University of
Portsmouth

1

2

3 **Evidence Based Surgery: evaluating outcome measurement,**

4 **methodology and reporting quality**

5

6

7 **Mr Ankur Khajuria**

8

9

10 **BSc (Hons) MBBS (Dist) AICSM FHEA FRSPH MRCS (Eng)**

11 **MAcadMEd. MSc (Oxon)**

12

13

14 **Research contribution for the Award of Doctor of Philosophy**

15 **by Publication**

16

17

18 **February 2020**

19

20 **Word count: 13,636 words**

DECLARATION

The results and conclusions in this thesis are the work of the named candidate and have been published in peer-reviewed journals. I declare that whilst studying for the degrees of Doctor of Philosophy by Publication of the University of Portsmouth, I have not been registered for any other award at another university.

ABBREVIATIONS

- ADM   Acellular Dermal Matrix
- BCS   Breast Conserving Surgery
- BREAST-Q   Breast Questionnaire
- BRR   Breast Reconstruction
- CDC   Clavien-Dindo Classification
- CENTRAL   Cochrane Controlled Register of Trials
- CI   Confidence Interval
- COS   Core Outcome Set
- DIEP   Deep Inferior Epigastric Perforator
- DTI   Direct to Implant
- EBCTCG   Early Breast Cancer Trialists Collaborative Group
- EORTC   European Organisation for Research and Treatment of Cancer
- FU   Follow-Up
- GRADE   Grading of Recommendations Assessment, Development and Evaluation
- ICER   Incremental Cost Effectiveness Ratio
- IDEAL   Idea, Development, Exploration, Assessment, Long-term
- IGAP   Inferior Gluteal Artery Perforator
- MD   Mean Difference
- MeSH   Medical Subject Headings (MeSH)
- MROC   Mastectomy and Breast Reconstruction Outcomes Collaborative
- NIHR   National Institute of Health Research
- OR   Odds Ratio

•	PMRT	Post mastectomy radiotherapy (adjuvant radiotherapy)
•	PRISMA	Preferred Reporting Items for Systematic Review And Meta-Analysis
•	PROMs	Patient-Reported Outcome Measures
•	QoL	Quality of Life
•	ROBINS-I	Risk Of Bias In Non-randomised Studies - of Interventions
•	RT	Radiotherapy
•	SGAP	Superior Gluteal Artery Perforator
•	SIEA	Superficial Inferior Epigastric Artery
•	STROBE	Strengthening the Reporting Of Observational Studies in Epidemiology
•	SUPREMO	Selective Use of Postoperative Radiotherapy AftEr MastectOmy trial
•	TEI	Tissue Expander Implant
•	TRAM	Transverse Rectus Abdominis Myocutaneous
•	UK	United Kingdom
•	USA	United States of America

ACKNOWLEDGEMENTS

ABSTRACT

The body of work in this thesis presents the research and publications under a general theme of outcome measurement, evidence synthesis and reporting quality in surgery. The work highlights the author's own personal contributions and publications under this theme, and collaborations with colleagues from Oxford University, Harvard Medical School, Imperial College London and University College London. This thesis provides an in depth commentary on evidence-based surgery, with discussion on the challenges of conducting randomised controlled trials in surgery. Systematic review and meta-analysis methodology is discussed, exploring the nuances and assumptions of random/fixed effect models, quality assessment using GRADE and assessment of risk of bias (RoB) using Cochrane's ROBINS-I tool. The author has evaluated reporting quality in surgery, identifying suboptimal compliance with the CONSORT-NPT checklist. This work formed basis for change of policy and the requirement for mandatory completion and uploading of a CONSORT statement by authors when submitting articles to the peer-reviewed journal, International Journal of Surgery (IJS), with significant improvement in compliance.

