52	Byakika-Kibwika, 2015[112]	not stated	Hospital inpatients	Not stated	57; 35.6, SD 15; 31M, 26F	Validation of consumer fitness in hospital	With effective hospital-patient partnerships, fitness trackers can be implemented for inpatients
53	Gordia, 2017[116]	Yamax Digi- Walker SW-200	[P] Obese (abdominal obesity)	Step-count	1044; 11.6, SD 3.3; 456M, 588F	Develop cut-off points for pedometer-determined step count, analyse the capacity of previous recommendations to discriminate abdominal obesity	Universal step-count recommendation for young people may not be adequate
54	Simpson, 2017[165]	Any	Eating Disorders (ED)	Health-tracking technology use (Y/N)	495; 345F, 20.3 SD 3.5 ; 148M, 21.0 SD 6.0	Associations between the use of calorie counting and fitness devices and eating disorder	fitness tracking was uniquely associated with ED symptomatology after adjusting for gender, and binging and purging behavior

55	Voss, 2017[117]	Fitbit Charge HR, Actigraph (GT3, GT9)	[P] Congenital Heart disease (CHD)	step-count, physical activity intensities	30; 13, SD 2.2, 14M, 16F	Validity of commercial trackers in children	Trackers enable remote monitoring of physical activity in CHD, but absolute values might differ from accelerometers
56	Wang, 2017[166]	Actical version B-1 (model 198-0200-03)	Diabetes	Physical activity levels using step/min (step-count)	1669; not stated; not stated	Association between activity levels and cardiovascular risk factor control	Reducing sedentary time but not moderate-to-vigorous activity is associated with improved CVD control
57	Dauriz, 2018[167]	SenseWear Armband	Type 2 Diabetes	Daily physical activity (step-count) Energy expenditure	41; 62 (52.5-67); 24M, 17F	To assess if daily physical activity (DPA) is associated with beta-cell function (BF) and/or insulin sensitivity (IS) in patients with T2DM at time of diagnosis	Moderate levels of DPA and total EE are independent predictors of IS after adjusting for age, sex and BMI
58	McKenna, 2018[168]	SenseWear Pro3 Armband	Rheumatoid arthritis	Total sleep time MVPA	75; see below; 28M, 47F Age groups - <50: 15, 50-59: 15 60-69: 30, 70-79: 15	To observe the relationship between total sleep time (TST) and physical activity (PA) in RA patients.	Higher PA positive associated with longer TST, but negatively correlated with functional limitations and CRP levels
Qualitative studies							
59	Nguyen, 2017[169]	Fitbit One, Jawbone Up 24, Garmin Vivofit 2, Garmin Vivosmart, Garmin Vivoactive, Polar A300	Breast Cancer Survivors	Qualitative study	14; not stated; not stated	Acceptability and usability of commercial activity tracker	Trackers increased self-awareness and motivation, were well accepted
60	Chum, 2017[170]	Fitbit One	Depression	Qualitative study	36; 53, SD 12.35; 18M, 18 F	Understand patients' perceived benefit from Fitbit and patient's experiences	Positive experiences: self-awareness, peer motivation & goal setting Negative themes: inconvenience, inaccuracies & disinterest
61	Randriambelonoro, 2017[171]	Fitbit One	Diabetes & Obese	Step-count	18; not stated; 7M, 11F	Patient's expectations, influence of lifestyle and long-term health decisions	Patient expectations of a fitness tracker change. If the device is able to meet expectations, it can help motivation
Retrospective studies							
62	Painter, 2016[172]	Fitbit	Overweight	Step-count	6-month group: 1387; not stated; 581M, 806F	Weight-loss outcomes	High performers were more likely to weigh in, wear the activity tracker and walk more
63	Painter, 2017[173]	Fitbit	Overweight	Step-count, Activity minute	2113; 44.5, SD 10.72; 860M, 1253F	Identify the significant contributors to weight loss	Regular weight checks, high activity minutes and regular food logs were associated with significant weight-loss

[P]Paediatric patient population
MVPA = Moderate to vigorous physical activity, LPA = Light physical activity, WOMAC = Western Ontario and McMaster Universities Arthritis Index, GLTEQ = Godin Leisure-Time Exercise Questionnaire

Table 2-1: Summary of published research studies including study type, devices used, inclusion criteria, metrics, sample characteristics, summary of objectives and conclusion

2.2.3.3 *Type and model of devices*

A total of 41 fitness trackers were used in the 63 studies and technical specifications are summarised in Table 2-2. The most frequently used provider of wearable devices for research was Fitbit Inc (San Francisco, California), with 27 (42.9%) of 63 studies using one of the following Fitbit models – One, Charge, Charge HR, Flex, Alta, Zip, and 2 unnamed models. Other device manufacturers included Actigraph, Activinsights, ActivPAL, BodyMedia, Dynaport, Garmin, Jawbone, Misfit, Omron, Ortho Innovations, Philips Respironics, Polar, ProMove, Sensewear, Weight Watchers & Philips, Withings and Yamax.

Step-count was the most commonly employed metric and used for 40 (63.5%) studies. Only three studies specifically mentioned the collection of heart rate, but studies such as Gordon *et al.* and Pérez-Alenda *et al.* reported exercise intensity calculated on the basis of heart-rate ranges[122,153]. It is important to consider that some studies collected sleep duration/quality [125,126,151,157,160,164] and calories burnt[122,149], but the devices have no reliable way to measure these data. Instead, they are surrogated estimates based on heart rate (photoplethysmogram) and movement (gyrometer and accelerometer) data.

No	Company	Device	Movement sensor Type	Steps on display?	Placement	Measurements (steps, heart rate, etc)	Size in mm (height x width x thickness)	Weight (g)	Software	Battery Life	Local data storage duration
1	Actigraph	GT3	3-axis Accelerometer	No	Hip, Wrist	Steps, heart rate, physical activity intensity, body position, energy expenditure, sleep time, active time, sedentary time, METS	46 x 33 x 15	19	Actigraph for iOS, android	Up to 25 days	180 days
2	Actigraph	wGT3X-BT	3-axis Accelerometer	No	Hip, Wrist, Ankle, Thigh	Steps, heart rate, physical activity intensity, body position, energy expenditure, sleep time, active time, sedentary time, METS	33 x 46 x 15	19	Actigraph for iOS, android	Up to 25 days	180 days
3	Actigraph	GT9	3-axis Accelerometer	Yes	Hip, Wrist	Steps, heart rate, physical activity intensity, body position, energy expenditure, sleep time, active and sedentary time, METS	35 x 35 x 10	14	ActiLife Mobile or PC connection	up to 14 days	180 days
4	Actigraph	GT3X+	3-axis Accelerometer		Hip, Wrist	Steps, heart rate, physical activity intensity, body position, energy expenditure, sleep time, active and sedentary time, METS	33 x 46 x 15	19	ActiLife Mobile or PC connection	Up to 31 days	Up to 42 days
5	Actigraph	Unspecified									
6	Activinsights	GENEActiv	3-axis Accelerometer	No	Wrist	Sleep, activity intensity, raw movement data	43 x 44 x 13	16	PC USB connector		Up to 60 days
7	ActivPAL	3 Micro	3-axis Accelerometer	No							
8	ActivPAL	Unspecified									

9	BodyMedia	FIT Core	3-axis accelerometer	No	Arm	Distance, skin temperature, heat flux, galvanic skin response, sleep data	25.4 x 76.2 x 50.8	113.4	Windows, Mac, Android, iOS	Up to 6 days	
10	Dynaport	Move Monitor	3-axis Accelerometer	No	Lower back	Steps, distance	85 x 58 x 11.5	55	USB link to PC	7 days	204 hours
11	Fitbit	Flex	3-axis Accelerometer	No	Wrist	Steps, distance, sleep, energy expenditure, hourly activity and stationary time	31.7 x 8.9 x 6.8	23.5	Fitbit for iOS, android and windows	up to 5 days	30 days
12	Fitbit	One	3-axis Accelerometer	Yes	Waist, Chest	Steps, distance, energy expenditure, active minutes, floor climbed	5.1 x 20.3 x 55.8	7.9	Fitbit for iOS, android and windows	up to 10 days	23 **days**
13	Fitbit	Charge	3-axis Accelerometer	Yes	Wrist	Track steps, distance, calories burned, floors climbed, active minutes, hourly activity & stationary time	21mm width	22.7	Fitbit for iOS, android and windows	Up to 10 days	30 days
14	Fitbit	Charge HR	3-axis Accelerometer	Yes	Wrist	Steps, distance, energy expenditure, floor climbed, active minutes, heart rate, hourly active time, stationary time	20.8 x 2.0 x 1.0	22.7	Fitbit for iOS, android and windows	up to 5 days	30 days
15	Fitbit	Atla	3-axis Accelerometer	Yes	Wrist	Steps, distance, energy expenditure, active minutes, hourly activity and stationary time	width 15mm	29	Fitbit for iOS, android and windows	up to 5 days	30 days
16	Fitbit	Unspecified									
17	Fitbit	Zip	3-axis Accelerometer	Yes	Waist, Chest	Steps, distance, sleep, energy expenditure	28 x 9.7 x 35.5	7.9	Fitbit for iOS, android and windows	4-6 months	23 days

18	Garmin	Vivofit 2	Accelerometer	Yes	Wrist	steps, calories, distance, sleep	25.5 x 10	25.5	Garmin app for iOS, Android	More than 1 year	
19	Garmin	Vivosmart	Accelerometer	Yes	Wrist	Steps, calories, distance	34.4 x 3.5	18.7	Garmin app for iOS, Android	7 days	4 weeks
20	Garmin	Vivoactive	GPS smartwatch with accelerometer	Yes	Wrist	Steps, sleep, calories, distance, time, goals	43.8 x 38.5 x 8.0	38	Garmin app for iOS, Android	10 hours to 3 weeks	14 days
21	Jawbone	Up 24	3-axis Accelerometer	No	Wrist	Steps, activity, calories, distance, sleep	139.7 x 50.8 x 50.8	20	iOS	14 days	
22	Misfit	Shine	3-axis Accelerometer	No	Wrist	Steps, distance, sleep, energy expenditure	30.5 x 30.5 x 8	8.5	Misfit for iOS (Companion)	Up to 6 months	4 weeks
23	Misfit	Flash	3-axis Accelerometer	No	Wrist	Steps, distance, sleep, energy expenditure	28 x 28 x 8	6	Misfit for iOS (Companion)	Up to 6 months	4 weeks
24	Omron	HJ 112	2-axis Accelerometer	Yes	Waist	Steps, distance, energy expenditure, active minutes, stride length	15.2 x 53.3x 35.5	82.2		Up to 6 months	7 days
25	Omron	HJ-320E	3-axis accelerometer	Yes	Hip	Step, count, distance, calories	75 x 31 x 8	19.8	None	Up to 6 months	7 days
26	Ortho Innovations	StepWatch3	Pedometer (unspecified)	No	Ankle	Steps	75 x 50 x 20	38	Windows, Mac with dock	Up to 7 years	2 months
27	Ortho Innovations	StepWatch	Pedometer (unspecified)	No	Ankle	Steps	75 x 50 x 20	38	Windows, Mac with dock	Up to 7 years	2 months
28	Philips Respironics	Actiwatch 2	Solid-state "Piezo-electric" accelerometer	No	Wrist	Sleep pattern, activity pattern	43 x 23 x 10	16	Actiwear for windows, USB compatible	Up to 30 days	30 days

29	Philips Respironics	Actical	omni-directional accelerometer	No	Waist, wrist, ankle	Steps, energy expenditure, physical activity	29 x 37 x 11	16 (w/o band), 22g with band	ActiReader for Windows, USB compatible		
30	Polar	A300	3-axis accelerometer	Yes	Wrist	steps, distance, burned calories, calories, sleep	12.7 thickness	48	Windows and Mac	4 weeks	60 days
31	Polar	A360	Unspecified accelerometer	Yes	Wrist	Steps, heart rate (active), distance, speed, calories	13.5 thickness	31.7 – 37.3	Polar Flow for iOS and Android	12 days	
32	ProMove	3D	Accelerometer, gyrometer	No	Not stated – versatile	Movement and 3D orientation			Intertia Studio (Windows, Mac, Linux)		
33	Sensewear	Pro	3-axis Accelerometer	No	Triceps	Temperature, steps			Innerview for Windows		
34	Sensewear	Armband	3-axis Accelerometer	No	Arm	Temperature, steps					
35	Sensewear	Mini	3-axis Accelerometer								
36	SIBET	CAS (in-house design)									
37	Terumo	MT-KT01	3-axis Accelerometer	Yes	Waist, Chest	steps, distance, active minutes	63.0 x 36.5 x 14.0	22			7 days
38	Weight watchers & Phillips	Activelink	3-axis Accelerometer	No	Waist, wrist, chest	Exercise intensity, exercise duration			USB link to PC	3 weeks	

39	Withings	Activité pop	3-axis Accelerometer	No	Wrist	Steps, distance, calories, sleep	36.3 x 11.5	35	Health Mate app for IOS and Android	up to 8 months	
40	Withings	Pulse	3-axis Accelerometer	Yes	Wrist	Steps, distance, elevation gain, calories, exercise intensity, heart rate, SpO2	43 x 22 x 8	8	Health Mate on iOS and Android	Up to 2 weeks	Up to 11 days
41	Yamax	SW200	Mechanical pedometer	Yes	Waist	Steps	50 x 38 x 14	21		up to 3 years	Up to 6 days

Table 2-2: Summary of fitness trackers used by healthcare studies and their technical specifications

2.2.3.4 *Summary of findings*

2.2.3.4.1 Randomised-controlled trials

Among the twenty-three randomised controlled trials, nine had a focus on patients with metabolic syndrome and associated conditions such as diabetes and obesity [102,113,118,119,121,123,128,136,138]. The overarching aims of these studies was either to test the value of adding wearable devices to lifestyle modification in improving outcomes, to use wearable devices to measure compliance with lifestyle modification regimens, or to assess change in activity-related metrics of an intervention. The results of the studies are not in consensus regarding the value of wearable devices in improving outcomes, possibly due to differing endpoint measures used in the trials: Tran *et al.*[118] reporting increase in activity intensity, Takahashi *et al.*[119] reporting step-count improvement, Miyauchi *et al.*[121] and Kooiman *et al*.[136] reporting change in HbA1C, Thomas *et al.*[123], Jakicic *et al.*[102] reporting weight loss, Smith *et al*. reporting change in 6MWT and quality of life . While Tran *et al.*[118] and Miyauchi *et al.*[121] reported the fitness tracking group had better outcomes, Takahashi *et al.*[119] and Thomas *et al.*[123] concluded that patients in the control group outperformed the intervention group with the fitness trackers. In RCTs which fitness trackers were used to measure a change instead of deliver an intervention[113,118,138], compliance was high and offered an objective comparison between the different trial arms.

Three RCTs assessed the value of adding fitness tracker with or without an additional intervention as a means of increasing physical activity – in patients with metabolic syndrome[118], in breast cancer survivors[120], patients with back pain[122] and knee osteoarthritis[124]. In all four studies, the fitness tracker group performed better than the control. However, the RCT by Gordon *et al.* [122] recruiting 19 participants with back pain reported that while there was significant difference in physical activity, no change in body composition was noted after six weeks. This finding was similar to the findings by Kooiman

et al.[136] which reported that an online self-tracking program with a fitness tracker significantly increased physical activity in 72 patients with type 2 diabetes mellitus, but this did not result in a significant change in BMI or HbA1c at 3 months. In contrast, a larger RCT by Miyauchi *et al*. reporting data from 187 patients reported that HbA1c in the tracker group was significantly lower in the fitness tracker group at 2 months. This difference in finding could be attributed to the larger sample size, or a more appropriate intervention in the treatment arm.

Two manuscripts described RCTs recruiting haemodialysis patients. Han *et al.*[126] compared haemodialysis patients who were given feedback on their fitness tracker-measured activity against patients who were asked to wear a device but not given any feedback. Their findings showed no difference, but patients in both arms reported that they considered physical activity more important after the study period. Williams *et al.* [125] also published the results of an RCT recruiting haemodialysis patients in urban and suburban neighbourhoods, and found that the suburban group had lower activity levels than the urban counterparts. Both manuscripts appear to be published by the same research group.

2.2.3.4.2 Prospective cohort studies

In total, twenty-one prospective cohort studies were identified. Of these, nine studies aimed to assess non-clinical aspects of fitness tracking: feasibility[140,148,153], patient acceptability [114,139,145,149], device accuracy[115], compliance[154], accuracy in measuring sleep[157]. All studies reported that patients were able and willing to use these devices and data collected was accurate, except Cook *et al*.[157] who reported that the Fitbit Flex overestimated sleep time and efficiency in 21 patients with major depressive disorder.

Shen *et al.*[150] reported the results of a study that recruited 40 patients with heart failure, and collected heart-rate as a metric from the ActiGraph GT3 tracker with an aim to use it to assess chronotropic response during exercise test. They concluded that heart rate

measurements could be used to identify heart failure patients with impaired chronotropic response. Kroll et al. performed another study focusing on heart rate detection, and reported that the Fitbit Charge HR had a sensitivity and specificity of 98.8% and 69.5% in detecting tachycardia respectively in 50 patients admitted to intensive care unit. Additionally, they reported that tracker-obtained sleep data corelated well with patient reported questionnaire data. In contrast, Sievi *et al.*[156] compared data collected from patient reported questionnaires and fitness trackers and concluded that there was no significant relationship between sleep and a patient reported questionnaire when quantifying physical activity.

Gardner et al.[155] recruited 244 patients with peripheral artery disease, in a study aiming to correlate amount and pace of walking with circulating antioxidant capacity based on blood-testing. Dividing the patients by daily strides into tertiles, they found that patient in the medium and high tertiles groups were associated with a higher circulating antioxidant capacity. They conclude that walking more than >2440 strides per day and walking at a cadence faster than 31.6/minutes for 30 minutes each day are both associated with greater antioxidant capacity in patients. These findings could be applied to an intervention to assess if this is a causal relationship, and can be tested in a future randomised controlled trial.

Hooke *et al.*[152] reported one of the five studies recruiting paediatric patients. Their patient population consisted of 16 patients with acute lymphoblastic anaemia. They were issued Fitbit One devices and provided coaching. After completing the coaching, children had a non-significant increase in daily steps. Heal et al. reported another study on a paediatric population, in which they recruited 31 patients with juvenile idiopathic arthritis (JIA) to assess if an adolescent population group would comply with fitness tracker use. They reported that all patients synced some data by wearing the smart watch and completed study measurements. In total, 72% of activity period was logged by the smart watch data.

2.2.3.4.3 Prospective Case-control studies

In total, six prospective case control-studies have been published. Two manuscripts reported results of activity monitoring studies recruiting patients with joint pain, with Glaviano *et al.*[161] recruiting patients with patellofemoral pain, and Jacquemin *et al.*[158] recruiting patients with rheumatoid arthritis and axial spondyloarthritis. Glaviano *et. al*[161] concluded that patients with patellofemoral pain are significantly less active than normal controls, and this relationship correlated well with their patient reported questionnaire scores. On the other hand, Jacquemin *et al.*[158] compared daily activity of rheumatoid arthritis, axial spondyloarthritis patients and healthy controls, and concluded that patients with both conditions had similar activity levels, which were reduced compared to the healthy controls. Van't Hul *et al.*[159] performed a similar experiment with bronchial asthma patients, and arrived at a similar conclusion: that patients with bronchial asthma have reduced physical activity when compared to a normal control cohort.

Two studies reported recovery after surgery, Peacock *et al.* [160](patients post-vertebroplasty) and Kuenze *et al*.[163], in patients undergoing vertebroplasty and anterior cruciate ligament reconstruction respectively. In the study by Peacock *et al.*, Patients were asked to wear an armband activity tracker, score their pain and fill out a disability questionnaire at baseline, and again at 30-days post-operatively. No correlation was identified between pain scores, disability scores and step-count. This could suggest that fitness tracking data is providing an additional dimension to surgical recovery than what is available from current patient-reported outcome measures and pain scores. In the study by Kuenze *et al.*, MVPA was the endpoint of choice. Across 47 male and 67 female patients, they found that female patients undergoing ACL reconstruction were less likely to meet their national physical activity guideline targets after surgery when compared with normal controls, but this finding was not noted in their male counterparts.

Colón-Semenza, *et al.*[162] published a study focusing on goal-setting in Parkinson's disease patients. They employed a remote peer coaching in a cohort of 5 patients, while offering standard of care to the control arm. They concluded that in this group of patients, remote peer coaching is well-accepted and feasible, and majority of patients had increased activity after peer-coaching. This effect was not observed in the control population.

2.2.3.4.4 Cross-sectional studies

Seven cross-sectional studies were conducted in patients with stroke[164], acute illness[112] (hospital inpatients), obesity[116], eating disorders[165], congenital heart disease[117], diabetes mellitus[166,167] or rheumatoid arthritis[168]. Ezeugwu *et al.*[164] monitored activity in patients following stroke and reported reduced activity and longer sleep duration by comparison to healthy controls, even after one month of rehabilitation. While it seems plausible that patients have less activity a month after having a cerebrovascular event, no comparison group was included to assess if rehabilitation had an effect on activity levels. Simpson *et al.*[165] surveyed patients with eating disorders (EDs), and reported that patients with ED are more likely to engage with a fitness-tracking device after adjusting for gender and ED-associated behaviours. Wang *et al.*[166] reported the results of the largest study in which 1669 diabetic patients provided fitness-tracking data, and showed that reducing sedentary time was associated with a reduction in HbA1c and triglyceride levels, which are associated with improving cardiovascular disease (CVD) risk. They also showed that increasing moderate-to rigorous activity did not affect CVD risk. Similarly, Dauriz *et al*.[167] reported that moderate levels of physical activity and total energy expenditure are independent predictors of insulin sensitivity in 41 diabetic patients. However, they did not study or report a relationship with CVD risk. McKenna *et al*.[168] reported sleep time and physical activity in 75 patients with rheumatoid arthritis. Their results showed that higher positive activity is associated with longer total sleep tight, but also correlates negatively with functional limitations and CRP levels. The remaining three studies did not have clinically

oriented aims[112,116,117], but tested feasibility or were conducted to develop activity thresholds which may be meaningful in future studies or tested validity of trackers in a specific patient population.

While cross-sectional studies are not often practice-changing, they can be hypothesis-generating. For instance, findings by Ezeugwu *et al.*[164] regarding stroke patients could be applied to a larger cohort of stroke patients in a randomised setting. Similarly, patients with rheumatoid arthritis can be set movement goals based on their functional limitation scores and CRP levels on the basis of the work by McKenna *et al*.[168].

2.2.3.4.5 Qualitative studies

Three qualitative studies were identified in the literature, recruiting breast cancer survivors[169], patients with depression[170] and obese diabetic patients[171]. Nguyen *et al.*[169] recruited 14 breast cancer survivors, to understand the acceptability of commercially available fitness trackers in this group of patients. The patients were issued two to three randomly assigned trackers from six available models, and the cohort reported an increase in self-awareness and motivation relating to physical activity after using fitness trackers. They concluded that patients had preferences over which trackers they preferred, and choosing an appropriate tracker is a key component in designing studies involving activity monitoring. Chum *et al.*[170] issued the Fitbit One tracker to 36 patients with depression, and reported that 23 (63.9%) patients found the device helpful in their physical activity, with increased self-awareness, peer motivation and goal-setting. However, negative themes regarding inconvenience, inaccuracies and disinterest were also noted. Of note, prior familiarity with technology and goal-setting were not associated with any perceived benefit from tracker use. Randriambelonoro *et al*.[171] conducted a qualitative study to understand patient expectations of fitness trackers and concluded that patient expectations of fitness trackers

change with tracker use, but that if these expectations are met, devices can be used to increase motivation to adhere to a more active lifestyle.

2.2.3.4.6 Retrospective studies

Both retrospective studies were performed by the Painter *et al.* [172,173] from Retrofit, Inc. and explored the use of a tracker within a weight-loss program in patients with body mass index (BMI) of over 25 kg/m^2. In the first study[172], the aim was to determine the effectiveness of a 6-month commercial weight-loss program which required participants to wear a step-tracker and weigh themselves regularly. Participation in a weight loss program itself was found to be effective, with 51.9% of participants losing 5.21% of weight. As a single arm retrospective study, it was not possible to account for confounding factors when assessing the causal relationship between fitness tracker use and weight-loss. In the second study[173], the authors attempted to identify factors contributing to weight loss and found that participants who weighed themselves regularly, had longer higher-activity time, engaged more with web-based coaching conversations were more likely to lose weight. The authors concluded that patients who are more motivated are more successful at losing weight, and using a fitness tracker may be an indicator of higher motivation.

2.2.4 Discussion

The lack of concordance between RCTs in the value of adding fitness trackers to lifestyle-modification programs is an important finding. Among the nine studies reporting a fitness-tracker paired lifestyle intervention for patients with metabolic syndrome or associated conditions [102,113,118,119,121,123,128,136,138], four reported a benefit to using fitness trackers[102,113,118,121], one found that patients performed better without a fitness tracker[123] and two found no difference[119,128]. In the two remaining studies, Kooiman et al found an improvement in physical activity in patients with type 2 diabetes mellitus but not a change in BMI or HbA1c, and Mitchell *et al*. reported that while a tracker-assisted intervention did increase physical activity, this change was not sustained 9 months after the completion of the 3-month programme. Furthermore, a large RCT by Finkelstein *et al.*[103] recruiting 800 healthy volunteers found that while tracker use did increase moderate-to-vigorous activity, this effect did not translate a change in outcomes such as blood pressure and heart rate at one year. In the same trial, giving financial incentive was found to be the most effective method in increasing physical activity, but this effect was not sustained after incentives were discontinued. These trial findings suggest that even in cohorts where trackers increased physical activity, the increase in physical activity did not have a direct impact on the change of a health-related metric such as BMI, glycated haemoglobin (HbA1c) or blood pressure.

The commonest patient group was overweight and obese patients, with ten studies recruiting these patients. The primary aims of eight studies were to test the role of fitness trackers and self-monitoring in weight-loss programs, but their objectives varied: assessing acceptability of tracker use, measuring changes in behaviour by measuring step-counts, and assessing weight loss during the study period. Of note, there were two studies recruiting obese patients as one of their inclusion criteria: adolescents[113] or patients undergoing total knee arthroplasty[128], and investigated the effect on involving patients/guardians in

using a fitness tracker and impact of a home-based training programme respectively. The next commonest group was patients with joint pain or recent orthopaedic surgery. Most studies had different aims to those monitoring overweight and obese patients, as the aims were to quantify physical activity and not to improve mobility. The study by Li *et al.* [124] was an exception, where patients with knee osteoarthritis were randomised to immediate intervention with pedometer use and physiotherapist-led goal-setting, compared with delayed intervention. Other conditions which were the focus of multiple studies were: cancer (9), respiratory conditions (6), substance abuse and mental health (5) and diabetes (4).

Four studies recruited patients who had undergone surgery[128,133,142,160,163], to monitor and track their recovery progress. Four of these studies were performed on patients undergoing Peacock *et al.*[160] recruited 15 patients who underwent vertebroplasty, and measured their activity levels at baseline and at 30 days post-operatively. Early mobilisation is an important part of the surgical recovery process, and fitness trackers provide an avenue to measure the recovery progress. During the hospital admission, patients have known to benefit from an enhanced recovery programs[174], but this goal-setting and monitoring stops when patients are discharged. Fitness trackers could provide this information, and it could serve as a valuable triage tool for patients who are not recovering as expected.

In addition to the studies reported in this systematic review, twenty-three study protocols were also identified for ongoing or unreported studies during our literature search. Seventeen (73.9%) of these protocols were for RCTs. In contrast, twenty-three (36.5%) of the sixty-three studies included in this systematic review are RCTs. While it is difficult to predict if RCT protocols are more likely to publish than non-randomised studies, this trend could be reflective of the increased uptake of fitness trackers, and therefore a shift in more robust study designs to evaluate their value in various settings.

2.2.5 Conclusions

Fitness trackers are an important technological advancement, which have given consumers the ability to directly track and monitor their own physiological output. While newer devices offer many different health-related features, they are only available at the expense of battery life. It is only a matter of time before these become an integral tool in the healthcare, particularly in the remote monitoring of patients outside the hospital setting. Despite the promising applications of fitness trackers in goal-setting, rehabilitation and remote monitoring, more evidence is required to prove their benefit in improving patient outcomes.

2.3 Systematic review and meta-analysis of Randomised Controlled Trials comparing Robotic Assisted Radical Cystectomy vs Open Radical Cystectomy

2.3.1 Introduction

Before embarking on a randomised controlled trial in the field of radical cystectomy, it is important to understand the work that has been done so far, and also to be familiar with the metrics which are used in these comparisons.

The primary objective is to review published randomised controlled trials comparing RARC and ORC that have been published in the peer-reviewed literature, and gain an understanding of the commonly used metrics used to offer a comparison. The overarching aim of my PhD is to setup and successfully recruit for a randomised clinical trial comparing a different technique of RARC (iRARC) that have been performed in previous trials. Therefore, in this section, I will discuss the findings of my systematic review. This is a modified version of a manuscript published in PLOS One journal[175]. The manuscript's 1st author is Dr Tan Wei Shen, and I (2nd author) was involved in both the data curation as well as the writing of the manuscript.

2.3.2 Methods

2.3.2.1 *Search Strategy*

A systemic search of the literature was performed in MEDLINE/PubMed, Embase, Web of Science and clinictrials.gov databases up till 10th March 2016. The following keywords and MeSH terms were used: (bladder cancer OR transitional cell carcinoma OR urothelial cell carcinoma OR urinary bladder cancer OR urinary bladder neoplasm OR urinary bladder tumor OR urinary bladder tumour OR urinary bladder carcinoma) AND (cystectomy OR cystoprostatectomy OR bladder resection) AND (robotic OR da vinci OR robotic-assisted OR robotic assisted) AND (open) AND (randomised OR randomized). Only studies published in English were included. All conference abstracts, review articles, editorials, comments, letters to the editor and duplicate records were excluded.

The inclusion criteria for eligible studies were: 1) RCTs and 2) comparisons between ORC and RARC for bladder cancer. The exclusion criteria were: 1) non-English studies and 2) conference abstracts, literature reviews, editorials, comments, and letters to the editor. Abstracts and full text articles for eligible studies were independently screened by two authors. When there was a discrepancy, the study was discussed with a third author. The PRISMA flowchart is shown in Figure 2-2. Risk of bias for each study was assessed by two authors independently using the Cochrane 'risk of bias table', which is included in Supplementary Table 10-8.

2.3.2.2 *Data extraction*

The following data were extracted from studies which met the inclusion criteria:

Patient demographics: Age, gender, body mass index (BMI), American Society of Anesthesiologists (ASA) score, type of urinary diversion, pathological T staging, previous pelvic or abdominal surgery and use of neo-adjuvant chemotherapy (NAC).

Perioperative variables: Estimated blood loss (EBL), blood transfusion requirement, operative time, length of hospital stay (LOS), quality of life (QoL) assessment and 90-day postoperative complications. Complications were classified according to the modified Memorial Sloan-Kettering Cancer Center (MSKCC) Clavien-Dindo (CD) system[35]. Minor and major complications were defined as CD I-II and CD III-IV respectively.

Oncological variables: Cystectomy histopathological tumour and nodal stage (according to 2002 TNM classification) [176], positive surgical margins (PSM), mean lymph node yield and positive lymph node status.

2.3.2.3 *Statistical analysis*

The meta-analysis was conducted using Review Manager software v.5.3 (Cochrane Collaboration, Oxford, UK). The weighted mean difference (WMD) and odds ratio (OR) were used to compare continuous and dichotomous variables respectively. For studies presenting continuous data as median and range or interquartile range (IQR), mean and standard deviation was calculated according to methodology described by Hozo et al.[177].

Study heterogeneity was assessed for each outcome using Cochrane's $\chi 2$ test, with $p<0.10$ indicating evidence of heterogeneity. Degree of heterogeneity was quantified using the I2 statistic, with I2 ≥25% indicating substantial heterogeneity. A random-effect model was used to attempt to account for significant heterogeneity. Statistical significance was set at $p<0.05$ in all tests.

2.3.3 Results

2.3.3.1 *Characterisation of studies*

One-hundred and seventy-six citations were identified from the database search (Figure 2-2). After screening of citations, 16 full text studies were reviewed and six manuscripts from five RCTs were met the inclusion criteria[39–42,99,178]. No published data was available for one RCT which closed early due to poor recruitment [178]. The remaining four RCTs contributed to 239 patients (RARC: 121, ORC: 118). Four RCTs reported perioperative complications [39–42], three studies reported QoL data [39,42,99], one study reported oncological outcomes [42] and one performed cost analysis [39]. One of the four studies had a third group treated with laparoscopic cystectomy and this group was not included in the analysis [42].

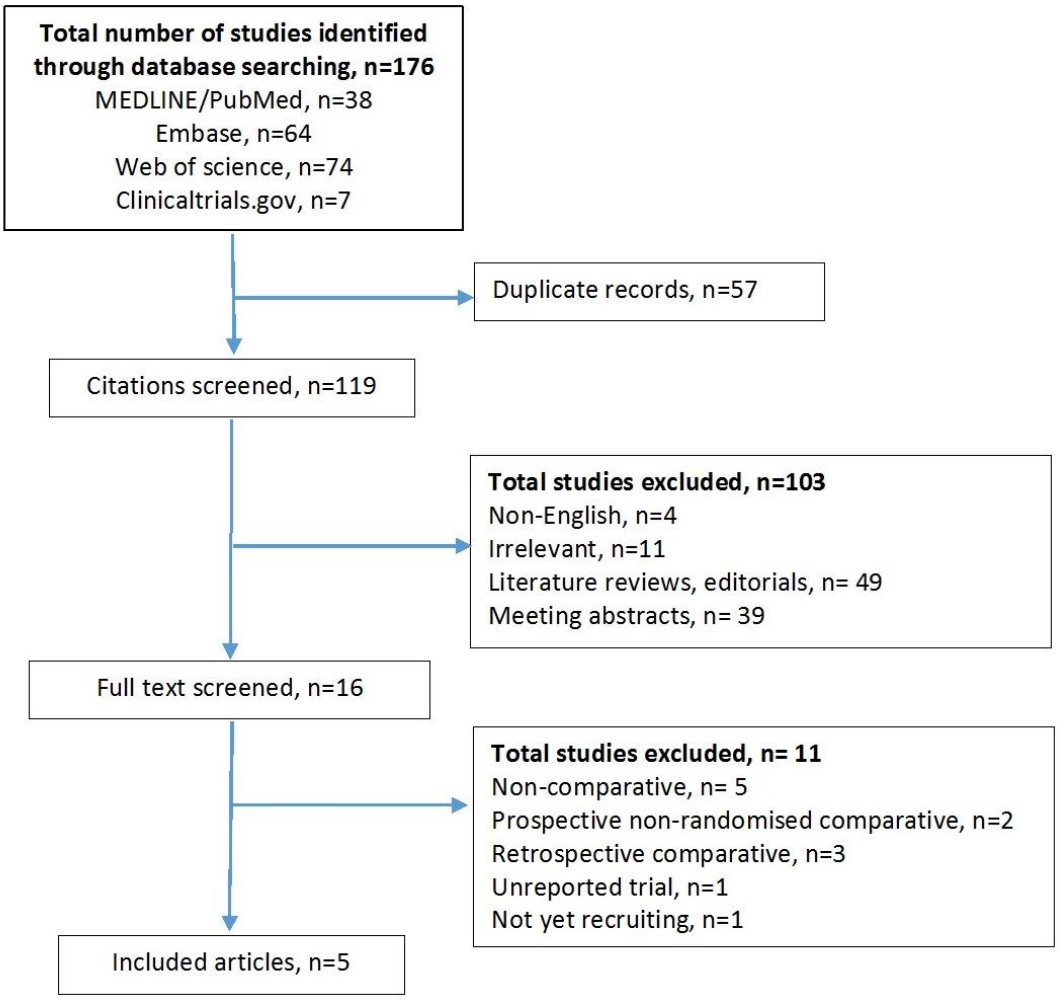


Figure 2-2: PRISMA diagram outlining selection of articles in the systematic review

2.3.3.2 *Patient demographics and clinical characteristics*

Patient demographics and clinical characteristics are shown in Table 2-3 & Table 2-4. There was no baseline difference for RARC and ORC patients in age, sex, BMI, ASA and T-stage in all four studies. Three studies excluded patients with extensive previous abdominal surgery and one study did not specify this [40]. Similarly, data from three studies reported no difference in NAC use and data was not available in one study [39].

Urinary diversion from the robotic group of all four RCTs were performed by an extracorporeal approach. More patients underwent ileal conduit urinary diversion (113 patients vs 86 patients) compared to neobladder, even though there were a similar number of neobladders were reconstructed between RARC and ORC groups (RARC: 42, ORC: 44). One study did not report type of urinary diversion constructed [41].

First author and reference	Recruitment	Country	Primary end point	Number of patients, ORC/ RARC	Male sex, ORC/ RARC	Age, median/ mean, ORC/ RARC	IC patients, ORC/ RARC	NB patients, ORC/ RARC	Match factors
Nix et al. 2010[40]	April 2008- Jan 2009	USA	Lymph node yield	20/ 21	17/ 14	69.2/ 67.4	14/ 14	6/ 7	1,2,3,4,7,8
Parekh et al. 2013[44]	July 2009- June 2011	USA	Feasibility study	20/ 20	16/ 18	64.5/ 69.5	NA	NA	1,2,3,4,5,6, 7
Bochner et al. 2015[39]	March 2010- March 2013	USA	Perioperative complication	58/ 60	42/ 51	65.0/ 66.0	23/ 27	35/ 33	1,2,3,4,5,6, 7,8
Khan et al. 2016[42]	March 2009- July 2012	UK	Perioperative outcomes	20/ 20	18/ 15	66.6/ 68.6	17/ 18	3/ 2	1,2,3,4,5,6, 7,8

Table 2-3: Characteristics of included studies

1 = age, 2 = gender, 3 = BMI, 4 = ASA, 5 = previous abdominal surgery, 6 = neoadjuvant chemotherapy, 7 = clinical stage, 8 = diversion type, ORC: open radical cystectomy, RARC: robotic assisted radical cystectomy, IC: ileal conduit, NB: neobladder

	Number of RARC/ ORC patients	WMD/ OR (95% CI)	P value	X^2	Study heterogeneity		
					df	I^2 (%)	P value
Age	121/ 118	1.14 [-0.70, 3.61]	0.19	2.82	3	0	0.42

Proportion of males	121/118	1.15 [0.61, 2.14]	0.67	6.51	3	54%	0.09
BMI	100/ 98	-0.65 [-2.01, 0.70]	0.34	0.54	2	0	0.76
ASA I-II	36/31	1.46 [0.65, 3.30]	0.36	0.08	1	0	0.78
ASA III-IV	44/47	0.68 [0.30, 1.54]	0.36	0.08	1	0	0.78
Previous NAC	100/98	1.22 [0.63, 2.34]	0.56	0.81	2	0	0.67
Pathological T stage: ≤pT2	85/85	0.75 [0.38, 1.49]	0.41	1.24	3	0	0.74
Pathological T stage: ≥pT3	36/33	1.36 [0.67, 2.75]	0.40	1.20	3	0	0.75

Table 2-4: Analysis of patient demographics and clinical variables comparing RARC vs ORC

BMI: body mass index, ASA: American Society of Anaesthetics, NAC: neoadjuvant chemotherapy, ORC: open radical cystectomy, RARC: robotic assisted radical cystectomy, WMD: weighted mean difference, OR: odds ratio, CI: confidence interval

2.3.3.3 *Perioperative outcomes*

2.3.3.3.1 Intra-operative outcomes: Estimated blood loss, blood transfusion rate, requirement and operating time.

Pooling data from 239 patients showed that EBL was significantly lower in RARC group compared to ORC (p<0.0001) (Figure 2-3). Only one RCT with 40 cases, reported blood transfusion rate and requirements and showed no significant difference in both median units of blood transfused (RARC: 0 vs ORC: 2, p=0.410) and requirements (RARC: 8/20 vs ORC: 10/20, p=0.410) [41]. Pooled data from all four studies suggested that RARC was associated with significantly longer operative times (WMD: 71.98 mins; 95% CI (15.89, 128.07); p=0.01) (Figure 2-4).

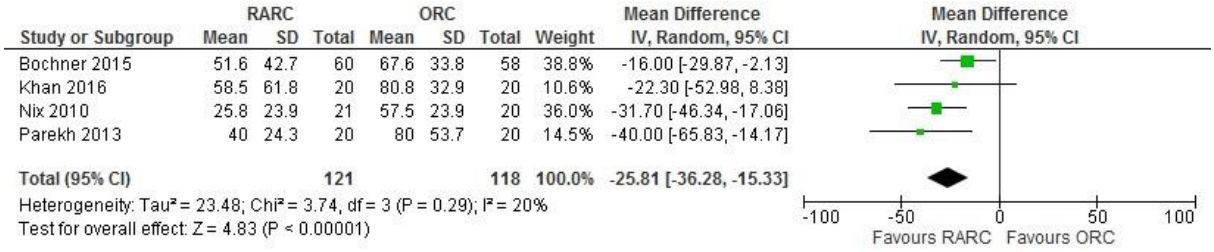


Figure 2-3: Forest plot and meta-analysis of blood loss (10ml)

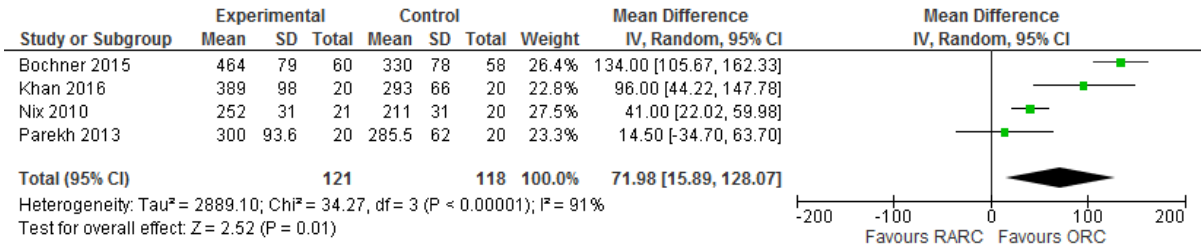


Figure 2-4: Forest plot and meta-analysis of operating time (mins)

2.3.3.3.2 Postoperative complications: Length of stay, 90-day all complications, 90-day major complications, 90-day mortality and complication type

Data extracted from all four studies did not show a significant difference between LOS when RARC was compared to ORC (WMD: -0.46 days; 95% CI (-1.34, 0.42); p=0.30) (Figure 2-5). Pooled data from 239 patients did not show a difference in all 90-day complications in the RARC and

ORC groups (OR: 0.75; 95% CI (0.44, 1.28); p=0.29) (Figure 2-6). Similarly, no significant difference was observed in 90-day major complications between both groups (OR: 1.11; 95% CI (0.56, 2.23); p=0.76) (Figure 2-7). No difference was observed in 90-day mortality between RARC and ORC (OR: 0.32; 95% CI (0.03, 3.00); p=0.32). Wound complication was the only complication which was significantly lower in RARC compared to ORC (OR: 0.23; 95% CI (0.03, 0.88); p=0.03) (Table 2-5).

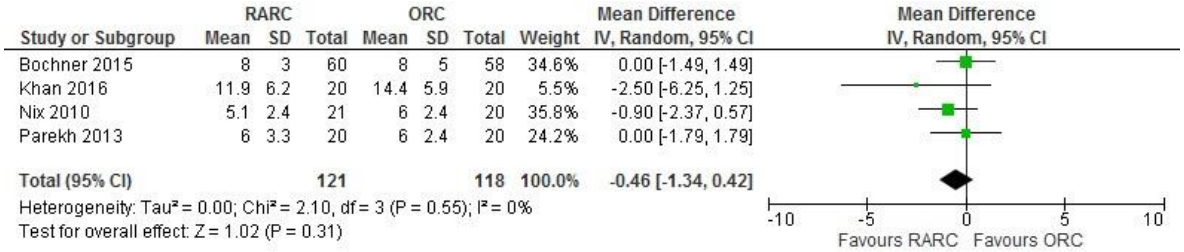


Figure 2-5: Forest plot and meta-analysis of length of stay

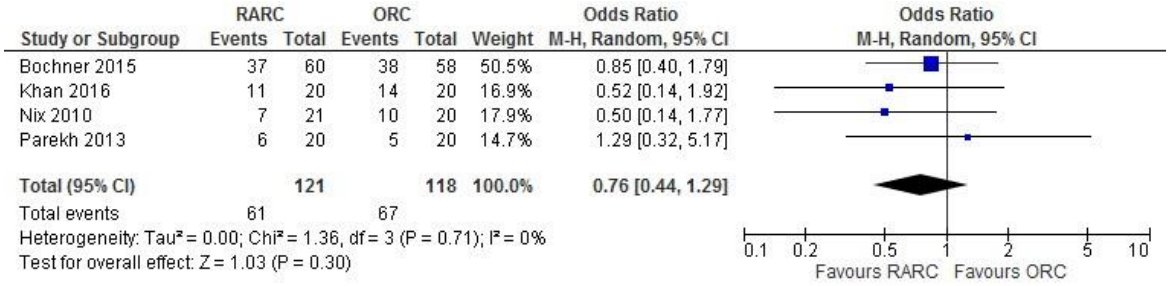


Figure 2-6: Forest plot and meta-analysis of all complications

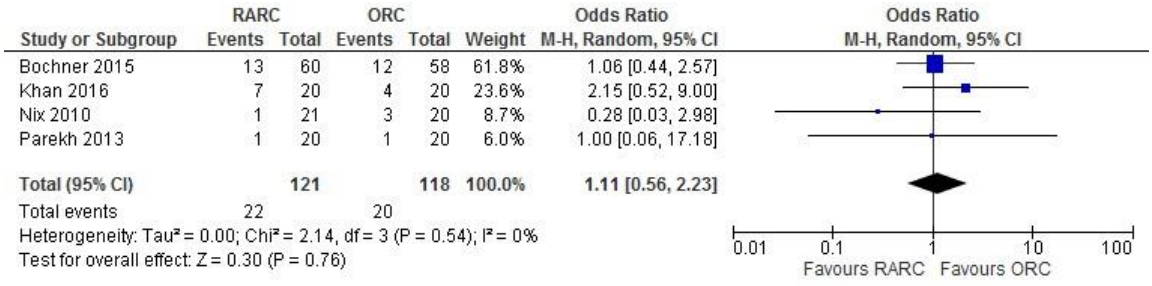


Figure 2-7: Forest plot and meta-analysis of major complications

Complications	Number of RARC/ ORC patients	WMD/ OR (95% CI)	p-value	X_2	Study heterogeneity		
					df	I^2 (%)	p-value
Bleeding	121/ 118	1.27 (0.30, 5.29)	0.75	0.41	1	0	
Cardiac	121/ 118	1.06 [0.48, 2.32]	0.88	0.99	3	0	0.80
Gastrointestinal	121/ 118	0.66 [0.40, 1.10]	0.11	1.34	3	0	0.72
Genitourinary	121/ 118	0.81 [0.27, 2.45]	0.71	4.92	3	39	0.18
Infectious	121/ 118	1.18 [0.80, 1.73]	0.40	0.80	3	0	0.85
Miscellaneous	121/ 118	0.55 [0.12, 2.52]	0.44	0.15	1	0	0.70
Neurologic	121/ 118	1.38 [0.42, 4.58]	0.60	2.30	3	0	0.51
Pulmonary	121/ 118	0.32 [0.03, 3.01]	0.32	N/A	N/A	N/A	N/A
Surgical	121/ 118	1.40 [0.23, 8.64]	0.72	1.22	2	0	0.54
Thromboembolic	121/ 118	1.24 [0.43, 3.52]	0.69	0.75	2	0	0.69
Wound	121/ 118	0.23 [0.06, 0.88]	0.03	0.02	1	0	0.89
Death	121/ 118	0.32 [0.03, 3.00]	0.32	0.00	1	0	1.00

Table 2-5: Analysis of perioperative complications according to Memorial classification

2.3.3.3.3 Histopathological variables: Positive surgical margin (PSM), lymph node count and positive lymph node status

Data from four studies that accessed PSM status showed no significant difference between the RARC and ORC groups (OR: 0.98; 95% CI (0.29, 3.23); p=0.97) (Figure 2-8). There was also no significant difference between lymph node yield (WMD: 3.89; 95% CI (-1.55, 9.33); p=0.16) (Figure 2-9) and positive lymph node status (WMD: 0.84; 95% CI (0.48, 1.47); p=0.54) (Figure 2-10) between RARC and ORC groups. Bochner et al. was the only study to divide lymph node dissection (LND) to standard and extended [39]. While only lymph node yield of standard dissection was used for meta-analysis to avoid introducing heterogeneity in the analysis, no difference in lymph node yield between RARC and ORC was observed in an extended LND (p=0.5).