The commentary also reviews plastic & reconstructive breast surgery, and provides an in-depth discussion on quality of life assessment, COSMIN, minimal important differences (MID), how to choose a questionnaire and particular domains in a study, and reporting using CONSORT and SPIRIT-PRO checklists. Health-utility measures for cost-utility analyses are also discussed. The authors' systematic reviews and meta-analyses are presented on clinical outcomes and PROs of DIEP versus implant-based reconstruction; and on immediate versus delayed reconstruction, in context of radiotherapy (RT). The former review provides a weak recommendation that DIEP reconstruction maybe more cost-effective and yield higher PRO scores, with suitable warnings in light of poor quality and serious risk of bias. The RT review identified no statistically significant difference in outcomes between immediate and delayed breast reconstruction (BRR), challenging dogma where majority of UK BRR are delayed, with significant heterogeneity in outcome measurement, suboptimal reporting of core outcome set and no grading of complications. The reviews have demonstrated paucity of high quality evidence and the need for future high quality studies. The PhD award will facilitate the author in establishing a research group and to undertake postdoctoral research. This will include conducting a national stream funded large prospective cohort study in evaluating immediate versus delayed autologous BRR, in context of radiotherapy, with robust reporting of BRR core outcome set and incorporation and measurement of disease-specific PROs and cost-effectiveness, addressing the limitations highlighted by the reviews.

CONFLICTS OF INTEREST

The author is an Associate Editor of Systematic Reviews Journal. There are no other disclosures or conflicts of interests to declare.

TABLE OF CONTENTS

187 1. **INTRODUCTION**

188

189 1.1 Evidence based Medicine (EBM)

190

191 EBM is defined as the integration of the best available evidence with clinical

192 expertise and patient preferences for optimal decision-making (Kang, 2016).

193 Surgical research has come a long way since Richard Horton likened it to

194 'comic opera' in a Lancet commentary (Horton, 1996). There has been rising

195 interest in surgeons conducting high quality prospective cohort studies and

196 randomised controlled trials (RCTs). Initiatives such as the implementation of

197 a nationwide surgical trials programme in the UK have been welcomed

198 (Khajuria and Agha, 2013). However, surgical research has inherent

199 challenges including issues related to blinding, inconsistent care provider

200 expertise, differential learning curves and centre's volume. Poor reporting of

201 outcomes is associated with bias in evaluating intervention effectiveness,

202 culminating in increasing inconsistency between conclusions and results and

203 precluding reliable critical appraisal and data interpretation by the readers.

204 Indeed surgeons have much work to do to enhance the quality of the scientific

205 basis on which their practice is based on, and a major issue to address is the

206 poor quality reporting of outcomes (Khajuria and Ahmed Agha, 2015).

207

208 Furthermore, it is well established that RCTs in Surgery are challenging to

209 conduct (McCulloch et al., 2002, Davies et al., 2020). Firstly, there maybe a

210 lack of clinician equipoise. The state of equipoise is, however, a prerequisite

211 for conducting RCTs (McCulloch et al., 2002). Lack of funding and

212 infrastructure has also been cited as a barrier, and funding bodies may be

213 influenced by poor quality of previous surgical research. Issues related to

214 blinding, inconsistent care provider expertise, differential learning curves and

215 centre's volume also impact surgical RCTs (Khajuria and Agha, 2013). During

216 learning curves, errors, adverse events and complications maybe more likely.

217 Randomising between a familiar and an unfamiliar operation may therefore

218 introduce bias against the latter. Moreover, the degree of acceptable technical

219 variation within a surgical procedure needs to be clearly defined a priori.

220 Imprecise definitions may lead to overlap in treatments with resultant bias.

221 Blinding is challenging in surgical trials, for both patients and surgeons,

222 although the outcome assessment can be blinded. 'Type 3 RCTs', i.e. those

223 that compare a surgical and a non-surgical treatment pose difficulties with

224 regards to equipoise of patients, as adverse events may differ greatly

225 (Solomon et al., 1994). For example, a surgical procedure is irreversible.

226 Patients may also perceive benefit for one technique over another; this can

227 hamper recruitment into surgical trials and/or make randomized allocation of

228 treatment challenging (Winters et al., 2015).