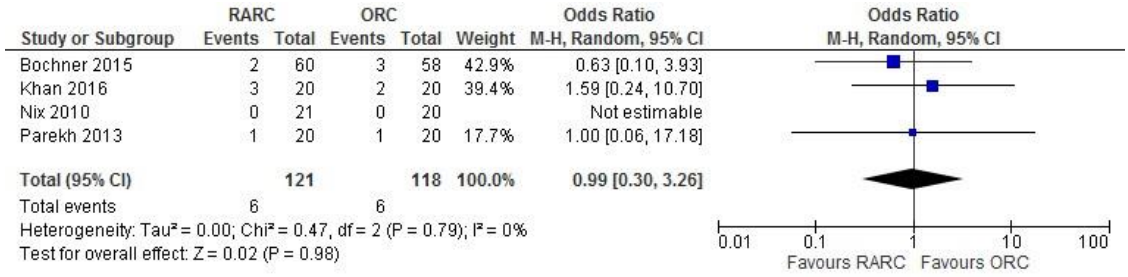


Figure 2-8: Forest plot and meta-analysis of positive surgical margin

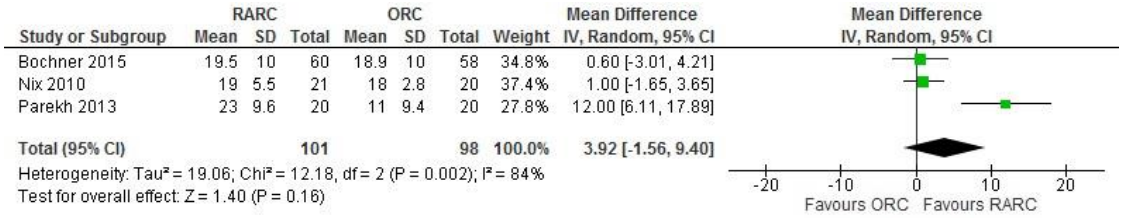


Figure 2-9: Forest plot and meta-analysis of lymph node yield

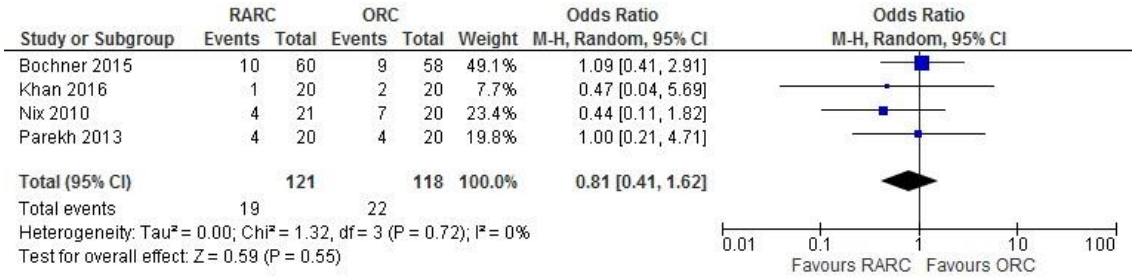


Figure 2-10: Forest plot and meta-analysis of lymph node positive status

2.3.3.3.4 Quality of life outcomes

Although three studies evaluated the QoL postoperatively, different questionnaires were used, hence pooled analysis of data was not possible [39,42,99]. Messer et al. used the Functional Assessment of Cancer Therapy–Vanderbilt Cystectomy which were completed 3-monthly for 12 months[99], Bochner et al. used the European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30 questionnaire which was completed at 3 and 6 months postoperatively [39], while Khan et al. used the Functional Assessment of Cancer Therapy-General, Functional Assessment of Cancer Therapy-Bladder and Trial Outcome Index questionnaire which was

completed at a mean of 8 months postoperatively [42]. However, all studies concluded that there was no significant difference in QoL between the RARC and ORC groups.

2.3.3.4 *Oncological outcomes*

Of the four studies, only one study reported oncological outcomes with no significant difference in recurrence free survival (RFS) (RARC: 73.6%; ORC: 89.0%; p=0.5), cancer specific survival (CCS) (RARC: 100%; ORC: 100%; p=1.0) and overall survival (OS) (RARC: 95%; ORC: 100%; p=0.1) [42].

2.3.3.5 *Cost analysis*

Only one study performed cost analysis based on Medicare reimbursement [39]. Patients who had RARC with neobladder reconstruction generated an average additional average cost of $3,920 compared to ORC patients ($p < 0.0001$) whereas patients who had an ileal conduit following RARC incurred an additional average cost of $1,740 compared to ORC ($p < 0.05$). Longer operating time attributed to 98% and 69% of additional cost in ileal conduit and neobladder patients respectively.

2.3.3.6 *Heterogeneity of studies*

Significant heterogeneity was detected between studies in lymph node yield and operating time. This is likely attributed to differences in surgical technique and experience between surgeons. Analysing pooled data using the random-effect model was performed to reduce the effect of between-study heterogeneity.

2.3.4 Discussion

This is the first systematic review and meta-analysis of randomised controlled trials comparing the outcomes of RARC and ORC. Previously, there have been four other systematic reviews on this topic, however these included both retrospective and prospective comparative studies which were at high risk of selection, reporting and publication bias[179–182]. These meta-analyses have concluded that RARC is associated with lower perioperative complications, reduced LOS, higher lymph node yield, lower transfusion requirement and equivocal PSM. The current meta-analysis comprising of pooled data with 239 patients from four RCTs does not support the conclusions from non-RCT meta-analysis[179–182]. The results of the current meta-analysis show that RARC is associated with lower EBL, lower wound complications rate and longer operating times. However, no significant difference is observed in 90-day perioperative complications, LOS, lymph node yield, PSM and QoL. A sensitivity analysis demonstrating that neither choice of statistical outcome measure nor any individual RCT impacted on the results supports the validity of the conclusions in this report.

Comparisons between morbidity rates reported for individual surgical series is often challenging due to significant variation in surgical technique, prior operative experience and documentation of complications [183]. 90-day complication rates of between 30% and 77% have been reported for RARC with extracorporeal urinary diversion [37]. To standardise reporting methodology for radical cystectomy, a modified Clavien-Dindo classification has been proposed [35]. All RCTs used either traditional Clavien-Dindo or modified classification system to standardise reporting.

In this analysis, we did not find a significant difference in 90-day perioperative complications between studies. A recent study analysed complications following RARC with intracorporeal urinary diversion in 134 cases and found that the majority of Clavien ≥III complications can be attributed to a surgical cause which may be related to surgeon experience [184]. In our meta-analysis of operating time, there was significant heterogeneity observed which may reflect a

variation in surgical experience in RARC. None of the RCTs reported prior surgical experience for either RARC or ORC, and therefore it was difficult to determine this. Although the learning curve to achieve minimal perioperative complications is yet to be defined, a minimum of 30 cases is suggested to achieve adequate lymph node yield and PSM [185] while experience of more than 100 cases has been put forward as a minimum to be considered very experienced [67]. In robotic assisted laparoscopic prostatectomy (RALP), perioperative complications continue to improve and plateau after 150 cases while improvements in urinary incontinence and sexual function outcomes were observed until after 600 cases[186,187]. Hence, these results may not be as heterogeneous if RARC was performed by experienced surgeons.

Patients undergoing radical cystectomy are often older, smoke tobacco and have co-morbidities such as cardiovascular and renal dysfunction, making them susceptible to perioperative complications. A single arm study in RARC with intracorporeal urinary diversion reported that poor cardiorespiratory fitness measured by cardiopulmonary exercise testing did not predict 30-day perioperative complications [47]. In colon cancer, a large RCT of minimally invasive versus open colectomy did not show differences in 60-day complications but did report significantly shorter LOS ($p \leq 0.001$) and lower use of opiate based analgesia ($p \leq 0.001$)[188]. Hence, it has been hypothesised that RARC will reduce perioperative morbidity or at the very least shorten LOS compared to ORC which is contrary to our findings. While there is no RCT comparing RALP with open radical prostatectomy (ORP), RALP has now succeeded ORP as the most common surgical approach for radical prostatectomy with excellent perioperative outcomes Perioperative outcomes of robot-assisted radical prostatectomy compared with open radical prostatectomy: results from the nationwide inpatient sample[189]. In comparison to previous meta-analyses, the current review did not show a reduction in LOS between RARC and ORC.

Urinary diversion reconstruction, accounts for the majority of complications following radical cystectomy [190]. All previous systematic review and meta-analyses included in this meta-

analysis performed urinary diversion reconstruction using an extracorporeal approach. The requirement for a mini laparotomy for the urinary diversion reconstruction has been postulated to negate potential perioperative benefits of a minimally invasive approach and with intracorporeal urinary diversion gaining popularity, the question remains whether the approach to diversion reconstruction will have an impact on perioperative outcomes.

All previous systematic reviews and meta-analyses including our current review consistently report that RARC is associated with a significantly lower EBL translating to a lower blood transfusion rate. This could be attributed to a more precise and controlled dissection using the robotic platform as well as pneumoperitoneum. No RCT has been designed to measure the effects of perioperative transfusion on either functional recovery or oncological outcome in cystectomy. Evidence that blood transfusion is associated with increased 30-day morbidity and mortality stems from the analysis of 10,100 patients who had non-cardiac surgery[93]. In radical cystectomy, a study of 1,490 consecutive cases showed that perioperative blood transfusion was associated with increased cancer specific mortality and overall mortality [191]. These small but highly significant effects may require a large sample size to uncover which would be very difficult to prove in an RCT and to alter practice would be based on inference.

PSM and lymph node yield are indicators of surgical quality. The presence of soft tissue PSM in particular reduces 5-year cancer specific survival to 32% (95% CI: 19-54) from 72% (95% CI: 69-75) [192]. In an analysis of 4,410 ORC patients with the overall incidence of a PSM was 6.3%, PSM was associated with higher pathological T stage; PSM for pT1, pT2, pT3 and pT4 was 1.8%, 2.3%, 7.6% and 24.0% respectively [193]. This meta-analysis shows no significant difference in PSM between RARC and ORC however only 18.0% of patients in the meta-analysis were ≥pT3 disease. In a series of 184 ORC and RARC cases, no difference in PSM have been reported between RARC and ORC [46].

Retrospective studies have shown that a higher lymph node yield of at least 8 is associated with cancer specific survival even in node negative disease [194]. Comparing lymph node yield is confounded by factors such as the use of NAC, pathological stage of disease, surgeon and method of pathological evaluation. None of the RCTs included an adjustment for case mix and the meta-analysis did not show a difference in lymph node yield between RARC and ORC. The Southwest Oncology Group (SWOG) S1011 (NCT01224665) trial is still ongoing and will address the issue if extended LND is necessary. Three of the four RCTs performed a standard template while Bochner et al. used both standard and extended LND with comparable lymph node yield suggesting that the quality of LND in RARC is equivocal to ORC [39].

It was not possible to pool QoL data for this analysis as QoL was assessed by different tools and at different time points. Among the three RCTs to date, there has been no difference in QoL reported for RARC compared to ORC. In the colorectal literature, patients treated with laparoscopic surgery showed better QoL in the early postoperative phase but this was no longer evident in longer term follow up [195]. However, a recent RCT comparing open retropubic prostatectomy with robotic assisted radical prostatectomy failed to show any significant difference between early functional outcomes as well as quality of life measured at 12 weeks postoperatively [196]. All three RCTs assessed QoL between 3- 8 months post-surgery. It is possible that any potential gain from a minimally invasive approach may have been undetected. A further limitation will be the sample size for individual studies. A health economic analysis has not been conducted by any of the RCTs however, one study did perform a cost analysis and attributed higher cost for RARC to longer operating time[39].

Limitations of this systematic review with meta-analysis include the small sample size for pooled data. In addition, each of the RCTs were conducted at a single institution. This is evident in operating time heterogeneity and might reflect individual surgeon experience rather than surgical technique. To date all RCTs have either been feasibility studies, have closed before

planned recruitment or were designed to measure surrogate endpoints. The pooled data set comprised 239 cases in total, and the systematic review with meta-analysis was not conducted on individual patient data and a test for heterogeneity has highlighted that surgical experience may have influenced the results. A further consideration is the conversion from a truly minimally invasive approach to open surgery for urinary diversion reconstruction which could confound the benefits of minimally invasive surgery.

2.3.5 Conclusions

This study is the first systematic review with meta-analysis to include data from only RCTs of ORC versus RARC. Unlike previous systematic reviews with meta-analyses, which have included observational data, our results do not show a benefit for RARC compared to ORC. There are significant issues with the trials which have been conducted in RARC which may influence the outcome and integrity of the meta-analysis at this time. RARC with intracorporeal urinary diversion remains an evolving technique and high quality RCTs will be required to determine benefit. In addition, RCTs should be performed by equally experienced ORC and RARC surgeons. For the present, the role of RARC and whether the technique can challenge ORC as the standard of care remains unanswered.

2.3.6 New evidence – the RAZOR trial

This section is a modified version of a peer-reviewed editorial I wrote and published on behalf of the investigators of the iROC trial after the RAZOR trial published their findings in 2018[34].

RAZOR is a multi-centre non-inferiority randomised controlled trial (RCT) comparing RARC to ORC[36]. In the trial 350 patients across 15 medical centres in the USA were recruited, including 150 and 152 patients who underwent RARC and ORC respectively. The authors reported a 2-year progression free survival of 72·3% (95% CI 64·3 to 78·8) and 71·6% (95% CI 63·6 to 78·2) in the RARC and ORC groups respectively (difference 0·7%, 95% CI –9·6% to 10·9%; p=0·001). This is an important statement as oncological equivalence is necessary to justify using the robotic platform. However, oncologic equivalence may be insufficient in causing large scale adoption of the robotic platform for RC.

In line with contemporary reports [175], RAZOR reported that RARC had significantly reduced blood loss and transfusion rates, but longer operating times. RAZOR did not provide a cost analysis for RARC vs ORC due to varying costs of RC across centres, but they reported a small albeit significant difference in length of stay between RARC and ORC (6 and 7 days, p=0.0216). There was no difference in complication rates. A question arising from these results is whether RAZOR reflects the full potential benefit of RARC. It is noted that the urinary diversion for all cases in robotic arm was performed extracorporeally (eRARC), which means that there was a conversion to open surgery for each case. Could conversion to open surgery negate many of the potential benefits of minimal access surgery? In contrast, intracorporeal RARC (iRARC) whereby both the extirpative and the diversion are performed robotically, is completely minimal access. As acknowledged by authors of the RAZOR trial[36], iRARC may improve peri-operative recovery when compared to eRARC.

Furthermore, there are no accepted definitions of surgical experience for RC. RAZOR required surgeons to have performed 10 RCs in the year prior to trial recruitment. In contrast, less than

5% of RCs in the UK are performed by surgeons undertaking <8 RCs/year, and the majority of RCs (56.5%) are performed by high-volume surgeons undertaking ≥30 RCs/year. While RAZOR's requirement of 10 RCs in the year prior is a minimum requirement and it is likely that most surgeons did significantly more than that, setting the bar this low – regardless of technique and outcomes – could potentially have allowed novice surgeons to operate on trial patients. Median RARC operating times in RAZOR were >7 hours, and the Pasadena consensus recommends that experienced surgeons should aim to complete the procedure between 5-6 hours[67]. According to the consensus definition, surgeons are on their learning curve for the first 30 cases – and it is possible that a subset of robotic surgeons in RAZOR were still on their learning curve. Indeed, the median time for eRARC is similar to retrospective data reported by Hussein et al.[197], which collected data from all surgeons without any criteria on experience, albeit with the possibility of selection bias, and reported a median operating time of 400 minutes across 1, 031 cases. However, it must be noted that there is also no standard definition for 'operating time'. In the RAZOR trial, "room in to room out" time was used, whereas the Pasadena consensus statement did not specify if they used the same definition, or console time, anaesthetic time, surgical time (difference between closure and knife-to-skin time).

2.3.6.1 *Updated meta-analysis*

After the publication of the RAZOR trial, an updated meta-analysis was performed by Satkunasivam *et al*.[198] to compare oncologic, peri-operative and complication-related outcomes in the comparison of RARC and ORC. RARC was associated with a significantly reduced blood loss, but a longer operative time compared to ORC. No difference in hospital length of stay, complication rates or major complication rates. No difference was identified between RARC and ORC for surgical margins rate, lymph node dissection yield, and both recurrence-free and progression-free survival. Patterns of recurrence for pelvic versus distant/abdominal sites was found to be significantly different for RARC and ORC. However, the authors acknowledge

that definitions of recurrence sites varied between RCTs, so this analysis should be interpreted with caution.

These findings by Satkunasivam *et al*. are consistent with our findings discussed in section 2.3.3.

2.3.7 Context

Even with the results of the RAZOR trial, which alone recruited more patients than the prior four RCTs combined, there is still no discernible benefit for RARC over ORC. Oncological equivalence is a reassuring finding, but does not provide a rationale for the comparatively expensive RARC if the only benefit is lower blood loss. These findings reaffirm the NHS England's clinical commissioning policy on robotic cystectomy, that NHS England will not routinely commission robotic assisted surgery for bladder cancer. It is likely that any benefit of RARC will be in the peri-operative recovery, and this needs to be compared in an RCT, with high volume surgeons in both arms, enhanced recovery and intracorporeal diversion.

To this end, a phase III multicentre randomised controlled trial to compare the efficacy of Robotically Assisted Radical Cystectomy (RARC) and intracorporeal urinary diversion with Open Radical Cystectomy (ORC) in patients with bladder cancer (the iROC trial) is currently recruiting in high volume centres across the UK[199]. Primary outcome of the iROC trial is to assess difference in days alive and out of hospital for patients undergoing RARC and ORC. The iROC will further help address key questions on the role of both RARC and intracorporeal urinary robotic diversion in current clinical practice from the perspective of a health economic analysis.

While RAZOR represents an important milestone for robotic surgeons, more evidence is required to understand if RARC should be adopted as the new definitive standard treatment for bladder cancer. Oncological equivalence is an important aspect to justify a rationale for RARC, but is not enough to sway policy decisions in favour of the relatively expensive procedure. We hope that results of the iROC trial will help inform the urological community of any difference in peri-operative outcomes between truly minimal access RARC and ORC.

2.4 Chapter conclusions

Section 2.2 outlines all reported clinical (i.e. patient recruiting) studies using fitness trackers and wearable devices. While studies have reported use of fitness trackers and they have been largely well accepted by patients, this technology has not been applied robustly to the peri-operative setting. In this setting, they offer a new way to monitor patients after hospital discharge. This could provide new insights into the peri-operative recovery of the individual patient, but also offer a further means to compare recovery after iRARC and ORC.

Section 2.3 has set out the research landscape meta-analysing currently completed RCTs comparing RARC and ORC, all of which focus on the extracorporeal approach of performing RARC (eRARC). As this technique is not minimally invasive, and can be more accurately described as a keyhole robotic cystectomy with an open urinary diversion, there in an opportunity to explore the comparison in a new trial – one that compares completely keyhole iRARC to ORC. Alongside this clinically important comparison, this provides an opportunity to explore the use of new technologies in monitoring health status and recovery in the peri-operative period of patients undergoing RC.

There are two research opportunities to explore in the peri-operative setting surrounding radical cystectomy – 1) the comparison between iRARC and ORC, and 2) the use of wearable devices to measure health status for patients before and after RC. While the former is likely a larger undertaking than the remit of this doctoral thesis, setting up an RCT comparing iRARC and ORC can provide a good pathway to explore the role of wearable devices and other metrics in measuring recovering recovery from RC.

Chapter 3 Testing research instruments for use in the peri-operative setting

3.1 Chapter summary

In this chapter, I will present the preliminary experiments required to test wearable devices and HRQOL questionnaires as research instruments before embarking on an RCT comparing ORC and iRARC. HRQOL questionnaires are considered a critical outcome measure of surgery, as they provide information about patients' post-operative quality of life[97]. This is important as it allows clinicians to measure the impact of surgery on patient lives in a holistic manner, as well as to guide patients pre-operatively about realistic expectations of life following surgery.

In the first section I will describe quantitative and qualitative comparisons between wearable devices. The quantitative component provides comparisons of device counted steps and manually-counted steps in healthy volunteers. The qualitative component assesses the device specifications and suitability for use in a prospective RCT.

In the second section, I will present the results of a small prospective pilot study collecting HRQOL data from a cohort of patients undergoing RARC as part of a service development within a cystectomy pathway. The EORTC QLQ-C30 and WHODAS 2.0 questionnaires were given to patients at baseline, 3 months and 6 months timepoint. Demographic and output data is collected alongside PROM questionnaire data and associations are explored.

3.2 Comparison of wrist-worn wearable devices

3.2.1 Introduction

Before deciding on which wearable devices to use in the trial setting, it is important to remember that the target patient population is older, and the technology available for use needs to be as simple for use as possible. The ideal tracker would be worn by the patient at home or with the assistance of a clinician in hospital, and not removed for charging, showering, etc. during the monitoring period. At the end of the tracking period, the tracker needs to be easy to remove, and easily returnable to the central receiving lab. The tracker should also have the ability to store this data locally, as opposed to a companion device such as a smartphone. Lastly, it is important to identify that the chosen device offers accurate counting on metrics collected, such as step-count. In this section I will describe the methodology for device selection and accuracy testing.

The results of this study are divided broadly into qualitative and quantitative comparisons between the wearable devices available through the UCL catalogue of devices. The hypothesis being tested in this section is that wearable devices can accurately measure step-counts under laboratory conditions. Some of these devices are included in the systematic review in section 2.2, and other readily available reference devices were included for comparison.

3.2.2 Methods

3.2.2.1 *Criteria for selecting devices*

Five main criteria were applied to select devices for testing:

1) Wrist-wearable
2) Waterproof
3) Daily step-count logging
4) Battery life of > 14 days
5) Local storage of > 14 days

Of devices available in the UCL purchasing catalogue, four met the criteria outlined above: 1) Misfit Shine, 2) Misfit Ray, 3) Jawbone UP Move, 4) Garmin Vivofit. Additional to these devices, I included three additional that were not available through the UCL purchasing, 5) Apple Watch Series 6) Fitbit Charge HR and 7)Fitbit Zip. The Apple Watch and Fitbit Charge HR were included despite not matching the criteria due to their commercial success, and to assess the accuracy of the shortlisted devices in comparison with the two most popular wearable devices in the market. The Apple Watch is the best-selling watch in the market with numerous activity and health tracking features, the Fitbit Charge HR is the flagship product of the largest fitness tracker company (Fitbit, Inc), and the Fitbit Zip is the only non-wrist worn tracker we tested (worn on a belt clip).

3.2.2.1.1 Sample

Eleven healthy volunteers (two males, nine females) volunteered to participate in the study. All 11 participants consented to the treadmill testing, and seven consented to the stairs testing. The participants were all undergraduate medical students at University College London (UCL), and they were recruited through a prospectively planned study. Participants were provided an information sheet, and informed consent was obtained. The study was formally approved by the

UCL Research Ethics Committee (ID: 12715/001). Participants had no known mobility limitations that could affect their ability to complete a 30-minute exercise.

3.2.2.1.2 Participant information

Prior to testing, height and weight was measured using a stadiometer and weighing scale respectively. Height and weight were recorded with lightweight clothing and no shoes. Additionally, their date of birth was recorded.

3.2.2.1.3 Device setup

For all the activities, participants wore all eight devices simultaneously. Six of the devices were worn on the left wrist, and the Fitbit Zip was worn at the participant's hip. On the wrist, all participants wore the devices in a fixed order, from closest to the hand to furthest as shown in Figure 3-1: Apple Watch, Misfit Shine 2, Fitbit Charge HR, Garmin Vivofit, Jawbone UP Move, Misfit Ray. Where possible, all devices were attached as close to the wrist as possible. Additionally, the Misfit Zip was attached to the collar of each participant. All devices were worn according to the manufacturer's recommendations, and no data about participants' age, height, weight or gender were entered into the companion mobile applications. Firmware for all device were updated to the latest available versions, and for devices that required a companion device, an iPhone with the latest software was used with the latest versions of the companion mobile applications downloaded from the iOS App Store. Firmware and software versions are listed in Table 3-1.

The true step-count was counted using a hand-held manual clicker held by a study investigator, and step-count for all devices were tallied after each activity by an investigator using either the on-device display, or the companion iPhone. To ensure time was provided for devices to sync, participants were required to stand still after each activity for a minute before recording the step-count. Only step-count was recorded from each exercise, as distance was pre-determined for the treadmill experiments and number of stairs was fixed for the stairs experiments.

Figure 3-1: Device placement for the seven wrist-worn devices

3.2.2.1.4 Treadmill tests

The aim of the treadmill tests was to compare the step-count on the 7 activity-trackers using a controlled, flat (incline was set to 0°) environment.

When mounting the treadmill, participants were told to stand with each leg to the side of the treadmill and hold the handles on the sides of the treadmill. At this point, a baseline step-count recording of each tracker was made. The treadmill was then sped up to the desired speed, and participants to start walking or jogging when the speed stabilised. Once the target distance was reached, participants were asked to stand on the sides of the treadmill, while holding the bars to keep the trackers from moving. After 1 minute of standing still, the step-count was recorded from the device displays, and companion applications for devices without a display.

Participants were then instructed to walk on the treadmill for 150 metres, at two different speeds, 3 km/h and 7 km/h. They were instructed to walk on the 3 km/h portion of the test, and to jog for the 7 km/h portion. They were instructed not to hold the handles on the sides of the

treadmill, but were left to choose if they needed to swing their arms as they walked or ran. A further recording was made from the wearable devices on-screen display or companion iPhone.

3.2.2.1.5 Stairs tests

The aim of the stairs tests was to compare the step-count on the 7 activity-trackers in a free-walking stepped environment. There were 95 stairs in the stair test, and participants started on the ground floor for this part of the exercise. Participants were instructed not to use the handrails, but were free to choose if they needed to swing their arms as they walked up the stairs. Additionally, participants were instructed not to skip steps while going up and down the stairs. An investigator walked with each participant, using hand-held clicker to manually count steps. Participants first walked down the stairs, and the results tabulated both before and after the exercise. Subsequently, participants walked up the stairs and the process was repeated.

3.2.2.1.6 Analysis

All analyses were performed in SPSS version 25. Each participant was assessed in two walking conditions on the treadmill tests at 3 km/h and 7 km/h, along with ascending and descending stairs in the stairs test. In all assessments, the difference between the device-measured step-count and the true count was recorded by the investigator. For each device, the median step-count across different participants was calculated, and the interquartile difference. Additionally, the percentage difference between the true count and the device-count was calculated for each activity. To assess correlation, the paired t-test was performed on the true step-count and measured step-count.

3.2.2.2 *Testing the specifications of trackers*

After the completion of accuracy testing, technical specifications listed by the manufacturer needed to be tested to ensure they were suitable for use as per the selection criteria outlined in section 2.2.1.1:

1) Wrist-wearable

2) Waterproof
3) Daily step-count logging
4) Battery life of > 14 days
5) Local storage of > 14 days

3.2.2.2.1 Wrist-wearable

Trackers were considered wrist-wearable if the manufacturer supplied a wrist-band in the official packaging for the product. Furthermore, trackers had to be easily removable to allow for patients to return them.

3.2.2.2.2 Waterproof

Five of each wearable device were left in a beaker of water at room temperature for 30 minutes, towel dried, and subsequently synced with the companion smartphone application to retrieve stored data. Devices were considered waterproof if they were functional after being dried. Functionality was defined as on-device LEDs being lit up, and the ability to pair successfully with a smartphone.

3.2.2.2.3 Daily step-count logging

Devices were tested by one participant, by wearing one tracker at a time, each worn from 7 pm on day 1 to 8 am on day 4. On day 3, the device was paired to a smartphone and step-count verified. If the device logged a non-zero number of steps on day 2, it was deemed to be able to log daily step-count. The device was unpaired from the smartphone on day 3. To ensure consistency, this process was repeated for five trackers for each model (Misfit Shine 2 and Misfit Ray).

3.2.2.2.4 Battery life of > 14 days

After the completion of 2.2.2.1.3, each device was kept static on a desk at room temperature, and tapped to re-activate on day 18. If the device LED or display re-activated, it was considered

to be activated successfully. Battery level was then checked in the mobile application on the smartphone.

3.2.2.2.5 Local storage of > 14 days

After the completion of 2.2.2.1.4, each tracker was re-paired with a smartphone on day 18, and the data from day 4 was checked. If a non-zero number of steps was shown in the companion application, it was deemed to have data storage of at least 14 days.

3.2.3 Results

3.2.3.1 *Qualitative comparison*

Table 3-1 summarises the comparison of these devices, including the technical specifications and placement location. It should be noted that all these devices are described by the manufacture as being waterproof.

The Garmin Vivofit has an on-device display, unlike the other three devices which require a companion device to get detailed step-count information. All four devices have a battery life of more than one month, with the Vivofit listing a battery life of over one year. While the Vivofit and Misfit devices do not have an on-device display, the Shine 2 and UP Move both have a clock-face pattern of lights that can be used to check time and provide feedback to the user regarding their progress of physical activity through the day, based on pre-defined targets.

All the devices have a 3-axis accelerometer, but the Apple Watch also has a gyrometer which collects data about the orientation of the watch and therefore the position of the wrist. Unlike the other devices, it is also marketed as a Smart Watch and not a fitness tracker. However, it has the shortest battery life at up to 18 hours. The Fitbit Charge HR is capable of measuring heart rate intermittently as well as steps, and utilises these measures to provide surrogate measurements such as energy, distance, sleep etc. It has a longer battery life of up to 5 days but is still insufficient for the criteria defined for use in a cystectomy patient trial. The Fitbit Zip was one of the earliest fitness trackers launched, and the first Fitbit with a removable battery. It is meant to be worn on the waist or collar, and like the Apple Watch and Fitbit Charge HR, comes with an in-built display.

Table 3-1 summarises the comparison of these devices, including the technical specifications and placement location. It should be noted that all these devices are waterproof.

Device	Type	Steps on display?	Placement	Measurement	Size (mm x mm x mm	Weight (g)	Software	Battery Life	Local data storage duration	Cost
Apple Watch	3-axis Accelerometer, gyroscope	Yes	Wrist	Steps, heart rate, sleep, energy expenditure, activity intensity	42.5 x 36.4 x 11.4	34.2	WatchOS (Device)	Up to 18 hours	Not listed	£399
Misfit Shine 2	3-axis Accelerometer	No	Wrist	Steps, distance, sleep, energy expenditure	30.5 x 30.5 x 8	8.5	Misfit for iOS (Companion)	**Up to 6 months**	**4 weeks**	£80
Misfit Ray	3-axis Accelerometer	No	Wrist	Steps, distance, sleep, energy expenditure	12 x 38 x 12	8	Misfit for iOS (Companion)	**Up to 4 months**	**4 weeks**	£79
Jawbone Up Move	3-axis Accelerometer	No	Wrist	Steps, distance, sleep, energy expenditure	23.6 x 23.6 x 6.9		UP for iOS (Companion)	Up to 6 months	Not listed	£39
Fitbit Charge HR	3-axis Accelerometer	Yes	Wrist	Steps, heart rate, sleep, energy expenditure, activity intensity	21 x 170* x 10	26	Fitbit Connect	Up to 5 days	**Up to 30 days**	£150
Garmin Vivofit	3-axis Accelerometer	Yes	Wrist	Steps, distance, sleep, energy expenditure	25.5 x 21 x 10	25.5	Garmin Connect	**More than 1 year**	**4 weeks**	£99
Fitbit Zip	3-axis Accelerometer	Yes	Collar	Steps, distance, sleep, energy expenditure	28 x 35.5 x 9.65	8	Fitbit Connect	4-6 months	23 days	£50
*Fitbit Charge HR device width is not listed in specifications, only strap length is listed by manufacturer.										

Table 3-1: Comparison of various wearable fitness trackers and wearable devices

3.2.3.2 *Quantitative comparisons*

3.2.3.2.1 Participant characteristics

Eleven healthy individuals (male: 9 and female: 2) of median age 21 (IQR 20.5 – 22) years were recruited into the study. The median height, weight and BMI of the group 168.2 cm (IQR 165.7 – 172.1), 68.5 kg (IQR 61.7 – 75.6) and 24.2 kg/m^2 (IQR 22.8 – 24.6) respectively. The median stride length was 52.7 cm (IQR 49.1 – 55.7).

3.2.3.2.2 Results of treadmill experiments

3.2.3.2.2.1 Walking speed (3 km/h)

Median % in Table 3-2 represents the median of the device measurement expressed as a percentage of the true step-count. A paired t-test was performed for the true step-count and each tracker measurement, and a statistically significant difference ($p<0.05$) was identified for the Misfit Shine 2, Misfit Ray, Jawbone Up Move and Fitbit Charge HR when compared with the true step-count. Overall, median step-count between all 7 trackers ranged from 84.4% to 109.8% when expressed as a percentage of the true value. The Apple Watch had the closest median step-count to the true measured steps. Of the four devices that met the criteria set out for use in a prospective trial to monitor patient activity, only the Garmin Vivofit did not have a significant difference with true step-count. Of note, the interquartile range of values for the Vivofit was narrower than both the Misfit trackers and the Jawbone Move UP. These results are summarised in Table 3-2, and Figure 3-2 shows the box and whisker plots of the tracker performance at the 3 km/h walking speed.

	3 km/h				
	Median	**IQR**	**Median %**	***t***	**Significance (2-tailed)**
True Steps	284	(269.25-305.25)	-	-	-
Apple Watch	278	(217.13-306.75)	97.9%	-0.096	0.925
Misfit Shine 2	261	(181.50-275.00)	91.8%	3.704	**0.004**
Misfit Ray	240	(169.88-276.00)	84.4%	2.743	**0.021**
Jawbone Up Move	274	(222.00-282.38)	96.3%	2.238	**0.049**
Fitbit Charge HR	312	(280.50-419.63)	109.8%	-3.420	**0.007**
Garmin Vivofit	274	(261.56-298.62)	96.3%	1.425	0.185
Fitbit Zip	292	(270.38-301.88)	102.6%	0.447	0.665
Median % refers to the median step-count measured from the device as a percentage of the true step-count					

Table 3-2: Summary of results of treadmill experiments comparing the accuracy of measurement of seven different fitness trackers at a walking speed (3km/h). Results of t test (significance and degrees of freedom 't' are included.

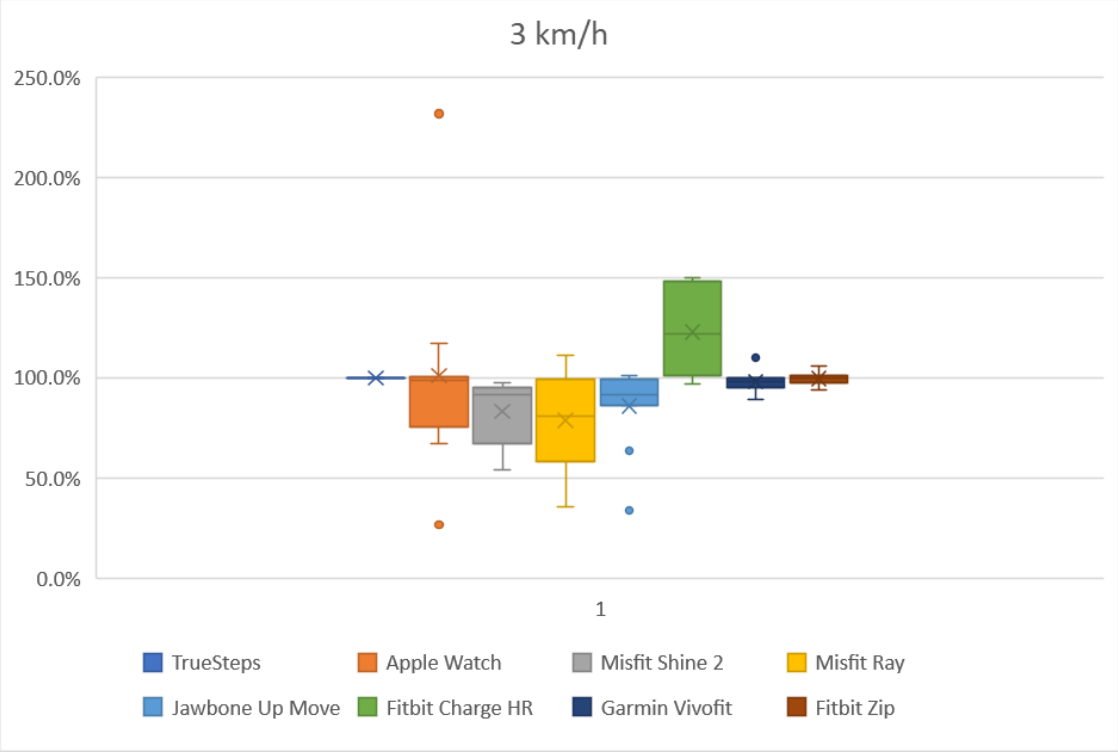


Figure 3-2: Box and whisker plots of the relative performance of the seven fitness trackers at the walking pace of 3 km/h.

3.2.3.2.2.2 Jogging speed (7 km/h)

None of the fitness trackers showed a significant correlation to true step-count at the 7 km/h speed. This is despite higher median step-count being closer to the true step-count at 95.5-100.2%, compared to the walking speed of 3 km/h (81.1-121.8%). These results are summarised in Table 3-3. Similarly, the IQR values for the 7 km/h values were smaller than those at 3 km/h, but multiple outliers were observed as shown in Figure 3-3.

Due to more pronounced arm movements and therefore a stronger signal for the accelerometer, it would have been reasonable to expect that devices would be more accurate at 7 km/h speed. Interestingly, median step-count was closer to the true step-count when compared with the 3 km/h speed despite no statistically significant correlation being identified. However, accuracy at this speed is less relevant than the 3 km/h speed for the iROC trial, as patients with bladder cancer, particularly in the peri-operative period following radical cystectomy, are unlikely to be walking or running at speeds close to 7 km/h for most of their daily activity.

	7 km/h				
	Median	**IQR**	**Median %**	***t***	**Significance (2-tailed)**
True Steps	380	(340.13-389.63)	-	-	-
Apple Watch	379	(333.00-387.00)	99.8%	-0.574	0.578
Misfit Shine 2	357	(276.00-376.50)	94.1%	0.909	0.385
Misfit Ray	315	(289.88-325.13)	83.0%	1.405	0.194
Jawbone Up Move	378	(366.75-394.88)	99.6%	-0.818	0.432
Fitbit Charge HR	378	(329.63-436.88)	99.6%	-1.961	0.078
Garmin Vivofit	377	(339.00-399.00)	99.4%	-0.969	0.355
Fitbit Zip	382	(341.25-392.25)	100.6%	-1.444	0.183
Median % refers to the median step-count measured from the device as a percentage of the true step-count					

Table 3-3: Summary of results of treadmill experiments comparing the accuracy of measurement of seven different fitness trackers at a jogging speed (7km/h). Results of t test (significance and degrees of freedom 't' are included.

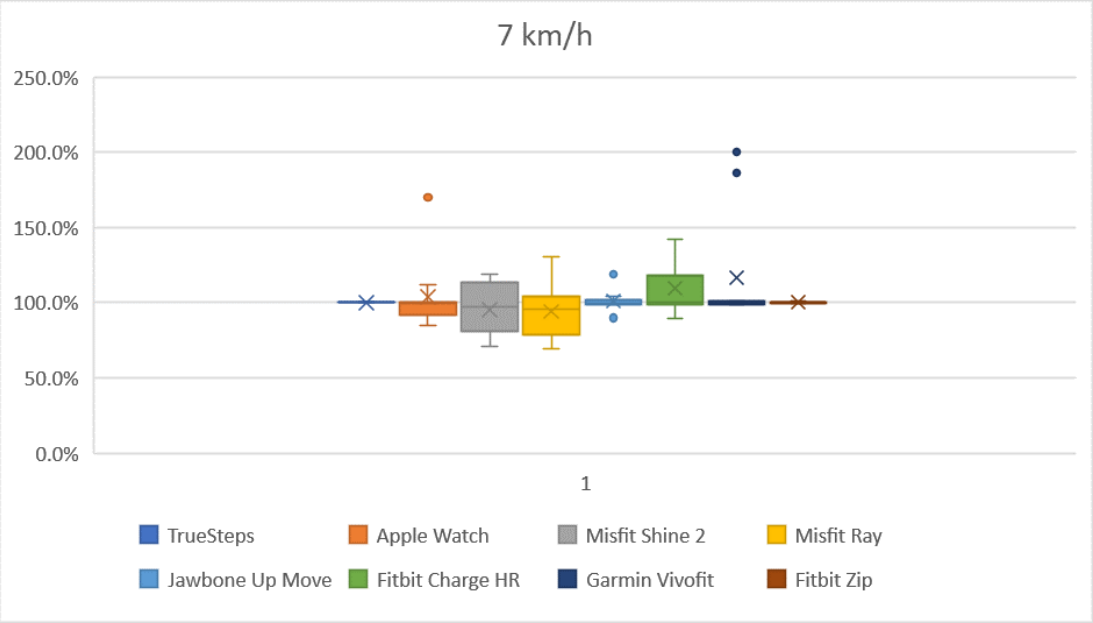


Figure 3-3: Box and whisker plots of the relative performance of the seven fitness trackers at the walking pace of 3 km/h

3.2.3.2.3 Results of stairs experiment

The stairs experiment was divided into two parts, descent followed by ascent.

3.2.3.2.3.1 Stairs Descent

During stairs ascent, none of the trackers were determined to have a statistically significant relationship with true step-count using the paired t-test. None of the trackers showed a significant correlation with the true step-count and he results are summarised in Table 3-4. As shown in the box-and-whisker plots in Figure 3-4, the Fitbit Charge HR and Fitbit Zip had the narrowest IQR.

3.2.3.2.3.2 Stairs Ascent

For stairs ascent, statistically significant correlations with true step-count were identified for the Apple Watch, Garmin Vivofit and Fitbit Zip. Of the three devices, the Fitbit Zip had the least wide IQR (106.50-111.50). The median percentage recorded by the Fitbit Zip was the closest to the true steps taken during stairs descent, with a 102.9%. However, only the Vivofit fit the criteria for device selection set out in 3.2.2.1. The Fitbit Charge HR had the closest median percentage to the true count, but this relationship was not found to be statistically significant due to outliers as shown in Figure 3-5.

	Stairs descent				
	Median	**IQR**	**Median %**	***t***	**Significance (2-tailed)**
True Steps	106	104.50-106.50	-	-	-
Apple Watch	111	101.00-118.00	104.7%	-0.258	0.805
Misfit Shine 2	84	63.00-104.00	79.2%	2.082	0.083
Misfit Ray	94	87.00-113.00	88.7%	1.084	0.320
Jawbone Up Move	110	108.25-113.25	103.8%	0.045	0.966
Fitbit Charge HR	110	107.00-112.50	103.8%	-1.178	0.283
Garmin Vivofit	122	116.50-136.50	115.1%	-2.227	0.068
Fitbit Zip	108	106.50-111.50	101.9%	-1.741	0.132

Table 3-4: Summary of results of stairs experiments comparing the accuracy of measurement of seven different fitness trackers during stairs decent. Results of t test (significance and degrees of freedom 't' are included.

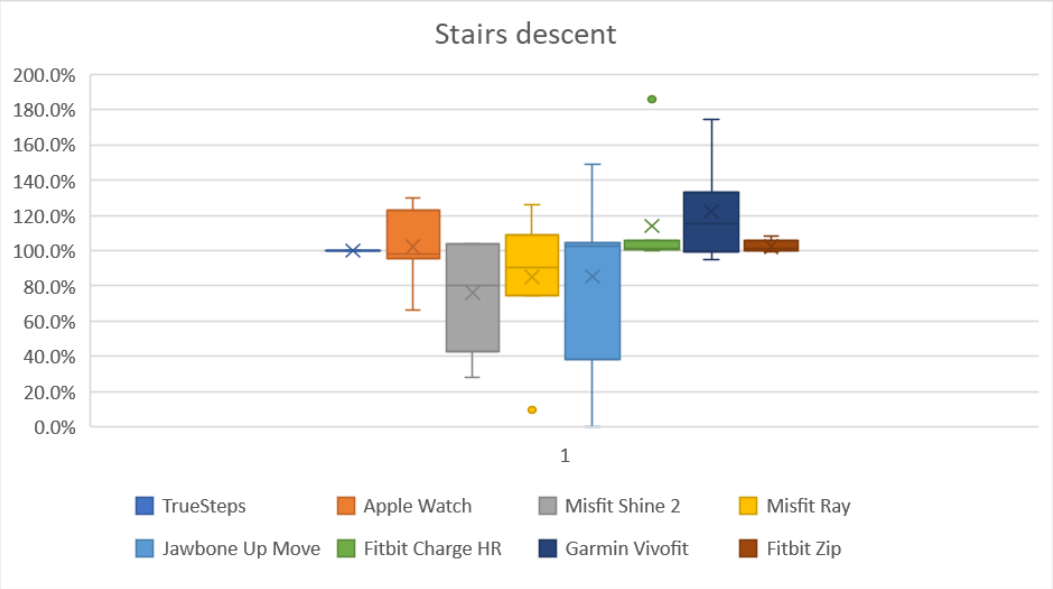


Figure 3-4: Box and whisker plots of the relative performance of the seven fitness trackers walking down 95 stairs

	Stairs ascent				
	Median	**IQR**	**Median %**	***t***	**Significance (2-tailed)**
True Steps	105	104-106.5	-	-	-
Apple Watch	127	124.5-135.5	121.0%	-4.526	**0.004**
Misfit Shine 2	49	22-133	46.7%	0.817	0.451
Misfit Ray	92	77-106	87.6%	0.868	0.419
Jawbone Up Move	101	98.5-109	96.2%	-0.485	0.645
Fitbit Charge HR	106	105.5-112.5	101.0%	-1.977	0.095
Garmin Vivofit	131	116.5-132	124.8%	-3.978	**0.007**
Fitbit Zip	108	105.5-109.5	102.9%	-3.092	**0.021**

Table 3-5: Summary of results of stairs experiments comparing the accuracy of measurement of seven different fitness trackers during stairs ascent. Results of t test (significance and degrees of freedom 't' are included.

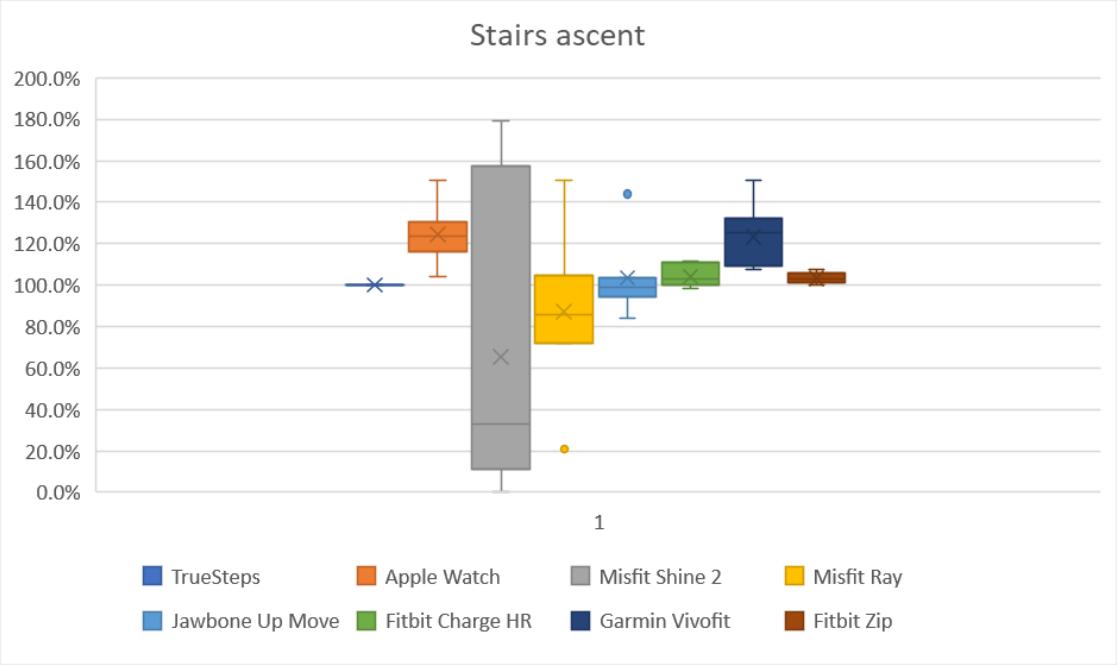


Figure 3-5: Box and whisker plots of the relative performance of the seven fitness trackers walking up 95 stairs

3.2.4 Discussion

No single tracker performed well on all of the four experiments (walking at 3 km/h, jogging at 7 km/h, stairs descent and stairs ascent). Overall, the trackers seemed to perform best at the 3 km/h walking speed, with four trackers showing a statistically significant correlation with true steps. A statistical correlation with stairs ascent was also identified for three trackers. Interestingly, there was no overlap in the two groups, with different trackers being accurate in both experiments.

Other experiments have shown similarly mixed results [117,200–202] for the accuracy of commercially available fitness trackers in variable conditions. For example, a study of 25 participants found that the Fitbit Ultra failed to measure any steps when walking at slow speed (3.24 km/h). This effect was not noted in my experiment, but it highlights the inter-device inaccuracy noted in different settings. In this experiment, trackers did not perform well at measuring steps while jogging at 7 km/h and stairs descent, with none of the trackers showing a statistically significant correlation to true step count.

Despite the noted inaccuracies, there is still merit in including a wearable device sub-study in the iROC trial. As four of the trackers tested correlated significantly with true-step-count at the walking pace, the majority of steps patients will take will be reflected in the measured step-count. The current experiment was performed to compare the accuracy of trackers in healthy volunteers, but we did not measure test-retest reliability or reproducibility. This is an important aspect to consider, and has been tested by other researchers in the field[203]. Their experiment included the vivofit 2 and Misfit Shine, and found that these devices step-count errors of 1% and 6% respectively.

While these experiments highlight the ‘bluntness’ of trackers as an instrument to take accurate and precise measurements of step-count, wearable devices still offer a unique opportunity to monitor patient activity passively outside the hospital environment.

3.2.5 Selecting a wearable device for prospective trial

While it would be ideal to use a device that performed well in all scenarios, this is unlikely given the data presented. For the purposes of our patient population undergoing radical cystectomy, the pragmatic tracker of choice would have to be from the four that performed well at the walking speed: Misfit Shine, Misfit Ray, Jawbone Up Move and Fitbit Charge HR. The Fitbit Charge HR does not fit the criteria set out due to battery life constraints. However, following the completion of our experiments, the company producing the Jawbone Up (Jawbone Inc.) announced that they were undergoing liquidation and will stop production of all products[204].

Therefore, we decided to use the Misfit Shine 2 tracker for our prospective trial, as it was more accurate than the Misfit Ray. However, Misfit had supply issues in the UK after supplying their first consignment of Shine 2 trackers, and the Misfit Ray tracker was used for the majority of the study.

As shown in section 3.2.3, the Misfit Shine 2 showed the strongest correlation ($p=0.004$) with true step counts at the walking speed (3 km/h). This was at the expense of accuracy, as it only captured 91.8% of steps in its count. Similarly, the Misfit Ray also showed a significant correlation ($p=0.021$) but only captured a median 84.4% of the total steps. All other scenarios (jogging speed 7 km/h, stairs descent, stairs ascent) showed mixed results, but are likely to be less important in patients recovering from a major operation such as radical cystectomy.

As both Misfit devices under-counted steps in this experiment, this result may be replicated in patients in the iROC trial. However, this will still be an important first step to activity tracking in this patient cohort which has not been done before. These limitations must be acknowledged when interpreting the step-counts collected from these wearable devices. In the future, we hope that devices will be improved with better hardware or new firmware that can improve their accuracy.

3.2.5.1 *Testing the specifications of trackers*

No additional ethics were required for this section the study, no subjects were recruited for this study. Members of the iROC protocol planning committee tested the devices within the group for the step-count logging stage. All other experiments were performed by me.

3.2.5.1.1 Wrist-wearable

Both the Misfit Shine 2and Misfit Ray come with a wrist-wearable strap in the box. In the case of the Shine 2, a removable strap is provided, alongside a circular fitness tracker. Additionally, a belt clip is also provided, and the fitness tracker can be slotted into either the removable wrist-strap or the belt clip. The Ray strap on the Ray is non-removable, and there is no belt clip provided. Both trackers use thermoplastic polyurethane straps, with grooves on one end of the strap, and a pin on the other end to fasten the tracker to the wrist.

3.2.5.1.2 Waterproof

After being left in a beaker of water at room temperature for 30 minutes, all trackers were still functional – the LEDs on both trackers lit up when tapped. Both devices were then connected wirelessly to the smartphone companion application and were able to complete data sync.

3.2.5.1.3 Daily step-count logging

All five Misfit Ray devices and five Misfit Shine 2 devices were able to complete data sync and provide data for day 2 on the first attempt. The ten trackers collected a median of 13628.5 steps (IQR 11847.5-16290.5) in one day. The mobile application also provided data on distance walked, calories and sleep, as shown in Figure 4-5. This data was not collected for the purpose of this study.

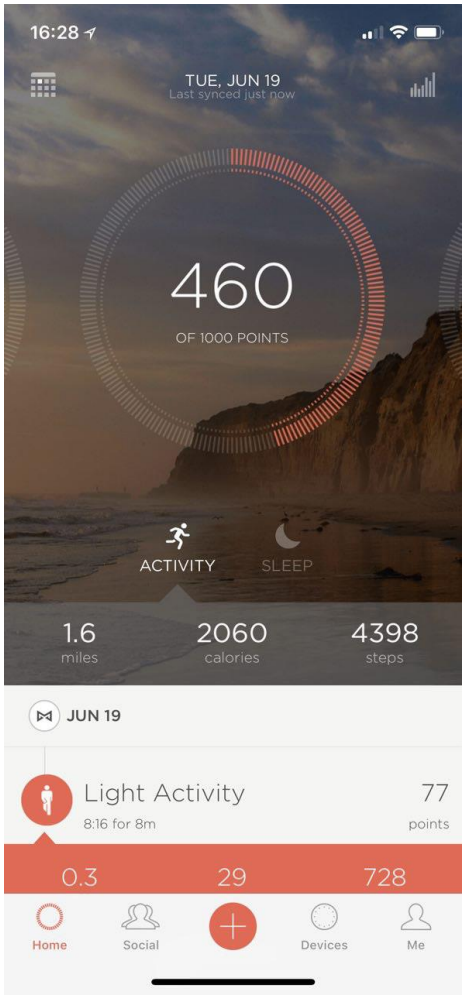


Figure 3-6: Screenshot of the Misfit mobile application on iOS showing data collected from one day including step-count, calories, distance in miles.

3.2.5.1.4 Battery life of > 14 days

On day 18, all ten devices were tapped to re-activate, and LEDs activated immediately. Following this, login was attempted on the mobile application to check battery level.

Misfit Ray devices were logged into on the first attempt. The Misfit Shine 2 required a mandatory software update which was performed after pairing the device to the smartphone. However, in all instances, this led to the mobile app becoming unresponsive after the update was marked as complete. After manually force-closing the application and restarting, four of the five devices

completed sync without any issues. For the last device, an additional step of un-pairing and re-pairing the tracker was needed. Following this process, data extraction was attempted.

Battery life on all ten devices was labelled “Full” by the Misfit application. Figure 4-6 displays the screenshot for a Misfit Ray device, along with how recently the battery level of the device was checked (3 minutes ago in Figure 4-6). Other device levels that can be displayed by the application are “High”, “Medium” and “Low”. These are the same levels used by the Misfit Shine devices. Batteries are user-replaceable in both devices, with three AG5 and one CR2032 required for the Misfit Ray and Misfit Shine respectively.

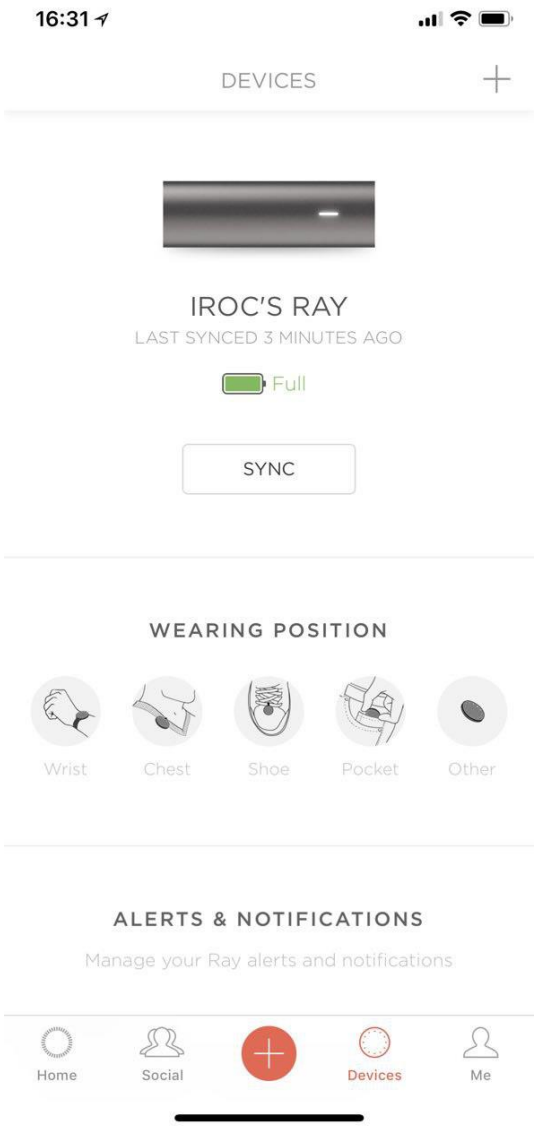


Figure 3-7: Misfit Mobile application screenshot illustration the battery level for a Misfit Ray device

3.2.5.1.5 Local storage of > 14 days

After re-pairing the devices as described in section 4.1.4.5, all ten devices showed a non-zero number in the companion application, and were deemed to have data storage of at least 14 days.

3.2.6 Conclusions

Of the four devices tested, only the Misfit Ray and Misfit Shine devices fit the criteria and were available. Both devices are suitable for use in a fitness tracking sub-study for the iROC trial. Both devices do not have an in-built display, which has the additional advantage of blinding patients to their own step-count, preventing patients from using the devices for goal-setting during the study. Since the Misfit Shine 2 had production issues at the beginning of the iROC trial, the Misfit Ray tracker was used for the wearable device sub-study.

3.3 Measurement of health status using PROMs prospectively in patients undergoing iRARC

3.3.1 Introduction

As discussed in the systematic review and meta-analysis of RCTs comparing RARC vs ORC (Section 2.3), patient reported outcome measures (PROMs) in the form of questionnaires are frequently used in RCTs, including those comparing ORC and RARC. As discussed in section 1.4.3, there are various different PROMs used to measure HRQOL for patients who have undergone RC, with the EORTC QLQ-C30 questionnaire being one of the two most commonly used[97]. Compared with the SF-36 (the other most commonly used questionnaire for patients undergoing RC), the EORC QLQ-C30 questionnaire has the advantage of being modular, which means that its data can be analysed by different domains - physical, role, emotional, cognitive and social function, as well as a global health score. The WHODAS 2.0 questionnaire is a validated questionnaire measuring health and disability used across all diseases. Due to these reasons, the WHODAS 2.0 and EORTC QLQ-C30 questionnaires were chosen to provide validated measurement of HRQOL for patients in the iROC study.

In this section, I will summarise the results of prospectively collected EORTC QLQ-C30 and WHODAS 2.0 questionnaires from patients undergoing RARC at a high-volume centre in the UK. The full methodology is outlined in section 3.3.2.

PROM questionnaires provide insight about patients' own assessment of their health status, which can be useful in assessing their return to normal function if a pre-operative baseline has been established. Unlike objective metrics like length of stay and complication rates, they also allow for measurement of the impact of surgery on patients' daily activities and lives. Additional to a global health status score, the EORTC QLQ-C30 questionnaire is a modularised questionnaire and offers measurement of functional scales and symptom scales. In contrast, the WHODAS 2.0 questionnaire provides an overall disability score only.

In this section, I will present the results of the EORTC QLQ-C30 questionnaire at three different timepoints (baseline, 3 months and 6 months) and assess for correlations between the two questionnaire scores. Additionally, I will compare the QoL questionnaire scores for patient groups who had complications during their peri-operative recovery.

3.3.2 Methods

3.3.2.1 *Sample*

Twenty-six patients undergoing intracorporeal robotic radical cystectomy (iRARC) at UCLH participated in this service development program to collect PROMs from patients in a clinical care pathway.

3.3.2.2 *Data collection*

3.3.2.2.1 Participant information

At time of recruitment, patient sex, age, BMI, diversion type and ASA was collected. Complications in the peri-operative period (90-days post-cystectomy) were graded according to the Clavien-Dindo Classification of surgical complications. Patients were divided into two groups: 1) major complications (CD≥3) and 2) minor or no complications.

3.3.2.2.2 Instruments

Health-related Quality of life questionnaires were given to patients at baseline, 3 months and 6 months post-operatively. The questionnaires used are: EORTC-QLQ-C30 and the WHODAS II questionnaires.

3.3.2.2.2.1 WHODAS 2.0

The World Health Organization Disability Assessment Schedule II (WHODAS 2.0) 36-item version is a validated[205] generic instrument for health and disability.

3.3.2.2.2.2 EORTC QLQ-C30

The European Organisation for Research and Treatment of Cancer has designed and validated[206] a modularised quality-of-life questionnaires (QLQ) to monitor patients with cancer (C30). The C30 has been validated for use in cancer patients, and offers modularised functional domains: physical, role, emotional, cognitive and social function, as well as a global health score.

3.3.2.3 *Analysis*

All statistical analyses were carried out using SPSS for Windows version 25.0. Patient characteristics of the cohort are reported using descriptive statistics. Completion rates for questionnaires at each timepoint are reported in percentages. As the EORTC QLQ-C30 questionnaire is modularised, scores were tabulated for individual domains of health and the global QoL score, and calculated using EORTC's scoring manual[207]. For the WHODAS 2.0 questionnaire, only the global QoL score was recorded. Wilcoxon signed ranked test was performed to compare quality of life measures collected from each questionnaire. Mann-Whitney U test was performed to compare non-parametrically distributed independent variables, and Wilcoxon Signed Ranks test was performed to compare non-parametrically distributed dependent variables.

3.3.3 Results

3.3.3.1 *Baseline characteristics*

26 patients undergoing RC participated in this observational study. Patient characteristics at baseline are describe in Table 3-6. All 26 participants had urothelial cell carcinoma (UCC) and underwent robot-assisted radical cystectomy with intracorporeal urinary diversion (iRARC). Overall, 69.2% of patients had MIBC (≥T2). In line with the incidence and prevalence of bladder cancer[5], the majority of participants were male (84.6%).

Characteristics	n	%
Sex		
Male	22	84.6%
Female	4	15.4%
Age *median (IQR)*	72 (66.25-75.75)	
BMI		
<18.5	0	-
18.5-24.9	7	26.9%
25-29.9	12	46.2%
>29.9	7	26.9%
Urinary diversion		
Ileal Conduit	24	92.3%
Continent diversion	2	7.7%
ASA		
1	0	-
2	15	57.7%
3	11	42.3%
Histology		
UCC	26	100.0%
SCC	0	-
Adenocarcinoma	0	-
Other	0	-

Table 3-6: Patient characteristics at baseline

3.3.3.2 *Completion rates*

Summary of completion rates are presented in Table 3-7. Completion rates for the EORTC QLQ-C30 questionnaire at baseline were higher than the WHODAS 2.0 (100% vs 61.5%). This is likely

to be attributed to the fact that the questionnaire bundle had the EORT QLQ-C30 questionnaire bundled in front of the WHODAS 2.0 36-item questionnaire. Patients who responded to the baseline questionnaires were asked to complete the post-operative questionnaires, and the completion rates for the WHODAS 2.0 questionnaire were higher at 3 and 6 months (68.8% and 93.8% respectively).

	WHODAS2.0	**EORTC QLQ-C30**
Baseline	16/26 (61.5%)	26/26 (100%)
3-month	11/16 (68.8%)*	23/26 (88.4%)
6-month	15/16 (93.8%)*	26/26 (100%)
*Only patients who completed questionnaires at baseline were offered questionnaires post-operatively		

Table 3-7: Completion rates for WHODAS 2.0 and EORTC QLQ-C30 questionnaires

3.3.3.3 *EORTC QLQ-C30 questionnaire scores*

A Wilcoxon signed-rank test showed that radical cystectomy elicited a statistically significant change in physical function (PF) at 3 months (Z = -2.254, p = 0.024) and 6 months (Z = -2.681, p = 0.007). This is despite the median PF score being similar (96.7, 93.3 and 93.3 at baseline, 3 months and 6 months respectively) both pre- and post-cystectomy.

No significant difference was identified in cognitive, emotional and social functions when compared to baseline. Figure 3-8 shows the PF and QoL scores of the patient cohort as box and whisker plots before and after RC. The majority of patients were below their pre-operative physical function scores even 6 months after RC. While the median QoL score remained relatively stable pre and post-RC, the confidence interval was wider at the lower margin, as shown in Figure 3-8(a). Like QoL, Role function (a measure of an individual's ability to perform tasks related to their daily routine) was also found to be significantly different at 6 months when compared to baseline, but this difference was not noted at 3 months.

In section 3.3.3.6, the correlation between major complications and change in quality of life is presented.

	Baseline Median (IQR)	3 months Median (IQR)	6 months Median (IQR)	Test statistics		
Physical Function	96.7 (86.6-100)	93.3 (86.7-100)	93.3 (71.7-100)	**p-value**	**0.024**	**0.007**
				Coefficient	-2.254	-2.681
Role Function	100 (100-100)	100 (83.3-100)	100 (79.2-100)	**p-value**	0.130	**0.032**
				Coefficient	-1.513	-2.140
Cognitive Function	100 (83.3-100)	100 (83.3-100)	83.3 (83.3-100)	**p-value**	1.000	0.330
				Coefficient	0.000	-0.975
Emotional Function	91.7 (75.0-100)	91.7 (75.0-100)	83.3 (72.9-100)	**p-value**	0.683	0.512
				Coefficient	-0.409	-0.655
Social Function	83.3 (66.7-100)	100 (83.3-100)	83.3 (66.7-100)	**p-value**	0.608	0.245
				Coefficient	-0.513	-1.163
Quality of Life (Global Health)	83.3 (70.8-83.3)	83.3 (75.0-83.3)	83.3 (47.9-93.7)	**p-value**	0.413	**<0.001**
				Coefficient	-0.819	-3.536
p-values and test statistics (z-value) are generated from Wilcoxon Signed Ranks test.						

Table 3-8: Scores of individual domains and overall quality of life measured by the EORTC QLQ-C30 questionnaire, and results of Wilcoxon signed-rank test comparing baseline to post-operative scores

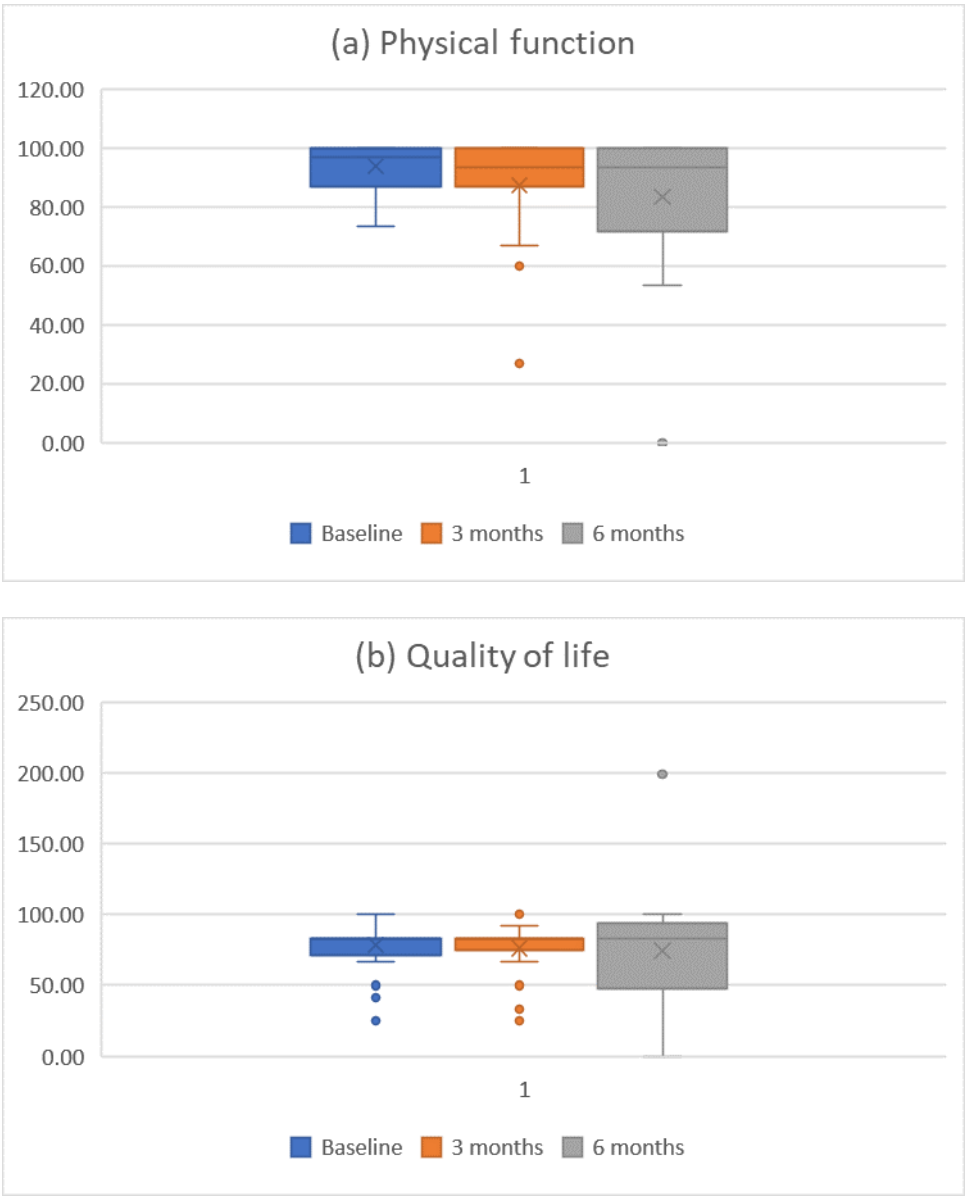


Figure 3-8: EORTC QLQ-C30 Physical function (a) and QoL (b) before and after RC, represented as box and whisker plots

3.3.3.4 *WHODAS 2.0 questionnaire scores*

Unlike the EORTC QLQ-C30 quality of life scores, there was a decline in the median quality of life score after RC from 5.20 at baseline to 4.16 at both 3 and 6 months respectively. However, there was no statistically significant difference between WHODAS 2.0 quality life scores at baseline and post-operatively.

In section3.3.3.5, correlations between WHODAS 2.0 disability and EORTC QLQ-C30 global health scores are explored.

	Baseline	**3 months**	**6 months**	**Test Statistics**		
Quality of Life *score (IQR)*	5.20 (4.14-9.89)	4.16 (2.08-39.58)	4.16 (1.56-32.81)	**p-value**	0.489	0.925
				Coefficient	-0.692	-0.094

Table 3-9: Global quality of life score as measured by the WHODAS 2.0 questionnaire

3.3.3.5 *Correlations between EORTC QLQ-C30 and WHODAS 2.0 quality of life scores*

Figure 3-8(b) and Figure 3-9 illustrates the quality of life scores measured by the EORTC QLQ-C30 and WHODAS 2.0 questionnaires as box and whisker plots. Compared to baseline, the median C30 global health score increases whereas the WHODAS 2.0 disability decreases. This is in line with clinical recovery as patients are recovering by three and six months after surgery.

Spearman rho correlation results are tabulated in Table 3-10. There is a statistically significant correlation between quality of life scores as assessed by the WHODAS 2.0 and EORC QLQ-C30 questionnaires at all three timepoints. As the WHODAS 2.0 score measured disability (low score = low disability) and the EORTC QLQ-C30 global health status measures level of functioning (low score = low level of functioning), the correlation coefficient at all three timepoints is negative.

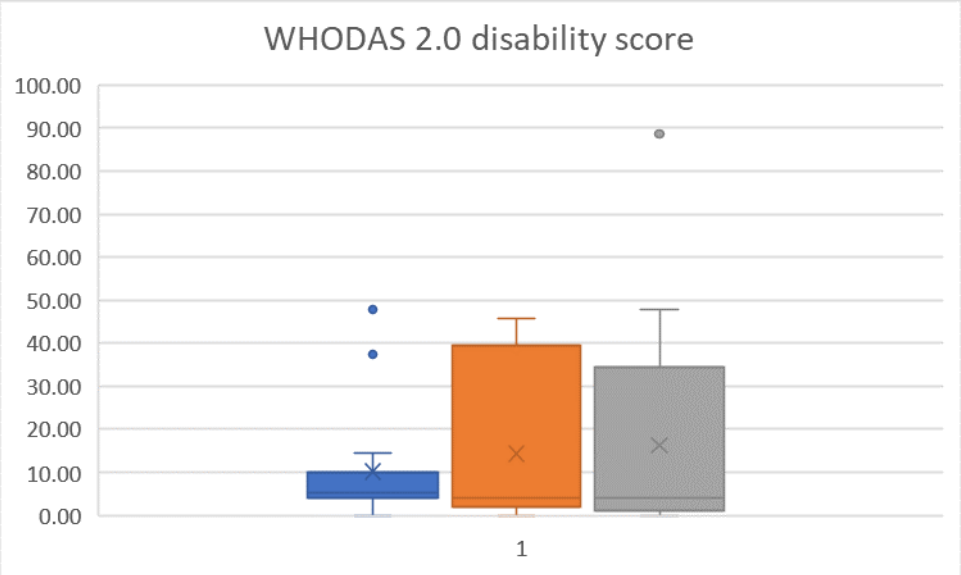


Figure 3-9: Box and whisper plots of quality of life scores measured by WHODAS 2.0 questionnaire at three different timepoints.

	WHODAS2.0	EORTC QLQ-C30		p-value
Baseline	5.20 (4.14-9.89)	83.3 (70.8-83.3)	p-value	**0.004**
			Coefficient	-0.692
3-month	4.16 (2.08-39.58)	83.3 (75.0-83.3)	p-value	**0.044**
			Coefficient	-0.526
6-month	4.16 (1.56-32.81)	83.3 (47.9-93.7)	p-value	**<0.001**
			Coefficient	-0.768

Table 3-10: Spearman Rho correlation between quality of life scores measured by WHODAS 2.0 and EORTC QLQ-C30 questionnaires at three different timepoints.

3.3.3.6 *Relationship between major complications and quality of life scores at 90 days*

No significant relationship was identified between major complications at 90 days and patient reported quality of life scores. Table 3-11 summarises the major complication (≥3) and minor complication (≤2) groups according to the Clavien-Dindo classification, and Table 3-12 summarises the test statistics. No statistically significant relationship was identified between major complications and quality of life.

	90 days Clavien-Dindo grade	**N**	**Mean Rank**	**Sum of Ranks**
EORTC QLQ-C30 QoL score at 3 months	≤2	16	11.28	180.50
	≥3	5	10.10	50.50
	Total	21		
WHODAS2.0 QoL score at 3 months	≤2	10	7.00	70.00
	≥3	4	8.75	35.00
	Total	14		

Table 3-11: Quality of life measures and 90-day Clavien-Dindo grade of complications, with patients grouped into major complications (≥3) and none or minor complications (≤2).

	EORTC QLQ-C30 QoL score at 3 months	**WHODAS2.0 QoL score at 3 months**
Mann-Whitney U	35.500	15.000
Wilcoxon W	50.500	70.000
Z	-0.388	-0.713
Asymp. Sig. (2-tailed)	0.698	0.476
a. Grouping Variable: 90 days complications CD ≥ 3		
b. Not corrected for ties.		

Table 3-12: Test statistics for the Mann-Whitney U test comparing quality of life measured by WHODAS 2.0 and EORTC QLQ C-30, with patients grouped into major complications (≥3) and none or minor complications (≤2).

3.3.4 Discussion

Compliance rates for the WHODAS 2.0 questionnaire (61.5%) at baseline are much lower than the EORTC QLQ-C30 Questionnaire (100%). As the EORTC QLQ-C30 questionnaire was presented first, this is likely to be due to questionnaire fatigue. Swapping the order of the questionnaires may also help assess if patients have a preference for one questionnaire over the other but this has not been assessed in this study. A shorter, validated version of the WHODAS 2.0 questionnaire is available, and can be used for future trials where multiple questionnaires need to be collected. Post-operative compliance improved to 68.8% and 93.3% at 3 months and 6 months respectively for the WHODAS 2.0 questionnaire, implying that patients who filled out the second questionnaire at baseline were more likely to continue responding. While this finding could indicate that patients are inherently less likely to complete the WHODAS 2.0 questionnaire, our sample size in this current study is limited so a bigger study would need to be done to assess this comprehensively. This conclusion is unlikely as the WHODAS 2.0 questionnaire is used in many different disease groups in the literature.

As expected, undergoing major surgery such as RC had a significant impact on the physical function post-operatively when compared to baseline. However, this effect was still noted 6 months post-operatively. Figure 3-8(a) shows the box and whisker plot for post-operative physical function measured by the EORTC QLQ-C30. Physical function does not recovery completely even 6 months after RC also shown in Table 3-8. The majority of patients reported lower physical function at 6-months post-operatively when compared with baseline. This is reflective of the prolonged recovery period following a major operation such as RC.

There was a statistically significant relationship between the QoL scores collected from both questionnaires at all timepoints. While the WHODAS 2.0 and EORTC QLQ-C30 are validated in chronic disease and cancer patients respectively, they have not been validated for use specifically in bladder cancer patients. Despite this, the EORTC QLQ-C30 questionnaire is

commonly used and well-accepted to measure QoL outcomes after radical cystectomy for bladder cancer[97]. The concordance between the QoL measurements between both questionnaires suggest that these instruments note similar trends in recovery in patients. This relationship will be tested more robustly after the completion of the iROC trial in a larger sample size.

Of note, no significant correlation was noted between major complications (CD≥3) and patient reported QoL scores for both the WHODAS 2.0 and the EORTC QLQ-C30. Given that major complications are complications requiring surgical intervention, admission to intensive care or death (there were no deaths noted in our current cohort), it is surprising that no difference was identified between the two groups. Previous studies have noted that PROM scores do not correlate with quality of care or surgery, as PROMs are patients' own assessment of their expectation of what their recovery should be like[208]. It could also be attributed to the fact that patients felt that their complications were dealt with that did not impact their overall quality of life at the post-op timepoint. However, this could be attributed to the small sample size, relatively few 'events' (major complications) or other delays in recovery. In the iROC trial, we will be able to explore the relationship between patients' perception of health measured through validated PROM instruments and objectively measured peri-operative outcomes in a larger group of patients. As the EORTC QLQ-C30 questionnaire is modularised and includes a physical function domain, it would be important to understand if the patient-reported score is consistent with an objective measure of mobility.

3.3.5 Conclusions

Health-related Quality of Life (HRQoL) measurement in the form of PROM questionnaires are an important tool in assessing recovery from RC. To improve compliance in the iROC trial, patients will be offered the WHODAS 2.0 12-item version, instead of the 36-item version that was used in this preliminary study. Both the WHODAS 2.0 and EORTC QLQ-C30 questionnaires are validated questionnaires in assessing overall QoL (in patients with chronic disease and cancer respectively), but the EORTC QLQ-C30 offers more modularised information divided into the various health domains. This allows for a more granular approach to analysing this data, as specific health domains can be interrogated as opposed to a single quality of life or disability score. An important health domain assessed in the EORTC QLQ-C30 is physical function, and it would be useful to compare patients' perception of their physical function to objectively measured physical activity using wearable devices.

3.4 Chapter conclusions

In this chapter, I tested various fitness trackers for accuracy, precision, and the criteria set out for their suitability for use in a trial comparing ORC and iRARC. None of the devices performed well in the different environments (walking at 3 km/h, 7 km/h, up and down stairs), overall devices performed the best at 3 km/h. Additionally, these errors in calculations maybe be more apparent when a smaller number of total steps are measured, and the percentage difference may be smaller when monitoring activity for a daily total. Despite these limitations, he Misfit Shine 2 and Misfit Ray were found to be the most appropriate wearable devices for use in the iROC trial. Aside from comparing recovery across the two arms in a trial, activity measurement could also be a surrogate for post-operative mobility, and provide information about the effects of complications on recovery in the peri-operative period.

PROMs such as the WHODAS 2.0 and EORTC QLQ-C30 questionnaires are validated measures of health status in patients with chronic disease, and could similarly be used to measure post-operative health and quality of life (QoL). The global QoL scores and mobility scores could be compared to the objectively measured physical activity recorded by wearable devices. Used in tandem, they may offer a more holistic understanding of the individual patient's journey through complex surgery.

Chapter 4 Designing a trial to compare open and robotic cystectomy with a sub-study using wearable devices to measure patient mobility

4.1 Chapter summary

The work undertaken in this thesis draws on data collected as part of the iROC trial – a prospective multi-centre randomised-controlled trial comparing open radical cystectomy vs robot-assisted radical cystectomy with intracorporeal diversion for bladder cancer. I was involved in the design and development of the iROC trial, and wrote the trial protocol under the guidance and supervision of the Chief Investigator (Professor James W.F. Catto) and Co-Investigator (Professor John D. Kelly). The full trial protocol is published in the BMJ Open Journal[199]. The full version of the protocol is included in **Error! Reference source not found.**.

Relevant portions of the trial protocol are summarised in this chapter to provide a skeletal overview of the trial, and it also draws from the protocol manuscript published in the BMJ Open Journal[199]. The trial visit schedule provides the structure for clinical follow up, and the fitness tracker sub-study was designed pragmatically to collect data on a similar schedule. As the remit of this thesis extends beyond the scope of the iROC primary outcome, later chapters provide detailed methodology for specific experiments.

4.2 The iROC trial

The iROC trial was designed to fill an unmet need described in section 2.3.4 – the lack of a randomised trial offering comparisons to RARC with intracorporeal urinary diversion (iRARC). All five RCTs offering comparisons between ORC and RARC have used eRARC for the robotic comparator. As discussed in section 2.3, all of these have failed to show a significant difference in recovery.

The results of the RAZOR trial have shown that the robotic approach is non-inferior to ORC[36] in 2-year oncological outcomes. As highlighted in section 1.2, The extirpative (cancer removal) component of eRARC and iRARC are similar, but the diversion is performed open and robotically respectively. As such, the pertinent unanswered question remains centred on early recovery. Based on IRCC data[48], gastrointestinal complication rates were significantly lower in the iRARC group compared with eRARC. The overall 90-day complication rates were not significantly different, but a trend favouring iRARC was identified (41% vs 49%, p = 0.05).

In the UK, 67.8% and 20.6% of RCs are performed as ORCs and RARCs respectively, according to the cystectomy national audit[29] data for 2014-2015. Only a sub-set of the RARCs are performed intracorporeally, and a further subset of those are performed by experienced surgeons in high-volume centres. Recommendations published by Pasadena Consensus panel suggests that a learning curve exists for RARC, and provides guidance for gaining experience safely to operate intracorporeally.

4.2.1 Endpoints

Endpoints were selected for the iROC trial based on the results presented in section 2.3. Measures of recovery such as length of stay and CD classification of complications measure aspects of recovery from surgery. While these are both important measures of surgical recovery, newer metrics like ‘days alive and out of hospital’ (DAOH) aim to capture the impact of the length of stay and any re-admissions to hospital on patient’s normal function at home. To measure

DAOH at 90 days, all days spent in hospital (index admission and any resulting admission nights) are subtracted from 90 days.

Primary endpoint: Days alive and out of hospital (DAOH) at 90 days following RC

The original primary outcome was log(DAOH) within the first 90 days from surgery. However, a review of blinded interim analysis (this data is presented in Chapter 1) suggested that this outcome would be left skewed and that other sample size assumptions would not hold. The primary outcome measure was subsequently modified to log(90-DAOH) within the first 90 days from surgery.

The secondary endpoint measures include:

1) **Quantified activity levels** (baseline, 5 days, 5 weeks, 3 months, 6 months, 12 months): Fitness tracking devices will record steps taken for 7 consecutive days at predetermined time points 39. The WHO 30 Second sit to stand test will also be administered at these time points 33.
2) **EORTC QLQ-BLM30 & QLQ-C30** (baseline, 5 weeks, 3 months, 6 months, 12 months): A 30-item questionnaire for patients with bladder cancer (T2, T3, T4a and T4b). This module is designed to be used in conjunction with QLQ-C30 and includes an assessment of urinary symptoms, bowel symptoms, sexual functioning, urostomy problems, problems associated with the use of a catheter, and body image 31 32.
3) **EQ-5D-5L** (baseline, 5 weeks, 3 months, 6 months, 12 months): a self-completion tool for patients which is applicable to a wide range of health conditions and treatments. Measured domains include mobility, self-care, usual activities, pain and anxiety or depression 29.
4) **WHODAS 2.0** (baseline, 5 weeks, 3 months, 6 months, 12 months): a generic assessment instrument for health and disability used across all diseases, including mental, neurological and addictive disorders directly linked at the level of the concepts to the

International Classification of Functioning, Disability and Health (ICF). WHODAS 2.0 covers 6 Domains of Functioning, including cognition, mobility, self-care, getting along, life activities and participation 30.

5) **Oncological outcomes** (3 months, 6 months, 12 months): The curative outcomes from the RC will be examined at 3, 6 and 12 months to determine local and distant recurrence, metastases, need for palliative treatment and survival (overall and cancer specific). CT scans will be undertaken at 12 months, according to usual practice, and if clinically indicated.
6) **Translational sample collection** (baseline, 1 month, 3 months, 6 months, 12 months): At baseline and during cystectomy, blood, urine and tissue (paraffin-embedded) will be collected. This is to test and validate a cell-free DNA based biomarker for both blood and urine in the detection of cancer mutations associated with bladder cancer.

The full table of events for the iROC trial is summarised in Table 4-1. This schedule serves as a guide for when all data was collected during the trial.

	Baseline	RC	Post-op	Visit 2	Visit 3	Visit 4	Visit 5
Time:	Pre-op	Day 0	POD 4-7	5 wks ±2 wks	12 wks ±2 wks	24 wks ±4 wks	1 yr ±4 wks
Informed consent & randomisation	X						
Demographic data, medical history etc.	X						
Physical examination, vital signs	X						
Fitness for surgery assessment	X						
12-lead ECG and CPET testing [1]	X						
Haematology & Biochemistry	X			X	X	X	X
Pregnancy test [2]	X						
Translational and research bloods [3]	X			X	X	X	X
Urinalysis	X						
Urine collection for research [3]	X						
Chest, abdomen and pelvis imaging[4]	X					X	X
Clavien-Dindo assessment			X	X	X		
Adverse events		X	X	X	X	X	X
Tumour sample		X					
Paraffin embedded tissue		X					
Survival and treatment data			X	X	X	X	X
EQ-5D-5L	X			X	X	X	X
WHODAS 2.0	X			X	X	X	X
EORTC QLQ-BLM30	X			X	X	X	X
30 Second Chair to Stand test	X		X	X	X	X	X
Quantified activity levels (steps tracker)	X		X	X	X	X	X

Abbreviations: RC = Radical Cystectomy; POD = Post-operative day;
1 CPET: Cardio Pulmonary Exercise Testing only in participating centres that routinely perform this test
2 Can be urine or blood-based pregnancy test, depending on site's standard of care
3 Home collection kit to be given
4 Imaging schedule is not mandated by the trial. However, if imaging studies have been conducted, these results are collected

Table 4-1: Timing of events and outcome collection for the iROC trial

4.2.2 Study population

The iROC trial will recruit patients referred to high-volume tertiary centres for radical cystectomy for bladder cancer. Participants will be recruited from NHS cancer centres undertaking both ORC

and iRARC. Recruiting centres will be invited by the Trial Management Group (TMG) as having well developed RARC programs with sufficient volume to recruit a reasonable number of patients to the trial (see section 4.2.4.1).

The full inclusion and exclusion criteria for patients are as follows:

4.2.2.1 *Inclusion Criteria*

i. Participants must be over 18 years of age.

ii. Histopathological confirmation of bladder cancer (UCC, SCC, adenocarcinoma or rare variant)

iii. CIS or stage pTa or pT1 or ≥pT2 or mobile bladder mass on bimanual examination under anaesthesia

iv. Node status ≤ N1 on imaging criteria or PET –ve outside pelvis

v. ECOG grade 1,2 or 3.

vi. Able to give informed written consent to participate.

4.2.2.2 *Exclusion Criteria*

i. Unwilling to undergo cystectomy.

ii. Previous abdominal surgery rendering them unsuitable for either iRARC or ORC.

iii. Patients with upper urinary tract disease.

iv. Concomitant disease that would render the patient unsuitable for the trial

v. Pregnant or lactating females

vi. Previous radiotherapy for bladder cancer

4.2.3 Sample size

For the complete trial, the sample size was initially set at 320, to be amended according to the results of the interim analysis. More information about the power calculation for the main iROC trial is included in the full protocol included in **Error! Reference source not found.**. An interim analysis was planned after the first 30 patients recruited completed 90 days of follow up

required for the primary outcome analysis. The purpose of the interim analysis is to serve as an internal feasibility study to test the practicality of recruitment and data collection. The primary outcome measure is the number of days alive and out of hospital within the first 90 days of follow-up, so a combination of the number of days in hospital and the number of days since death, if applicable.

4.2.4 Setting

4.2.4.1 *Surgeon and unit accreditation*

It is recognised that variations in surgical team performance and practice produce wide differences in morbidity and mortality from RC[209]. Previous UK based surgical RCTs have required audited data on surgical outcomes for all surgeons undertaking radical surgery within the trial (e.g. NIHR funded ProtecT RCT[210]). This approach proved successful at reducing variability in outcomes. Therefore, surgeons and surgical teams undertaking radical surgery within this study require accreditation from the Trial Management Group before undertaking such surgery.

Surgeon accreditation is achieved through the submission of outcomes data from the last consecutive 30 RCs or preferably an export of consecutive contemporaneously collected outcomes from the BAUS RC complex dataset. Important measures used to assess accreditation include length of stay and pathological outcomes (node yields and positive margin rates). Accredited surgeons will have undertaken more than 10 RCs per year for the last 2 years as primary surgeon, have a median length of stay under 14 days and will have 90-day post-RC mortality rate of less than 5%. Individual surgeon data will act as surrogate measures for the entire surgical team.

All centres recruiting for the iROC trial will be following the NHS England Chief Commissioning Policy, that there is no proven benefit for RARC over ORC[211]. Therefore, these centres will

withdraw RARC from standard of care, and only offer the treatment as part of the randomisation in the iROC trial.

4.2.4.2 *Radical cystectomy*

Within this RCT, RC will be performed as is standard of care throughout the NHS. In females this includes anterior pelvic exenteration (with the uterus, fallopian tubes and a component of the anterior vaginal wall). The urethra will be excised in females choosing an ileal conduit. In males, this includes the prostate and seminal vesicles. Nerve sparing to the prostatic neurovascular bundles should be attempted as per typical practice in that unit. Oophorectomy is optional, as per local practice, and individualised for each patient.

Pelvic lymphadenectomy should be included in all cases, unless contraindicated clinically. The lymphadenectomy template should include the external iliac, obturator and internal iliac nodes, with a proximal extension to the level of the ureteric crossing of the common iliac vessels. A more extended lymphadenectomy is acceptable. Excised lymphatic tissue should be submitted for histological analysis.

Urinary tract reconstruction within this trial is limited to either i). ileal conduit or ii). Orthotopic neobladder (by whichever design is practiced by that unit).

4.2.4.3 *Enhanced recovery pathways*

Centres will be expected to have an Enhanced Recovery programme in place locally. It should be based on the recently published BAUS Guidelines[62]. It is anticipated that there will be minor variation of practice *between* centres, according to local expertise, but each local protocol should be consistently applied within the *individual* centre and agreed by all surgeons performing RC. A baseline assessment will be made at each centre using a self-reported questionnaire and centres will also be asked to describe the process by which longitudinal

compliance with the protocol is assessed (examples could include a snapshot audit or SPC charts for length of stay).

4.2.5 Data Collection

All data collection will be online as this study will use an electronic Case Report Form(eCRF).

All data will be entered in the approved iROC database by a member of the iROC study team and protected using established procedures. Access to the eCRF system will only be provided to staff with relevant authority delegated to them on the site's delegation log.

4.2.5.1 *Objective measures of performance status*

4.2.5.1.1 Cardiopulmonary exercise testing results

There is some evidence that cardiopulmonary exercise testing (CPET) can be used to identify patients who are at higher risk of complications[86]. As an optional sub-study, recruiting sites were asked to provide data from CPET, if this test was performed on patients recruited into the iROC trial.

4.2.5.1.2 30-second chair-to-stand (30-CtS) test

The 30-second chair-to-stand test is used as a quick test of frailty in various healthcare settings[212–214]. Patients will be asked to perform the 30-second chair to stand test at four timepoints – baseline and 3 post-operative timepoints: day 5, 1 month and 3 months. The test involves patients being seated in a chair without armrests, folding their arms across their chest, standing up and sitting back down repeatedly. The number of times patients are able to stand up during the 30-second period is recorded as the score.

4.2.5.1.3 Quantified activity levels using wearable devices

Patients were consented to wear a wrist-worn wearable devices with a 3-axis accelerometer for seven consecutive days at the same timepoints as the 30-CtS testing. The tracker was issued to

patients in clinic, and monitoring started at midnight at the end of the day of clinic appointment. After the seven-day monitoring period, patients mailed the tracker back to the central receiving lab in a pre-stamped envelope, with a data label containing: Subject ID, patient initials or identifier, date of birth, cystectomy date, data tracker attached, and date tracker mailed. All fields except the date the tracker was mailed were pre-filled by the research nurse prior to the kit being issued. A photograph of the wearable devices kit is displayed in Figure 4-1. Daily step-count data was extracted from wearable devices upon return to the central receiving lab. Each tracker was assigned a unique serial number, and was wiped with disinfectant before being re-issued to a different patient.