229

230 Recruitment challenges in surgery and breast reconstruction have been

231 demonstrated by the QUEST (quality of life after mastectomy and breast

232 reconstruction) randomization trials (Winters et al., 2015). These were

233 designed for the primary aim of determining the acceptability of a RCT of

234 breast reconstruction among patients and clinicians. Those not needing

235 PMRT were randomised to extended autologous LD or implant-assisted LD

236 reconstruction (QUEST A). Those needing PMRT were randomized to either

immediate autologous LD or staged–delayed autologous LD procedures (QUEST B). However, after 18 months of recruitment, only 17 and 8 patients respectively were recruited to QUEST A and B, and acceptance rates of 19 and 22% respectively. The trials failed to reach target recruitment in a timely manner. The challenges to recruitment identified were misperceptions of clinical equipoise by patients, and patients expressing strong preferences for breast reconstruction types and timings, despite provision of adequate trial information (Winters et al., 2015).

Finally, even when trials meet target recruitment, challenges remain. This was demonstrated by the BRIOS RCT comparing immediate one-stage ADM-DTI versus 2-stage IBR (Negenborn et al., 2018). This demonstrated increased complications in the single stage group, but despite this, no difference in QOL outcomes, with several potential methodological issues accounting for the discrepancy (Winters and Khajuria, 2018b). Despite adding to the evidence-base, uncertainty still remains regarding the role of ADM in IBR, with future studies, such as the iBRA study (Potter et al., 2016), underway to further inform the debate.

1.1.1 Systematic Reviews

Systematic reviews seek to collate evidence using pre-defined eligibility criteria in order to answer a specific research question. They aim to minimize bias by using explicit, systematic methods documented a priori in a protocol. The research question is clearly defined and often the Patient, Intervention,

Comparator and Outcome (PICO) framework is utilised. The components include pre-specified eligibility criteria (inclusion/exclusion criteria), a systematic search strategy, assessment of the methodological quality and risk of bias in the studies, interpretation and presentation of the results, with or without a meta-analysis. The PRISMA guidance is followed.

Conversely, a literature review qualitatively summarises evidence using informal or subjective methods. It provides a summary and overview of a topic, as opposed to answering a specific research question. Unlike in a systematic review, there is no a priori defined eligibility criteria, systematic search strategy, assessment of methodological quality and risk of bias or any meta-analysis.

For meta-analysis, two popular statistical models exist, Random effects and Fixed effects models. The assumption for the Fixed effects model is that there is one true effect size that underlies all studies in the analysis; any differences in observed effects are secondary to sampling error (Borenstein et al., 2010). Conversely, for the Random effects model, one allows the true effect to differ, i.e. the effect sizes may vary from study to study. The goal is to estimate the mean of a distribution of true effect sizes. Since each study provides information about a different effect size, small studies cannot be discounted by giving them small weights, nor can large studies be given too much weight (Schmidt et al., 2009). The width of the confidence intervals (CIs) for a meta-analysis is influenced by the individual study estimates and number of studies combined. Moreover, for the random effects model, the precision of the

287 estimate will decrease with increasing heterogeneity, and the confidence

288 interval width will widen. As more and larger studies are entered in the meta-

289 analysis, one would expect the confidence intervals to decrease, based on

290 overall greater sample size. However, if the additional studies increase the

291 heterogeneity, the confidence interval width may increase.

292

293 **1.2 Breast cancer reconstruction**

294

295 Breast cancer is the commonest malignancy and the primary cause of cancer-

296 associated mortality in women (Ginsburg et al., 2017, Winters et al., 2017).

297 Risk factors include: number of relatives affected, menstrual status, advancing

298 age, family history and presence of bilateral disease (Howell et al., 2014,

299 Singletary, 2003). Upto 10 percent of cases are attributed to hereditary

300 malignancy, primarily due to BRCA gene mutations. Presence of BRCA1

301 confers a 50-85% chance of developing breast malignancy (King et al., 2003).

302 Another risk factor is length of oestrogen exposure, associated with early

303 menarche, late menopause and late age at first full-term pregnancy.