The 7-day monitoring period was selected pragmatically based on the findings of preliminary experiments presented in section 0 based on battery life and local data storage limits of affordable and available wearable devices and fitness trackers. Gretebeck and Montoye[215] have previously reported that at least 5-6 days of pedometer data were needed to accurately describe the activity pattern.

Figure 4-1: Photograph of the wearable device and kit issued to patients

4.2.5.2 *Peri-operative complications*

4.2.5.2.1 Adverse events recorded using the Clavien-Dindo classification

The Clavien-Dindo (CD) classification is used to grade surgical complications in the peri-operative period (see section 1.4.2.3). A grade of 0 refers to no complications during peri-operative recovery, and a grade of 5 refers to death. The full CD classification is outlined in Table 1-3. Complication data will be collected for 30 and 90 days post-operatively and reported according to the CD classification.

4.2.5.2.2 Re-admission to hospital within 90 days of surgery.

As noted in the systematic review in section 1.4.2.1, re-admission to hospital is often overlooked during data collection for RCTs, despite being as important as the length of stay during the index hospital admission. Re-admission days, and re-interaction with healthcare (GP visits, A&E visits) will be recorded during the first 90 days after RC.

4.2.5.2.3 30 and 90-day mortality rate

Mortality rates for RC of 3-6%[75,216] have been reported in the literature. Therefore, high mortality rates are not expected in the iROC trial. Nonetheless, all deaths during the peri-operative period will be reported in the 30 days and 90 days after RC.

4.2.5.2.4 Translational sample collection

Baseline blood, paraffin blocks and urine samples and sequential blood samples will be collected from patients participating in this study. Sequential blood samples will be collected at the time points outlined in section 4.2.1. An additional snap frozen tumour tissue sample will be collected at the time of cystectomy for a subset of patients at sites that are able process and store frozen samples. All blood and urine samples will be posted to the central receiving lab for storage and analysis. A summary of the samples to be collected and the time points is outlined in the study schedule and table of assessments (Table 4-1).

This sub-study is included in the iROC trial because of a pilot project I worked on during my doctoral time. I bio-banked over 1,000 blood, urine and tissue from patients undergoing radical cystectomy for bladder cancer. Blood was collected from patients in the first year of follow-up, plasma separated and DNA extracted. DNA samples were interrogated for 20 bladder cancer-related mutations using 50 primers. As part of this work, I published a systematic review of blood-based genomic and tumour-cell biomarkers[217], received The Urology Foundation Research Scholarship Award, won the Best Poster Prize at the American Urological Association 2017 conference and was chosen for the Best Academic Papers session at the British Association of Urological Surgeons session.

This work is not presented in this thesis, as the content is too divergent from the included material.

4.2.6 Analysis Plan

The statistical analysis plan for the trial will be decided before the completion of the iROC trial.

An interim analysis of the data on the first 30 patients will be done after completion of their 90-day assessments. This analysis will be done on the 30 patients as a single cohort, and allocation to treatment arm (open vs robotic) will not be revealed. The purpose of this analysis is to adjust the power calculation of the iROC trial, but it will also provide data trends for secondary outcomes collected in the trial that are relevant to this doctoral thesis.

Additionally, baseline and peri-operative recovery data collected for patients in the first year of recruitment will be analysed. This includes patient reported outcome measures, complications, step-count etc. For baseline data, correlations will be explored between daily step-count with objective measurements of health status (demographic data, cardiopulmonary exercise test, etc) and patient reported outcome measures (PROMs). For post-operative data, correlations between daily step-count and complications as well as 1-year survival will be explored. Detailed methodology for each section is provided before each results section in subsequent chapters.

4.2.7 Trial registration details

The iROC trial protocol received Health Research Authority: London- NRES Committee North East – Newcastle & North Tyneside 1 Research Ethics Committee approval on the 18th of January 2017 (IRAS project ID: 211187; REC reference: 16/NE/0418, **Error! Reference source not found.**). This trial is registered on clinicaltrials.gov (NCT03049410) and ISRCTN (ISRCTN13680280).

Subsequent amendments were submitted and approved and are included in **Error! Reference source not found.**, **Error! Reference source not found.** and **Error! Reference source not found.**. Their changes are reflected in the final trial protocol in **Error! Reference source not found.**.

Chapter 5 The iROC trial interim analysis and recruitment

5.1 Chapter summary

In this chapter, I will present the results of our pre-planned interim analysis, which was performed after the completion of 90-day follow-up of the first 30 patients. The purpose of this analysis is to test if recruitment for the iROC trial is feasible in the designated time duration. Further, to assess the feasibility and compliance with prospective sensor and QL data collection. The data collected remains blinded and will be presented in accordance with the CONSORT statement and reported descriptively. This chapter is a modified version of the manuscript published in European Urology[218], I am a joint-first author for this publication.

5.2 Pre-planned interim analysis and recruitment progress for the year

5.2.1 Introduction

As reported in section 2.3, all prior RCTs have been single institution studies, and have either been feasibility studies, closed before planned recruitment or designed to measure surrogate endpoints. The pooled dataset from prior trials only added up to 239 cases in total, which will be eclipsed by the iROC trial with its target recruitment of 320 patients. Of note, the recently completed RAZOR trial also has a sample size of 320 patients[36]. Furthermore, due to differences in PROM tools used by various studies, quality of life data could not be aggregated by prior studies.

When designing the iROC trial, an interim analysis was included when sufficient follow-up had been reached (90 days) for the first 30 patients recruited into the trial. It is important to test the recruitment rate of patients into the trial to understand if the target recruitment of 320 patients is achievable with the trial milestones and funding available. Additionally, it was to test data collection for both primary and secondary endpoints.

5.2.2 Methods

5.2.2.1 *Patient population*

Patients included in this analysis fit the inclusion and exclusion criteria described in Section 4.2.1. An interim analysis was planned as per the trial protocol (outlined in Section 4.2.6) after the first 30 randomised patients completed 90 days of follow up post-RC. The 30th patient randomised underwent RC on 2nd August 2017.

5.2.2.2 *Data extraction*

With the approval of the trial management group and the data monitoring committee, data from the electronic case report form (eCRF) was downloaded on 30th November 2017. Data was extracted from the following forms: baseline characteristics, intra-operative details, post-operative recovery and outpatient follow-up forms. Randomisation arm (ORC or iRARC) data and any data that could reveal randomisation arm was excluded from data extraction, so all analysis will be reported as a single consolidated cohort undergoing RC.

5.2.2.3 *Outcomes*

Complications data was collected as per the Clavien-Dindo classification described in section 1.4.2.3. Length of stay, re-admission, complications, primary care engagement (General Practice or Accident and Emergency department) and post-operative histology were all collected. Days alive and out of hospital at 90 days were calculated using the formula DAOH = 90 – index admission days – re-admission days. Primary care engagement is not included in the DAOH metric which accounts for hospital admissions.

5.2.3 Results

5.2.3.1 *Patient characteristics*

The CONSORT diagram of patients recruited into the iROC trial is shown in Figure 5-1, and patient demographics and baseline clinicopathological characteristics are summarised in Table 5-1. In total, 51 patients were screened for eligibility, 38 were approached for consent and 36 consented to the iROC trial. Six patients were withdrawn from the iROC trial prior to randomisation, so a total of 36 patients were recruited before 30 were successfully randomised. Of the 6 patients who were withdrawn prior to randomisation, two were deemed to be screening failures (recruited despite failing to meet inclusion & exclusion criteria stated in 4.2.2), and four patients were subsequently withdrawn for the following reasons: one due to patient preference for radical radiotherapy instead of RC, two due to inoperable disease being identified at the time of pre-operative examination under anaesthesia and one being subsequently unfit for surgery due to co-morbidities.

An additional patient was identified to have metastatic disease in theatre at time of RC, and diversion was performed as per the randomisation arm but no cystectomy was performed. This case was discussed with the trial management group, and the consensus was to include this patient in the interim analysis.

The overall recruitment sample was predominantly male – 73.3%. As expected, the commonest histology type was UCC 26/30 (86.7%) followed by SCC 2/30 (6.7%) and other non-adenocarcinoma histology 2/30 (6.7%). Three and sixteen (total: nineteen) patients were current or ex-smokers respectively, making up 63% of 30 patients consented into the trial. Five patients were due to undergo a continent diversion (orthoptic neobladder) formation, and the remaining 25 patients were due to undergo ileal conduit formation.

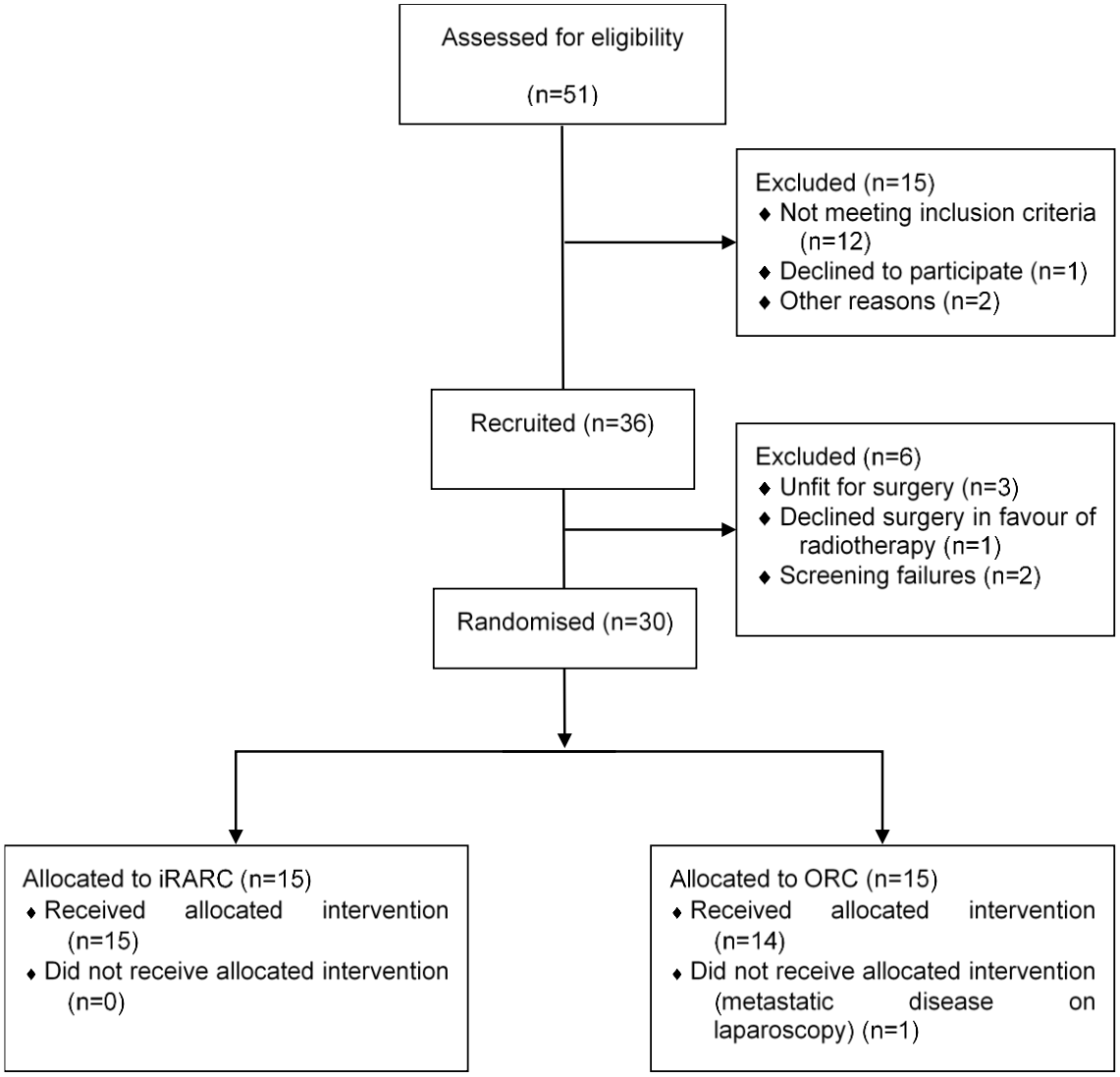


Figure 5-1: CONSORT diagram of patients consented for the iROC trial (interim analysis)

Gender		Male (%)	22 (73.3)
		Female (%)	8 (26.7)
Age		Median (IQR)	69 (61.0-75.75)
ECOG		0 (%)	4 (13.3)
		1 (%)	17 (56.7)
		2 (%)	8 (26.7)
		3 (%)	1 (3.3)
Neoadjuvant	**Chemotherapy**	n (%)	15 (50.0)
	Immunotherapy	n (%)	1 (3.3)
Histology		UCC (%)	26 (86.7)
		SCC (%)	2 (6.7)
		Adenocarcinoma	-
		Other (%)	2 (6.7)
BMI		Median (IQR)	26.6 (25.3-29.6)
Smoking		Current smoker (%)	3 (10)
		Ex-smoker (%)	16 (53)
		Non-smoker (%)	11 (37)
Diversion type		Ileal Conduit	25 (83.3)
		Neobladder	5 (16.7)

Table 5-1: Baseline characteristics of patients recruited for the iROC trial

5.2.3.2 *Surgical and Index admission outcomes*

Index admission outcomes of all 29 patients that underwent RC as part of the interim analysis cohort are summarised in Table 5-2. One intra-operative injury was identified – rectal injury that was repaired intra-operatively. Surgery times and intra-operative transfusion rates are not included for analysis, due to the risk of unblinding. All patients underwent the type of procedure they were randomised to receive. Re-interventions were defined as patients being taken back to the operating theatre or interventional radiology for a procedure under local or general anaesthetic. Only one patient underwent a re-intervention – for a reduction of an incisional hernia under general anaesthetic. Four patients had blood transfusions during their post-operative hospital stay, with two receiving one unit and another two receiving two units.

Length of stay *median (IQR)*		10 (6-15)
Intra-operative injury *n (%)*		1 (3.4)
Re-intervention *n (%)*		1 (3.4)
Post-operative transfusions	1 unit *n (%)*	2 (6.9)
	2 units *n (%)*	2 (6.9)
Tumour stage	≤pT1 *n (%)*	16 (55)
	pT2 *n (%)*	9 (31)
	pT3 *n (%)*	2 (6.9)
	pT4 *n (%)*	2 (6.9)
Node positive	N+ *n (%)*	3 (10)
Positive surgical margins *n (%)*		1 (3.4)
Lymph Node count *median (IQR)*		13.5 (10.25-18.25)

Table 5-2: Surgical and index admission outcomes collected as part of the iROC trial interim analysis

Of the patients who underwent RC, 16 (55%) had ≤T1 disease, and 9 (31%), 2 (7%) and 2 patients had T2, T3 and T4 disease based on post-operative histology. One patient had a positive surgical margin, and three patients had node positive disease – all four patients had MIBC.

5.2.3.3 *Peri-operative outcomes*

At 30 days post-operatively, no post-operative complications were observed in 10 (34%) of the 29 patients, with 17 (59%) patients having minor complications (CD≤2) and 2 (6.9%) patients experiencing major complications (CD≥3). At 90 days, no post-operative complications were observed in 10 (34%) of the 29 patients, with 15 (52%) patients having minor complications

(CD≤2) and 4 (14%) patients experiencing major complications (CD≥3), the full summary of Clavien-Dindo Classifications is provided in Table 5-3. A total of 20 (69%) of patients accessed primary care (defined as GP and A&E visits) during the 90 days post-operatively, but only 5 (17%) patients were readmitted to hospital in this period. The median days alive and out of hospital was 80 days (IQR 71.75-83 days). Conversely, this means that the median days in hospital was 10 (90 – DAOH), which is the same duration as the median length of stay.

Primary care engagement *n (%)*		20 (69)
Number of re-admissions *n (%)*		5 (17)
DAOH at 90 days *median (IQR)*		80 (71.75-83)
30-day CD classification	0 *n (%)*	10 (34.5)
	1 *n (%)*	5 (17.2)
	2 *n (%)*	12 (41.3)
	3 *n (%)*	2 (6.9)
	4 *n (%)*	0 (0)
	5 *n (%)*	0 (0)
90-day CD classification	0 *n (%)*	10 (34.5)
	1 *n (%)*	3 (10.3)
	2 *n (%)*	12 (41.3)
	3 *n (%)*	3 (10.3)
	4 *n (%)*	1 (3.4)
	5 *n (%)*	0 (0)

Table 5-3: Utilisation of healthcare services after primary discharge from hospital, Clavien-Dindo classification complications and days alive and out of hospital at 90 days.

5.2.3.4 *Secondary outcomes*

5.2.3.4.1 Baseline measures

Secondary outcomes for the iROC trial were to measure recovery using PROM questionnaires and activity trackers. Baseline compliance for each measure varied from 22/28 (79%) for activity trackers, 24/28 (86%) for CST30, 27/28 (96%) for WHODAS 2.0, 27/28 (96%) for QLQ-C30, to 28/28 (100%) for EQ-5D-5L. Two patients were not asked to complete the baseline measures (deviation from trial protocol logged). Four patients refused CST30. Two trackers could not be successfully synced to retrieve step-count, and four additional patients were not issued trackers due to supply issues at the respective sites. The observed values (Figure 5-2) for the first 30

patients in the iROC trial at baseline were similar to comparable populations. The WHODAS 2.0 score (average 15%) was approximately at the 78th percentile of general population[219], CTS30 (average 13) was similar to that for >65 year old males and >60 year old females[213], and step-count (average 5,750 steps) were slightly reduced compared to an a Canadian cohort of men and women who walked 7,869 and 6,970 steps/day, respectively [220]).

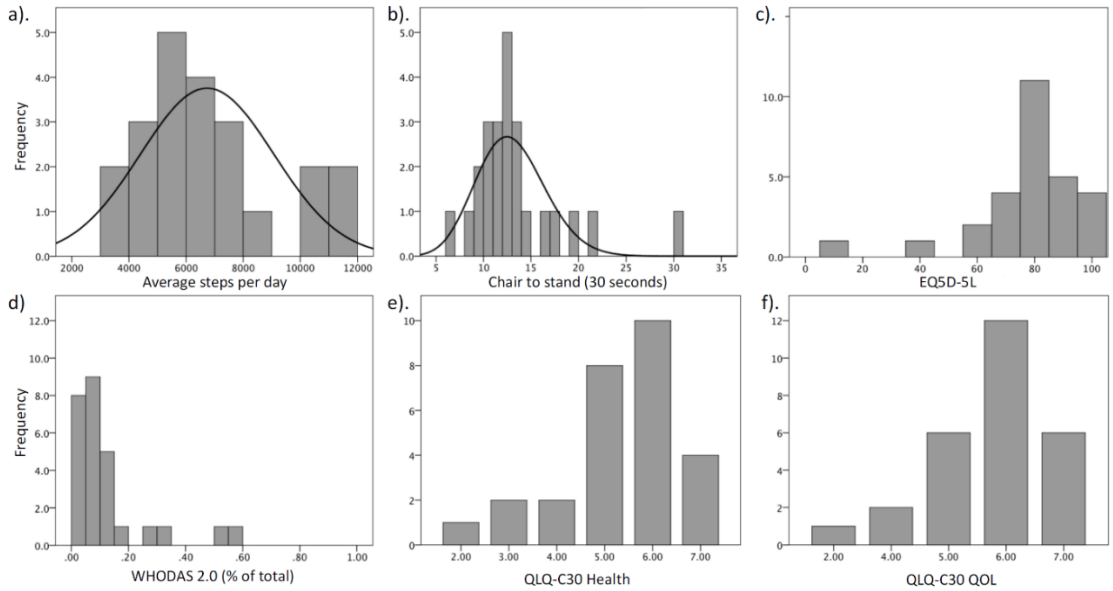


Figure 5-2: Baseline distribution of multimodal metrics in patients undergoing radical cystectomy.

5.2.3.4.2 Post-operative recovery

As expected, each measure deteriorated after surgery (Figure 5-3). At day 5 (POD5) the average number of daily steps was 1840 ± 1348 (32±22% of baseline) and CTS30 was 8.3±5.3 (62.0±38% baseline). Activities levels improved such that by week 5 walking reached 74±32% of the baseline (4294±2370 steps/day) and CTS30 reached 96±35% baseline (12±4.3/30 seconds). By week 12 many patients had returned to their baseline level of activity (average steps/day 6375±3246, 99±47% baseline and CTS30 13±5, 108±33%). Patient reported qualitative disability scores contrasted activity levels. At week 5, WHODAS 2.0 disability reached 26±22% (which was 2.9±3.3 fold higher than at baseline), before returning to pre-operative levels in most patients by week

12 (0.9±1.1 fold baseline). Changes in EQ-5D-5L scores rating 'health today'(Q6) and QLQ-C30 (Q29: overall health and Q30: QOL in past week) questionnaires mirrored activity levels with lower scores in week 5 (EQ-5D-5L 84±17%, QLQ-C30(Q29) 80±22% and QLQ-C30(Q30) 78±23% of baseline) that recovered to baseline by week 12 (93±17%, 98±16% and 93±16%, respectively). Patients seeking medical review after discharge (GP, A&E or hospital admission) averaged fewer daily steps at week 5 (medical review: 4069±2526 vs. no review: 4743±2132) and week 12 (5535±1786 vs. 6724±3703), and had lower absolute CTS30 numbers at the same times (week 5: 11.2±4.3 vs. 13.0±4.4 and week 12: 13.2±5.5 vs. 13.5±3.1), although the low sample size precluded meaningful statistical comparison. We hypothesised that multiple domains are needed to robustly measure recovery after RC and that accurate measurement will allow a meaningful comparison between open RC and RARC. Correlation of baseline data revealed no significant associations between measures of activity, qualitative disability or QOL data (Pearson correlation all $p>0.08$). Average daily steps did not correlate with CTS30 ($r=-0.08$, $p=0.7$ in 20 patients) and was closest to the QLQ-C30 quality of life domain ($r=0.41$, $p=0.08$). In this small sample size, one could hypothesise that average daily steps quantify actual activity whilst CTS30 is a measure of lower limb strength and exercise capacity (which may not be used).

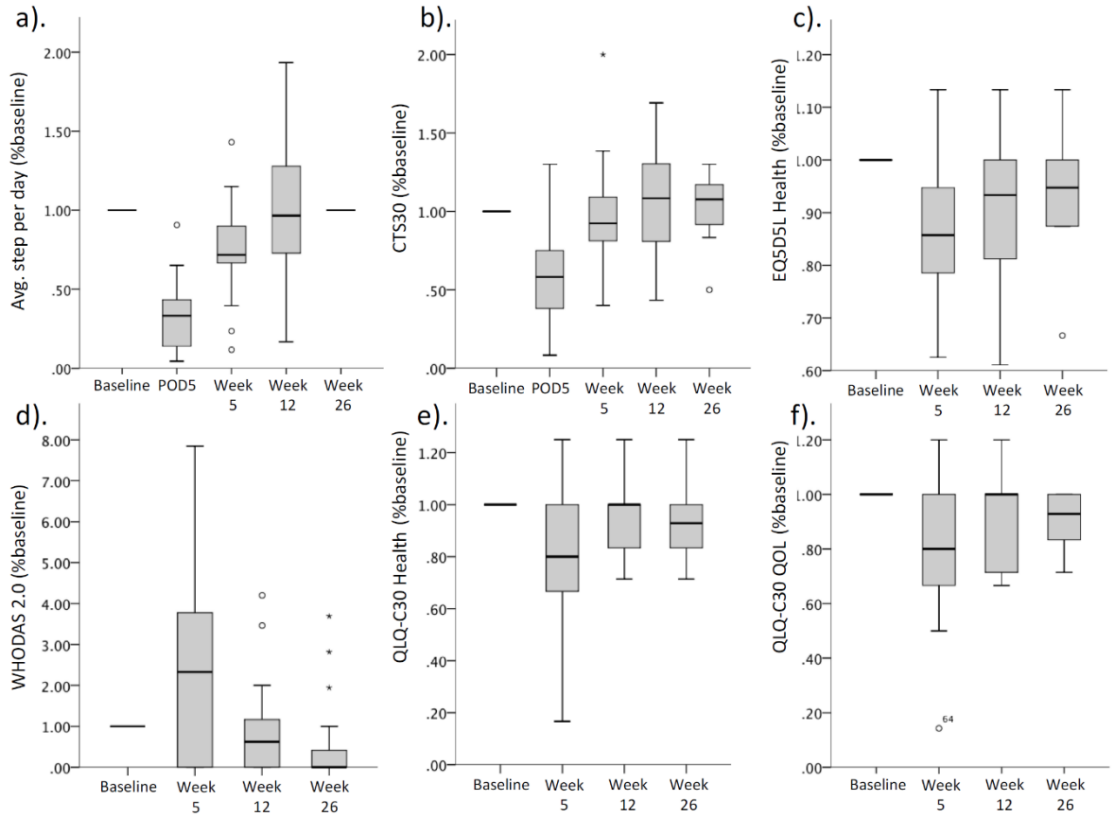


Figure 5-3: Changes in quantified activity levels, exercise capacity and patient reported disability and

5.2.3.5 *Feasibility & Recruitment in the first year of recruitment*

The interim analysis of the first 30 patients was completed from patients recruited at two NHS sites. However, for the trial to be completed on schedule, it was important to observe recruitment trends over a longer period. Trial progress was monitored through pre-defined checkpoints agreed with the funder before starting the trial in March 2017. In our agreed recruitment targets with the funder (The Urology Foundation), our recruitment milestone was set as 50 patients and open in three sites in the UK by the end of Month 12 (February 2018).

By the end of Month 12, the iROC trial opened in five different UK teaching hospitals:

1) University College London Hospital
2) Sheffield Teaching Hospitals NHS Foundation Trust
3) North Bristol NHS Trust
4) Royal Berkshire Hospital

5) Guys and St Thomas' NHS Foundation Trust.

At the end of 12 months, 142 patients were screened, and 93 patients recruited – 86% ahead of recruitment milestone for the Month 12 timepoint. The expected and actual recruitment curves are displayed in Figure 5-4. A total of 18 patients were recruited into the iROC trial who were not eventually undergo RC as part of the trial, resulting in 75 patients being randomised in the first year of the trial. Most patients were acceptable to their allocated treatment arm, and complied with trial procedures until the completion of the primary endpoint measure. In total, only two (2.1%) patients exited the trial after randomisation before surgery. An additional eight (8.6%) patients exited the trial early due to having to travel long distances for follow-up. As a result of this, the trial protocol was amended to allow for telephonic follow up to facilitate partial data collection in patients who wished to exit the trial early.

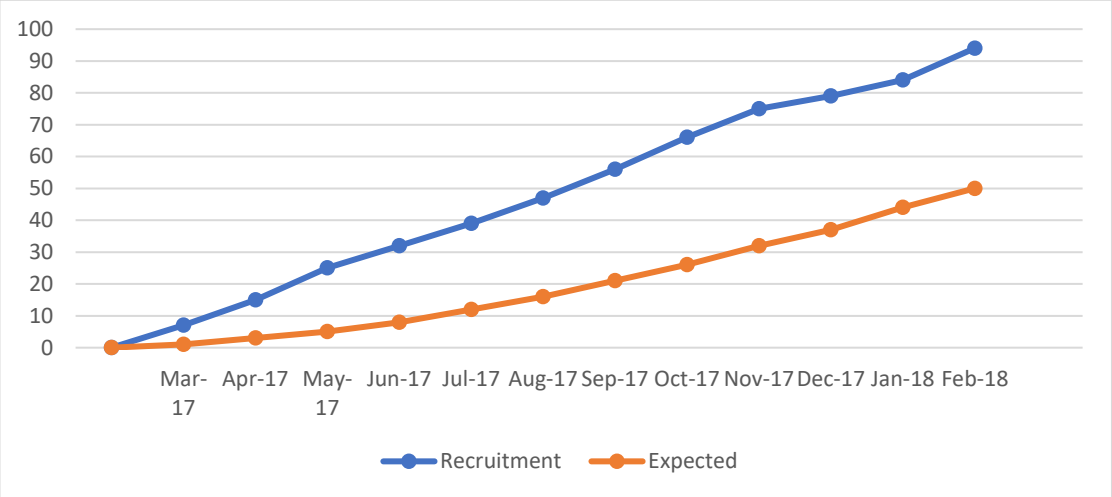


Figure 5-4: Recruitment progress for the iROC trial

5.2.4 Discussion

Recruitment into the iROC trial during the first year was much faster than initially projected. This was largely due to the trial opening successfully at the initial two sites, and subsequent opening of additional sites earlier than projected when milestones were being discussed with the funder. However, there was a high attrition rate after patients were recruited into the trial before randomisation was performed (18/93 patients). An internal analysis highlighted that this was due to limitations of the clinical pathways in some centres leaving a narrow window for recruitment, often before the treatment plan for patients have been finalised. Despite this limitation, 75 patients were randomised in the first year which is still 50% higher than the target of 50 patients.

The overall strategy for recruitment in the iROC trial was to open initially at University College London Hospital NHS Trust and Sheffield University Hospital NHS Trust, as members of the TMG were principle investigators at both sites and feedback could be collected easily to improve the standard operating procedures (SOPs) and frequently asked questions (FAQs) documentation to enable for easier rollout to future sites. These SOPs and FAQs were then tested when the trial was formally opened at Guys & St Thomas Hospitals NHS Trust, and further iterative changes were made to reference documents. After the opening of the 5th site (Royal Berkshire Hospital), the TMG meeting was opened up to research staff at all hospitals to dial in and troubleshoot problems. This was instrumental in preparing the three subsequent substantial amendments to make trial assessments easier to deliver in different hospitals with different clinical pathways.

Surgical and post-operative outcomes were similar to those reported in contemporary literature. The national cystectomy audit data showed a length of stay of 8 and 11 days for RARC and ORC in high volume centres respectively[31], which is similar to a median of 10 days (IQR 6-15 days) reported in our interim analysis. Major complication rate in iROC (14.7% CD≥3) was higher than the 8.5% rate for high volume centres nationally, but this could be attributed to our

small sample size for the interim analysis. The trial statistician submits data regularly to the data monitoring committee to ensure that morbidity, mortality or any other aspect of patient safety are not compromised in either arm of the trial. This data is not available to members of the trial management group.

While primary outcome data was collected successfully for 100% of patients, secondary outcome data collection was mixed. Questionnaires data was well-collected (96-100%), but tracker data and CST30 collection was at 79% and 86% respectively. CST30 data was incomplete due to patient preference not to perform the test. Tracker data was not available due to initial supply issues, and technical issues with the Fitbit app not syncing data across successfully. More devices were purchases and a feedback system was set in place to ensure that sites are always well-stocked with devices to give patients during enrolment and follow-up appointments. We contacted Misfit about an alternative way to collect data from devices that may be more reliable, but the suggestion they provided was discontinued shortly after (the Misfit Link app).

Patient reported quality of life and health scores suggested a similar trajectory for patient recovery than objective mobility measurement from wearable devices. By five weeks post-operatively, PROM scores and mobility were both below baseline levels (WHODAS 2.0 score was higher as it is a disability score), but were closer to baseline by three months post-operatively. This suggested that an aspect of recovery could be collected passively from wearable devices. As uptake of smartphones and wearable devices increases in healthcare, this will allow wearable data to be collected continuously instead of seven-day snapshots. Current smartphone paired trackers are capable of collecting heart rate and even ECG data as discussed in section 2.2, but these are expensive and require regular charging by patients. When these limitations have been overcome, wearable devices could serve as a regular stream of data which could inform clinicians about patient's return to baseline function.

5.3 Chapter Conclusions

The interim analysis of the iROC trial shows that completing recruitment in the three-year period agreed with the funder is feasible. In the first year, the trial recruited ahead of target and this trend is expected to continue as we continue to open more sites if the current sites continue to recruit at their initial pace. Peri-operative outcomes reported are similar to those reported in the national cystectomy audit data, but it was not possible to compare data by treatment arm as all analysis in this chapter and all subsequent chapters was presented single arm. The randomisation result will only be available for analysis after the recruitment of all patients into the trial.

Data collected from secondary outcomes provides information about recovery beyond length of stay and complications. The health-related quality of life questionnaires will allow for measurement of health status at various time points during the peri-operative and post-operative period, but also provide a comparator for mobility data collected using wearable devices. Furthermore, this can be contextualised with re-admission and complication data to better understand the trajectory of patients with different paces of recovery.

In addition to the data reported in this chapter, the results of the interim analysis were used to adjust the power calculation for the trial. Our initial recruitment target was set to be 320 patients, and this was changed to 340 patients following the interim analysis. The details of the statistical methods are included in the full trial protocol included in the **Error! Reference source not found.**.

Chapter 6 Mobility and quality of life during the perioperative period

6.1 Chapter summary

This chapter will describe the trends in mobility and quality of life for patients enrolled into the iROC trial. I will first describe the baseline mobility data collected from wearable devices, and then explore associations with other baseline variables collected such as age, BMI and performance status. These baseline metrics are part of the standard clinical pathway for pre-operative assessments across all NHS hospitals. Additionally, I will report mobility trends for patients during the pre and peri-operative periods up to 90 days following RC.

In the second part of the chapter, I will build on the pilot work presented in section 3.3 by reporting the results of QoL questionnaires. Unlike section 3.3, only the EORTC QLQ-C30 questionnaire will be used as it offers modularised quality of life domains. The EORTC QLQ-C30 scores at 5 weeks and 3 months timepoints will be reported. In particular, the patient-reported physical functioning scale and global health status will be compared with wearable device-derived mobility data collected at the same timepoints.

6.2 The role of wearable devices in measuring patient mobility at baseline and during the peri-operative period following radical cystectomy in the iROC trial

6.2.1 Introduction

Mobility trends in patients with bladder cancer have not been reported in the literature, as evidenced by my systematic review in section 2.2. In this chapter I will provide an overview of mobility data collected from patients at baseline, alongside metrics traditionally used to assess pre-operative fitness for surgery. Associations between mobility data and 'clinical' data collected will be explored. Additionally, I will discuss mobility trends in the first three months following surgery. Whilst mobility trends were described briefly in Chapter 5 for the first 30 patients, this section will present mobility data for a larger cohort of patients. Similar to previous analysis, all data is described as single arm and blinded to the arm of the study patients were in (open vs robotic).

The aim of this section is to describe step-count as a variable in terms of its descriptive statistics so that the most appropriate statistical tests can be applied to the data collected. Additionally, correlations and comparisons with baseline demographic and clinical data are explored to understand how these variables affect step-count.

6.2.2 Methods

6.2.2.1 *Sample*

Patients undergoing radical cystectomy as part of the iROC trial were consented to complete the EORTC-QLQ C30 questionnaire and wear a wrist-worn wearable device at timepoints specified in Table 4-1.

6.2.2.2 *Data collection*

6.2.2.2.1 Patient demographic data

At time of recruitment, patient sex, age, BMI, urinary diversion type and ECOG performance status were collected.

6.2.2.2.2 Fitness tracking data

The Misfit Ray wrist-worn wearable device (described in section 0 and pictured in Figure 3-1) was provided for patients to wear for seven consecutive days at three timepoints: baseline, 5 days, 5 weeks and 3 months post-operatively. Patients were provided with the trackers during their hospital clinic appointments, and were also issued pre-stamped envelopes to return the wearable devices to a central receiving laboratory. Upon receiving the wearable devices, daily step-count data was extracted using the iOS Misfit mobile application.

6.2.2.3 *Statistical analysis*

Statistical analysis was performed in SPSS version 25.0.

All analysis described is blinded to the arm of the trial, so data will be presented as single arm. Patient characteristics of the cohort are reported using descriptive statistics. Three metrics were computed from the step-count data extracted from each wearable device: average step-count (ASC), maximum step-count (MSC) and minimum step-count (MiSC). Maximum and minimum step-counts were taken as the highest and lowest total steps in a day for the 7-day study duration respectively. Relative recovery of mobility was calculated as 3-month average divided by baseline average for each patient.

Kolmogorov-Smirnova test was used to assess the distribution of variables. Spearman rho was used to assess associations between non-parametrically distributed variables. Mann-Whitney U test was performed to compare non-parametrically distributed independent variables.

6.2.3 Results

197 patients had completed three-month follow-up at the time of data download from the electronic case report form (eCRF), and 179 patients provided data at any timepoint. Randomisation in the trial was performed accounting for diversion type: 150 and 24 patients respectively were randomised in the ileal conduit and neobladder groups respectively. The baseline characteristics of included patients are described in Table 6-1.

Gender	Male (%)	142 (79.3)
	Female (%)	37 (20.6)
Age	Median (IQR)	70 (63-75)
ECOG	0 (%)	140 (78.2)
	1 (%)	30 (16.8)
	2 (%)	6 (3.4)
	3 (%)	3 (1.7)
Diversion type	Ileal Conduit *n(%)*	151 (84.4)
	Neobladder *n(%)*	25 (14.0)
	No cystectomy performed	3 (1.7)
Neoadjuvant	Chemotherapy *n (%)*	59 (33.0)
	Immunotherapy *n (%)*	21 (11.7)
Histology	UCC (%)	158 (88.3)
	SCC (%)	9 (5.0)
	Adenocarcinoma	4 (2.2)
	Other (%)	8 (4.5)
BMI	Median (IQR)	26.7 (24.2-30.1)
Smoking	Current smoker (%)	20 (11.2)
	Ex-smoker (%)	103 (57.5)
	Non-smoker (%)	56 (31.3)

Table 6-1: Baseline characteristics of patients who provided fitness tracking data in the iROC trial

In total, 143 (79.9%), 132 (73.7%), 124 (69.2%) and 106 (59.2%) patients provided step-count data at baseline, post-operatively at day 5, 1 month and 3 months respectively. Table 6-2 shows the results of the Kolmogorov-Smirnova test of normality of daily step-count collected at each of the four time points (baseline, day 5, 5 weeks, 12 weeks) and combined across all four time periods. The p-value for all five test results was <0.001, strongly suggesting that step-count data in this patient group is not normally distributed. The histograms visually representing this data are presented in Supplementary Figure 10-1 to Supplementary Figure 10-5. All further analysis in using step-count data will use non-parametric tests.

Figure 6-1 displays the box and whisker plots for the three step-count derived metrics at the four different timepoints. Across all three metrics, a similar trend was observed: patients have reduced step-count in the immediate post-operative period, and this increases in the first three months following RC. Median values for average step-count at baseline, 1 week, 1 month and 3 months were 5821, 1525, 3819 and 5774 respectively. In total 86 patients provided tracking data at baseline and the three-month timepoint and relative recovery of mobility was calculated. In this patient group, 26 (30.2%) patients achieved 100% or more of their baseline activity, 46 (53.5%) patients were between 50-100% of their baseline and 14 (16.3%) patients were below 50% of their baseline activity.

Baseline characteristics were then compared with other variables collected at baseline (age, BMI, smoking status, gender, chemotherapy, immunotherapy). Correlations for continuous variables are presented in Table 6-3. Age was inversely correlated with maximum step (p=0.037, coeff=-0.162), but no other significant correlations were identified. To explore the differences of these variables in the extreme groups, bounds of the upper (75-100) and lower (0-25) quartiles of these variables were used to as cut-off values for further analysis as categorical variables. None of the step-count variables were significantly different in patients with higher

BMI, but average (p=0.041, z=-2.04) and maximum step-count (p=0.008, z=-2.655) were significantly higher in younger patients.

Furthermore, no significant difference was detected between step-count in patients grouped by smoking status, chemotherapy or immunotherapy. Male patients had a significantly higher maximum step-count than female step-count, but this was not the case for average or minimum steps. These results are summarised in Table 6-4. Difference for extremes in ECOG performance status (score 0 vs 3) were not explored due to insufficient patients (n=3 or 1.7%) having an ECOG of 3.

Kolmogorov-Smirnova test			
	Statistic	df	Sig.
Baseline	0.065	1098	<0.001
Week 1	0.127	961	<0.001
Week 5	0.104	848	<0.001
Week 12	0.065	703	<0.001
Combined	0.100	3610	<0.001

Table 6-2: Kolmogorov-Smirnov test of normality

		Average steps	Maximum steps	Minimum steps
Age	Correlation Coefficient	-0.114	-0.162	-0.029
	Sig. (2-tailed)	0.143	**0.037**	0.711
BMI	Correlation Coefficient	-0.118	-0.139	-0.073
	Sig. (2-tailed)	0.148	0.087	0.373

Table 6-3: Spearman's rho correlation for continuous baseline variables

		Average steps	Maximum steps	Minimum steps
BMI	Z	-1.651	-1.537	-1.191
	Sig. (2-tailed)	0.099	0.124	0.234
Age	Z	-2.04	-2.655	-0.547
	Sig. (2-tailed)	**0.041**	**0.008**	0.585
Ever smoked	Z	-1.624	-1.418	-2.08
	Sig. (2-tailed)	0.104	0.156	0.038
Current smoker	Z	-0.407	-0.283	-1.203
	Sig. (2-tailed)	0.684	0.777	0.229
Gender	Z	-1.669	-2.173	-1.355
	Sig. (2-tailed)	0.095	**0.03**	0.175
Chemotherapy	Z	-1.24	-0.903	-1.213
	Sig. (2-tailed)	0.215	0.367	0.225
Immunotherapy	Z	-1.607	-1.357	-1.079
	Sig. (2-tailed)	0.108	0.175	0.281

Table 6-4: Mann-Whitney U test results comparing step-count for different baseline variables

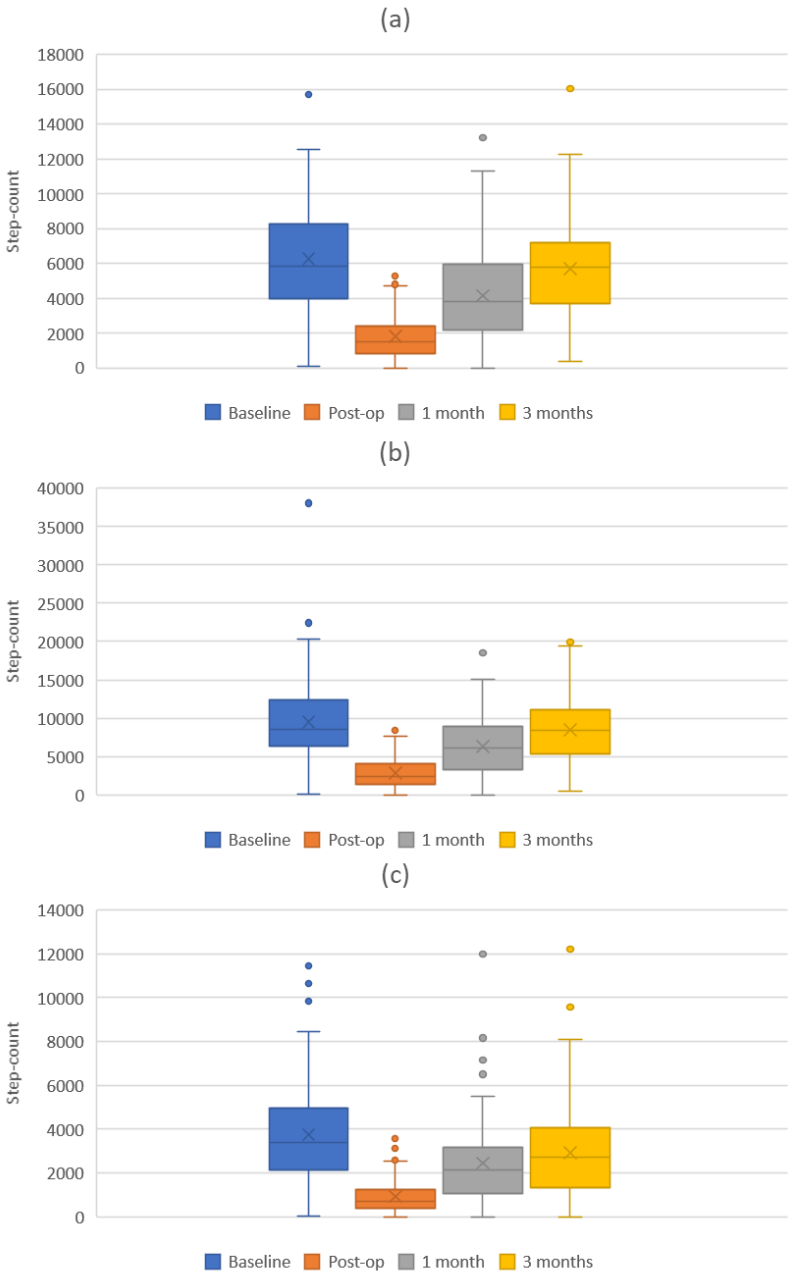


Figure 6-1:Box and whisker plots for baseline, post-op (5 days), 1 month and 3 months using the three different step-count based metrics derived from wearable devices. (a) average, (b) maximum steps, (c) minimum steps

6.2.4 Discussion

The three different metrics of average steps, maximum steps and minimum steps were included separately instead of total steps in our analysis. Average steps were used as a direct replacement of total steps to include patients who provided less than 7 days of data (to a minimum of 4 days of data). Maximum steps represent a patient's most active day during the study period, and are likely to be more reflective of an individual's exercise capacity than average or total activity. Lastly, minimum steps were included as it represents the patient's lowest possible activity required to complete daily tasks.

No difference in activity levels was identified for patients divided into chemotherapy and immunotherapy groups when compared with patients who had no neoadjuvant treatment. As patients provided fitness tracking data during their pre-surgery appointment, which could be scheduled any time prior to starting neoadjuvant chemotherapy up until after the completion of the final cycle of treatment, this could have contributed to the lack of a difference being identified. A future study is under development to explore activity levels during chemotherapy, particularly since newer wearable devices offer longer data storage, as well as monitoring of other metrics such as heart rate, sleep etc.

Surprisingly, smoking status did not have any effect on activity levels when comparing current smokers as well as current or ex-smokers vs non-smokers. A recent publication by Lee *et al*.[221] reported step volume (step-count) in 17,466 women over the age of 45 for seven consecutive days and found that patients in the lowest activity quartile were more likely to be smokers when compared with the highest activity quartile. While this result contradicts our findings, this could be attributed to having a different cohort of patients: any women over the age of 45 vs patients undergoing major pelvic surgery for bladder cancer. Additionally, our patient group consists of patients who have been deemed fit to undergo major surgery, and there may be a selection bias

for patients with a higher physiological reserve. Lastly, this could be attributed to sample size and this analysis will be repeated once the iROC trial has been completed.

As the wearable device sub-study was included in the trial as an exploratory objective, compliance with wearable device was not as high as the primary objective data completion (100%). Additionally, the higher attrition rate at 3 months can be partially attributed to the delay in trackers being returned to the central receiving lab for analysis for patients that recently attended their 3-month follow-up appointment.

Only 30% (26/86) of patients surpassed their baseline mobility at 3 months after RC. When patients are counselled prior to RC, a complete recovery time between 6 weeks and 12 weeks is quoted[222–224]. Whilst other domains of recovery than mobility exist, this data suggests that mobility does not return to baseline within 3 months following surgery. Furthermore, the baseline data captured in the trial is when patients have been diagnosed with cancer, and this step-count data could be lower than their true disease-free baseline. It would be of interest to analyse patient activity levels at the 6 months and 12 months post RC to understand when majority of patients return to true baseline activity levels, particularly older patients and those who experience major or prolonged complications. Recovery in other domains will be explored in section 6.3 using patient reported outcome measures (PROMs).

While our current study did not explore ECOG as a predictor of performance status, Gresham *et al*.[225] have previously reported correlations between performance status and daily step-count. However, their cohort of patients consisted mostly of patients with inoperable stage 4 disease. In contrast, our cohort consists entirely of patients who have lower stage disease and have been determined to be fit for surgery. After the completion of the iROC trial, it would be useful to analyse step-count in the ECOG performance status groups once sufficient numbers have been recruited into each ECOG group.

6.2.5 Conclusions

Wearable devices offer an opportunity to collect objectively measured mobility data in an unsupervised environment (such as patient homes). Surprisingly, only age and gender correlated with activity levels while smoking status, BMI and neoadjuvant treatment did not. This could be due to selection bias, as patients recruited into the study were already deemed fit for surgery and are therefore more likely to be more physically active. Data over the three-month peri-operative period also enables the assessment of patient recovery to baseline mobility. While post-operative step-count data is described in this section, the relationship of this data with clinical outcomes has not been explored. This work will be presented in chapters 7 and 8, relating specifically to peri-operative complications and 1-year outcomes respectively.

6.3 Comparing quality of life measured by the EORTC QLQ-C30 questionnaire and mobility measured by wearable devices during the peri-operative period following radical cystectomy

6.3.1 Introduction

As discussed in section 3.3, the EORTC QLQ-C30 correlates well with the WHODAS 2.0 in their assessment of global health and disability score status. However, the EORTC QLQ-C30 questionnaire offers modularised functional domains and symptom domains that the WHODAS 2.0 does not. In this section, data from the EORTC QLQ-C30 will be presented as the five functional scales (physical, role, emotional, cognitive and social functioning) at baseline and during the first three months of recovery.

Additionally, mobility data extracted from wearable devices from the same timepoints (baseline, 5 weeks and 3 months) will be compared with the physical functioning domain of the EORTC QLQ-C30 questionnaire. While the questionnaire scoring manual[207] does not provide definitions for what constitutes physical functioning, the questions are directed at measuring mobility. Questions 1-5 of the questionnaire make up the physical functioning score; the full EORTC QLQ-C30 questionnaire is included in **Error! Reference source not found.**.

The main aim of this section is to assess if objectively measured recovery of mobility using a wrist-worn wearable device correlates well with patient-reported recovery as measured by the EORTC QLQ-C30 questionnaire.

6.3.2 Methods

6.3.2.1 *Sample*

Patients undergoing radical cystectomy as part of the iROC trial were consented to complete the EORTC-QLQ C30 questionnaire and wear a wrist-worn wearable device at various timepoints.

6.3.2.2 *Data collection*

6.3.2.2.1 Demographic information

At time of recruitment, patient sex, age, BMI, diversion type and ECOG performance status was collected.

6.3.2.2.2 Quality of Life tool

The EORTC-QLQ C30 questionnaire[206] was given to patients at baseline, 1 month and 3 months post-operatively. This has been previously described in section 3.3.2.2.2.2. Patients were mailed the questionnaire with the confirmation of their clinic appointment, and asked to return the questionnaire in clinic. For patients who did not bring a completed questionnaire to their appointment, a new set was provided for the patient to complete before leaving clinic to improve compliance. Only patients who returned the completed questionnaire at baseline and one other timepoint are included in this analysis.

6.3.2.2.3 Wearable device

The Misfit Ray wrist-worn wearable device (described in section 0) was provided to patients to wear for seven consecutive days at four timepoints: baseline, 5 days, 5 weeks and 3 months post-operatively. Patients were provided with the trackers during their hospital clinic appointments and were also issued pre-stamped envelopes to return the wearable devices to a central receiving laboratory. Upon receiving the wearable devices, daily step-count data was extracted using the iOS Misfit mobile application.

6.3.2.3 *Statistical Analysis*

Patient characteristics of the cohort are reported using descriptive statistics. Completion rates for questionnaires at each timepoint are reported in percentages. As the EORTC QLQ-C30 questionnaire is modularised, scores were tabulated for individual domains of health and the global QoL score, and calculated using EORTC's scoring manual[207]. As the manual does not provide guidance on how to interpret scores or changes in score, absolute and relative changes in scores are computed.

As previously described, three metrics were computed from the step-count data extracted from each wearable device: average step-count, maximum step-count and minimum step-count. Maximum and minimum step-count were the highest and lowest total steps in a day for the 7-day study duration. Peri-operative complications were excluded from the data downloaded from the electronic case report form (eCRF) due to ongoing data monitoring for more recently recruited patients.

Mann-Whitney U test was performed to compare non-parametrically distributed independent variables, and Wilcoxon Signed Ranks test was performed to compare non-parametrically distributed dependent variables. Spearman correlation was performed to compare non-parametrically distributed variables.

All statistical analyses were carried out using SPSS for Windows version 25.0.

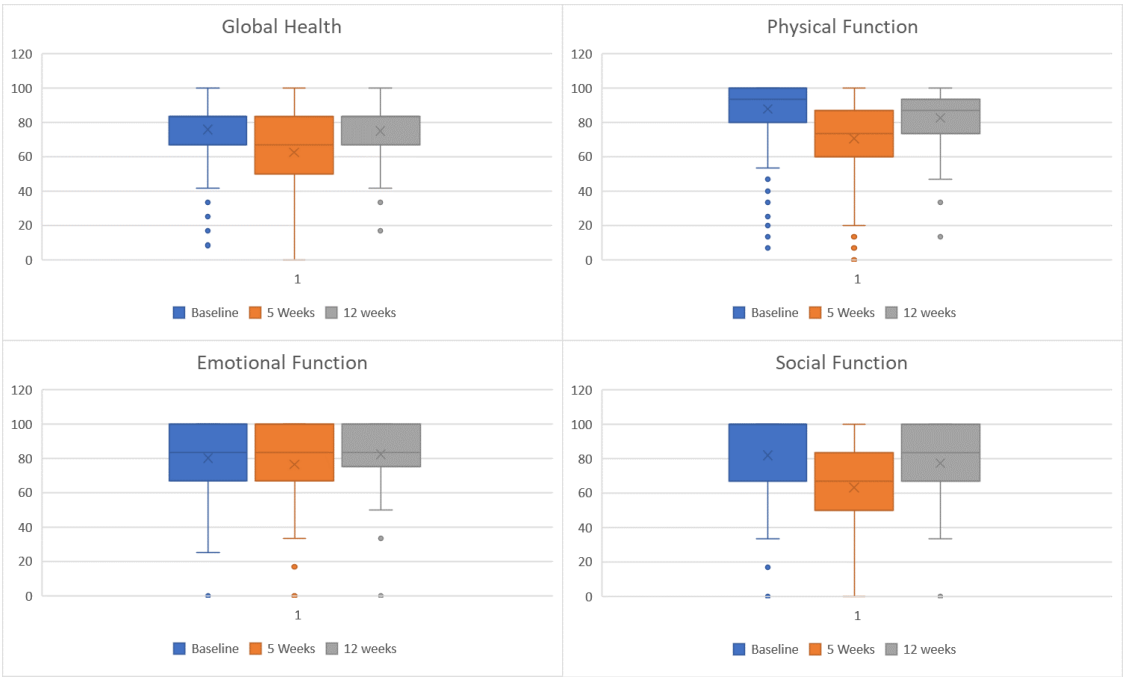
6.3.3 Results

At the time of the data download from the electronic Case Report Form (eCRF), 197 patients had completed three-month follow-up. 174 (88.3%) patients provided baseline data, and 134 (77.0%) and 113 (64.9%) of these patients returned their questionnaires at 5 weeks and 3 months respectively. Baseline characteristics of this patient group are summarised in Table 6-5. As reported in section 6.2.3, tracker return rates at these time points were 143 (79.9%) 124 (69.2%) and 106 (59.2%) respectively.

Gender	Male (%)	137 (78.7)
	Female (%)	37 (21.3)
Age	Median (IQR)	70 (63-75)
ECOG	0 (%)	135 (77.6)
	1 (%)	29 (16.7)
	2 (%)	7 (4.0)
	3 (%)	3 (1.7)
Diversion type	Ileal Conduit *n(%)*	152 (87.4)
	Neobladder *n(%)*	22 (12.6)
Neoadjuvant	Chemotherapy *n (%)*	60 (34.5)
	Immunotherapy *n (%)*	18 (10.3)
Histology	UCC (%)	157 (90.2)
	SCC (%)	10 (5.7)
	Adenocarcinoma	2 (1.1)
	Other (%)	5 (2.9)
BMI	Median (IQR)	26.7 (24.0-30.2)
Smoking	Current smoker (%)	17 (9.8)
	Ex-smoker (%)	99 (56.9)
	Non-smoker (%)	58 (33.3)

Table 6-5: Baseline characteristics of patients in the iROC trial who completed the EORTC QLQ-C30 questionnaires during the first 3 months post-RC

The five health domains and the global health score of the EORTC QLQ-C30 questionnaire during the peri-operative period following RC are represented as box and whisker plots in Figure 6-2. Additionally, a Wilcoxon signed-rank test was performed comparing the six health scores at 5 weeks and 3 months post-RC to baseline scores. These results are summarised in Table 6-6. At the 5 weeks post-RC timepoint, there was a significant decrease (significant correlation with negative coefficient) in scores for global health ($p<0.001$), physical function ($p<0.001$), role function ($p<0.001$), cognitive function ($p=0.025$) and social function ($p<0.001$), but no significant change in emotional function ($p=-0.055$) score when compared with the baseline scores. At the 3 months post-RC timepoint to baseline scores, there was no longer significant difference global health, cognitive function and social function scores. Difference in emotional function was still not significant. There is, however, still a significant difference in physical function ($p<0.001$) and role function ($p=0.030$) scores.



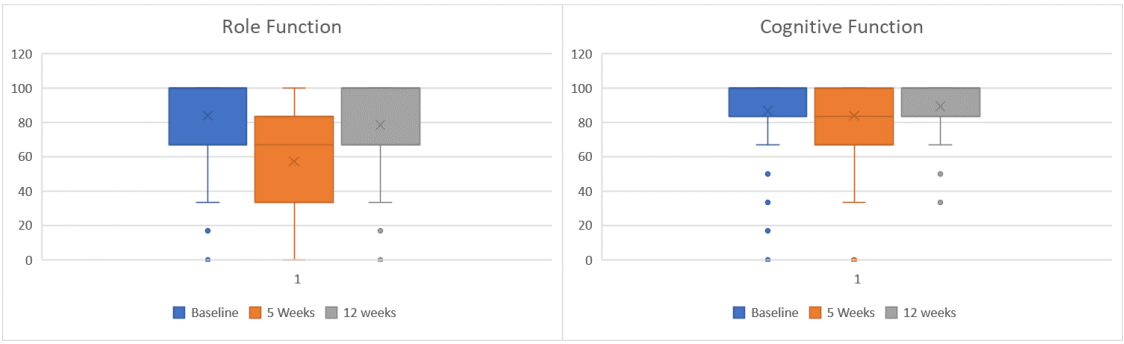


Figure 6-2: Box and whisker plots representing the six quality of life domains as measured by the EORTC QLQ-C30 questionnaire

As discussed in section 6.2.3, activity data patients have reduced step-count at 5 days post-operatively, with activity for the cohort increasing in the first three months following RC. Box and whisker plots representing this data are shown in Figure 6-1 in section 6.2.3. A Wilcoxon-signed rank test (Table 6-7) showed that there was a statistically significant change in mobility at all three post-operative timepoints when compared with mobility at baseline. The z-score (coefficient) at each timepoint becomes less negative, implying that the size of the difference in mobility is decreasing and returning to baseline levels with time. Of note, this result is in consistent with the change in physical function as measured by the physical function domain of the EORTC QLQ-C30. This implies that both wearable device measured mobility and the physical functioning domain of the EORTC QLQ-C30 questionnaire show similar trends in recovery of function.

Next, correlations between the change in physical function as measured by the EORTC QLQ-C30 and the mobility metrics as measured by the wearable devices before (baseline) and after cystectomy at the two different timepoints (5 weeks and 3 months) were assessed. Both physical function score and average steps at 1 month and 3 months were represented as fractions of each patient's baseline score. These results are represented as scatter plots in Figure 6-3 (1 month) and Figure 6-4 (3 months). Spearman correlation was performed on this dataset, and no correlation was observed between patient reported and objectively measured mobility at both

1 month (p=0.075) and 3 months (p=0.317). This finding suggests that the measurement of recovery to baseline according to the physical functioning domain of the EORTC QLQ-C30 questionnaire is different to the measurement of recovery to baseline according to step-count measured by the Misfit Ray wearable device.

	Baseline Median (IQR)	5 weeks Median (IQR)	3 months Median (IQR)	Test statistics		
					5 weeks	3 months
Global Health	83.3 (66.7-83.3)	66.7 (50.0-83.3)	83.3 (66.7-83.3)	**p-value**	**<0.001**	0.13
				Coefficient	-6.174	-1.513
Physical Function	93.3 (80.0-100)	73.3 (60.0-86.7)	86.7 (73.3-93.3)	**p-value**	**<0.001**	**<0.001**
				Coefficient	-7.865	-3.947
Role Function	100 (66.7-100)	66.7 (33.3-83.3)	100 (66.7-100)	**p-value**	**<0.001**	**0.030**
				Coefficient	-6.746	-2.164
Cognitive Function	100 (83.3-100)	83.3 (66.7-100)	100 (83.3-100)	**p-value**	**0.025**	0.464
				Coefficient	-2.239	-0.732
Emotional Function	83.3 (66.7-100)	83.3 (66.7-100)	83.3 (75.0-100)	**p-value**	0.055	0.212
				Coefficient	-1.920	-1.249
Social Function	100 (66.7-100)	66.7 (50.0-83.3)	83.3 (66.7-100)	**p-value**	**<0.001**	0.070
				Coefficient	-5.986	-1.812

Table 6-6: Wilcoxon signed-rank test results for the 6 functional domains of the EORTC QLQ-C30 questionnaire at 5 weeks and 3 months post-operatively compared with baseline

	Baseline Median (IQR)	1 week Median (IQR)	5 weeks Median (IQR)	3 months Median (IQR)	Test statistics			
						1 week	5 weeks	3 months
Average steps	5921 (3966-8203)	1525 (847-2409)	3529 (2152-5600)	5796 (3740-7219)	**p-value**	**<0.001**	**<0.001**	**0.006**
					Coefficient	-8.485	-6.207	-2.753
Maximum steps	8554 (6348-12436)	2401 (1405-4009)	5498 (3234-8571)	8422 (5452-11212)	**p-value**	**<0.001**	**<0.001**	**<0.021**
					Coefficient	-8.447	-6.164	-2.314
Minimum steps	3442 (2136-5046)	742 (394-1257)	2032 (1072-3162)	2688 (1272-3870)	**p-value**	**<0.001**	**<0.001**	**0.030**
					Coefficient	-8.323	-4.830	-2.591

Table 6-7: Wilcoxon signed-rank test results for the 3 step-count variables extracted from wearable devices worn at 1 week, 5 weeks and 3 months post operatively compared with baseline

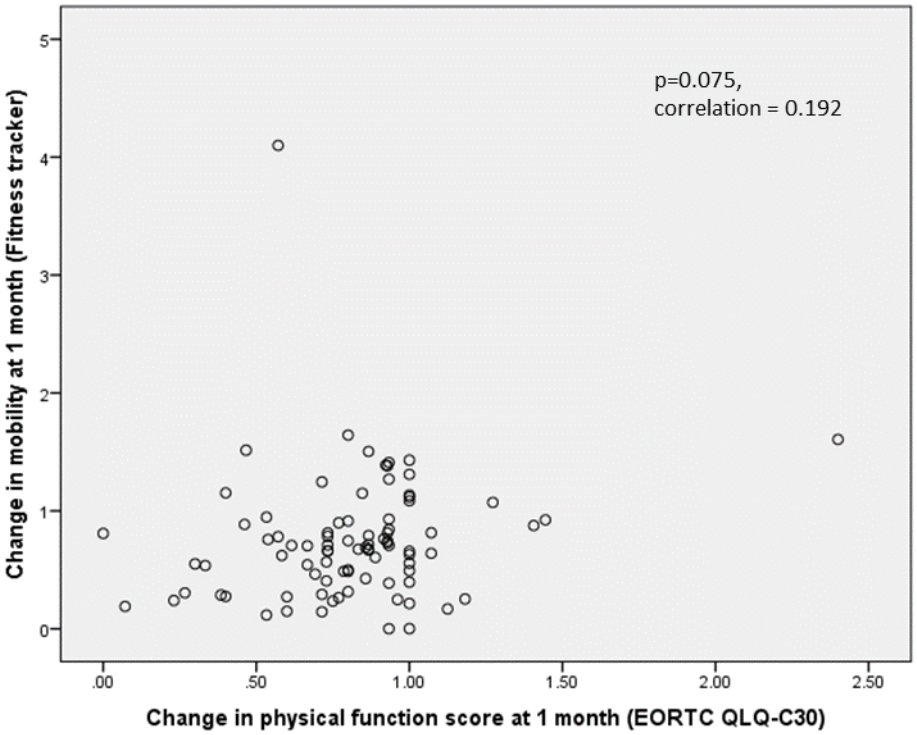


Figure 6-3: Scatter plot illustrating change (1 month/baseline) in mobility as measured by wearable device vs physical function score as measured by the EORTC QLQL-C30.

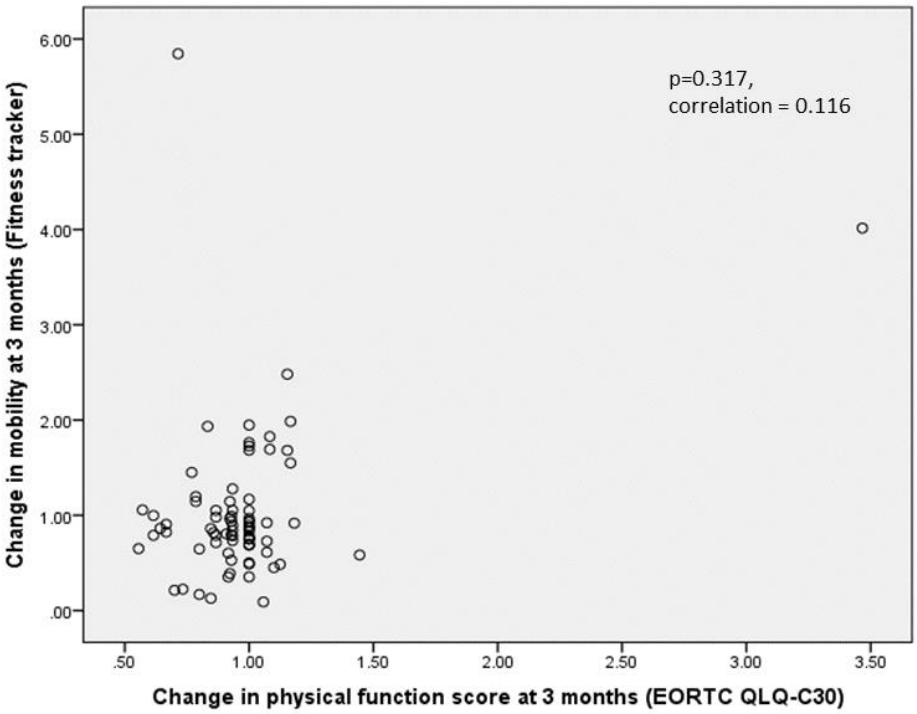


Figure 6-4: Scatter plot illustrating change (3 month/baseline) in mobility as measured by wearable device vs physical function score as measured by the EORTC QLQL-C30.

6.3.4 Discussion

Four (physical, role, cognitive and social functioning) of the five functioning domains and global health were significantly reduced at 5 weeks post-operatively. By three months post-operatively, only two domains (physical and role functioning) were still significantly reduced compared to baseline while global health, cognitive health and social function had recovered. These findings are consistent with prior reports that RC is a morbid procedure with a lengthy recovery[226], regardless of the diversion type performed. These results are different to those presented in section 3.3, in which I presented data from the same questionnaire at the 3 months and 6 months timepoints. In that patient cohort, only physical function was different at 3 months post-operatively, whereas global health and role function were not significantly different. This could partly be attributed to the small sample size in the previous study, which comprised of only 26 patients. Furthermore, this could also be due to the fact that the iROC cohort includes both iRARC and ORC patients, whereas the data in section 3.3 included only patients undergoing iRARC.

Objectively measured step-count showed a similar trend in terms of physical functioning. All three metrics derived from raw step-count data (average steps, maximum steps and minimum steps) were significantly reduced post-operatively when compared to baseline. However, the coefficients became numerically smaller with time, suggesting that there was an improvement in mobility with time. Despite the similar trends in data change at 1 and 3 months compared with baseline that was noted in both EORTC QLQ-C30 and step-count data, no correlation was identified between the two instruments. This suggests that objectively measured step-count from wearable devices are providing different and potentially new information about recovery that is not captured by the EORTC QLQ-C30 questionnaire.

Wearable devices are quickly evolving from simple pedometers to being able to offer new health features such as fall detection, ECG tracking and temperature measurement. With smartphone

pairing, these devices offer the ability to collect data passively from patients, and with modern trackers could also offer the ability to collect data continuously as opposed to 7-day periods like the iROC study design. A future study could use these more advanced devices to measure patient activity trends in greater detail, as well as measure clinical parameters that could predict clinical decline.

While this section highlights the potential advantage of adding wearable devices to measure recovery in addition to PROMs, it does not interrogate the relationship of these scores with outcome data such as length of stay, complications or even oncological recurrence. At the time of analysis, post-operative data is not available for analysis for this cohort. However, complications and outcomes data are available for a smaller cohort of patients recruited in the first year, and will be discussed in 7.3 and Chapter 8.

6.3.5 Conclusions

Objectively measured mobility data from wearable devices offers an opportunity to collect recovery of physical function passively after RC. While PROM tools such as the EORTC QLQ-C30 questionnaire seem to show similar trends in the data, the recovery of post-operative scores do not correlate statistically. This suggests that wearable device data could offer a new dimension of information regarding post-operative recovery. In the next chapters, I will build on this work by comparing the value of fitness tracking data in predicting outcomes following RC, both using pre-operative data and post-operative recovery data.

6.4 Chapter conclusions

This chapter provides an overview of the trends in mobility and their associations with baseline data, as well as PROMs during the peri-operative period. In the first part of the chapter, I provided descriptive data about step-count data as a metric and showed that it is not normally distributed. Additionally, I compared tracking data at baseline with demographic data collected at baseline to assess for any difference in activity that could be associated with patient factors. Only age and gender have a statistically significant relationship with step-count data, while BMI, neoadjuvant chemotherapy and smoking status did not. Tracking data trends over the three-month peri-operative period are also described, and their association with clinical outcomes will be explored in the next two chapters.

In the second part of the chapter I compared baseline data gathered from the patient-filled EORTC QLQ-C30 questionnaire and wearable devices with post-operative data at one and three months from the same patient population. While both the questionnaire and wearable devices showed similar trends in recovery – which is that patients have reduced mobility or physical functioning at one month but this improves by three months, although still significantly reduced compared to baseline – no correlation was identified in the percentage difference of both measures when compared to baseline.

Chapter 7 Comparing step-count from wearable devices with other metrics in predicting risk of complications following radical cystectomy.

7.1 Chapter summary

The main aim of this chapter is to compare step-count data from wearable devices to data obtained from cardiopulmonary exercise testing (CPET). The chapter is divided into two parts, associations of step-count data from wearable devices CPET data, and using different baseline data including step-count and CPET risk stratification to predict outcomes after RC. As CPET is part of the standard-of-care investigations in many major surgical centres, the significance of this chapter is to consider utility of the relatively expensive CPET test by investigating its comparability to step-count collected from cheaper wearable devices.

In the first part, physiological variables from CPET will be compared with baseline wearable device-derived step-count. As CPET is used for risk-stratification of patients based on their physiological reserve, this section of the chapter compares key physiological measures from CPET results with step-count data collected pre-RC. In the second part of the chapter, step-count at baseline will be compared with other metrics such as CPET collected at baseline as predictors for complications in the peri-operative period following RC.

7.2 Correlations between tracker-derived activity data to cardiopulmonary exercise testing

7.2.1 Introduction

Cardiopulmonary exercise testing (CPET) is a non-invasive method used to assess the performance of the heart and lungs at rest and during exercise. It has become standard of care[85] in many pre-operative pathways to measure the physiological reserve of patients prior to major surgery, and to risk stratify them for post-operative complications (and pre-emptive high dependency unit or intensive care unit admission). For cystectomy, Tolchard *et al*[86]. reported that low Anaerobic Threshold (<11 mL/kg/min) and high VE/VC02 (≥33) were predictive of significant complications in patients undergoing RC. These thresholds are applied to risk stratify patients undergoing CPET as part of the pre-operative pathway prior to RC in our centre.

In this section, data derived from wearable devices will be compared with CPET variables to explore correlations between them. During CPET, patients are pushed to their physiological limits in a controlled environment. Conversely, wearable devices measure patient activity in the patient passively and non-invasively in the home environment. Step-count data from wearable devices therefore may be reflective of a patient's performance status. The value of step-count in measuring performance status has been reported in cancer patients by Gresham *et al*.[227] in a cohort of cancer patients with ECOG as the reference standard.

The analysis in this section will focus on CPET as the reference standard, and report on associations between step-count and key CPET variables used for risk stratification, as well as any significant difference in step-count for patients who are considered high risk by previously reported standards[86].

7.2.2 Methods

7.2.2.1 *Sample*

Patients undergoing radical cystectomy as part of the iROC trial were consented to complete wear a wrist-worn wearable device at pre-determined timepoints. Cardiopulmonary exercise testing (CPET) data was collected from hospital sites that used the test as part of their standard pre-operative investigations.

7.2.2.2 *Data collection*

7.2.2.2.1 Demographic data

At time of recruitment, patient sex, age, BMI, diversion type and ECOG performance status was collected.

7.2.2.2.2 Wearable device

The Misfit Ray wrist-worn wearable device (described in section 0) was provided for patients to wear for seven consecutive days at three timepoints: baseline and 3 months post-operatively. Patients were provided with the trackers during their hospital clinic appointments, and were also issued pre-stamped envelopes to return the wearable devices to a central receiving lab. Upon receiving the wearable devices, daily step-count data was extracted using the iOS Misfit mobile application.

7.2.2.2.3 Cardiopulmonary exercise testing (CPET)

CPET was conducted on a cycle ergometer (Lode Corival) with continuous side stream gas exchange analysis (Cortex Metalyzer 3B). Three minutes of rest preceded testing, during which oxygen and carbon dioxide concentration and gas flow were measured. Three minutes of unloaded cycling at 60-65 revolutions/minute was then undertaken. Work rate was increased continuously thereafter until the test was terminated due to symptoms, volitional fatigue, or ECG changes. All CPET variables were determined (as previously described[47]) on the day of the test and then independently verified by a Consultant Anaesthetist with CPET expertise.

Previously published thresholds[86] for key variables of CPET (AT <11 and VE/VCO_2 ≥33) were applied as predictors of major complications.

7.2.2.3 *Statistical analysis*

All statistical analyses were carried out using SPSS for Windows version 25.0.

Non-parametric tests were used to analyse step-count as explained in section 6.2.3. Three metrics were derived from the step-count data extracted from each wearable device: average step-count (ASC), maximum step-count (MSC) and minimum step-count (MiSC). MSC and MiSC were defined as the highest and lowest total steps respectively in any one day during the 7-day study period respectively. Mann-Whitney U test was performed to compare non-parametrically distributed independent variables.

Peri-operative complications were excluded from the data downloaded from the eCRF due to ongoing data monitoring for more recently recruited patients. Complications and outcomes data are available for a smaller cohort of patients recruited in the first year. Analysis of this cohort will be discussed in detail in section 7.3 and Chapter 8.

7.2.3 Results

A total of 71 patients provided CPET data as well as fitness tracking data at baseline. Their baseline characteristics are summarised in Table 7-1.

Gender		Male (%)	58 (81.7)
		Female (%)	13 (18.3)
Age		Median (IQR)	71 (63.5-75.5)
ECOG		0 (%)	54 (76.1)
		1 (%)	14 (19.7)
		2 (%)	2 (2.8)
		3 (%)	1 (1.4)
Neoadjuvant	**Chemotherapy**	n (%)	34 (47.9)
	Immunotherapy	n (%)	13 (18.3)
Histology		UCC (%)	63 (88.7)
		SCC (%)	4 (5.6)
		Adenocarcinoma	1 (1.4)
		Other (%)	3 (4.2)

BMI	Median (IQR)	26.6 (25.1-29.0)
Smoking	Current smoker (%)	6 (8.5)
	Ex-smoker (%)	52 (73.2)
	Non-smoker (%)	13 1(8.3)

Table 7-1: Baseline characteristics of patients who provided fitness tracking data at baseline and had a cardiopulmonary exercise test (CPET) prior to radical cystectomy in the iROC trial

From the activity tracking data, average step-count (ASC), maximum step-count (MSC) and minimum step-count (MiSC) were derived. A Spearman Rank correlation was performed with the two CPET variables (AT and VE/VCO_2) used to risk-stratify patients for having a high risk for complications. Significant correlations were identified for MSC with AT (p= 0.005, coeff= 0.339) and VE/VCO2 (p=0.002, -0.357). ASC also correlated significantly with AT (p= 0.014, coeff= 0.299) and VE/VCO2 (p= 0.006, coeff= -0.326) but with lower correlation coefficients, and MiSC correlated only with AT (p= 0.050, coeff= 0.241). These results are displayed in Table 7-2.

Thresholds of AT<11 and $VE/VCO_2 \geq 33$ were then applied as high risk for post-operative complications. Since MSC correlated more significantly with AT and VE/VCO2 than ASC or MiSC, patient were divided into high and low risk groups by CPET variable thresholds and their MSC at baseline compared. Figure 7-1 and Figure 7-2 show the box and whisker plots for AT risk (<11 mL/kg/min) and VE/VCO_2 (≥33). Patients considered high risk by either threshold had reduced MSC, and both were statistically significant (p=0.002 and p=0.0005 for AT and VE/VCO2 respectively). Lastly, Figure 7-3 shows the box and whisker plot for combined CPET risk (high CPET risk defined as AT risk <11 mL/kg/min or $VE/VCO_2 \geq 33$), and this was also statistically significant (p<0.001).

		AT	**VE/VCO_2**
Maximum Steps (MSC)	**Coefficient**	0.339	-0.357
	Sig. (2-tailed)	**0.005**	**0.002**
Average steps (ASC)	**Coefficient**	0.299	-0.326
	Sig. (2-tailed)	**0.014**	**0.006**
Minimum steps (MiSC)	**Coefficient**	0.241	-0.139
	Sig. (2-tailed)	**0.050**	0.252

Table 7-2: Spearman correlations for CPET variables and step-count

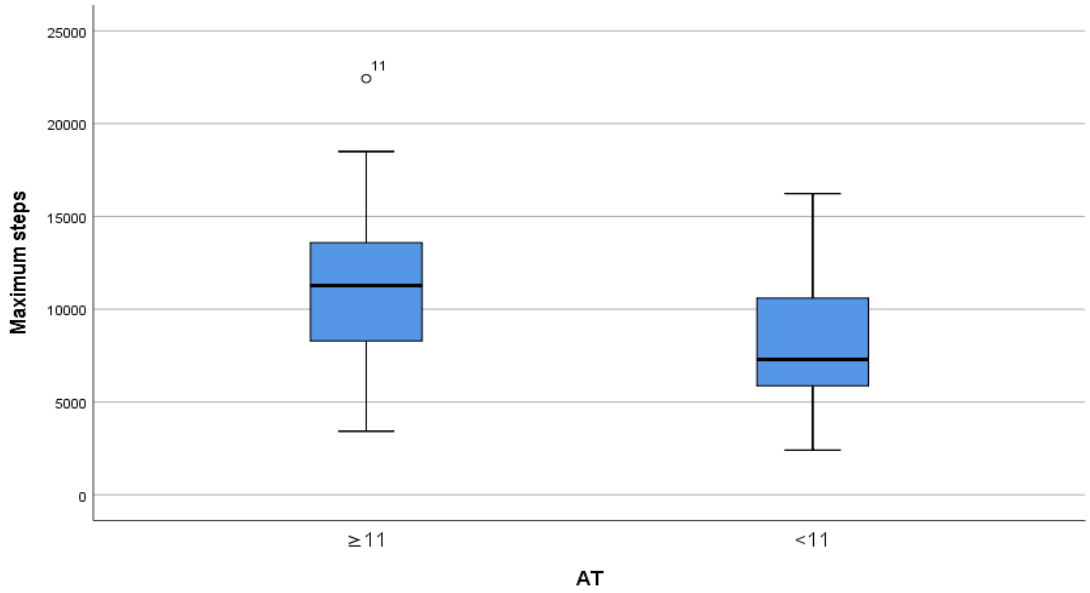


Figure 7-1: Box and whisker plots of maximum step-count (MSC) for patients risk-stratified by anaerobic threshold (AT) over <11 as high risk

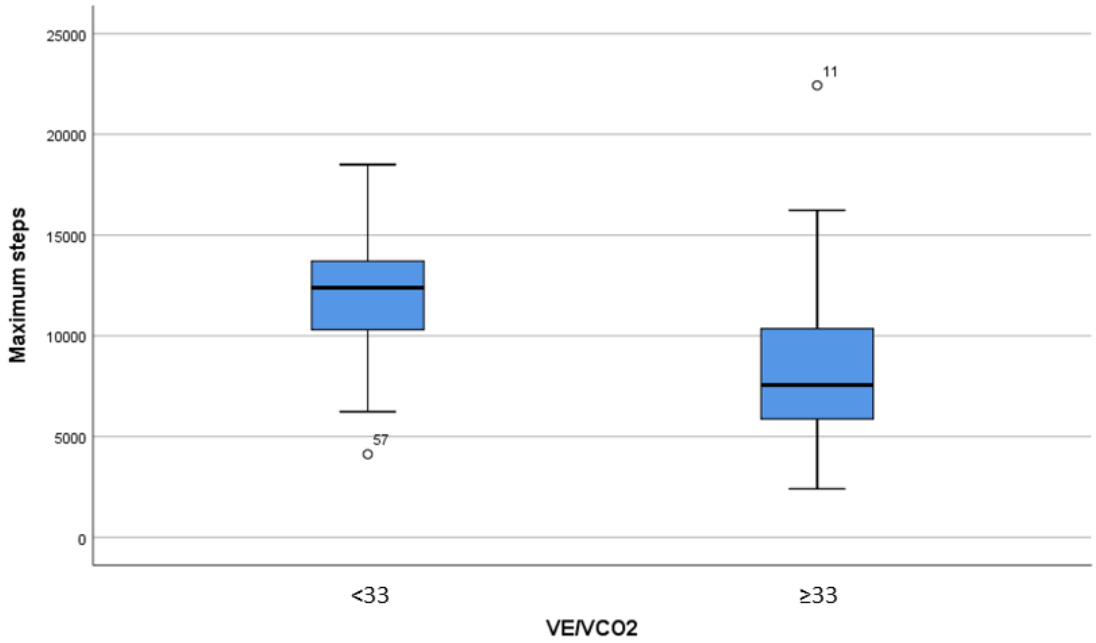


Figure 7-2: Box and whisker plots of maximum step-count (MSC) for patients risk-stratified by VE/VCO_2≥33 as high risk

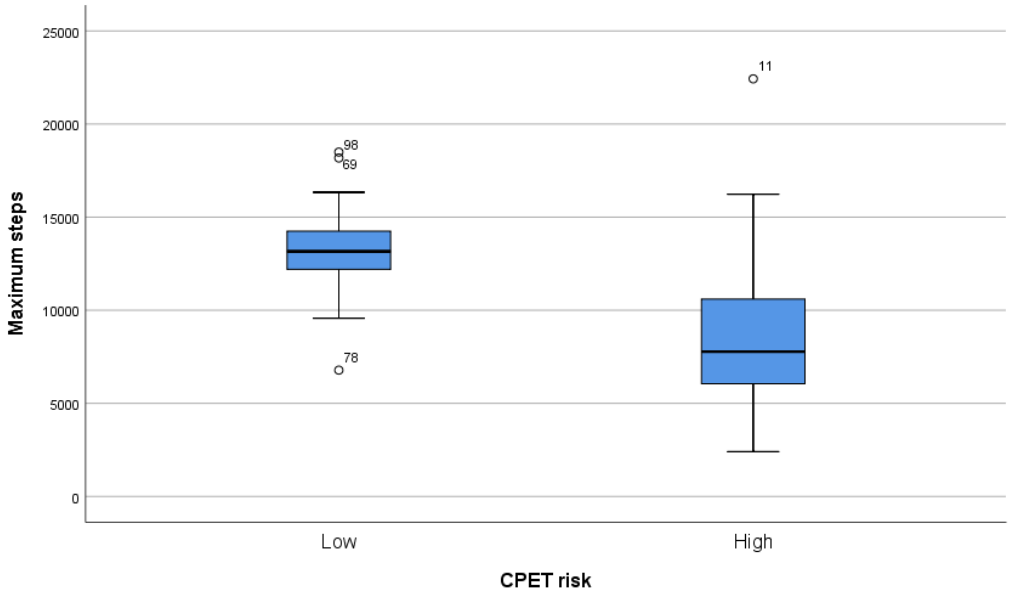


Figure 7-3: Box and whisker plots of maximum step-count (MSC) for patients with a combined high CPET risk ($VE/VCO_2 \geq 33$ or AT<11)

	AT risk	**VE/VCO_2 risk**	**CPET combined risk**
Mann-Whitney U	300	245.5	108.5
Wilcoxon W	966	1373.5	1486.5
Z	-3.09077	-3.49593	-4.007670756
Asymp. Sig. (2-tailed)	0.001996	0.000472	0.000061

Table 7-3: Mann-Whitney U test for maximum steps in patients considered high risk of complications by AT(<11), $VE/VCO_2(\geq 33)$ and CPET combined (AT<11 or $VE/VCO_2 \geq 33$).

7.2.4 Discussion

CPET estimates physiological reserve by getting participants to reach peak exercise levels on a bike or treadmill[88]. There are two theories on why this measurement of physiological reserve can predict post-operative outcome: 1) patients with higher fitness levels cope better with surgery without outpacing their anaerobic threshold. 2) Regular exercise creates a similar effect to ischaemic preconditioning, lessening the impact of deficit in oxygen demand[228]. Wearable devices offer an opportunity to collect a surrogate measures of exercise capacity while tracking patients in their activities of daily living, offering an alternative method of measuring patients' physiological reserves. Some wearable devices are able to estimate certain CPET variables such as maximal oxygen uptake ($VO2_{max}$) and anaerobic threshold (AT) [229,230], but as yet none have been validated for clinical use.

Maximum step-count correlated more significantly than average step-count with both anaerobic threshold and VE/VCO_2, the two variables predominantly used clinically to risk-stratify patients for complications following RC and other surgery[85]. We hypothesise that this could be because maximum step-count captures a patient's most active day, which is more likely to be reflective of their exercise capacity than average step-count, in the same way that CPET captures physiological parameters at peak exercise levels. The average step-count includes patients' inactive days, which can have a disproportionate effect on the average given the relatively short number of total days monitored in this study.

Maximum step-count data is significantly different when stratifying patients by risk according to AT ($p=0.002$) and VE/VCO2 ($p=0.000472$) individually. This relationship was even stronger when the two variables were combined for risk stratification ($p=0.000061$). Passive step-counting could therefore offer a cheaper alternative to CPET in risk-stratifying patients prior to RC and other major surgeries. The Misfit Ray tracker retails for a price of £79 and can be reused (Table 3-1), whereas a CPET costs over £200 for a single assessment[47].

7.2.5 Conclusions

This study highlights that data obtained from a relatively inexpensive wearable device correlates well with physiological variables computed using CPET. As CPET is used as a standard pre-operative assessment tool in many centres, it is reasonable to hypothesise that step-count data from wearable devices at baseline may offer similar utility in predicting post-operative complications following RC. This hypothesis will be tested in section 7.3.

7.3 Evaluation of baseline step-count as measured by wrist-worn wearable devices to predict major complications following radical cystectomy

7.3.1 Introduction

Based on the data presented in 0 and section 7.2, step-count from wearable devices offer objective measurement of health status during the peri-operative period that offer an additional dimension compared to PROM questionnaires, and correlate well with CPET data. However, post-operative outcomes have not been reported in the analysis so far. In this section, 3-month complications will be an endpoint measure to explore the value of pre-operative parameters such as fitness tracking data and CPET in predicting major complications. Various studies have reported the value of CPET in predicting complications and other post-operative outcomes for RC and other major surgery[86,89,231]. A systematic review by Moran *et al*.[88] concluded that CPET is a useful pre-operative risk stratification tool in various surgery types, but a separate systematic review by Lam *et al*.[89] concluded that the use of CPET prior to major cancer surgery did not yield sufficient accuracy to predict post-operative morbidity except in lung cancer patients. Despite such contrasting evidence, the number of CPETs performed in the UK have more than doubled since 2011, with more than 30,000 patients undergoing CPET annually[85].

Wearable devices offer a less invasive and cheaper method to measure performance status compared to CPET. As mentioned in section 7.2.1, Gresham *et al*. reported that step-count correlate well with ECOG status as an indicator of performance status for patients with cancer. However, the relationship between either step-count or ECOG and any outcome measure in their study was not assessed. Furthermore, their cohort consisted largely of patients with late stage cancer who are comparatively less fit than our current cohort.

The main aim of this section is to explore associations and the predictive value of baseline step-count and other baseline metrics in predicting major complications (Clavien-Dindo classification of surgical complications ≥3) following RC.

7.3.2 Methods

7.3.2.1 *Sample*

During a 12-month period, patients undergoing radical cystectomy (RC) for bladder cancer were recruited as part of the iROC trial (NCT03049410) [232] across five high-volume centres across the UK. Patients were consented to complete wear a wrist-worn wearable device at pre-determined timepoints. Cardiopulmonary exercise testing (CPET) data was collected from hospital sites that used the test as part of their standard pre-operative investigations.

7.3.2.2 *Data collection*

7.3.2.2.1 Demographic data

At time of recruitment, patient sex, age, BMI, diversion type and ECOG performance status was collected.

7.3.2.2.2 Wearable device

The Misfit Ray wrist-worn wearable device (described in section 0) was provided for patients to wear for seven consecutive days at three timepoints: baseline and 3 months post-operatively. Patients were provided with the trackers during their hospital clinic appointments, and were also issued pre-stamped envelopes to return the wearable devices to a central receiving lab. Upon receiving the wearable devices, daily step-count data was extracted using the iOS Misfit mobile application.

7.3.2.2.3 Cardiopulmonary exercise testing (CPET)

CPET was conducted on a cycle ergometer (Lode Corival) with continuous side stream gas exchange analysis (Cortex Metalyzer 3B). Three minutes of rest preceded testing, during which oxygen and carbon dioxide concentration and gas flow were measured. Three minutes of unloaded cycling at 60-65 revolutions/minute was then undertaken. Work rate was increased continuously thereafter until the test was terminated due to symptoms, volitional fatigue, or ECG changes. All CPET variables were determined (as previously described[47]) on the day of

the test and then independently verified by a Consultant Anaesthetist with CPET expertise. Previously published thresholds[86] for key variables of CPET (AT <11 and VE/VCO_2 ≥33) were applied as predictors of major complications.

7.3.2.2.4 Study outcomes measured

All 30-day and 90-day complications were classified according to the modified Memorial Sloan-Kettering Cancer Center (MSKCC) Clavien–Dindo (CD) system[233]. Based on peri-operative outcomes, patients were divided into two groups:

1) patients who had major complications following RC (CD≥3)
2) patients who had no complications or minor complications (CD≤2)

7.3.2.3 *Statistical methods*

All statistical analysis was performed using the SPSS software version 25.0. All continuous data such as mean, median, interquartile range (IQR) and 95% confidence interval (CI) were reported using descriptive statistics. The Mann-Whitney *U* test was performed to explore associations between non-parametrically distributed variables. For variables with skewed distributions, a log-transformation was performed.

7.3.3 Results

Of the 79 patients with CPET data who underwent RC as part of the iROC trial, 57 (72.2%) participated in the wearable device sub-study. Patient baseline characteristics, diversion type, and histopathological outcomes are shown in Table 7-4. Overall, 10 patients (17.5%) had major complications (CD≥3) following RC.

	All patients	No or minor complications (CD≤2)	Major complications (CD≥3)
n (%)	57 (-)	47 (82.5)	10 (17.5)
Sex (%)			
Male	42	35 (74)	7 (70)
Female	15	12 (26)	3 (30)
Age			
BMI (%)			
<25	14	12 (26)	2 (20)
25.1-30.0	30	25 (53)	5 (50)
>30	13	10 (21)	3(30)
ECOG Performance status (%)			
0	19	18 (38)	1(10)
1	22	18 (38)	4 (40)
2	10	6 (13)	4 (40)
3	6	5(11)	1 (10)
Neoadjuvant chemotherapy (%)			
Yes	33	28(60)	5(50)
No	24	19(40)	5
Diversion type (%)			
Ileal Conduit	48	40(85)	8 (80)
Neobladder	9	7 (15)	2 (20)
Histology (%)			
UCC	49	40(85)	9 (90)
SCC	4	3(6)	1 (10)
Adenocarcinoma	1	1(2)	0 (0)
Other	3	3(6)	0 (0)

Table 7-4: Baseline characteristics for the 57 patients that provided fitness tracking data at baseline during the first year of follow-up

The median ASC and MSC for the cohort (n=57) were 5493 (IQR: 4007-7612) and 8626 (IQR: 6561-12358) steps respectively. The distribution of step-count is represented in Figure 7-4. There is a big difference of 3133 steps (57%) between the median and average steps, and this is reflective of the variations in daily activity of patients and the need to measure activity for a longer time period. At the pre-operative timepoint, this difference is likely to be more reflective of a difference in daily routine rather than an acute decline in health status.

7.3.3.1 *Associations between step-count and major complications*

Figure 7-4 shows a box and whisker plot of ASC and MSC distribution in the two groups: 1) none or minor complications (CD≤2), 2) major complications (CD≥3). For both ASC and MSC, patients who went on to have none or minor complications group (n=47) had higher counts at baseline compared to the patients who had major complications (n=10).

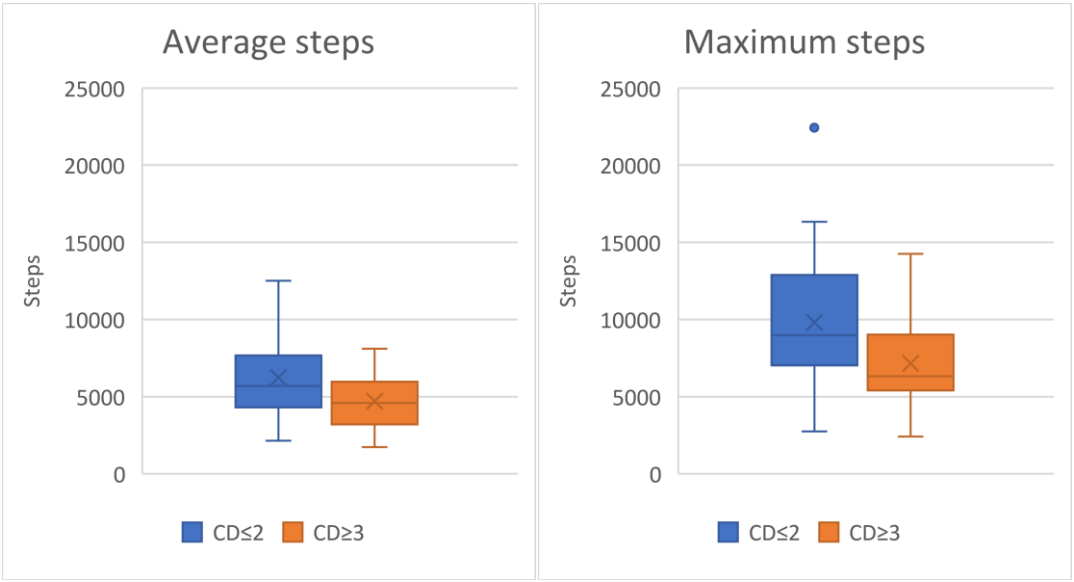


Figure 7-4: Baseline average step-count and maximum step-count for patients who went on to have none or minor complications (CD≤2) or major complications (CD≥3).

A Mann–Whitney U test was performed comparing the baseline activity metrics for both groups, with results shown in Table 7-5. Median MSC in the CD≤2 and CD≥3 were 8980 and 6311 respectively, and the distribution of MSC differ significantly in the two groups (p=0.031, z=-2.161). Median ASC in the CD≤2 and CD≥3 were 5692 and 4592 respectively. However, the

distribution of ASC did not differ significantly (p=0.085, z=-1.72) between the two groups. This result suggests that major complications group had significantly lower maximum steps prior to surgery when compared with patients who had no or minor complications. The Mann-Whitney U test was then performed in other baseline characteristics including risk determined by CPET, neoadjuvant chemotherapy (NAC), age, body mass index (BMI) and the Eastern Cooperative Oncology Group (ECOG) performance status were also compared across the major and none or minor complications patient groups. No statistically significant difference was found between these baseline characteristics. And complications These results are presented in Table 7-6.

Pre-operative step-count		
	Average steps	Maximum steps
Mann-Whitney U	153	132
Wilcoxon W	208	187
Z	-1.720	-2.161
Asymp. Sig. (2-tailed)	0.085	**0.031**
Grouping Variable: CD≥3		

Table 7-5: Mann-Whitney-U test for baseline step-count across two patient groups: 1) patients who had major complications (CD≥3) 90 days post RC, 2) patients who had none or minor complications (CD<3) 90 days post RC

Test Statistics					
	CPET risk	NAC	Age	BMI	ECOG
Mann-Whitney U	225.5	212.5	168.0	177.000	152.500
Wilcoxon W	291.5	1340.5	1296.0	1212.000	1280.500
Z	-0.409	-0.552	-1.407	-0.592	-1.825
Asymp. Sig. (2-tailed)	0.682	0.581	0.159	0.554	0.068
Grouping Variable: CD≥3 NAC = Neoadjuvant chemotherapy, BMI = Body Mass Index, ECOG = Eastern Cooperative Oncology Group performance status					

Table 7-6: Mann-Whitney-U test for baseline variables across two patient groups: 1) patients who had major complications (CD≥3) 90 days post RC, 2) patients who had none or minor complications (CD<3) 90 days post RC

7.3.3.2 *Regression statistics*

Since ASC was not significantly different in patients who had major complications vs patients who did not, MSC was used as a covariate for regression statistics. Binary logistic regression was performed using MSC, CPET risk, NAC, age, BMI and ECOG. Since there were only ten total events (patients who had major complications post-cystectomy), univariate analysis was undertaken with each covariate. The results of the logistic regression are presented in Table 7-7. MSC is the only significant baseline predictor of major complications in the first 90 days following RC in this patient cohort, with an odds ratio of 0.025. This finding means that patients with lower MSC at baseline are more likely to have major complications in the first 90 days following RC. Of note, risk stratification based on CPET results (AT <11mL/kg/min or VE/VCO_2) of ≥33 being high risk) was not a predictor of major complications.

Variable	**B**	**S.E**	***p***	**O.R.**
MSC	-3.7	1.819	**0.042**	0.025 (0.001-0.874)
CPET risk	0.087	0.793	0.913	1.091 (0.231-5.161)
NAC	0.388	0.699	0.579	1.474 (0.375-5.797)
Age	0.060	0.047	0.201	1.062 (0.968-1.165)
BMI	0.026	0.019	0.166	1.026 (0.989-1.065)
ECOG	0.557	0.354	0.116	1.746 (0.872-3.495)
MSC = Maximum step-count, CPET = Cardiopulmonary Exercise Testing, NAC = Neoadjuvant chemotherapy, BMI = Body Mass Index, ECOG = Eastern Cooperative Oncology Group performance status				

Table 7-7: Binary logistic regression for baseline variables in predicting 90-day major complications (CD≥3)

7.3.4 Discussion

Patients who experienced major complications (n=10) following RC had significantly lower maximum-step-count (MSC) than their counterparts who did not have major complications (n=47). However, no significant difference was identified in ASC. As discussed in section 0, MSC is likely to be a better predictor of physiological reserve in a similar way to CPET than ASC as it reports maximal capacity for activity as opposed to an aggregate of the patient's routine. Additionally, the close correlations between CPET and MSC reported in 7.2.3 suggested that MSC could be associated with major complications in a similar way to CPET[86,87]. Unlike prior contemporary publications, no significant difference was identified in CPET risk stratification for patients who had major complications and those who did not. Although the role of CPET in risk-stratification is relatively well-established, there is some ongoing debate regarding its true utility[234]. An earlier publication by our centre [47] reported that poor cardiopulmonary fitness as measured by CPET did not predict major complications for patients undergoing iRARC, which supports my findings that CPET risk stratification is not predictive of major complications in patients undergoing cystectomy in this mixed iRARC and ORC cohort.

Due to the limited sample size and low event rate of major complications (n=10), I was restricted to running a univariate logistic regression model to avoid overfitting. ASC was not included as a variable in the regression model as MSC was significantly associated with major complications (p=0.031) while ASC was not (p=0.085), and both variables draw from the same raw data. Of the six variables in the logistic regression (MSC, CPET risk, NAC, age, BMI, ECOG), only MSC was a significant predictor of major complications (p=0.042, O.R. = 0.025). These results are consistent with the results of the Mann-Whitney U test reported in Table 7-5 and Table 7-6.

7.3.5 Conclusions

The findings in this small cohort suggest that MSC is a predictor of major complications, unlike CPET and other baseline variables reported. CPET testing is much more expensive than the cheap wearable devices described, and if this work can be replicated in a larger cohort this would provide a cheaper avenue to measure physiological reserve prior to radical cystectomy and other major surgery. In the current study, these findings could be attributed to a small sample size and a low event rate. Once the iROC trial is completed, this analysis will be repeated with the entire cohort of 340 patients to assess if these findings can be replicated in a larger sample size.

This current study reports the value of step-counts in predicting major complications. However, wearable devices and fitness trackers are becoming more advanced with new features such as heart rate measurement, pulse oximetry and ECG monitoring. These new metrics can be combined with step-counts to estimate physiological parameters such as VO_2Max, heart rate variability and moderate to vigorous physical activity (MVPA). Composite metrics that capture multiple physiological signals have the potential to outperform CPET and step-counts. However, these devices currently have short battery lives and require a companion smartphone device. As wearable devices become more accepted and the technology improves, newer studies must utilise more advanced devices to compare their performance in predicting complications in RC and other major surgery groups.

7.4 Chapter conclusions

The two experiments presented in this chapter examine the utility of the step-count data collected from the Misfit Ray tracker in patients undergoing RC as part of the iROC trial. Section 7.2 compared wearable device data with CPET variables used to risk stratify patients undergoing RC. My analysis showed that both MSC (maximum step-count recorded in a single day during a 7-day period) and ASC (average step-count/day during a 7-day period) correlated significantly with both AT and $VE/VC0_2$ while MiSC(minimum step-count recorded in a single day during a 7-day period only correlated with AT. Since MSC correlated most significantly with both CPET risk-stratification variables, a Mann-Whitney U test was performed to compare MSC in patients who were in the high-risk group with patients in the low-risk group by CPET. MSC was significantly different in the high-risk and low-risk groups by each CPET variable, but most significant when the risk stratification variables were combined (p=0.000061).

Section 7.3 assessed the utility of step-count data and other metrics collected at baseline in predicting major complications following RC. First, a Mann-Whitney U test was performed comparing patients who experienced major complications with patients who did not. Only MSC at baseline was found to be significantly different, while other variables such as CPET risk stratification, ECOG, BMI, age or NAC were not. Next, univariate multiple logistic regression was performed and MSC was the only significant predictor (p=0.042, O.R. = 0.025) of major complications following RC in this patient cohort.

The findings of this chapter suggest that pre-operative measurement of daily step-count using wearable devices could be used as a risk stratification tool for major complications following RC and other major surgery groups. The next chapter will explore the utility of post-operative step-count in predicting longer term outcomes (1 year).

Chapter 8 Evaluating recovery of mobility following radical cystectomy as a predictor of survival at 1 year

8.1 Chapter summary

As discussed in section 6.2, there is a sharp reduction in daily step-count immediately following RC and a steady recovery in step-count in the cohort by 3 months post operatively. Even at 3 months, the majority of patients have not returned to their baseline mobility. While patients who experience late complications could have a slower and prolonged recovery period, the number of patients who did not reach their baseline mobility (69.8%) was significantly higher than major complication rates or all complication rates (Section 6.2.3). It is important to consider why patients have such varying recovery patterns and assess any relationships with outcomes during the first year.

To explore reasons for failure to return to baseline activity, I investigated whether there was an association between development of complications and failure to return to activity. Although a factor, the occurrence of a major complication was not an independent predictor of failure to achieve more than 50% of baseline activity at 3 months. The main cause of death in patients who undergo RC for MIBC is cancer recurrence, and nearly 90% of all metastatic recurrences are detectable in the first 2 years (as shown in the Kaplan Meier curve shown in Figure 1-1). Up to 20% of patients undergoing RC develop metastatic disease in the first year, which is associated with a 5-year survival of approximately 10%[5].

In this chapter, I set out to explore the relationship between failure to recover activity and long-term outcomes (measured at 12 months). To understand the potential for metastatic disease to impact return to baseline activity, I used survival (overall survival and cancer-specific were the same in this cohort) survival as endpoint measures.

8.2 Introduction

A recently published study by Lee *et al*.[221] reported that all-cause mortality in a cohort of 16, 741 women >45 years of age was significantly lower in participants who had higher daily step-count. Similarly, a meta-analysis by Ekelund *et al*.[235] reported findings across 8 studies comprising 36, 363 subjects and concluded that higher levels of activity at any intensity substantially reduced risk for mortality. However, most studies only had a single monitoring period so changes in physical activity in patients could not be assessed longitudinally. It is reasonable to expect that most patients would have a steady decline in overall function preceding their death and therefore a reduction in physical activity, particularly for patients with worsening chronic diseases. However, this hypothesis has not been tested in any currently reported studies.

In this section, I will report mobility trends at baseline and 90 days post-RC and assess the relationship of recovery in terms of mobility with 1-year survival outcomes in this patient cohort.

8.3 Methods

8.3.1 Patient population

Data for patients who had undergone RC as part of the iROC trial was requested in May 2019 from the iROC trial management committee. As with all previous analysis in this thesis, all data provided was blinded to procedure type (open or robotic). Patients were included for analysis if data was available for the following two criteria:

1) Baseline and 3-month step-count as measured by the Misfit Ray (MR) device.
2) Completed 1-year follow up or died before reaching 1-year follow up.

8.3.2 Data collection

8.3.2.1 *Study outcomes measured*

The primary outcome measure for the study is 1-year overall survival, and data on cancer-specific survival is also collected. Additionally, 90-day complications are also reported using the modified Memorial Sloan-Kettering Cancer Center (MSKCC) Clavien–Dindo (CD) system[233].

8.3.2.2 *Step-count data*

Step-count data was included in the analysis if more than 5 days of continuous data was available. The following parameters were computed: average daily step-count (ASC) per day and maximum step-count (MSC) which is highest step-count in a single day. Step-count indices at baseline and 3 months following surgery were included. Based on my results in section 6.2, step-count measures are analysed as non-parametrically distributed variables.

8.3.2.3 *Statistical methods*

All statistical analysis was performed using the SPSS software version 25.0. Continuous data such as mean, median, interquartile range (IQR) and 95% confidence interval (CI) were reported using descriptive statistics. The Mann-Whitney *U* test was performed to explore associations between non-parametrically distributed variables. For variables with skewed distributions, a log-

transformation was performed. The Kaplan-Meier estimator in SPSS is used to estimate the survival function.

8.4 Results

8.4.1 Demographic data

52 patients met the criteria outlined in section 8.3.1. The baseline demographics of this patient cohort are included in Table 8-1. Of these patients, 7 (13.5%) died during the first year of follow-up, with a median overall survival (OS) of 260 (IQR 181-279) days post-RC. All 7 deaths were due to CT-proven metastatic cancer recurrence, so this reflects cancer-specific survival (CSS) in the cohort as well. Of note, all 7 patients had CT scans at baseline and at 3 months post-operatively, and all had been reported to have no visible metastatic disease at these timepoints.

Gender	Male (%)	38 (73.1)
	Female (%)	14 (26.9)
Age	Median (IQR)	70 (63-75)
ECOG	0 (%)	40 (76.9)
	1 (%)	8 (15.4)
	2 (%)	3 (5.8)
	3 (%)	1 (1.9)
Diversion type	Ileal Conduit *n(%)*	44 (84.6)
	Neobladder *n(%)*	8 (15.4)
Neoadjuvant	Chemotherapy *n (%)*	16 (30.8)
	Immunotherapy *n (%)*	6 (11.5)
Histology	UCC (%)	47 (90.4)
	SCC (%)	2 (3.8)
	Adenocarcinoma (%)	0 (0)
	Other (%)	3 (5.8)
BMI	Median (IQR)	26.7 (24.0-30.2)
Smoking	Current smoker (%)	3 (5.8)
	Ex-smoker (%)	35 (67.3)
	Non-smoker (%)	14 (26.9)

Table 8-1: Baseline characteristics of patients in the iROC trial with fitness tracking data and 1-year outcome data.

Table 8-2 summarises the 90-day complications for these patients using the Clavien-Dindo classification of surgical complications. In summary, 50%, 36.5%, 13.4% of patients experienced no complications (CD= 0), low grade complications (CD= 1 to 2) and major complications (CD= 3 to 5) respectively.

Clavien-Dindo Grade	**n (%)**
0	26 (50.0)
1	5 (9.6)
2	14 (26.9)
3	2 (3.8)
4	5 (9.6)
5	0 (0)

Table 8-2: 90-day Clavien-Dindo Classification of complications for patients undergoing RC as part of the iROC trial

8.4.2 Step-count data

Average (ASC) and maximum step-count (MSC) at baseline and 3 months are presented in Figure 8-1. A Wilcoxon signed-rank test was performed on each pair of values for ASC and MSC for these patients. These data show that ASC is significantly different ($p=0.025$, $z=2.240$) at 3 months when compared with baseline while no significant difference was observed in MSC ($p=0.089$, $z=1.703$). This result contrasts the findings described in section 6.3.3, which noted that both MSC and ASC were significantly different at 3 months when compared to baseline. Similar to the data presented in section 6.3.3, 10 (19.2%) and 9 (17.3%) patients from our current cohort had an ASC and MSC of under 50% of their baseline at 3 months respectively.

No significant association was identified between complications (all or major) and reduced ASC or MSC at 3 months. This suggests that post-operative complications did not have a significant effect on 3 months post-operative mobility.

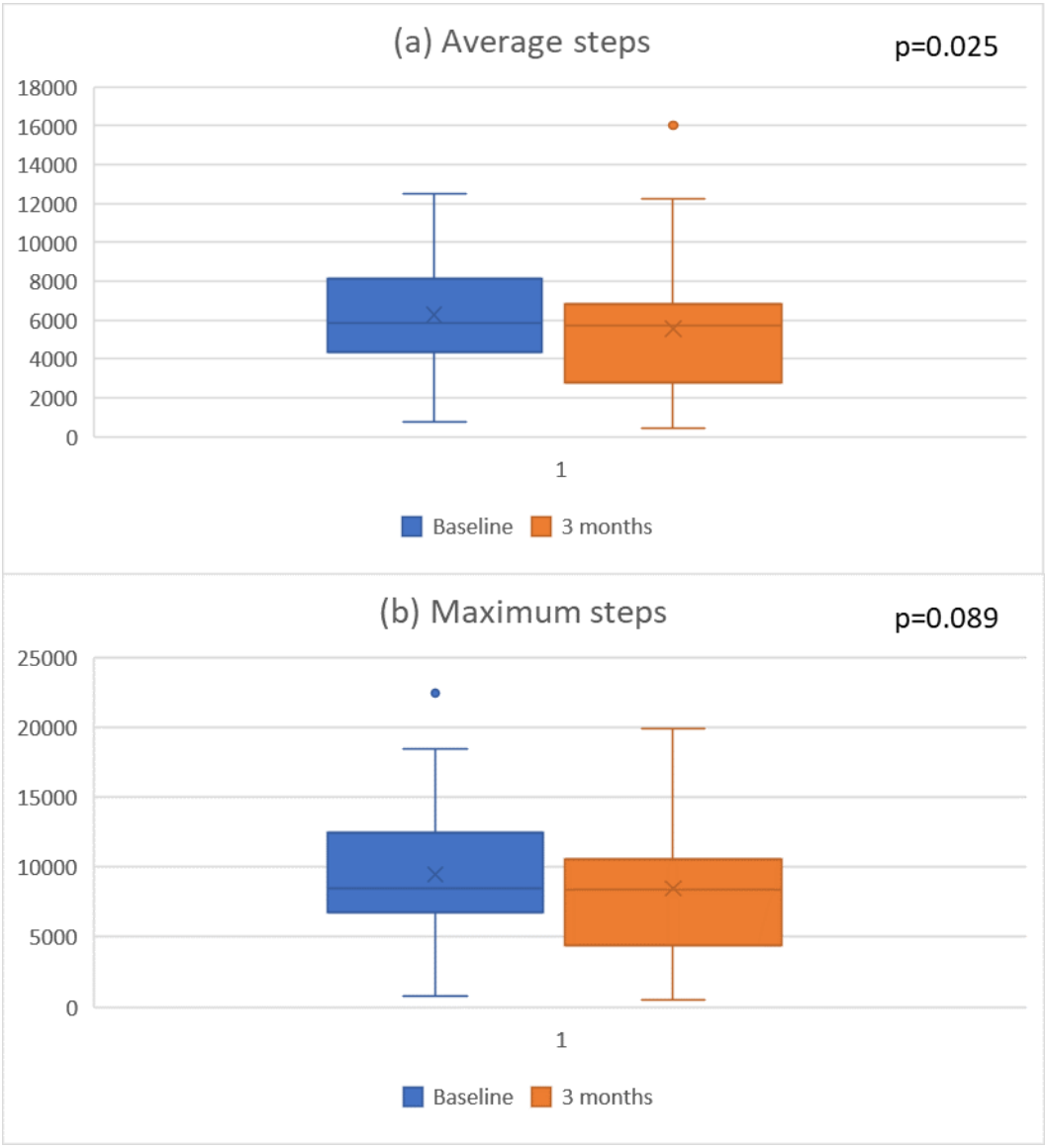


Figure 8-1: Box and whisker plots of (a) average and (b) maximum step-count at baseline and 3 months

8.4.3 Receiver operating characteristic (ROC) curves

ROC analysis was performed for percentage recovery of maximum and average steps at 3 months (compared to baseline) to assess their diagnostic value for predicting disease-free survival (DFS) at 1 year following RC. The area under the curve for maximum steps and average steps is 0.863 and 0.990 respectively. Next, Youden's J statistic was used to identify the best sensitivity specificity pairing to choose a best-fit threshold before plotting Kaplan-Meier curves for both metrics. Youden's J statistic for maximum and average steps is 44.2% and 49.43% respectively. The ROC curves are presented in Figure 8-2. Tables summarizing coordinates of both ROC curves are included in Supplementary Table 10-10 and Supplementary Table 10-11.

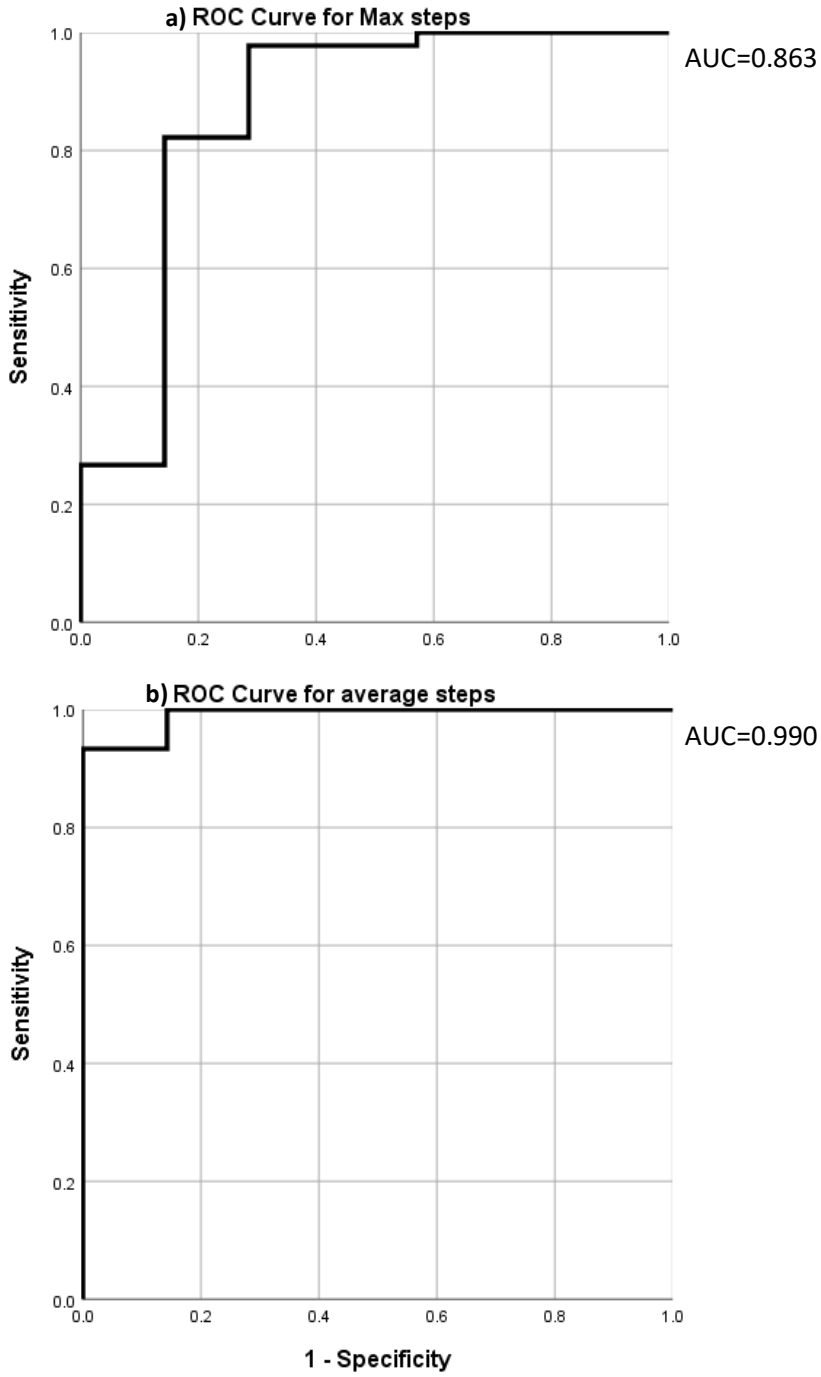


Figure 8-2: ROC analysis for percentage recovery at 3 months post-RC compared with baseline for (a) maximum steps and (b) average steps predicting disease-free survival at 1 year following RC

8.4.4 Survival analysis and Kaplan-Meier curves

In the next part of the analysis, survival analysis is presented, dividing patients into two groups:

1) Patients who had >50% of their baseline activity levels according to their step-count
2) Patients who had ≤50% of their baseline activity levels according to their step-count

A threshold of 50% was chosen as both MSC and ASC had their optimal cut-offs close to 50%, as shown in Supplementary Table 10-10 and Supplementary Table 10-11, and patients being able to recover 50% of their baseline mobility by three months is a more clinically justifiable. This was repeated for both ASC and MSC. Figure 8-3 shows the Kaplan Meier survival curves for (a) ASC and (b) MSC. Of the 10 patients with ≤50% baseline ASC at 3 months, 7 (70%) died following oncological recurrence. All 42 patients who had >50% ASC at 3 months were alive and cancer-free at 12 months post-RC. Of the 9 patients with ≤50% baseline MSC at 3 months, 5 (55.6%) died following an oncological recurrence and 41 of 43 patients who had >50% MSC at 3 months were alive at cancer-free at 12 months.

Using 50% as the threshold for failed recovery following surgery, the diagnostic values of step-count is presented in Table 8-3. Overall, ASC was a better predictor of survival (overall and cancer specific since all deaths were cancer related) with sensitivity and specificity of 100% and 93% respectively, compared with MSC which had a sensitivity and specificity of 71% and 91% respectively.

Supplementary Table 10-12 presents the results of MSC and ASC at their optimal cut-off values of 44.2% and 49.43% respectively. The sensitivity and specificity for ASC improved to 71% and 98% respectively, while the pairings for MSC remained the same. While both results need to be validated in a larger cohort, this suggests that ASC may offer better sensitivity while MSC offers better specificity for OS and CSS at one year.

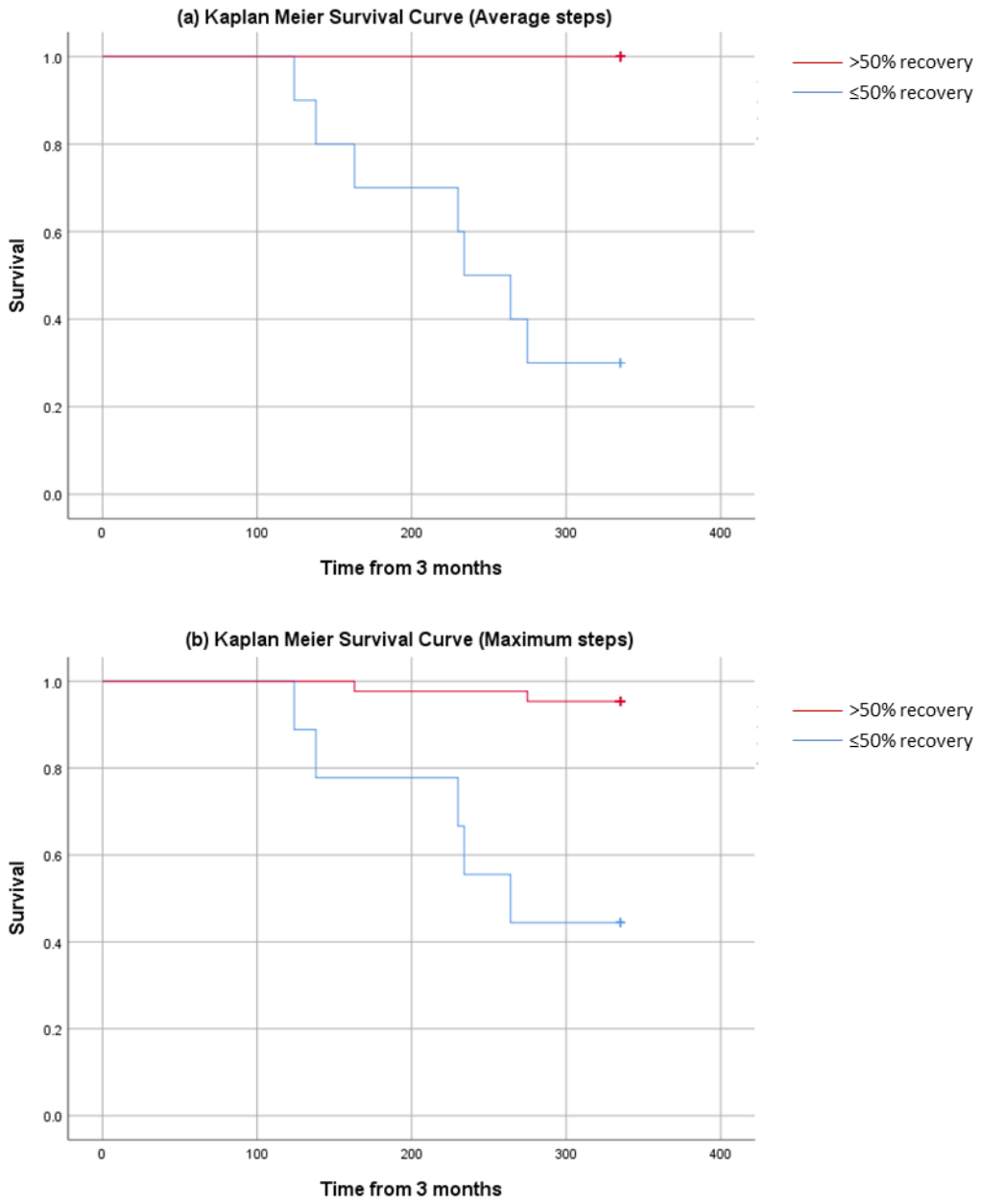


Figure 8-3: Kaplan Meier Survival curves during the first year of follow-up for (a) average steps and (b) maximum steps recovery with a threshold of 50% recovery by 3 months.

	ASC	**MSC**
Sensitivity	100%	71%
Specificity	93%	91%
Positive predictive value	70%	56%
Negative predictive value	86%	89%

Table 8-3: Diagnostic value of ASC and MSC at 3 months being ≤50% of baseline at predicting overall survival at the end of 1-year post RC.

8.5 Discussion

The results presented in this chapter suggest that step-count may have a prognostic value for survival at 1 year following RC. Previous studies have reported that patients with chronic conditions such as asthma and patellofemoral pain have reduced step-count than healthy individuals[159,161], so it would stand to reason that a patient with metastatic recurrence of bladder cancer should have significantly reduced step-count compared to a patient that has been cured of bladder cancer through radical surgery. These results suggest that differences in step-count may be apparent by three-months post operatively even before metastatic disease is detectable on CT scans (these patients had a CT scan at 3 months showing no evidence of metastases).

Of note, no correlation was identified between recovery of average or maximum step-count (or absolute step-count) at 90 days post-operatively and all complications or major complications. This suggests that 90-day complications have no significant impact on mobility at the end of this peri-operative time period. Once the iROC trial is completed, it would be informative to analyse this data with the re-admission dates for all patients as not enough patients who experienced late complications may have been included in the current analysis, which could account for these results. Hospital admission data was not available for this analysis because the trial management committee considers it part of primary outcome of the iROC trial which is embargoed until its completion.

All deaths reported in this cohort of patients were cancer-related. It is therefore difficult to distinguish if recovery of mobility at 90 days post-operatively was predictive of overall survival or cancer-specific survival. This could be partly attributed to the small sample size of 52 patients, and a low event-count of 7 deaths (13.5%). This is a limitation of the current study, and a similar analysis will be performed after the completion of 1-year follow-up for all patients recruited into the iROC trial.

8.6 Conclusions

Average step-count outperform maximum step-count as a metric in predicting one-year survival, suggesting that measuring a patient's mobility in the routine home setting could provide an early warning sign regarding cancer-specific or overall survival. To my knowledge, this relationship has not previously been reported in the literature. While the sample size of this study is relatively small, these results are hypothesis-generating and will be validated after the completion of the trial, which is expected to finish data collection in 2021. If these results are consistent in the larger iROC cohort (post-operative step counts can predict metastatic recurrence), this approach could be used alongside CT scanning and other new tests such as genomic biomarkers. This functional 'biomarker' could be used to select patients for early adjuvant chemotherapy or immunotherapy, which may directly impact patient prognosis. This relationship would need to be tested in a multi-arm randomised trial, with patients randomised to additional therapy or standard of care if they are below the 50% recovery threshold. Such a trial would need to be a large sample size given that in our current cohort, only 13.5% of patients died of recurrence in the first year.

Chapter 9 Overall conclusion

The results in this thesis suggest that wearable devices offer a new method to collect health status information on patients. These devices require little to no interaction from the patient, and offer a continuous stream of data that offers new insight into real time health status. Wearable devices may offer different utility at different stages in the patient journey: pre-operative assessment, post-operative measurement of recovery and predicting survival.

The iROC trial was an ideal trial to embed the fitness tracker sub-study, as it recruits patients who undergo a major operation with an aim to measure post-operative morbidity associated with the two different approaches (ORC and RARC). At the time of submission, 283 of 340 patients (83%) have been randomised. Recruitment is expected to finish in 2020, with full one year analysis of secondary outcomes to be completed in 2021.

PROMs such as the EORTC QLQ-C30 questionnaire are commonly used validated research instruments that offer insight into patient physical functioning in the home setting. In my analysis, I compared step-count with EORTC QLQ-C30 questionnaire scores during the peri-operative period. As expected, patient mobility was significantly reduced immediately post-RC and approached baseline mobility by the end of 90 days, which shows a similar trend to the EORTC QLQ-C30 questionnaire scores. However, the changes in questionnaire scores and step-count from baseline do not correlate significantly. This finding suggests that the objective measurement of step-count offers a new dimension into patient recovery that can supplement information gathered through PROM questionnaires.

CPET is a widely adopted pre-operative assessment tool used to risk stratify patients undergoing RC. My analysis showed that step-count correlated significantly with key CPET variables. However, logistic regression showed that only step-count was an independent predictor of major complications post-RC, and not other variables including CPET. These results suggest that an inexpensive wearable device may be of greater utility in risk stratification of patients undergoing RC, compared to traditional pre-operative metrics including CPET.

In current clinical practice, surveillance computer tomography scanning is performed three times in the first year (3, 6, 12 months) and yearly thereafter to detect any metastatic recurrences. I evaluated recovery of mobility measured by a wearable device in predicting survival at 1 year post-RC, and reported that none of the patients who recovered at least 50% of their baseline activity levels by 3 months-RC went on to develop metastatic disease. Recovery of average daily step-count by three months had a sensitivity and specificity of 100% and 93% respectively when a 50% recovery threshold was applied. With new therapies such as immunotherapy being available to patients with bladder cancer, wearable devices may offer an avenue for patient selection. Certain cancers such as lung cancers use genomic testing prior to therapy selection for patients, and this functional biomarker of physical activity could be used in tandem with such laboratory tests to better select patients and directly impact patient outcomes.

Since embarking on the iROC trial and all the undertaken experiments described, the technology in wearable devices has vastly improved, with additional health related features, longer battery life and less expensive devices. Additionally, the increased uptake of smartphones among patients will also allow for continuous synchronisation of data that can be uploaded to a cloud server in real time, negating data storage as a limiting factor for data collection. As we gather large amounts of data on patient mobility, heart rate trends, temperature variability and other metrics, there is no doubt that wearable devices will be an integral part of healthcare in the next decade.

Chapter 10 Future work

The cumulative work undertaken in this thesis highlights the value wearable devices offer in the peri-operative period for patients undergoing RC or other major surgery. Step-count data from wearable devices correlates with standard pre-operative assessments such as CPET, has been associated with major complications in the peri-operative period as well as 1-year survival. An important aspect of my future work will be to repeat the analysis offered in this thesis on the entire iROC cohort after the completion of the study with a larger sample size. Once the trial has completed recruitment, it would also be valuable to compare mobility for ORC and RARC using wearable device-measured step-count.

Wearable devices have become more affordable and additional health features have been included in newer devices. The increased uptake of smartphones also means that patients have access to their own smartphones that can directly synchronise data into the cloud at regular intervals, which means data could be reviewed and actioned more quickly. An important application of this advancement is monitoring patients remotely after their discharge from hospital following RC and other major surgery. To this end, I have setup a new prospective observational study called Domiciliary recovery after medicalisation Pathway[236] (DREAMPath) which is funded by The Urology Foundation and the St Peter's Trust charities. DREAMPath is a prospective observational study to measure patient compliance with remote monitoring following discharge from hospital after major surgery using the Apple Watch Series 4, Bluetooth enabled devices (sphygmomanometer, thermometer and pulse oximeter) and an iPhone. These devices were chosen to mimic the standard measures used in hospital to monitor patients as part of an early warning score[237]. All data is collected on a cloud-based platform in real time. The primary objective of the study is to measure patient compliance with these measures at home during the first 30 days after discharge. Additionally, an important secondary objective is to measure if the physiological and PROM measures collected can predict hospital re-admissions during this high-risk period. In an early interim analysis, nearly 6,000 data points

were collected per patients and a remote early warning score could offer a 48-hour lead time over patient-led Accident & Emergency department attendance.

After the completion of DREAMPath, the next step will be to assess if an easily delivered intervention could rescue these patients from failure by triaging patients for hospital attendance based on the early warning score. I am working closely with Professor John Kelly and Professor James Catto to develop the trial protocol and apply for grant funding. This study will reduce the number of devices that patients will engage with, and test standard of care patient-led return to hospital with a remotely delivered clinician-led return to hospital.

While the wearable-device sub-study in the iROC trial is likely to allude to any differences in mobility between ORC and RARC, another independent prospective case-control study is planned to compare mobility at baseline and for a period of 30 consecutive days with a wearable device that can measure step-count, heart rate, sleep and other health data that patients will synchronise to their smartphones and this will be uploaded and logged automatically for a continuous 30 day period, as opposed to the seven day period collected in the iROC trial. A grant proposal for this work was submitted to Intuitive Surgical in June 2019, and I have been shortlisted for the final round of applications with a decision due in November 2019.

While wearable devices offer exciting opportunities in peri-operative recovery, I believe they have the potential to benefit patients in other specialties as well. For example, wearable devices that can measure temperature changes could be useful in monitoring patients undergoing chemotherapy who are at high risk of neutropenic sepsis. To this end, there is ongoing work to embed wearable device data into patient electronic healthcare records at University College London Hospital. I am working closely with Professor John Kelly and Professor Ramani Moonesinghe to develop this project in the coming years.

My time as a doctoral student has enabled me to explore ideas about applying wearable devices in the peri-operative pathway in a systematic way, and generate hypothesis that will be tested

in future studies. As the field of wearable devices grows and patient acceptance to such technologies improves, I have no doubt that wearable devices will provide new insight into patient wellness in peri-operative and other settings in healthcare.

References

[1] Yeung C, Dinh T, Lee J. The Health Economics of Bladder Cancer: An Updated Review of the Published Literature. Pharmacoeconomics 2014;32:1093–104. doi:10.1007/s40273-014-0194-2.

[2] Leal J, Luengo-Fernandez R, Sullivan R, Witjes JA. Economic Burden of Bladder Cancer Across the European Union. Eur Urol 2016;69:438–47. doi:10.1016/j.eururo.2015.10.024.

[3] Botteman MF, Pashos CL, Redaelli A, Laskin B, Hauser R. The health economics of bladder cancer: a comprehensive review of the published literature. Pharmacoeconomics 2003;21:1315–30.

[4] Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. Eur Urol 2017;71:96–108. doi:10.1016/j.eururo.2016.06.010.

[5] Cancer Research UK. Bladder cancer mortality Statistics n.d. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bladder-cancer/survival (accessed June 24, 2016).

[6] Freedman ND, Silverman DT, Hollenbeck AR, Schatzkin A, Abnet CC. Association Between Smoking and Risk of Bladder Cancer Among Men and Women. JAMA 2011;306:737. doi:10.1001/jama.2011.1142.

[7] Cumberbatch MG, Rota M, Catto JWF, La Vecchia C. The Role of Tobacco Smoke in Bladder and Kidney Carcinogenesis: A Comparison of Exposures and Meta-analysis of Incidence and Mortality Risks. Eur Urol 2016;70:458–66. doi:10.1016/j.eururo.2015.06.042.

[8] Rushton L, Hutchings SJ, Fortunato L, Young C, Evans GS, Brown T, et al. Occupational cancer burden in Great Britain. Br J Cancer 2012;107:S3–7. doi:10.1038/bjc.2012.112.

[9] Ploeg M, Aben KKH, Kiemeney LA. The present and future burden of urinary bladder cancer in the world. World J Urol 2009;27:289–93. doi:10.1007/s00345-009-0383-3.

[10] Vaidya A, Soloway MS, Hawke C, Tiguert R, Civantos F. De novo muscle invasive bladder cancer: Is there a change in trend? J Urol 2001;165:47–50. doi:10.1097/00005392-200101000-00012.

[11] May M, Stief C, Brookman-May S, Otto W, Gilfrich C, Roigas J, et al. Gender-dependent cancer-specific survival following radical cystectomy. World J Urol 2012;30:707–13. doi:10.1007/s00345-011-0773-1.

[12] Johansson SL, Cohen SM. Epidemiology and etiology of bladder cancer. Semin Surg Oncol 1997;13:291–8. doi:10.1002/(SICI)1098-2388(199709/10)13:5<291::AID-SSU2>3.0.CO;2-8.

[13] KANTOR AF, HARTGE P, HOOVER RN, NARAYANA AS, SULLIVAN JW, FRAUMENI JF. URINARY TRACT INFECTION AND RISK OF BLADDER CANCER. Am J Epidemiol 1984;119:510–5. doi:10.1093/oxfordjournals.aje.a113768.

[14] Tan WS, Sarpong R, Khetrapal P, Rodney S, Mostafid H, Cresswell J, et al. Can renal and bladder ultrasound replace CT urogram in patients investigated for microscopic hematuria? J Urol 2018. doi:10.1016/J.JURO.2018.04.065.

[15] National Institute for Health and Care Excellence. Urological cancers - recognition and referral - NICE CKS 2015. https://cks.nice.org.uk/urological-cancers-recognition-and-referral#!references (accessed August 16, 2018).

[16] The Union for International Cancer Control. TNM classification of malignant tumors. n.d. https://www.uicc.org/resources/tnm (accessed January 9, 2018).

[17] Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part B: Prostate and Bladder Tumours. Eur Urol 2016;70:106–19. doi:10.1016/j.eururo.2016.02.028.

[18] Oosterlinck W, Lobel B, Jakse G, Malmström PU, Stöckle M, Sternberg C, et al. Guidelines on non-muscle invasive Bladder Cancer. Eur Urol 2015;41:105–112.

[19] Shanks JH, Chandra A, McWilliam L, Varma M. Standards and datasets for reporting cancers 2013. https://www.rcpath.org/uploads/assets/e2c11ff6-780a-471e-a21a4dc48788d35b/dataset-for-tumours-of-the-urinary-collecting-system-renal-pelvis-ureter-urinary-bladder-and-urethra.pdf.

[20] Babjuk M, Böhle A, Burger M, Compérat E, Kaasinen E, Palou J, et al. EAU Guidelines on Bladder Cancer 2016:1–45.

[21] Ghoneim MA, Abdel-Latif M, El-Mekresh M, Abol-Enein H, Mosbah A, Ashamallah A, et al. Radical cystectomy for carcinoma of the bladder: 2,720 consecutive cases 5 years later. J Urol 2008;180:121–7. doi:10.1016/j.juro.2008.03.024.

[22] Stein BJP, Lieskovsky G, Cote R, Groshen S, Feng A, Boyd S, et al. Radical Cystectomy in the Treatment of Invasive Bladder Cancer : Long-Term Results in 1, 054 Patients 2001;19:666–75.

[23] Advanced Bladder Cancer (ABC) Meta-analysis Collaboration. Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration. Eur Urol 2005;48:202–5; discussion 205-6. doi:10.1016/j.eururo.2005.04.006.

[24] Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar A V, Necchi A, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet (London, England) 2016;387:1909–20. doi:10.1016/S0140-6736(16)00561-4.

[25] Shelley M, Barber J, Wilt TJ, Mason M. Surgery versus radiotherapy for muscle invasive bladder cancer. Cochrane Database Syst Rev 2001. doi:10.1002/14651858.CD002079.

[26] BLOOM HJG, HENDRY WF, WALLACE DM, SKEET RG. Treatment of T3 Bladder Cancer: Controlled Trial of Pre-operative Radiotherapy and Radical Cystectomy Versus Radical Radiotherapy: Second Report and Review (for the Clinical Trials Group, Institute of Urology). Br J Urol 1982;54:136–51. doi:10.1111/j.1464-410X.1982.tb13537.x.

[27] Sell A, Jakobsen A, Nerstrøm B, Sørensen BL, Steven K, Barlebo H. Treatment of advanced bladder cancer category T2 T3 and T4a. A randomized multicenter study of preoperative irradiation and cystectomy versus radical irradiation and early salvage cystectomy for residual tumor. DAVECA protocol 8201. Danish Vesical Cancer Group. Scand J Urol Nephrol Suppl 1991;138:193–201.

[28] Miller LS. Bladder cancer. Superiority of preoperative irradiation and cystectomy in clinical stages B2 and C. Cancer 1977;39:973–80. doi:10.1002/1097-0142(197702)39:2+<973::AID-CNCR2820390737>3.0.CO;2-O.

[29] Cresswell J, Mariappan P, Thomas SA, Khan MS, Johnson MI, Fowler S. Radical cystectomy: Analysis of trends in UK practice 2004–2012, from the British Association of Urological Surgeons' (BAUS) Section of Oncology Dataset. J Clin Urol 2016;9:48–56. doi:10.1177/2051415815595325.

[30] Sánchez de Badajoz E, Gallego Perales JL, Reche Rosado A, Gutierrez de la Cruz JM,

Jimenez Garrido A. Laparoscopic cystectomy and ileal conduit: case report. J Endourol 1995;9:59–62. doi:10.1089/end.1995.9.59.

[31] Khadhouri S, Miller C, Fowler S, Hounsome L, McNeill A, Adshead J, et al. The British Association of Urological Surgeons (BAUS) radical prostatectomy audit 2014/2015 – an update on current practice and outcomes by centre and surgeon case-volume. BJU Int 2018;121:886–92. doi:10.1111/bju.14156.

[32] Murphy DG, Challacombe BJ, Elhage O, O'Brien TS, Rimington P, Khan MS, et al. Robotic-assisted Laparoscopic Radical Cystectomy with Extracorporeal Urinary Diversion: Initial Experience. Eur Urol 2008;54:570–80. doi:10.1016/j.eururo.2008.04.011.

[33] Pruthi RS, Nix J, McRackan D, Hickerson A, Nielsen ME, Raynor M, et al. Robotic-assisted laparoscopic intracorporeal urinary diversion. Eur Urol 2010;57:1013–21. doi:10.1016/j.eururo.2009.12.028.

[34] Khetrapal P, Kelly JD, Catto JWF, Vasdev N. Does the robot have a role in radical cystectomy? BJU Int 2018. doi:10.1111/bju.14579.

[35] Shabsigh A, Korets R, Vora KC, Brooks CM, Cronin AM, Savage C, et al. Defining Early Morbidity of Radical Cystectomy for Patients with Bladder Cancer Using a Standardized Reporting Methodology. Eur Urol 2009;55:164–76. doi:10.1016/J.EURURO.2008.07.031.

[36] Parekh DJ, Reis IM, Castle EP, Gonzalgo ML, Woods ME, Svatek RS, et al. Robot-assisted radical cystectomy versus open radical cystectomy in patients with bladder cancer (RAZOR): an open-label , randomised , phase 3 , non-inferiority trial. Lancet 2018;391:2525–36. doi:10.1016/S0140-6736(18)30996-6.

[37] Novara G, Catto JWF, Wilson T, Annerstedt M, Chan K, Murphy DG, et al. Systematic review and cumulative analysis of perioperative outcomes and complications after robot-assisted radical cystectomy. Eur Urol 2015;67:376–401. doi:10.1016/j.eururo.2014.12.007.

[38] Khadhouri S, Miller C, Cresswell J, Rowe E, Fowler S, Hounsome L, et al. The BAUS radical cystectomy audit 2014/2015 - an update on current practice and an analysis of the effect of centre and surgeon case volume. Eur Urol Suppl 2017;16:e473–4. doi:10.1016/S1569-9056(17)30337-8.

[39] Bochner BH, Dalbagni G, Sjoberg DD, Silberstein J, Keren Paz GE, Donat SM, et al. Comparing Open Radical Cystectomy and Robot-assisted Laparoscopic Radical Cystectomy: A Randomized Clinical Trial. Eur Urol 2015;67:1042–50. doi:10.1016/j.eururo.2014.11.043.

[40] Nix J, Smith A, Kurpad R, Nielsen ME, Wallen EM, Pruthi RS. Prospective Randomized Controlled Trial of Robotic versus Open Radical Cystectomy for Bladder Cancer : Perioperative and Pathologic Results. Eur Urol 2010;57:196–201. doi:10.1016/j.eururo.2009.10.024.

[41] Parekh DJ, Messer J, Fitzgerald J, Ercole B, Svatek R. Perioperative outcomes and oncologic efficacy from a pilot prospective randomized clinical trial of open versus robotic assisted radical cystectomy. J Urol 2013;189:474–9. doi:10.1016/j.juro.2012.09.077.

[42] Khan MS, Gan C, Ahmed K, Ismail AF, Watkins J, Summers JA, et al. A Single-centre Early Phase Randomised Controlled Three-arm Trial of Open, Robotic, and Laparoscopic Radical Cystectomy (CORAL). Eur Urol 2016;69:613–21. doi:10.1016/j.eururo.2015.07.038.

[43] Bochner BH, Dalbagni G, Marzouk KH, Sjoberg DD, Lee J, Donat SM, et al. Randomized Trial Comparing Open Radical Cystectomy and Robot-assisted Laparoscopic Radical Cystectomy : Oncologic Outcomes. Eur Urol 2018:1–7. doi:10.1016/j.eururo.2018.04.030.

[44] Smith ND, Castle EP, Gonzalgo ML, Svatek RS, Weizer AZ, Montgomery JS, et al. The RAZOR (randomized open vs robotic cystectomy) trial: Study design and trial update. BJU Int 2015;115:198–205. doi:10.1111/bju.12699.

[45] Tyritzis SI, Hosseini A, Collins J, Nyberg T, Jonsson MN, Laurin O, et al. Oncologic, functional, and complications outcomes of robot-assisted radical cystectomy with totally intracorporeal neobladder diversion. Eur Urol 2013;64:734–41. doi:10.1016/j.eururo.2013.05.050.

[46] Tan WS, Sridhar A, Ellis G, Lamb B, Goldstraw M, Nathan S, et al. Analysis of open and intracorporeal robotic assisted radical cystectomy shows no significant difference in recurrence patterns and oncological outcomes. Urol Oncol 2016:1–9. doi:10.1016/j.urolonc.2016.02.010.

[47] Lamb BW, Tan WS, Eneje P, Bruce D, Jones A, Ahmad I, et al. Benefits of robotic cystectomy with intracorporeal diversion for patients with low cardiorespiratory fitness: A prospective cohort study. Urol Oncol Semin Orig Investig 2016;34:1–7. doi:10.1016/j.urolonc.2016.04.006.

[48] Ahmed K, Khan SA, Hayn MH, Agarwal PK, Badani KK, Derya Balbay M, et al. Analysis of intracorporeal compared with extracorporeal urinary diversion after robot-assisted radical cystectomy: Results from the international robotic cystectomy consortium. Eur Urol 2014;65:340–7. doi:10.1016/j.eururo.2013.09.042.

[49] Bansal SS, Dogra T, Smith PW, Amran M, Auluck I, Bhambra M, et al. Cost analysis of open radical cystectomy versus robot-assisted radical cystectomy. BJU Int 2018;121:437–44. doi:10.1111/bju.14044.

[50] Smith A, Kurpad R, Lal A, Nielsen M, Wallen EM, Pruthi RS. Cost Analysis of Robotic Versus Open Radical Cystectomy for Bladder Cancer. J Urol 2010;183:505–9. doi:10.1016/J.JURO.2009.09.081.

[51] Richards KA, Kader K, Pettus JA, Smith JJ, Hemal AK. Does initial learning curve compromise outcomes for robot-assisted radical cystectomy? A critical evaluation of the first 60 cases while establishing a robotics program. J Endourol 2011;25:1553–8. doi:10.1089/end.2010.0630.

[52] Hollenbeck BK, Miller DC, Taub DA, Dunn RL, Khuri SF, Henderson WG, et al. The Effects of Adjusting for Case Mix on Mortality and Length of Stay Following Radical Cystectomy. J Urol 2006;176:1363–8. doi:10.1016/j.juro.2006.06.015.

[53] Leow JJ, Reese S, Trinh QD, Bellmunt J, Chung BI, Kibel AS, et al. Impact of surgeon volume on the morbidity and costs of radical cystectomy in the USA: A contemporary population-based analysis. BJU Int 2015;115:713–21. doi:10.1111/bju.12749.

[54] Ghaferi AA, Birkmeyer JD, Dimick JB. Hospital volume and failure to rescue with high-risk surgery. Med Care 2011;49:1076–81. doi:10.1097/MLR.0b013e3182329b97.

[55] Kulaksizoglu H, Toktas G, Kulaksizoglu IB, Aglamis E, Ünlüer E. When Should Quality of Life be Measured after Radical Cystectomy? Eur Urol 2002;42:350–5. doi:10.1016/S0302-2838(02)00351-2.

[56] Raza SJ, Wilson T, Peabody JO, Wiklund P, Scherr DS, Al-Daghmin A, et al. Long-term Oncologic Outcomes Following Robot-assisted Radical Cystectomy: Results from the International Robotic Cystectomy Consortium. Eur Urol 2015;68:4–11. doi:10.1016/j.eururo.2015.04.021.

[57] Hu JC, Chughtai B, O'Malley P, Halpern JA, Mao J, Scherr DS, et al. Perioperative Outcomes, Health Care Costs, and Survival After Robotic-assisted Versus Open Radical Cystectomy: A National Comparative Effectiveness Study. Eur Urol 2016:1–8. doi:10.1016/j.eururo.2016.03.028.

[58] Afshar M, Goodfellow H, Jackson-Spence F, Evison F, Parkin J, Bryan RT, et al. Centralisation of radical cystectomies for bladder cancer in England, a decade on from the 'Improving Outcomes Guidance': the case for super centralisation. BJU Int 2018;121:217–24. doi:10.1111/bju.13929.

[59] NHS England. Enhanced Recovery Programme - NHS Institute for Innovation and Improvement n.d. http://www.institute.nhs.uk/quality_and_service_improvement_tools/quality_and_service_improvement_tools/enhanced_recovery_programme.html (accessed June 24, 2016).

[60] Dutton TJ, Daugherty MO, Mason RG, McGrath JS. Implementation of the Exeter enhanced recovery programme for patients undergoing radical cystectomy. BJU Int 2014;113:719–25. doi:10.1111/bju.12533.

[61] Smith J, Meng ZW, Lockyer R, Dudderidge T, McGrath J, Hayes M, et al. Evolution of the Southampton Enhanced Recovery Programme for radical cystectomy and the aggregation of marginal gains. BJU Int 2014;114:375–83. doi:10.1111/bju.12644.

[62] British Association of Urological Surgeons. Radical Cystectomy n.d. http://www.baus.org.uk/_userfiles/pages/files/Publications/Cystectomy.pdf (accessed June 24, 2016).

[63] Collins JW, Patel H, Adding C, Annerstedt M, Dasgupta P, Khan SM, et al. Enhanced Recovery After Robot-assisted Radical Cystectomy: EAU Robotic Urology Section Scientific Working Group Consensus View. Eur Urol 2016;70:649–660. doi:10.1016/j.eururo.2016.05.020.

[64] NICE. Improving outcomes in urological cancers 2002. https://www.nice.org.uk/guidance/csg2.

[65] Hounsome LS, Verne J, McGrath JS, Gillatt DA. Trends in operative caseload and mortality rates after radical cystectomy for bladder cancer in England for 1998-2010. Eur Urol 2015;67:1056–62. doi:10.1016/j.eururo.2014.12.002.

[66] Goossens-Laan CA, Gooiker GA, Van Gijn W, Post PN, Bosch JLHR, Kil PJM, et al. A systematic review and meta-analysis of the relationship between hospital/surgeon volume and outcome for radical cystectomy: An update for the ongoing debate. Eur Urol 2011;59:775–83. doi:10.1016/j.eururo.2011.01.037.

[67] Wilson TG, Guru K, Rosen RC, Wiklund P, Annerstedt M, Bochner BH, et al. Best practices in robot-assisted radical cystectomy and urinary reconstruction: Recommendations of the pasadena consensus panel. Eur Urol 2015;67:363–75. doi:10.1016/j.eururo.2014.12.009.

[68] Trinh VQ, Trinh QD, Tian Z, Hu JC, Shariat SF, Perrotte P, et al. In-hospital mortality and failure-to-rescue rates after radical cystectomy. BJU Int 2013;112:20–7. doi:10.1111/bju.12214.

[69] Brooke BS, Goodney PP, Kraiss LW, Gottlieb DJ, Samore MH, Finlayson SRG. Readmission destination and risk of mortality after major surgery: an observational cohort study. Lancet 2015;386:884–95. doi:10.1016/S0140-6736(15)60087-3.

[70] Ahmed K, Khan R, Mottrie A, Lovegrove C, Abaza R, Ahlawat R, et al. Development of a standardised training curriculum for robotic surgery: a consensus statement from an international multidisciplinary group of experts. BJU Int 2015;116:93–101. doi:10.1111/bju.12974.

[71] Volpe A, Ahmed K, Dasgupta P, Ficarra V, Novara G, van der Poel H, et al. Pilot Validation Study of the European Association of Urology Robotic Training Curriculum. Eur Urol 2015;68:292–9. doi:10.1016/j.eururo.2014.10.025.

[72] Fernandes E, Elli E, Giulianotti P. The role of the dual console in robotic surgical training. Surg (United States) 2014;155:1–4. doi:10.1016/j.surg.2013.06.023.

[73] Walliczek-Dworschak U, Mandapathil M, Förtsch A, Teymoortash A, Dworschak P, Werner JA, et al. Structured training on the da Vinci Skills Simulator leads to improvement in technical performance of robotic novices. Clin Otolaryngol 2017;42:71–80. doi:10.1111/coa.12666.

[74] Collins JW, Levy J, Stefanidis D, Gallagher A, Coleman M, Cecil T, et al. Utilising the Delphi Process to Develop a Proficiency-based Progression Train-the-trainer Course for Robotic Surgery Training. Eur Urol 2019;75:775–85. doi:10.1016/j.eururo.2018.12.044.

[75] Djaladat H, Katebian B, Bazargani ST, Miranda G, Cai J, Schuckman AK, et al. 90-Day complication rate in patients undergoing radical cystectomy with enhanced recovery protocol: a prospective cohort study. World J Urol 2017;35:907–11. doi:10.1007/s00345-016-1950-z.

[76] Stimson CJ, Chang SS, Barocas DA, Humphrey JE, Patel SG, Clark PE, et al. Early and Late Perioperative Outcomes Following Radical Cystectomy: 90-Day Readmissions, Morbidity and Mortality in a Contemporary Series. J Urol 2010;184:1296–300. doi:10.1016/J.JURO.2010.06.007.

[77] Weiser TG, Regenbogen SE, Thompson KD, Haynes AB, Lipsitz SR, Berry WR, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. Lancet 2008;372:139–44. doi:10.1016/S0140-6736(08)60878-8.

[78] Weiser TG, Haynes AB, Molina G, Lipsitz SR, Esquivel MM, Uribe-Leitz T, et al. Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. Lancet 2015;385:S11. doi:10.1016/S0140-6736(15)60806-6.

[79] Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in Hospital Mortality Associated with Inpatient Surgery. N Engl J Med 2009;361:1368–75. doi:10.1056/NEJMsa0903048.

[80] Dimick JB, Pronovost PJ, Cowan JA, Lipsett PA, Stanley JC, Upchurch GR. Variation in postoperative complication rates after high-risk surgery in the United States. Surgery 2003;134:534–40. doi:10.1016/S0039-6060(03)00273-3.

[81] Donabedian A. Evaluating the Quality of Medical Care. Milbank Q 2005;83:691. doi:10.1111/J.1468-0009.2005.00397.X.

[82] Emick DM, Riall TS, Cameron JL, Winter JM, Lillemoe KD, Coleman J, et al. Hospital Readmission After Pancreaticoduodenectomy. J Gastrointest Surg 2006;10:1243–53. doi:10.1016/j.gassur.2006.08.016.

[83] Hendren S, Morris AM, Zhang W, Dimick J. Early discharge and hospital readmission after colectomy for cancer. Dis Colon Rectum 2011;54:1362–7. doi:10.1097/DCR.0b013e31822b72d3.

[84] Hopkins TJ, Raghunathan K, Barbeito A, Cooter M, Stafford-Smith M, Schroeder R, et al. Associations between ASA Physical Status and postoperative mortality at 48 h: a contemporary dataset analysis compared to a historical cohort. Perioper Med 2016;5:29. doi:10.1186/s13741-016-0054-z.

[85] Reeves T, Bates S, Sharp T, Richardson K, Bali S, Plumb J, et al. Cardiopulmonary exercise testing (CPET) in the United Kingdom—a national survey of the structure, conduct, interpretation and funding. Perioper Med 2018;7:1–8. doi:10.1186/s13741-018-0087-6.

[86] Tolchard S, Angell J, Pyke M, Lewis S, Dodds N, Darweish A, et al. Cardiopulmonary reserve as determined by cardiopulmonary exercise testing correlates with length of stay and predicts complications after radical cystectomy. BJU Int 2015;115:554–61. doi:10.1111/bju.12895.

[87] Prentis JM, Trenell MI, Vasdev N, French R, Dines G, Thorpe A, et al. Impaired

cardiopulmonary reserve in an elderly population is related to postoperative morbidity and length of hospital stay after radical cystectomy. BJU Int 2013;112:13–9. doi:10.1111/bju.12219.

[88] Moran J, Wilson F, Guinan E, McCormick P, Hussey J, Moriarty J. Role of cardiopulmonary exercise testing as a risk-assessment method in patients undergoing intra-abdominal surgery: a systematic review. Br J Anaesth 2016;116:177–91. doi:10.1093/bja/aev454.

[89] Lam S, Hart A. Cardiopulmonary Exercise Testing for Predicting Early Outcomes after Major Cancer Resection: A Systematic Review. J Anesth Perioper Med 2018;5:136–48. doi:10.24015/japm.2018.0052.

[90] Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001;56:M146-56.

[91] Musallam KM, Tamim HM, Richards T, Spahn DR, Rosendaal FR, Habbal A, et al. Preoperative anaemia and postoperative outcomes in non-cardiac surgery: a retrospective cohort study. Lancet 2011;378:1396–407. doi:10.1016/S0140-6736(11)61381-0.

[92] Whitlock EL, Kim H, Auerbach AD. Harms associated with single unit perioperative transfusion: retrospective population based analysis. BMJ 2015;350:h3037–h3037. doi:10.1136/bmj.h3037.

[93] Glance LG, Dick AW, Mukamel DB, Fleming FJ, Zollo RA, Wissler R, et al. Association between Intraoperative Blood Transfusion and Mortality and Morbidity in Patients Undergoing Noncardiac Surgery. Anesthesiology 2011;114:283–92. doi:10.1097/ALN.0b013e3182054d06.

[94] Tan WSWS, Lamb BWBW, Khetrapal P, Tan M-YMEMEM-Y, Tan M-YMEMEM-Y, Sridhar A, et al. Blood Transfusion Requirement and Not Preoperative Anemia Are Associated with Perioperative Complications Following Intracorporeal Robot-Assisted Radical Cystectomy. J Endourol 2017;31:end.2016.0730. doi:10.1089/end.2016.0730.

[95] Xia L, Guzzo TJ. Preoperative Anemia and Low Hemoglobin Level Are Associated With Worse Clinical Outcomes in Patients With Bladder Cancer Undergoing Radical Cystectomy: A Meta-Analysis. Clin Genitourin Cancer 2017;15:263-272.e4. doi:10.1016/j.clgc.2016.08.017.

[96] Clavien P a, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg 2009;250:187–96. doi:10.1097/SLA.0b013e3181b13ca2.

[97] Yang LS, Shan BL, Shan LL, Chin P, Murray S, Ahmadi N, et al. A systematic review and meta-analysis of quality of life outcomes after radical cystectomy for bladder cancer. Surg Oncol 2016;25:281–97. doi:10.1016/j.suronc.2016.05.027.

[98] Lauridsen SV, Tønnesen H, Jensen BT, Neuner B, Thind P, Thomsen T. Complications and health-related quality of life after robot-assisted versus open radical cystectomy: a systematic review and meta-analysis of four RCTs. Syst Rev 2017;6:150. doi:10.1186/s13643-017-0547-y.

[99] Messer JC, Punnen S, Fitzgerald J, Svatek R, Parekh DJ. Health-related quality of life from a prospective randomised clinical trial of robot-assisted laparoscopic vs open radical cystectomy. BJU Int 2014;114:896–902. doi:10.1111/bju.12818.

[100] P&S Market Research. Wearable Fitness Trackers Market Size, Share, Forecast to 2023 2018. https://www.psmarketresearch.com/market-analysis/wearable-fitness-trackers-market (accessed May 11, 2018).

[101] Farr C. You can get an Apple Watch for only $25 ... with one small catch. CNBC 2017. https://www.cnbc.com/2017/10/23/apple-watches-offered-to-all-john-hancock-life-

insurance-customers.html (accessed November 3, 2018).

[102] Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of Wearable Technology Combined With a Lifestyle Intervention on Long-term Weight Loss: The IDEA Randomized Clinical Trial. Jama 2016;316:806–14. doi:10.1001/JAMA.2016.12858.

[103] Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EEK, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016;4:983–95. doi:10.1016/S2213-8587(16)30284-4.

[104] Polzien KM, Jakicic JM, Tate DF, Otto AD. The Efficacy of a Technology-based System in a Short-term Behavioral Weight Loss Intervention*. Obesity 2007;15:825–30. doi:10.1038/oby.2007.584.

[105] Fearon KCH, Ljungqvist O, Von Meyenfeldt M, Revhaug A, Dejong CHC, Lassen K, et al. Enhanced recovery after surgery: A consensus review of clinical care for patients undergoing colonic resection. Clin Nutr 2005;24:466–77. doi:10.1016/J.CLNU.2005.02.002.

[106] Shrime MG, Dare A, Alkire BC, Meara JG. A global country-level comparison of the financial burden of surgery. Br J Surg 2016;103:1453–61. doi:10.1002/bjs.10249.

[107] Ariti CA, Cleland JGF, Pocock SJ, Pfeffer MA, Swedberg K, Granger CB, et al. Days alive and out of hospital and the patient journey in patients with heart failure: Insights from the Candesartan in Heart failure: Assessment of Reduction in Mortality and morbidity (CHARM) program. Am Heart J 2011;162:900–6. doi:10.1016/j.ahj.2011.08.003.

[108] Statista. Wearable device revenue worldwide 2016-2022 (in billion U.S. dollars). Statista 2017:1. https://www.statista.com/statistics/610447/wearable-device-revenue-worldwide (accessed January 19, 2019).

[109] PROSPERO. Systematic review of the use of fitness trackers in healthcare studies 2018. http://www.mcia.co.uk/page/motorcycle-industry-association-reaction-to-brexit (accessed January 5, 2019).

[110] Cochrane. RoB 2.0: A revised tool to assess risk of bias in randomized trials [webinar] | Cochrane Training n.d. https://training.cochrane.org/resource/rob-20-webinar (accessed March 30, 2019).

[111] Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919. doi:10.1136/bmj.i4919.

[112] Byakika-Kibwika P, Muwonge M, Watts W, Kange J, Watts R. Lessons Learned from Implementing a Rapid Test of a Technology Device in a Tertiary Hospital in Uganda. Ann Glob Heal 2015;81:725–30. doi:10.1016/j.aogh.2015.12.009.

[113] Phan TLT, Barnini N, Xie S, Martinez A, Falini L, Abatemarco A, et al. Feasibility of using a commercial fitness tracker as an adjunct to family-based weight management treatment: Pilot randomized trial. JMIR MHealth UHealth 2018;6:1–13. doi:10.2196/10523.

[114] Heale LD, Dover S, Goh YI, Maksymiuk VA, Wells GD, Feldman BM. A wearable activity tracker intervention for promoting physical activity in adolescents with juvenile idiopathic arthritis: a pilot study. Pediatr Rheumatol 2018;16:66. doi:10.1186/s12969-018-0282-5.

[115] Klassen TD, Semrau JA, Dukelow SP, Bayley MT, Hill MD, Eng JJ. Consumer-Based Physical Activity Monitor as a Practical Way to Measure Walking Intensity during Inpatient Stroke Rehabilitation. Stroke 2017;48:2614–7. doi:10.1161/STROKEAHA.117.018175.

[116] Gordia AP, de Quadros TMB, Mota J, Silva LR. Number of Daily Steps to Discriminate Abdominal Obesity in a Sample of Brazilian Children and Adolescents. Pediatr Exerc Sci

2017;29:121–30. doi:10.1123/pes.2016-0078.

[117] Voss C, Gardner RF, Dean PH, Harris KC. Validity of Commercial Activity Trackers in Children With Congenital Heart Disease. Can J Cardiol 2017;33:799–805. doi:10.1016/j.cjca.2016.11.024.

[118] Tran VD, Lee AH, Jancey J, James AP, Howat P, Mai LTP. Physical activity and nutrition behaviour outcomes of a cluster-randomized controlled trial for adults with metabolic syndrome in Vietnam. Trials 2017;18:18. doi:10.1186/s13063-016-1771-9.

[119] Takahashi PY, Quigg SM, Croghan IT, Schroeder DR, Ebbert JO. Effect of pedometer use and goal setting on walking and functional status in overweight adults with multimorbidity: A crossover clinical trial. Clin Interv Aging 2016;11:1099–106. doi:10.2147/CIA.S107626.

[120] Valle CG, Deal AM, Tate DF. Preventing weight gain in African American breast cancer survivors using smart scales and activity trackers: a randomized controlled pilot study. J Cancer Surviv 2017;11:133–48. doi:10.1007/s11764-016-0571-2.

[121] Miyauchi M, Toyoda M, Kaneyama N, Miyatake H, Tanaka E, Kimura M, et al. Exercise Therapy for Management of Type 2 Diabetes Mellitus: Superior Efficacy of Activity Monitors over Pedometers. J Diabetes Res 2016;2016:1–7. doi:10.1155/2016/5043964.

[122] Gordon R, Bloxham S. Influence of the Fitbit Charge HR on physical activity, aerobic fitness and disability in non-specific back pain participants. J Sports Med Phys Fitness 2017;57:1669–75. doi:10.23736/S0022-4707.17.06688-9.

[123] Thomas JG, Raynor HA, Bond DS, Luke AK, Cardoso CC, Foster GD, et al. Weight loss in Weight Watchers Online with and without an activity tracking device compared to control: A randomized trial. Obesity 2017;25:1014–21. doi:10.1002/oby.21846.

[124] Li LC, Sayre EC, Xie H, Clayton C, Feehan LM. A Community-Based Physical Activity Counselling Program for People With Knee Osteoarthritis: Feasibility and Preliminary Efficacy of the Track-OA Study. JMIR MHealth UHealth 2017;5:e86. doi:10.2196/mhealth.7863.

[125] Williams S, Han M, Ye X, Zhang H, Meyring-Wösten A, Bonner M, et al. Physical Activity and Sleep Patterns in Hemodialysis Patients in a Suburban Environment. Blood Purif 2017;43:235–43. doi:10.1159/000452751.

[126] Han M, Williams S, Mendoza M, Ye X, Zhang H, Calice-Silva V, et al. Quantifying Physical Activity Levels and Sleep in Hemodialysis Patients Using a Commercially Available Activity Tracker. Blood Purif 2016;41:194–204. doi:10.1159/000441314.

[127] Lynch BM, Nguyen NH, Moore MM, Reeves MM, Rosenberg DE, Boyle T, et al. A randomized controlled trial of a wearable technology-based intervention for increasing moderate to vigorous physical activity and reducing sedentary behavior in breast cancer survivors: The ACTIVATE Trial. Cancer 2019:cncr.32143. doi:10.1002/cncr.32143.

[128] Smith WA, Zucker-Levin A, Mihalko WM, Williams M, Loftin M, Gurney JG. A Randomized Study of Exercise and Fitness Trackers in Obese Patients After Total Knee Arthroplasty. Orthop Clin North Am 2019;50:35–45. doi:10.1016/J.OCL.2018.08.002.

[129] McNeil J, Brenner DR, Stone CR, O'Reilly R, Ruan Y, Vallance JK, et al. Activity Tracker to Prescribe Various Exercise Intensities in Breast Cancer Survivors. Med Sci Sports Exerc 2019;51:930–40. doi:10.1249/MSS.0000000000001890.

[130] Falck RS, Best JR, Li LC, Chan PCY, Feehan LM, Liu-Ambrose T. Can we improve cognitive function among adults with osteoarthritis by increasing moderate-to-vigorous physical activity and reducing sedentary behaviour? Secondary analysis of the MONITOR-OA study. BMC Musculoskelet Disord 2018;19:447. doi:10.1186/s12891-018-2369-z.

[131] Varas AB, Córdoba S, Rodríguez-Andonaegui I, Rueda MR, García-Juez S, Vilaró J.

Effectiveness of a community-based exercise training programme to increase physical activity level in patients with chronic obstructive pulmonary disease: A randomized controlled trial. Physiother Res Int 2018;23:e1740. doi:10.1002/pri.1740.

[132] Duscha BD, Piner LW, Patel MP, Craig KP, Brady M, McGarrah RW, et al. Effects of a 12-week mHealth program on peak VO2 and physical activity patterns after completing cardiac rehabilitation: A randomized controlled trial. Am Heart J 2018;199:105–14. doi:10.1016/j.ahj.2018.02.001.

[133] Van der Walt N, Salmon LJ, Gooden B, Lyons MC, O'Sullivan M, Martina K, et al. Feedback From Activity Trackers Improves Daily Step Count After Knee and Hip Arthroplasty: A Randomized Controlled Trial. J Arthroplasty 2018;33:3422–8. doi:10.1016/j.arth.2018.06.024.

[134] Orme MW, Weedon AE, Saukko PM, Esliger DW, Morgan MD, Steiner MC, et al. Findings of the chronic obstructive pulmonary disease-sitting and exacerbations trial (COPD-SEAT) in reducing sedentary time using wearable and mobile technologies with educational support: Randomized controlled feasibility trial. JMIR MHealth UHealth 2018;6:1–15. doi:10.2196/mhealth.9398.

[135] Katz P, Margaretten M, Gregorich S, Trupin L. Physical Activity to Reduce Fatigue in Rheumatoid Arthritis: A Randomized Controlled Trial. Arthritis Care Res (Hoboken) 2018;70:1–10. doi:10.1002/acr.23230.

[136] Kooiman TJM, De Groot M, Hoogenberg K, Krijnen WP, Van Der Schans CP, Kooy A. Self-tracking of Physical Activity in People with Type 2 Diabetes: A Randomized Controlled Trial. CIN - Comput Informatics Nurs 2018;36:340–9. doi:10.1097/CIN.0000000000000443.

[137] Kanai M, Izawa KP, Kobayashi M, Onishi A, Kubo H, Nozoe M, et al. Effect of accelerometer-based feedback on physical activity in hospitalized patients with ischemic stroke: a randomized controlled trial. Clin Rehabil 2018;32:1047–56. doi:10.1177/0269215518755841.

[138] Mitchell BL, Smith AE, Rowlands A V, Fraysse F, Parfitt G, Lewis NR, et al. Promoting physical activity in rural Australian adults using an online intervention. J Sci Med Sport 2019;22:70–5. doi:10.1016/j.jsams.2018.07.002.

[139] Rossi A, Frechette L, Miller D, Miller E, Friel C, Van Arsdale A, et al. Acceptability and feasibility of a Fitbit physical activity monitor for endometrial cancer survivors. Gynecol Oncol 2018;149:470–5. doi:10.1016/j.ygyno.2018.04.560.

[140] Marthick M, Dhillon HM, Alison JA, Cheema BS, Shaw T. An Interactive Web Portal for Tracking Oncology Patient Physical Activity and Symptoms: Prospective Cohort Study. JMIR Cancer 2018;4:e11978. doi:10.2196/11978.

[141] Xue F, Wang F-Y, Mao C-J, Guo S-P, Chen J, Li J, et al. Analysis of nocturnal hypokinesia and sleep quality in Parkinson's disease. J Clin Neurosci 2018;54:96–101. doi:10.1016/J.JOCN.2018.06.016.

[142] Thijs I, Fresiello L, Oosterlinck W, Sinnaeve P, Rega F. Assessment of Physical Activity by Wearable Technology During Rehabilitation After Cardiac Surgery: Explorative Prospective Monocentric Observational Cohort Study. JMIR MHealth UHealth 2019;7:e9865. doi:10.2196/mhealth.9865.

[143] Knight A, Bidargaddi N. Commonly available activity tracker apps and wearables as a mental health outcome indicator: A prospective observational cohort study among young adults with psychological distress. J Affect Disord 2018;236:31–6. doi:10.1016/J.JAD.2018.04.099.

[144] Champ CE, Ohri N, Klement RJ, Cantor M, Beriwal S, Glaser SM, et al. Assessing Changes in the Activity Levels of Breast Cancer Patients During Radiation Therapy. Clin Breast

Cancer 2018;18:e1–6. doi:10.1016/j.clbc.2017.08.009.

[145] Nyrop KA, Deal AM, Choi SK, Wagoner CW, Lee JT, Wood A, et al. Measuring and understanding adherence in a home-based exercise intervention during chemotherapy for early breast cancer. Breast Cancer Res Treat 2018;168:43–55. doi:10.1007/s10549-017-4565-1.

[146] van Leuteren RW, Dijkhuis S, de Jongh FHC, van der Valk PDLPM, Tabak M, Brusse-Keizer MGJ. The Mozart study: a relation between dynamic hyperinflation and physical activity in patients with chronic obstructive pulmonary disease? Clin Physiol Funct Imaging 2018;38:409–15. doi:10.1111/cpf.12430.

[147] Osadnik CR, Loeckx M, Louvaris Z, Demeyer H, Langer D, Rodrigues FM, et al. The likelihood of improving physical activity after pulmonary rehabilitation is increased in patients with COPD who have better exercise tolerance. Int J Chron Obstruct Pulmon Dis 2018;13:3515–27. doi:10.2147/COPD.S174827.

[148] Le A, Mitchell H-R, Zheng DJ, Rotatori J, Fahey JT, Ness KK, et al. A home-based physical activity intervention using activity trackers in survivors of childhood cancer: A pilot study. Pediatr Blood Cancer 2017;64:387–94. doi:10.1002/pbc.26235.

[149] Wilson M, Ramsay S, Young KJ. Engaging Overweight Adolescents in a Health and Fitness Program Using Wearable Activity Trackers. J Pediatr Heal CARE 2017;31:E25–34. doi:10.1016/j.pedhc.2017.03.001.

[150] Shen H, Zhao J, Zhou X, Li J, Wan Q, Huang J, et al. Impaired chronotropic response to physical activities in heart failure patients. BMC Cardiovasc Disord 2017;17:4–11. doi:10.1186/s12872-017-0571-9.

[151] Kroll RR, McKenzie ED, Boyd JG, Sheth P, Howes D, Wood M, et al. Use of wearable devices for post-discharge monitoring of ICU patients: a feasibility study. J INTENSIVE CARE 2017;5. doi:10.1186/s40560-017-0261-9.

[152] Hooke MC, Gilchrist L, Tanner L, Hart N, Withycombe JS. Use of a Fitness Tracker to Promote Physical Activity in Children With Acute Lymphoblastic Leukemia. Pediatr Blood Cancer 2016;63:684–9. doi:10.1002/pbc.25860.

[153] Pérez-Alenda S, Carrasco JJ, Megías-Vericat JE, Poveda JL, Bonanad S, Querol F. Quantification of physical activity in adult patients with haemophilic arthropathy in prophylaxis treatment using a fitness tracker. Haemophilia 2018;24:e28–32. doi:10.1111/hae.13388.

[154] Abrantes AM, Blevins CE, Battle CL, Read JP, Gordon AL, Stein MD. Developing a Fitbit-supported lifestyle physical activity intervention for depressed alcohol dependent women. J Subst Abuse Treat 2017;80:88–97. doi:10.1016/j.jsat.2017.07.006.

[155] Gardner AW, Montgomery PS, Zhao YD, Silva-Palacios F, Ungvari Z, Csiszar A, et al. Association between daily walking and antioxidant capacity in patients with symptomatic peripheral artery disease. J Vasc Surg 2017;65:1762–8. doi:10.1016/j.jvs.2016.12.108.

[156] Sievi NA, Brack T, Brutsche MH, Frey M, Irani S, Leuppi JD, et al. Accelerometer- versus questionnaire-based assessment of physical activity and their changes over time in patients with COPD. Int J Chron Obstruct Pulmon Dis 2017;12:1113–8. doi:10.2147/COPD.S130195.

[157] Cook JD, Prairie ML, Plante DT. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: A comparison against polysomnography and wrist-worn actigraphy. J Affect Disord 2017;217:299–305. doi:10.1016/j.jad.2017.04.030.

[158] Jacquemin C, Servy H, Molto A, Sellam J, Foltz V, Gandjbakhch F, et al. Physical Activity Assessment Using an Activity Tracker in Patients with Rheumatoid Arthritis and Axial Spondyloarthritis: Prospective Observational Study. JMIR MHealth UHealth 2018;6:e1.

doi:10.2196/mhealth.7948.

[159] Van't Hul AJ, Frouws S, Van Den Akker E, Van Lummel R, Starrenburg-Razenberg A, Van Bruggen A, et al. Decreased physical activity in adults with bronchial asthma. Respir Med 2016;114:72–7. doi:10.1016/j.rmed.2016.03.016.

[160] Peacock JG, Vine RL, McDonald JS, Novotny PJ, Kallmes DF. Armband activity monitor data do not correlate with reported pain scores in patients receiving vertebroplasty. J Neurointerv Surg 2017;9:905–9. doi:10.1136/neurintsurg-2015-012174.

[161] Glaviano NR, Baellow A, Saliba S. Physical activity levels in individuals with and without patellofemoral pain. Phys Ther Sport 2017;27:12–6. doi:10.1016/j.ptsp.2017.07.002.

[162] Colón-Semenza C, Latham NK, Quintiliani LM, Ellis TD. Peer Coaching Through mHealth Targeting Physical Activity in People With Parkinson Disease: Feasibility Study. JMIR MHealth UHealth 2018;6:e42. doi:10.2196/mhealth.8074.

[163] Kuenze C, Lisee C, Pfeiffer KA, Cadmus-Bertram L, Post EG, Biese K, et al. Sex differences in physical activity engagement after ACL reconstruction. Phys Ther Sport 2019;35:12–7. doi:10.1016/J.PTSP.2018.10.016.

[164] Ezeugwu VE, Manns PJ. Sleep Duration, Sedentary Behavior, Physical Activity, and Quality of Life after Inpatient Stroke Rehabilitation. J Stroke Cerebrovasc Dis 2017;26:2004–12. doi:10.1016/j.jstrokecerebrovasdis.2017.06.009.

[165] Simpson CC, Mazzeo SE. Calorie counting and fitness tracking technology: Associations with eating disorder symptomatology. Eat Behav 2017;26:89–92. doi:10.1016/j.eatbeh.2017.02.002.

[166] Wang X, Strizich G, Hua S, Sotres-Alvarez D, Buelna C, Gallo LC, et al. Objectively measured sedentary time and cardiovascular risk factor control in US Hispanics/Latinos with diabetes mellitus: Results from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). J Am Heart Assoc 2017;6. doi:10.1161/JAHA.116.004324.

[167] Dauriz M, Bacchi E, Boselli L, Santi L, Negri C, Trombetta M, et al. Association of free-living physical activity measures with metabolic phenotypes in type 2 diabetes at the time of diagnosis. The Verona Newly Diagnosed Type 2 Diabetes Study (VNDS). Nutr Metab Cardiovasc Dis 2018;28:343–51. doi:10.1016/J.NUMECD.2017.12.011.

[168] McKenna S, Tierney M, O'Neill A, Fraser A, Kennedy N. Sleep and physical activity: a cross-sectional objective profile of people with rheumatoid arthritis. Rheumatol Int 2018;38:845–53. doi:10.1007/s00296-018-4009-1.

[169] Nguyen NH, Hadgraft NT, Moore MM, Rosenberg DE, Lynch C, Reeves MM, et al. A qualitative evaluation of breast cancer survivors' acceptance of and preferences for consumer wearable technology activity trackers. Support Care Cancer 2017;25:3375–84. doi:10.1007/s00520-017-3756-y.

[170] Chum J, Kim MS, Zielinski L, Bhatt M, Chung D, Yeung S, et al. Acceptability of the Fitbit in behavioural activation therapy for depression: a qualitative study. Evid Based Ment Health 2017;20:128+. doi:10.1136/eb-2017-102763.

[171] Randriambelonoro M, Chen Y, Pu P. Can Fitness Trackers Help Diabetic and Obese Users Make and Sustain Lifestyle Changes? Computer (Long Beach Calif) 2017;50:20–9. doi:10.1109/MC.2017.92.

[172] Painter S, Ditsch G, Ahmed R, Hanson NB, Kachin K, Berger J. Retrofit Weight-Loss Outcomes at 6, 12, and 24 Months and Characteristics of 12-Month High Performers: A Retrospective Analysis. JMIR MHealth UHealth 2016;4:e101. doi:10.2196/mhealth.5873.

[173] Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What matters in weight loss? An in-depth analysis of self-monitoring. J Med Internet Res 2017;19. doi:10.2196/jmir.7457.

[174] Wainwright T, Middleton R. An orthopaedic enhanced recovery pathway. Curr Anaesth Crit Care 2010;21:114–20. doi:10.1016/j.cacc.2010.01.003.

[175] Tan WS, Khetrapal P, Tan WP, Rodney S, Chau M, Kelly JD. Robotic Assisted Radical Cystectomy with Extracorporeal Urinary Diversion Does Not Show a Benefit over Open Radical Cystectomy: A Systematic Review and Meta-Analysis of Randomised Controlled Trials. PLoS One 2016;11:e0166221. doi:10.1371/journal.pone.0166221.

[176] Greene FL, American Joint Committee on Cancer., American Cancer Society. AJCC cancer staging manual. n.d.

[177] Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 2005;5:13. doi:10.1186/1471-2288-5-13.

[178] Harrop E, Kelly J, Griffiths G, Casbard A, Nelson A. Why do patients decline surgical trials? Findings from a qualitative interview study embedded in the Cancer Research UK BOLERO trial (Bladder cancer: Open versus Lapararoscopic or RObotic cystectomy). Trials 2016;17:35. doi:10.1186/s13063-016-1173-z.

[179] Fonseka T, Ahmed K, Froghi S, Khan SA, Dasgupta P, Shamim Khan M. Comparing robotic, laparoscopic and open cystectomy: a systematic review and meta-analysis. Arch Ital Di Urol Androl Organo Uff [Di] Soc Ital Di Ecogr Urol e Nefrol 2015;87:41–8. doi:10.4081/aiua.2015.1.41.

[180] Xia L, Wang X, Xu T, Zhang X, Zhu Z, Qin L, et al. Robotic versus open radical cystectomy: An updated systematic review and meta-analysis. PLoS One 2015;10:1–20. doi:10.1371/journal.pone.0121032.

[181] Tang K, Xia D, Li H, Guan W, Guo X, Hu Z, et al. Robotic vs. open radical cystectomy in bladder cancer: A systematic review and meta-analysis. Eur J Surg Oncol 2014;40:1399–411. doi:10.1016/j.ejso.2014.03.008.

[182] Li K, Lin T, Fan X, Xu K, Bi L, Duan Y, et al. Systematic review and meta-analysis of comparative studies reporting early outcomes after robot-assisted radical cystectomy versus open radical cystectomy. Cancer Treat Rev 2013;39:551–60. doi:10.1016/j.ctrv.2012.11.007.

[183] Martin RCG, Brennan MF, Jaques DP. Quality of complication reporting in the surgical literature. Ann Surg 2002;235:803–13.

[184] Tan WS, Lamb BW, Tan M-Y, Ahmad I, Sridhar A, Nathan S, et al. In-depth Critical Analysis of Complications Following Robot-assisted Radical Cystectomy with Intracorporeal Urinary Diversion. Eur Urol Focus 2016:1–7. doi:10.1016/j.euf.2016.06.002.

[185] Hayn MH, Hussain A, Mansour AM, Andrews PE, Carpentier P, Castle E, et al. The learning curve of robot-assisted radical cystectomy: Results from the international robotic cystectomy consortium. Eur Urol 2010;58:197–202. doi:10.1016/j.eururo.2010.04.024.

[186] Thompson JE, Egger S, Böhm M, Haynes A-MM, Matthews J, Rasiah K, et al. Superior quality of life and improved surgical margins are achievable with robotic radical prostatectomy after a long learning curve: A prospective single-surgeon study of 1552 consecutive cases. Eur Urol 2014;65:521–31. doi:10.1016/j.eururo.2013.10.030.

[187] Ou Y-C, Yang C-R, Wang J, Yang C-K, Cheng C-L, Patel VR, et al. The learning curve for reducing complications of robotic-assisted laparoscopic radical prostatectomy by a single surgeon. BJU Int 2011;108:420–5. doi:10.1111/j.1464-410X.2010.09847.x.

[188] Clinical Outcomes of Surgical Therapy Study Group, Nelson H, Sargent DJ, Wieand HS, Fleshman J, Anvari M, et al. A Comparison of Laparoscopically Assisted and Open Colectomy for Colon Cancer. N Engl J Med 2004;350:2050–9. doi:10.1056/NEJMoa032651.

[189] Trinh Q-D, Sammon J, Sun M, Ravi P, Ghani KR, Bianchi M, et al. Perioperative Outcomes of Robot-Assisted Radical Prostatectomy Compared With Open Radical Prostatectomy: Results From the Nationwide Inpatient Sample. Eur Urol 2012;61:679–85. doi:10.1016/j.eururo.2011.12.027.

[190] Hautmann RE, de Petriconi RC, Volkmer BG. Lessons learned from 1,000 neobladders: the 90-day complication rate. J Urol 2010;184:990–4; quiz 1235. doi:10.1016/j.juro.2010.05.037.

[191] Moschini M, Bianchi M, Gandaglia G, Cucchiara V, Luzzago S, Pellucchi F, et al. The Impact of Perioperative Blood Transfusion on Survival of Bladder Cancer Patients Submitted to Radical Cystectomy: Role of Anemia Status. Eur Urol Focus 2016;2:86–91. doi:10.1016/j.euf.2015.03.002.

[192] Dotan ZA, Kavanagh K, Yossepowitch O, Kaag M, Olgac S, Donat M, et al. Positive Surgical Margins in Soft Tissue Following Radical Cystectomy for Bladder Cancer and Cancer Specific Survival. J Urol 2007;178:2308–13. doi:10.1016/j.juro.2007.08.023.

[193] Novara G, Svatek RS, Karakiewicz PI, Skinner E, Ficarra V, Fradet Y, et al. Soft tissue surgical margin status is a powerful predictor of outcomes after radical cystectomy: a multicenter study of more than 4,400 patients. J Urol 2010;183:2165–70. doi:10.1016/j.juro.2010.02.021.

[194] Herr HW, Bochner BH, Dalbagni G, Donat SM, Reuter VE, Bajorin DF. Impact of the number of lymph nodes retrieved on outcome in patients with muscle invasive bladder cancer. J Urol 2002;167:1295–8.

[195] Braga M, Frasson M, Vignali A, Zuliani W, Civelli V, Di Carlo V. Laparoscopic vs. open colectomy in cancer patients: long-term complications, quality of life, and survival. Dis Colon Rectum 2005;48:2217–23. doi:10.1007/s10350-005-0185-7.

[196] Yaxley JW, Coughlin GD, Chambers SK, Occhipinti S, Samaratunga H, Zajdlewicz L, et al. Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study. Lancet 2016;388:1057–66. doi:10.1016/S0140-6736(16)30592-X.

[197] Hussein AA, May PR, Jing Z, Ahmed YE, Wijburg CJ, Canda AE, et al. Outcomes of Intracorporeal Urinary Diversion after Robot-Assisted Radical Cystectomy: Results from the International Robotic Cystectomy Consortium. J Urol 2018;199:1302–11. doi:10.1016/j.juro.2017.12.045.

[198] Satkunasivam R, Tallman CT, Taylor JM, Miles BJ, Klaassen Z, Wallis CJD. Robot-assisted Radical Cystectomy Versus Open Radical Cystectomy: A Meta-analysis of Oncologic, Perioperative, and Complication-related outcomes. Eur Urol Oncol 2018;2:443–447. doi:10.1016/j.euo.2018.10.008.

[199] Catto JWF, Khetrapal P, Ambler G, Sarpong R, Khan MS, Tan M, et al. Robot-assisted radical cystectomy with intracorporeal urinary diversion versus open radical cystectomy (iROC): Protocol for a randomised controlled trial with internal feasibility study. BMJ Open 2018;8:e020500. doi:10.1136/bmjopen-2017-020500.

[200] Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. JMIR MHealth UHealth 2018;6:e94. doi:10.2196/mhealth.9754.

[201] An HS, Jones GC, Kang SK, Welk GJ, Lee JM. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017;17:360–8. doi:10.1080/17461391.2016.1255261.

[202] Kendall B, Bellovary B, Gothe NP. Validity of wearable activity monitors for tracking steps and estimating energy expenditure during a graded maximal treadmill test. J Sports Sci 2018:1–8. doi:10.1080/02640414.2018.1481723.

[203] Maganja SA, Clarke DC, Lear SA, Mackey DC. Formative evaluation of consumer-grade activity monitors worn by older adults: Test-retest reliability and criterion validity of step counts. JMIR Form Res 2020;4. doi:10.2196/16537.

[204] Zoe Kleinman. Fitness tracker firm Jawbone faces liquidation - BBC News n.d. https://www.bbc.co.uk/news/technology-40530310 (accessed October 26, 2017).

[205] Garin O, Ayuso-Mateos JL, Almansa J, Nieto M, Chatterji S, Vilagut G, et al. Validation of the "World Health Organization Disability Assessment Schedule, WHODAS-2" in patients with chronic diseases. Health Qual Life Outcomes 2010;8. doi:10.1186/1477-7525-8-51.

[206] Groenvold M, Klee MC, Sprangers MA, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. J Clin Epidemiol 1997;50:441–50. doi:10.1016/S0895-4356(96)00428-3.

[207] EORTC. EORTC QLQ-C30 Scoring Manual 2001. https://www.eortc.be/qol/files/SCManualQLQ-C30.pdf.

[208] Shirley ED, Sanders JO. Measuring Quality of Care with Patient Satisfaction Scores. J Bone Joint Surg Am 2016;98:e83. doi:10.2106/JBJS.15.01216.

[209] Mata DA, Groshen S, Von Rundstedt FC, Skinner DG, Stadler WM, Cote RJ, et al. Variability in surgical quality in a phase III clinical trial of radical cystectomy in patients with organ-confined, node-negative urothelial carcinoma of the bladder. J Surg Oncol 2015;111:923–8. doi:10.1002/jso.23903.

[210] Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med 2016;375:1415–24. doi:10.1056/NEJMoa1606220.

[211] NHS England. Clinical Commissioning Policy : Robotic Assisted Surgery for Bladder Cancer n.d. https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2016/07/16033_FINAL.pdf (accessed July 27, 2016).

[212] Unver B, Kahraman T, Kalkan S, Yuksel E, Karatosun V, Gunal I. Test-retest reliability of the 50-foot timed walk and 30-second chair stand test in patients with total hip arthroplasty. Acta Orthop Belg 2015;81:435–41.

[213] Jones CJ, Rikli RE, Beam WC. A 30-s Chair-Stand Test as a Measure of Lower Body Strength in Community-Residing Older Adults. Res Q Exerc Sport 1999;70:113–9. doi:10.1080/02701367.1999.10608028.

[214] Hansen Bruun I, Mogensen CB, Nørgaard B, Schiøttz-Christensen B, Maribo T. Validity and Responsiveness to Change of the 30-Second Chair-Stand Test in Older Adults Admitted to an Emergency Department. J Geriatr Phys Ther 2017:1. doi:10.1519/JPT.0000000000000166.

[215] Gretebeck RJ, Montoye HJ. Variability of some objective measures of physical activity. Med Sci Sports Exerc 1992;24:1167–72.

[216] Konety BR, Allareddy V, Herr H. Complications after radical cystectomy: Analysis of population-based data. Urology 2006;68:58–64. doi:10.1016/J.UROLOGY.2006.01.051.

[217] Khetrapal P, Lee MWL, Tan WS, Dong L, de Winter P, Feber A, et al. The role of circulating tumour cells and nucleic acids in blood for the detection of bladder cancer: A systematic review. Cancer Treat Rev 2018;66:56–63. doi:10.1016/j.ctrv.2018.03.007.

[218] Catto JWF, Khetrapal P, Ambler G, Sarpong R, Potyka I, Khan MS, et al. Multidomain Quantitative Recovery Following Radical Cystectomy for Patients Within the Robot-assisted Radical Cystectomy with Intracorporeal Urinary Diversion Versus Open Radical Cystectomy Randomised Controlled Trial: The First 30 Patients. Eur Urol 2018;74:531–4. doi:10.1016/j.eururo.2018.05.008.

[219] Üstün TB, Chatterji S, Kostanjsek N, Rehm J, Kennedy C, Epping-Jordan J, et al. Developing the world health organization disability assessment schedule 2.0. Bull World Health Organ 2010;88:815–23. doi:10.2471/BLT.09.067231.

[220] Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian adults: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. Heal Reports 2011;22:7–14.

[221] Lee I-M, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of Step Volume and Intensity With All-Cause Mortality in Older Women. JAMA Intern Med 2019;02215. doi:10.1001/jamainternmed.2019.0899.

[222] Guy's and St Thomas' NHS Foundation Trust. Having your bladder removed – a robotic-assisted laparoscopic cystectomy 2016. https://www.guysandstthomas.nhs.uk/resources/patient-information/urology/bladder/robotic-assisted-cystectomy-for-men.pdf (accessed January 24, 2019).

[223] Healthwise Staff. Cystectomy With Ileal Conduit: What to Expect at Home 2018. https://myhealth.alberta.ca/Health/aftercareinformation/pages/conditions.aspx?hwid=ud1605 (accessed January 24, 2019).

[224] Mayo Clinic. Bladder removal surgery (cystectomy) n.d. https://www.mayoclinic.org/tests-procedures/cystectomy/about/pac-20385108 (accessed January 24, 2019).

[225] Gresham G, Hendifar AE, Spiegel B, Neeman E, Tuli R, Rimel BJ, et al. Wearable activity monitors to assess performance status and predict clinical outcomes in advanced cancer patients. Npj Digit Med 2018;1:27. doi:10.1038/s41746-018-0032-6.

[226] Erber B, Schrader M, Miller K, Schostak M, Baumunk D, Lingnau A, et al. Morbidity and Quality of Life in Bladder Cancer Patients following Cystectomy and Urinary Diversion: A Single-Institution Comparison of Ileal Conduit versus Orthotopic Neobladder. ISRN Urol 2012;2012:1–8. doi:10.5402/2012/342796.

[227] Gresham G, Hendifar AE, Gong J, Asher A, Walsh CS, Rimel B, et al. Digitally captured step counts for evaluating performance status in advanced cancer patients: A single cohort, prospective trial (Digi-STEPS). J Clin Oncol 2019;37:TPS6651–TPS6651. doi:10.1200/JCO.2019.37.15_suppl.TPS6651.

[228] Smith TB, Stonell C, Purkayastha S, Paraskevas P. Cardiopulmonary exercise testing as a risk assessment method in non cardio-pulmonary surgery: a systematic review. Anaesthesia 2009;64:883–93. doi:10.1111/j.1365-2044.2009.05983.x.

[229] Johnson D. Aerobic and Anaerobic training with wearables, what you should know n.d. https://gadgetsandwearables.com/2019/06/05/aerobic-or-anaerobic/ (accessed June 20, 2019).

[230] Maslakovic M. Wearables and lactate threshold, stamina training for runners n.d. https://gadgetsandwearables.com/2019/05/07/lactate-threshold/ (accessed July 4, 2019).

[231] Snowden CP, Prentis JM, Anderson HL, Roberts DR, Randles D, Renton M, et al. Submaximal cardiopulmonary exercise testing predicts complications and hospital length of stay in patients undergoing major elective surgery. Ann Surg 2010;251:535–41. doi:10.1097/SLA.0b013e3181cf811d.

[232] U.S. National Library of medicine. Trial to Compare Robotically Assisted Radical Cystectomy With Open Radical Cystectomy (iROC) 2017. https://clinicaltrials.gov/ct2/show/NCT03049410 (accessed February 28, 2017).

[233] Johar RS, Hayn MH, Stegemann AP, Ahmed K, Agarwal P, Balbay MD, et al. Complications

after robot-assisted radical cystectomy: Results from the international robotic cystectomy consortium. Eur Urol 2013;64:52–7. doi:10.1016/j.eururo.2013.01.010.

[234] Guazzi M, Bandera F, Ozemek C, Systrom D, Arena R. Cardiopulmonary Exercise Testing: What Is its Value? J Am Coll Cardiol 2017;70:1618–36. doi:10.1016/J.JACC.2017.08.012.

[235] Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ 2019;366:l4570. doi:10.1136/bmj.l4570.

[236] ISRCTN. ISRCTN62293620: DREAMPath – A prospective observational study to measure patient compliance with remote monitoring following discharge from hospital after major surgery. ISRCTN 2018. https://www.isrctn.com/ISRCTN62293620 (accessed February 27, 2019).

[237] Burch VC, Tarr G, Morroni C. Modified early warning score predicts the need for hospital admission and inhospital mortality. Emerg Med J 2008;25:674–8. doi:10.1136/emj.2007.057661.

Supplementary data

I. Tables

T - Primary Tumour		
Tx		Primary tumour cannot be assessed
T0		No evidence of primary tumour
Ta		Non-invasive papillary carcinoma
Tis		Carcinoma in situ: “flat tumour”
T1		Tumour invades subepithelial connective tissue
T2		Tumour invades muscle T2a Tumour invades superficial muscle (inner half) T2b Tumour invades deep muscle (outer half)
T3	3a	Tumour invades perivesical tissue microscopically
	3b	Tumour invades perivesical tissue macroscopically (extravesical mass)
T4	4a	Tumour invades prostate stroma, seminal vesicles, uterus, or vagina
	4b	Tumour invades pelvic wall or abdominal wall
N - Regional Lymph nodes		
Nx		Regional lymph nodes cannot be assessed
N0		No regional lymph-node metastasis
N1		Metastasis in a single lymph node in the true pelvis (hypogastric, obturator, external iliac, or presacral)
N2		Metastasis in multiple lymph nodes in the true pelvis (hypogastric, obturator, external iliac, or presacral)
N3		Metastasis in common iliac lymph node(s)
M - Distant Metastasis		
M0		No distant metastasis
M1		Distant metastasis

Supplementary Table 10-1: TNM classification of urinary bladder cancer

1973 WHO grading system	
Grade 1	Well differentiated
Grade 2	Moderately differentiated
Grade 3	Poorly differentiated
2004 WHO grading system (papillary lesions)	
Papillary urothelial neoplasm of low malignant potential (PUNLMP)	
Low-grade (LG) papillary urothelial carcinoma	
High-grade (HG) papillary urothelial carcinoma	

Supplementary Table 10-2: WHO grading in 1973 and 2004

No.	Author, year RCT Study design	Domain 1: a) Randomisation process b) [Cluster only] Timing of identification & recruitment of participants in relation to timing of randomisation	Domain 2: Deviation from the intended interventions	Domain 3: Missing outcome data	Domain 4: Measurement of the outcome	Domain 5: Selection of the reported result	Overall risk of bias
1	Valle, 2017[120] Standard	Low	Low	Low	High	Some concerns	**HIGH**
2	Miyauchi, 2016[121] Standard	Some concerns	Some concerns	Low	Low	Some concerns	**SOME CONCERNS**
3	Gordon, 2017[122] Standard	Some concerns	Low	High	Low	Some concerns	**HIGH**
4	Thomas, 2017[123] Standard	Low	Low	Low	Low	Some concerns	**SOME CONCERNS**
5	Li, 2017[124] Standard	Some concerns	Low	Low	Low	Some concerns	**SOME CONCERNS**
6	Jakicic, 2016[102] Standard	Low	Low	Low	Low	Low	**LOW**
7	Williams 2017[125]	High	Low	Low	Low	Some concerns	**HIGH**
8	Han 2016[126]	High	High	High	Low	High	**HIGH**
9	Tran, 2017[118] Cluster	a) Low b) Low	Low	Low	Low	Low	**LOW**
10	Takahashi, 2016[119] Crossover	Low	Low	Low	Low	Low	**LOW**
11	Lynch, 2019[127]	Some concerns	Some concerns	Low	Low	Low	**SOME CONCERNS**
12	Smith, 2019[128]	Some concerns	High	Low	Low	Some concerns	**HIGH**
13	McNeil, 2019[129]	High	Low	Low	Low	Some concerns	**HIGH**
14	Falck, 2018[130]	Some concerns	Low	Low	Low	Some concerns	**SOME CONCERNS**
15	Varas, 2018[131]	Some concerns	Some concerns	Low	Low	Some concerns	**SOME CONCERNS**

16	Duscha, 2018[132]	High risk	High risk	Low	Low	Some concerns	**HIGH**
17	Phan, 2018[113]	Low	Low	Low	Low	Low	**LOW**
18	Van der Walt, 2018[133]	Low	Low	Low	Low	Low	**LOW**
19	Orme, 2018[134]	Some concerns	Low	High	Low	Some concerns	**SOME CONCERNS**
20	Katz, 2018[135]	Low	Low	Low	Low	Some concerns	**LOW**
21	Kooiman, 2018[136]	Some concerns	Low	Some concerns	Low	Some concerns	**SOME CONCERNS**
22	Kanai, 2018[137]	Low	Low	Low	Low	Low	**LOW**
23	Mitchell, 2019[138]	a) Some Concerns b) Low	Low	Low	Low	Some Concerns	**SOME CONCERNS**
Supplementary Table 1: Risk of bias assessment for included randomised controlled studies using the Cochrane Collaboration's risk of bias-2 tool							

Supplementary Table 10-3: Risk of bias assessment for included randomised controlled studies using the Cochrane Collaboration's risk of bias-2 tool

Author, Year	Valle, 2017	Miyauchi, 2016	Gordon, 2017	Thomas, 2017	Li, 2017	Jakicic, 2016	Williams 2017	Han 2016
Doman 1: Risk of bias arising from the randomisation process								
1.1 Was the allocation sequence random?	Yes	No information	No information	Yes	Yes	Yes	No information	No information
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Yes	No information	No information	Yes	No information	Yes	No information	No information
1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	No	No	No	No	Yes	No	Yes	Yes
Risk-of-bias judgement	Low	Some concerns	Some concerns	Low	Some concerns	Low	High	High
Domain 2: Risk of bias due to deviations from the intended interventions								
2.1. Were participants aware of their assigned intervention during the trial?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	Yes	No information	No information	Yes	No information	No information	Yes	Yes
2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the experimental context?	No	No	Probably no	Probably no	Probably no	No	No	No
2.4. If Y/PY to 2.3: Were these deviations from intended intervention balanced between groups?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2.5 If N/PN/NI to 2.4: Were these deviations likely to have affected the outcome?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	Yes	No	Yes	Yes	Yes	Yes	Yes	No
2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on	N/A	Probably no	N/A	N/A	N/A	N/A	N/A	Probably yes

the result) of the failure to analyse participants in the group to which they were randomized?								
Risk-of-bias judgement	Low	Some concerns	Low	Low	Low	Low	Low	High
Domain 3: Missing outcome data								
3.1 Were data for this outcome available for all, or nearly all, participants randomized?	Yes	Yes	No information	No	Yes	No	Yes	No
3.2 <u>If N/PN/NI to 3.1</u>: Is there evidence that result was not biased by missing outcome data?	N/A	N/A	Probably no	Yes	N/A	Probably no	N/A	No
3.3 <u>If N/PN to 3.2</u>: Could missingness in the outcome depend on its true value?	N/A	N/A	No information	N/A	N/A	No	N/A	No information
3.4 <u>If Y/PY/NI to 3.3</u>: Do the proportions of missing outcome data differ between intervention groups?	N/A	N/A	No information	N/A	N/A	N/A	N/A	No information
3.5 <u>If Y/PY/NI to 3.3</u>: Is it likely that missingness in the outcome depended on its true value?	N/A	N/A	No information	N/A	N/A	N/A	N/A	No information
Risk-of-bias judgement	Low	Low	High	Low	Low	Low	Low	High
Domain 4: Risk of bias in measurement of the outcome								
4.1 Was the method of measuring the outcome inappropriate?	No	No	No	No	No	No	No	No
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	Probably yes	No	No	No	No	No	No	No
4.3 <u>If N/PN/NI to 4.1 and 4.2</u>: Were outcome assessors aware of the intervention received by study participants?	Yes	No information	No information	No	No information	No information	Yes	Yes
4.4 <u>If Y/PY/NI to 4.3</u>: Could assessment of the outcome have been	Probably yes	No	Probably no	N/A	No	No	Probably no	Probably no

influenced by knowledge of intervention received?								
4.5 <u>If Y/PY/NI to 4.4</u>: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	Probably yes	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Risk-of-bias judgement	High	Low	Low	Low	Low	Low	Low	Low
Domain 5: Risk of bias in selection of the reported result								
5.1 Was the trial analysed in accordance with a pre-specified plan that was finalized before unblinded outcome data were available for analysis?	No information	No information	No information	No information	No information	Yes	No information	No information
Is the numerical result being assessed likely to have been selected, on the basis of the results, from...								
5.2. ... multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	Probably no	No	Probably no	No	No	No	No	No
5.3 ... multiple analyses of the data?	Probably no	No	No	No	No	No	No	No
Risk-of-bias judgement	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	Low	Some concerns	Some concerns
OVERALL RISK OF BIAS	**High**	**Some concerns**	**High**	**Some concerns**	**Some concerns**	**Low**	**High**	**High**

Supplementary Table 10-4: Detailed risk of bias assessment for RCT studies (standard)

Author, Year	Lynch, 2019	Smith 2019	McNeil 2019	Falck 2018	Beatriz-Varas 2018	Duscha 2019
Doman 1: Risk of bias arising from the randomisation process						
1.1 Was the allocation sequence random?	Yes	Yes	No information	Yes	Yes	No information
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Yes	No information	No information	No information	No information	No information
1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	Probably yes	No	Probably yes	Probably no	No	Probably yes
Risk-of-bias judgement	Some concerns	Some concerns	High risk	Some concerns	Some concerns	High risk
Domain 2: Risk of bias due to deviations from the intended interventions						
2.1. Were participants aware of their assigned intervention during the trial?	Yes	Yes	Yes	Yes	Yes	Yes
2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	No information	Yes	Yes	Yes	Yes	Yes
2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the experimental context?	No	Probably no	Probably no	Probably no	No	No
2.4. If Y/PY to 2.3: Were these deviations from intended intervention balanced between groups?	N/A	N/A	N/A	N/A	N/A	N/A
2.5 If N/PN/NI to 2.4: Were these deviations likely to have affected the outcome?	N/A	N/A	N/A	N/A	N/A	N/A
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	No	No	Yes	Yes	No	No
2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse	Probably no	Probably yes	N/A	N/A	Probably no	Probably yes

participants in the group to which they were randomized?						
Risk-of-bias judgement	Some concerns	High risk	Low risk	Low risk	Some concerns	High risk
Domain 3: Missing outcome data						
3.1 Were data for this outcome available for all, or nearly all, participants randomized?	No	No	No	No	No	No
3.2 If N/PN/NI to 3.1: Is there evidence that result was not biased by missing outcome data?	Yes	No	Yes	Probably yes	No	Probably no
3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	N/A	Probably no	N/A	N/A	Probably no	Probably no
3.4 If Y/PY/NI to 3.3: Do the proportions of missing outcome data differ between intervention groups?	N/A	N/A	N/A	N/A	N/A	N/A
3.5 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	N/A	N/A	N/A	N/A	N/A	N/A
Risk-of-bias judgement	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Domain 4: Risk of bias in measurement of the outcome						
4.1 Was the method of measuring the outcome inappropriate?	No	No	No	No	No	No
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	No	No	No	No	No	No
4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?	Yes	No information	No	Yes	No	No information
4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	No	No	N/A	No	N/A	No

4.5 <u>If Y/PY/NI to 4.4</u>: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	N/A	N/A	N/A	N/A	N/A	N/A
Risk-of-bias judgement	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Domain 5: Risk of bias in selection of the reported result						
5.1 Was the trial analysed in accordance with a pre-specified plan that was finalized before unblinded outcome data were available for analysis?	Yes	No information	No information	No information	No information	No information
5.2. ... multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	No	No	No	No	No	No
5.3 ... multiple analyses of the data?	No	No	No	No	No	No
Risk-of-bias judgement	Low risk	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns
OVERALL RISK OF BIAS	**Some concerns**	**High**	**High**	**Some concerns**	**Some concerns**	**High**

Supplementary Table 14-4: Detailed risk of bias assessment for RCT studies (standard) continued [1]

Author, Year	Phan, 2018[113]	Van der Walt, 2018[133]	Orme, 2018[134]	Katz, 2018[135]	Kooiman, 2018[136]	Kanai, 2018[137]	Mitchell, 2019[138]
Doman 1: Risk of bias arising from the randomisation process							
1.1 Was the allocation sequence random?	Yes	Yes	Yes	Yes	Yes	Yes	No information
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	No information	Yes	No	Yes	No information	Yes	No information
1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	No	No	Yes	No	Yes	No	No
Risk-of-bias judgement	Low	Low	Some concerns	Low	Some concerns	Low	Some concerns
Domain 2: Risk of bias due to deviations from the intended interventions							
2.1. Were participants aware of their assigned intervention during the trial?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the experimental context?	No	No	No	No	No	No	No
2.4. If Y/PY to 2.3: Were these deviations from intended intervention balanced between groups?	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2.5 If N/PN/NI to 2.4: Were these deviations likely to have affected the outcome?	N/A	N/A	N/A	N/A	N/A	N/A	N/A

2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomized?	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Risk-of-bias judgement	Low	Low	Low	Low	Low	Low	Low
Domain 3: Missing outcome data							
3.1 Were data for this outcome available for all, or nearly all, participants randomized?	Yes	Yes	No	Yes	Yes	Yes	Yes
3.2 If N/PN/NI to 3.1: Is there evidence that result was not biased by missing outcome data?	N/A	N/A	Probably no	N/A	N/A	N/A	N/A
3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	N/A	N/A	No information	N/A	N/A	N/A	N/A
3.4 If Y/PY/NI to 3.3: Do the proportions of missing outcome data differ between intervention groups?	N/A	N/A	Yes	N/A	N/A	N/A	N/A
3.5 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	N/A	N/A	No information	N/A	N/A	N/A	N/A
Risk-of-bias judgement	Low	Low	High	Low	Low	Low	Low
Domain 4: Risk of bias in measurement of the outcome							
4.1 Was the method of measuring the outcome inappropriate?	No	No	No	No	No	No	No
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	No	No	No	No	No	No	No
4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the	N/A	Yes	No information	No	No information	No information	Yes

intervention received by study participants?							
4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	N/A	No	Probably no	N/A	No	No	No
4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Risk-of-bias judgement	Low	Low	Low	Low	Low	Low	Low
5.1 Was the trial analysed in accordance with a pre-specified plan that was finalized before unblinded outcome data were available for analysis?	No	No information	No information	No information	No information	Yes	No information
5.2. ... multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	No	No	Probably no	No	No	No	No
5.3 ... multiple analyses of the data?	No	No	No	No	No	No	No
Risk-of-bias judgement	Low	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
OVERALL RISK OF BIAS	**Low**	**Low**	**Some concerns**	**Low**	**Some concerns**	**Low**	**Some concerns**

Supplementary Table 14-4: Detailed risk of bias assessment for RCT studies (standard) continued [2]

Author, Year	Tran 2017	Mitchell, 2019[138]
Domain 1a: Bias arising from the randomisation process		
1a.1 Was the allocation sequence random?	Probably yes	Probably yes
1a.2 Is it likely that the allocation sequence was subverted?	Probably no	Probably no
1a.3 Were there baseline imbalances that suggest a problem with the randomization process?	No	No
Risk of bias judgement	Low	Low
Domain 1b: Bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomization		
1b.1 Were all the individual participants identified before randomization of clusters (and if the trial specifically recruited patients were they all recruited before randomization of clusters)?	Probably yes	Yes
1b.2 <u>If N/PN/NI to 1b.1:</u> Is it likely that selection of individual participants was affected by knowledge of the intervention?	N/A	N/A
1b.3 Were there baseline imbalances that suggest differential identification or recruitment of individual participants between arms?	No	No
Risk of bias judgement	Low	Low
Domain 2: Bias due to deviations from intended interventions		
2.1a Were participants aware that they were in a trial?	No information	Yes
2.1b <u>If Y/PY/NI to 2.1a:</u> Were participants aware of their assigned intervention during the trial?	Yes	Yes
2.2. Were carers and trial personnel aware of participants' assigned intervention during the trial?	Yes	Yes
2.3. <u>If Y/PY/NI to 2.1 or 2.2:</u> Were there deviations from the intended intervention beyond what would be expected in usual practice?	Probably no	No
2.4. <u>If Y/PY to 2.3:</u> Were these deviations from intended intervention unbalanced between groups and likely to have affected the outcome?	N/A	N/A
2.5a Were any clusters analysed in a group different from the one to which they were assigned?	No	No

2.5b Were any participants analysed in a group different from the one to which their original cluster was randomized?	No	No
2.6 If Y/PY/NI to 2.5: Was there potential for a substantial impact (on the estimated effect of intervention) of analysing participants in the wrong group?	N/A	N/A
Risk of bias judgement	Low	Low
Domain 3: Bias due to missing outcome data		
3.1a Were outcome data available for all, or nearly all, clusters randomized?	No	Yes
3.1b Were outcome data available for all, or nearly all, participants within clusters?	No	Yes
3.2 If N/PN/NI to 3.1a or 3.1b: Are the proportions of missing outcome data and reasons for missing outcome data similar across intervention groups?	Yes	N/A
3.3 If N/PN/NI to 3.1a or 3.1b: Is there evidence that results were robust to the presence of missing outcome data?	No information	N/A
Risk of bias judgement	Low	Low
Domain 4: Bias in measurement of the outcome		
4.1a Were outcome assessors aware that a trial was taking place?	No information	No information
4.1b If Y/PY/NI to 4.1: Were outcome assessors aware of the intervention received by study participants?	Yes	Probably Yes
4.2 If Y/PY/NI to 4.1: Was the assessment of the outcome likely to be influenced by knowledge of intervention received?	Probably no	Probably No
Risk of bias judgement	Low	Low
Domain 5: Bias in selection of the reported result		
Are the reported outcome data likely to have been selected, on the basis of the results, from...		
5.1. ... multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	No	No
5.2 ... multiple analyses of the data?	No	No

Risk of bias judgement	Low	Low
OVERALL RISK OF BIAS	**Low**	**Low**

Supplementary Table 10-5: Detailed risk of bias assessment for RCT studies (cluster)

Author, Year	Takahashi 2016
Domain 1: Bias arising from the randomisation process	
1.1 Was the allocation sequence random?	Yes
1.2 Was the allocation sequence concealed until participants were recruited and assigned to interventions?	Yes
1.3 Were there baseline imbalances that suggest a problem with the randomization process?	No
1.4 Is a roughly equal proportion of participants allocated to each of the two groups?	Yes
1.5 <u>If N/PN/NI to 1.4:</u> Are period effects included in the analysis?	N/A
Risk of bias judgement	Low
Domain 2: Bias due to deviations from intended interventions	
2.1. Were participants aware of their assigned intervention during each period of the trial?	Yes
2.2. Were carers and trial personnel aware of participants' assigned intervention during each period of the trial?	Yes
2.3. <u>If Y/PY/NI to 2.1 or 2.2:</u> Were there deviations from the intended interventions beyond what would be expected in usual practice?	No
2.4. <u>If Y/PY to 2.3:</u> Were these deviations from intended interventions unbalanced between the two interventions and likely to have affected the outcome?	N/A
2.5 Was there sufficient time for any carry-over effects to have disappeared before outcome assessment in the second period?	N/A*
Risk of bias judgement	Low
Domain 3: Bias due to missing outcome data	
3.1 Were outcome data available for all, or nearly all, participants randomized?	No
3.2 <u>If N/PN/NI to 3.1:</u> Are the proportions of missing outcome data and reasons for missing outcome data similar across interventions?	No
3.3. <u>If N/PN/NI to 3.1:</u> Is there evidence that results were robust to the presence of missing outcome data?	Yes
Risk of bias judgement	Low
Domain 3: Bias in measurement of the outcome	

4.1 Were outcome assessors aware of the intervention received by study participants?	Yes
4.2 <u>If Y/PY/NI to 4.1:</u> Was the assessment of the outcome likely to be influenced by knowledge of intervention received?	No
Risk of bias judgement	Low
Domain 5: Bias in selection of the reported result	
Are the reported outcome data likely to have been selected, on the basis of the results, from...	
5.1. ... multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	No
5.2 ... multiple analyses of the data?	No
5.3 … the outcome of a statistical test for carry-over?	N/A*
Risk of bias judgement	Low
OVERALL RISK OF BIAS	**Low**

Supplementary Table 10-6: Detailed risk of bias assessment for RCT studies (crossover)

No.	Author, year	Bias due to confounding	Bias in selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Risk of bias judgement
1	Jacquemin, 2018[158]	No	Potentially Yes	No	No	Potentially Yes	No	No	**MODERATE**
2	Van't Hul, 2016[159]	Yes	Yes	No	No	No	No	No	**SERIOUS**
3	Peacock, 2017[160]	Potentially Yes	No	No	No	No	No	No	**LOW**
4	Glaviano, 2017[161]	No	No	No	No	No	No	No	**LOW**
5	Colón-Semenza[162]	No	No	No	No	Potentially Yes	Potentially Yes	No	**MODERATE**
6	Kuenze, 2019[163]	No	No	No	No	No	No	No	**LOW**

Supplementary Table 10-7: Risk of bias assessment for included non-randomised studies using the ROBINS-I assessment tool

Author	Bias	Author's judgment	Support for judgment
Bochner et al.	Random sequence generation (selection bias)	Low risk	Patients were stratified by age (≤64 vs ≥64 yr) and American Society of Anaesthesiologist score (1-2 vs 3-4), then randomly assigned 1:1 using randomly permuted blocks of random length.
	Allocation concealment (selection bias)	Low risk	Randomisation performed by independent office where allocation concealment was ensured by a password-protected database
	Blinding of participants and researchers (performance bias)	High risk	Blinding not possible
	Blinding of outcome assessment (detection bias)	High risk	Blinding not possible
	Incomplete outcome data (attrition bias)	Low risk	No patients lost to follow-up
	Selective reporting (reporting bias)	Low risk	All pre-specified outcomes were reported
	Other bias	Low risk	None
Nix et al.	Random sequence generation (selection bias)	High risk	The randomisation schema was performed with five sequential patients undergoing an approach before alternating surgical modality
	Allocation concealment (selection bias)	High risk	Randomisation schema of five sequential patients may allow the investigator to predict allocation concealment
	Blinding of participants and researchers (performance bias)	High risk	Blinding not possible
	Blinding of outcome assessment (detection bias)	High risk	Low risk
	Incomplete outcome data (attrition bias)	Low risk	No patients lost to follow-up
	Selective reporting (reporting bias)	Low risk	All pre-specified outcomes were reported
	Other bias	Low risk	None
Parekh et al.	Random sequence generation (selection bias)	Low risk	Computerised randomisation
	Allocation concealment (selection bias)	Low risk	Each assignment was placed in a sealed envelope with the corresponding slot number written on the outside. At the time of consent, the lowest numbered envelope remaining was opened and the patient was assigned to the surgical procedure listed on the piece of paper inside the envelope.

	Blinding of participants and researchers (performance bias)	High risk	Blinding not possible
	Blinding of outcome assessment (detection bias)	High risk	Blinding not possible
	Incomplete outcome data (attrition bias)	Low risk	No patients were lost to follow-up
	Selective reporting (reporting bias)	Low risk	Peri-operative pathological outcomes were not reported for one patient. Reasons for missing outcome data unlikely to result in bias
	Other bias	Low risk	None
Khan et al.	Random sequence generation (selection bias)	Low risk	Simple randomisation
	Allocation concealment (selection bias)	Low risk	Allocation envelopes were opened by the patient in the presence of three members of the research team to ensure that no changes were made to allocation
	Blinding of participants and researchers (performance bias)	High risk	Blinding not possible
	Blinding of outcome assessment (detection bias)	High risk	Blinding not possible
	Incomplete outcome data (attrition bias)	Low risk	One patient lost to follow-up. Reasons for missing outcome data unlikely to be related to true outcome
	Selective reporting (reporting bias)	Low risk	All pre-specified outcomes were reported
	Other bias	Low risk	None

Supplementary Table 10-8: Risk of bias using the Cochrane RoB tool for RCTs comparing ORC and RARC

		N	Median	IQR
Baseline	Average steps	143	5921.4	4000.9-8270.8
	Maximum steps	143	8572.0	6348-12436
	Minimum steps	143	3442.0	2136-5006
day 5 post-op	Average steps	132	1525.4	833.6-2409.1
	Maximum steps	132	2401.0	1338.5-4009.5
	Minimum steps	132	717.0	394-1257
1 month post-op	Average steps	124	3819.9	2170.8-5926.8
	Maximum steps	124	6115.0	3274-8892.5
	Minimum steps	124	2129.0	1070-3180
3 months post-op	Average steps	106	5774.3	3698.4-7186.9
	Maximum steps	106	8374.0	5381-11058.5
	Minimum steps	106	2702.0	1326-4056

Supplementary Table 10-9: Median and IQR values for average, day 5 post-op, 1 month and 3 month step-count

Percentage cut-off	Sensitivity	1 - Specificity	Youden's J statistic
7.69%	1	1	0
9.87%	1	0.857	0.143
15.62%	1	0.714	0.286
23.46%	1	0.571	0.429
31.53%	0.978	0.571	0.407
40.11%	0.978	0.429	0.549
44.24%	0.978	0.286	0.692
45.07%	0.956	0.286	0.67
47.30%	0.933	0.286	0.647
50.37%	0.911	0.286	0.625
54.91%	0.889	0.286	0.603
59.13%	0.867	0.286	0.581
62.05%	0.844	0.286	0.558
64.51%	0.822	0.286	0.536
66.83%	0.822	0.143	0.679
69.12%	0.8	0.143	0.657
70.72%	0.778	0.143	0.635
72.34%	0.756	0.143	0.613
74.25%	0.733	0.143	0.59
75.83%	0.711	0.143	0.568
77.43%	0.689	0.143	0.546
81.76%	0.667	0.143	0.524
85.56%	0.644	0.143	0.501
87.90%	0.622	0.143	0.479
89.75%	0.6	0.143	0.457
90.73%	0.578	0.143	0.435
92.41%	0.556	0.143	0.413
94.43%	0.533	0.143	0.39
96.96%	0.511	0.143	0.368
99.49%	0.489	0.143	0.346
100.70%	0.467	0.143	0.324
101.95%	0.444	0.143	0.301
102.95%	0.422	0.143	0.279
103.07%	0.4	0.143	0.257
103.82%	0.378	0.143	0.235
104.66%	0.356	0.143	0.213
104.88%	0.333	0.143	0.19
107.38%	0.311	0.143	0.168
110.90%	0.289	0.143	0.146
113.38%	0.267	0.143	0.124
115.01%	0.267	0	0.267
116.60%	0.244	0	0.244
119.67%	0.222	0	0.222
123.89%	0.2	0	0.2

131.55%	0.178	0	0.178
146.61%	0.156	0	0.156
159.26%	0.133	0	0.133
171.76%	0.111	0	0.111
204.51%	0.089	0	0.089
250.41%	0.067	0	0.067
331.48%	0.044	0	0.044
631.49%	0.022	0	0.022
874.36%	0	0	0

Supplementary Table 10-10: Percentage cut-off values and Youden's J statistic for maximum step counts in predicting disease-free 1 year survival

Percentage cut-off	Sensitivity	1 - Specificity	Youden's J statistic
6.58%	1	1	0
585.52%	0	0	0
492.98%	0.022	0	0.022
300.00%	0.044	0	0.044
195.96%	0.067	0	0.067
183.06%	0.089	0	0.089
171.03%	0.111	0	0.111
168.82%	0.133	0	0.133
8.34%	1	0.857	0.143
163.39%	0.156	0	0.156
147.01%	0.178	0	0.178
131.66%	0.2	0	0.2
123.63%	0.222	0	0.222
118.18%	0.244	0	0.244
116.73%	0.267	0	0.267
12.96%	1	0.714	0.286
115.56%	0.289	0	0.289
110.09%	0.311	0	0.311
105.41%	0.333	0	0.333
102.45%	0.356	0	0.356
99.39%	0.378	0	0.378
97.86%	0.4	0	0.4
96.64%	0.422	0	0.422
24.88%	1	0.571	0.429
94.38%	0.444	0	0.444
91.96%	0.467	0	0.467
90.56%	0.489	0	0.489
87.84%	0.511	0	0.511
85.12%	0.533	0	0.533
83.70%	0.556	0	0.556
34.06%	1	0.429	0.571
82.61%	0.578	0	0.578
81.13%	0.6	0	0.6
79.99%	0.622	0	0.622
79.21%	0.644	0	0.644
78.64%	0.667	0	0.667
78.16%	0.689	0	0.689
77.00%	0.711	0	0.711
36.39%	1	0.286	0.714
75.64%	0.733	0	0.733
74.07%	0.756	0	0.756
68.96%	0.778	0	0.778
48.62%	0.933	0.143	0.79

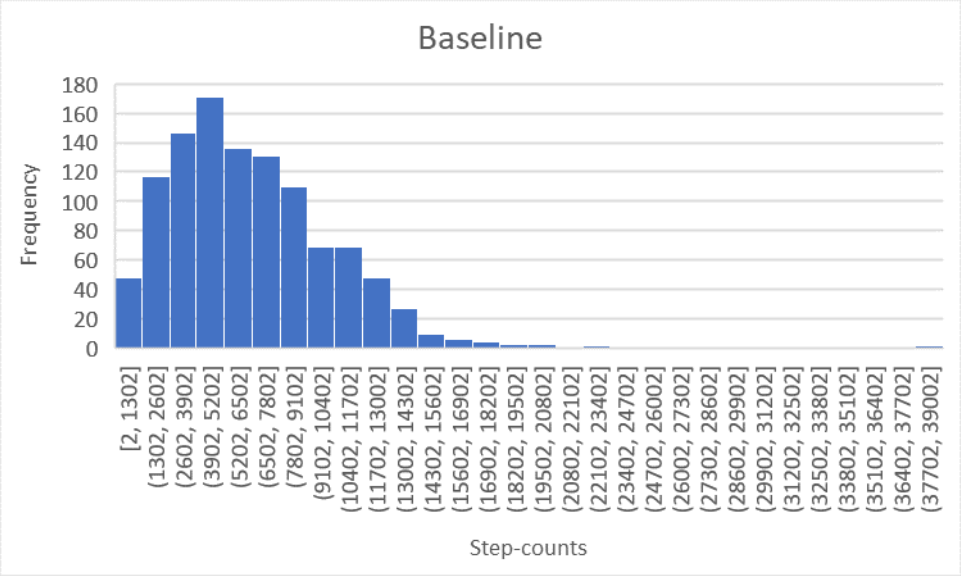
64.84%	0.8	0	0.8
46.79%	0.956	0.143	0.813
62.87%	0.822	0	0.822
41.99%	0.978	0.143	0.835
60.67%	0.844	0	0.844
38.23%	1	0.143	0.857
59.29%	0.867	0	0.867
55.59%	0.889	0	0.889
51.46%	0.911	0	0.911
49.43%	0.933	0	0.933

Supplementary Table 10-11: Percentage cut-off values and Youden's J statistic for average step counts in predicting disease-free 1-year survival

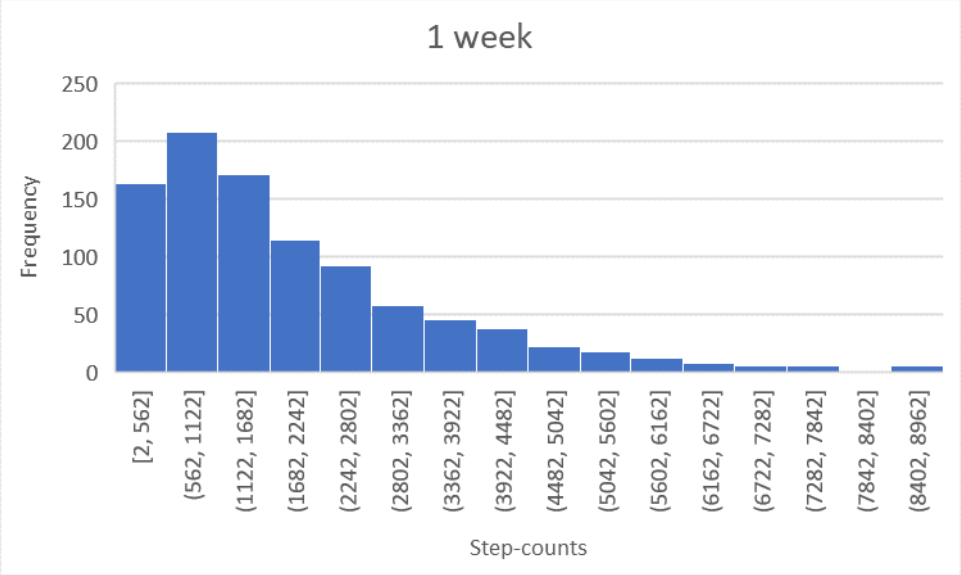
	ASC (cut-off = 49.43%)	**MSC (cut-off = 44.2%)**
Sensitivity	100%	83%
Specificity	93%	90%
Positive predictive value	70%	71%
Negative predictive value	86%	98%

Supplementary Table 10-12: Diagnostic value of ASC and MSC at 3 months using optimal sensitivity and specificity pairings to predict overall survival at the end of 1-year post RC.

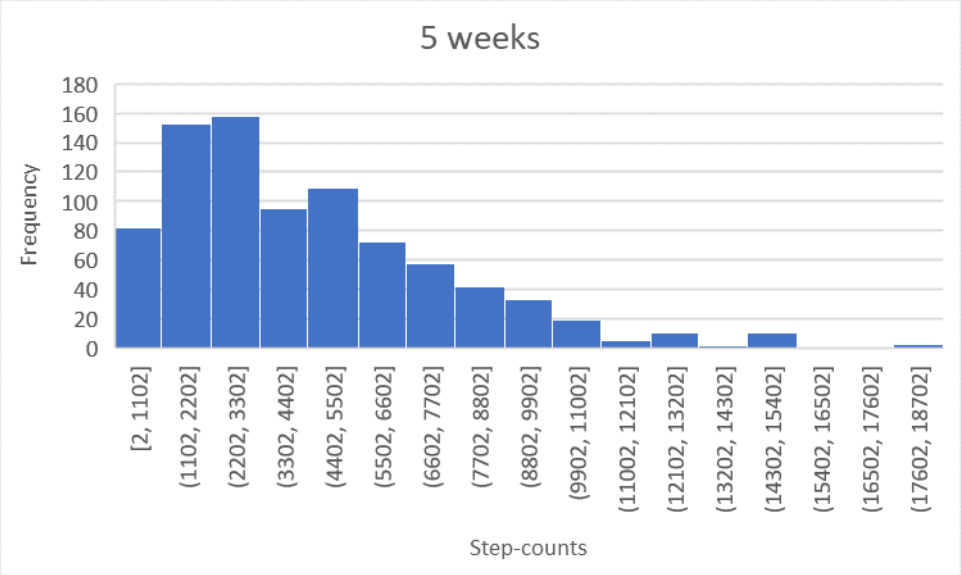
II. Figures



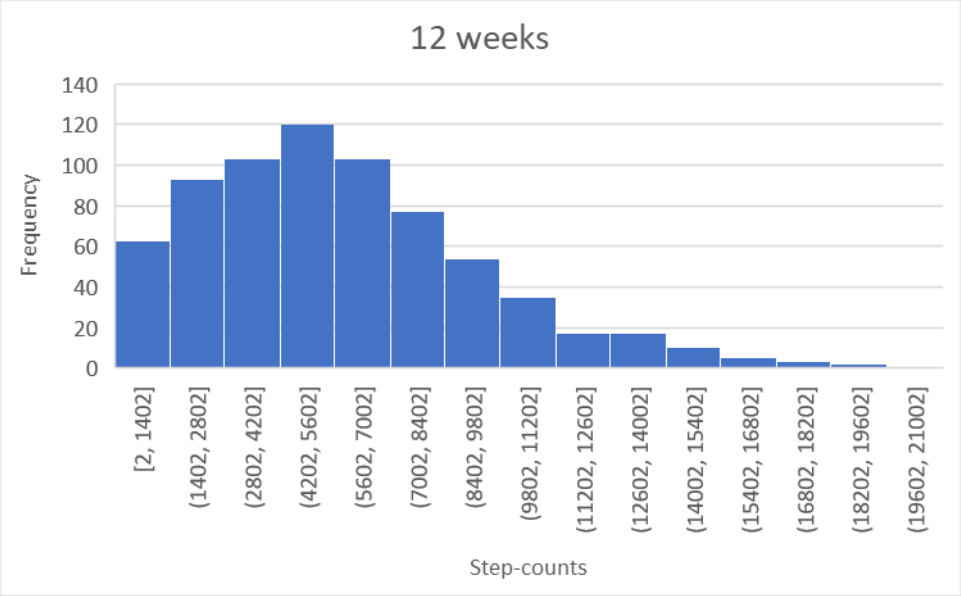
Supplementary Figure 10-1: Histogram representing all-step-count collected from patients at the baseline timepoint.



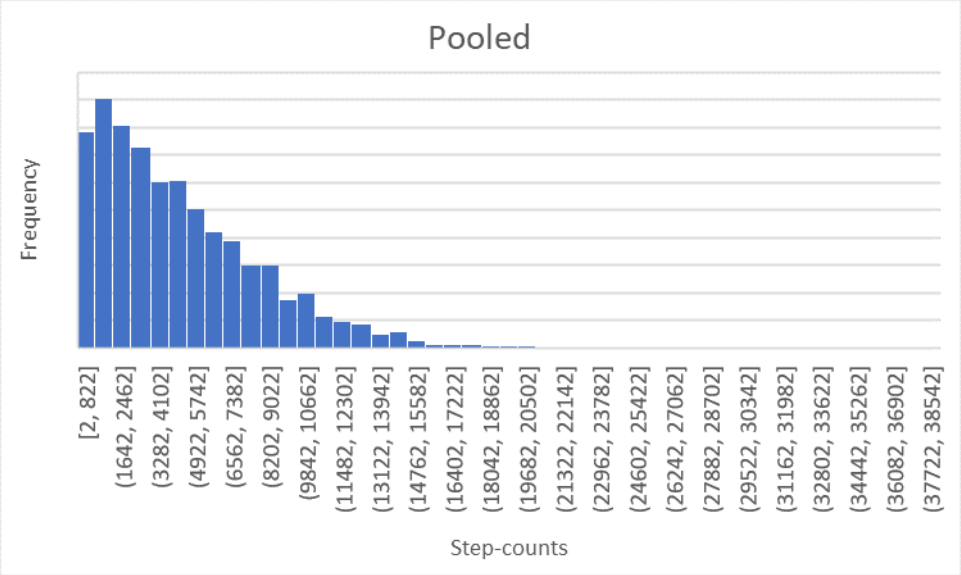
Supplementary Figure 10-2: Histogram representing all-step-count collected from patients at the 5-day post-op timepoint.



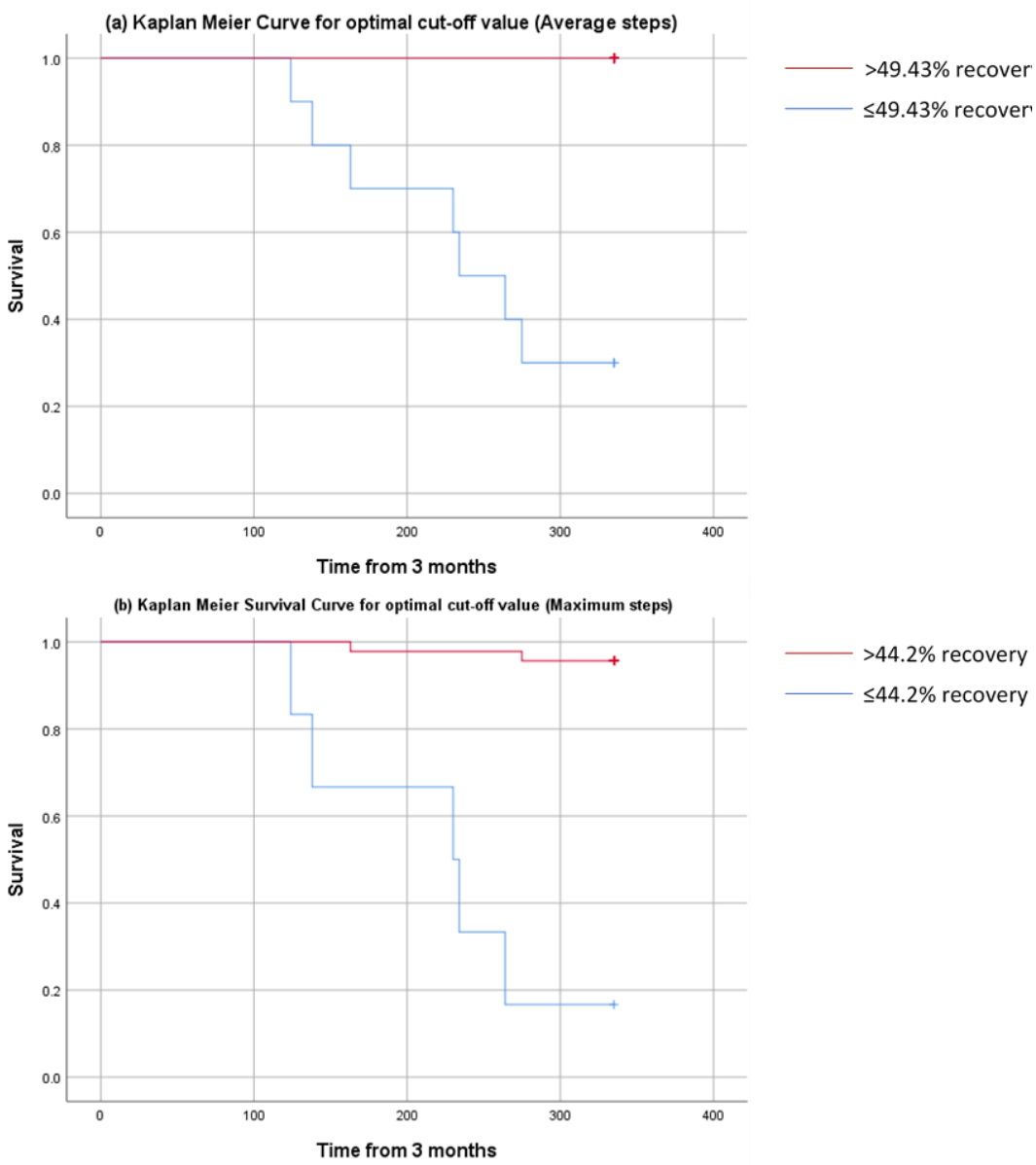
Supplementary Figure 10-3: Histogram representing all-step-count collected from patients at the 5 weeks timepoint.



Supplementary Figure 10-4: Histogram representing all-step-count collected from patients at the 3 months timepoint.



Supplementary Figure 10-5: Histogram representing all-step-count collected from patients pooled from all timepoints.



Supplementary Figure 10-6: Kaplan Meier Survival curves during the first year of follow-up for (a) average steps and (b) maximum steps recovery with optimal thresholds for highest sensitivity and specificity pairing by 3 months.