304

305 Breast conserving surgery (BCS) with radiotherapy, or mastectomy is

306 normally offered as management options, with comparable oncological

307 outcomes (Veronesi et al., 2002, van Maaren et al., 2019). Autologous

308 abdominal flaps and implant-based procedures are the most frequently

309 employed breast reconstruction (BRR) approaches in the United Kingdom

310 (UK) and the United States of America (USA) (Ho et al., 2017). Autologous

311 BRR involves utilising the patient's own tissues taken from a different parts of

312 the body, utilising excess skin and fat to restore breast volume and excised

313 skin following mastectomy. Skin and fat are taken with intact blood vessels,

314 which are anastomosed to blood vessels in the chest, establishing blood flow

315 and flap survival. Different donor sites can be used, most commonly the

316 abdomen, compared to thigh or buttocks (O'Halloran et al., 2018). Commonly

317 used abdominal-based flaps include the Deep Inferior Epigastric Perforator

318 (DIEP) flap; Transverse Rectus Abdominus Myocutaneous (TRAM) flap and

319 the Superficial Inferior Epigastric Artery (SIEA) flap (Figure 1).
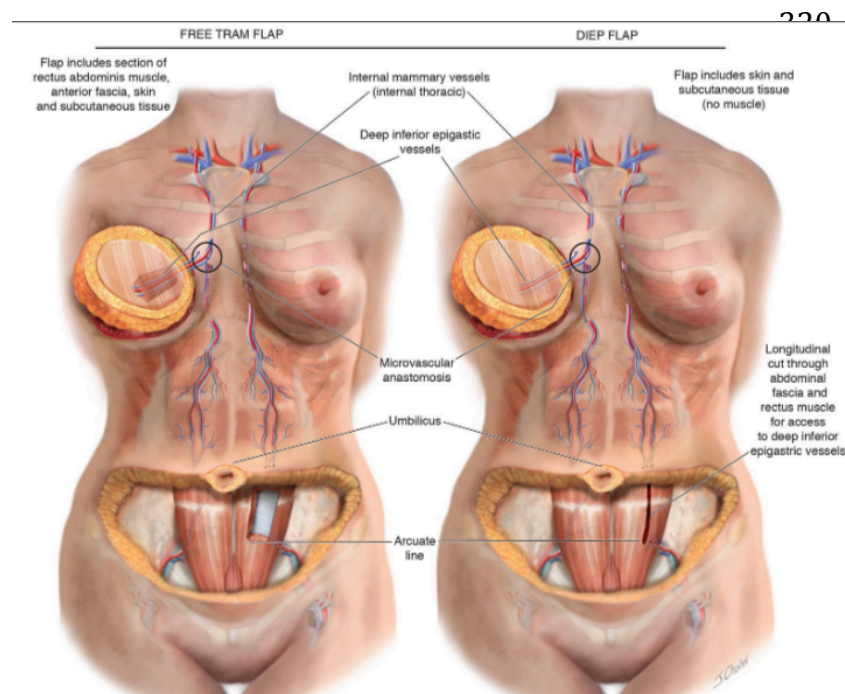
330

331 Figure 1. Figure depicting DIEP and TRAM donor sites; DIEP flap does not

332 contain section of the rectus abdominis muscle (Chovan, 2019)

333

334 Koshima and Soeda first described the DIEP flap in 1989 (Koshima and

335 Soeda, 1989), with Allen and Treece popularising its use in breast

336 reconstruction in 1994 (Allen and Treece, 1994). The Deep Inferior Epigastric

337 Perforator artery originates from the external iliac artery and advances to the

338 lateral edge of the rectus muscle, travelling towards the arcuate line on its

339 deep surface. The flap consists of skin, fat and the perforator vessel only;

340 hence no muscle is harvested unlike for the TRAM flap. The flap is transferred

341 to the chest where the DIEP artery is anastomosed to the recipient vessels,

342 most commonly the internal mammary vessels. Conversely, the TRAM flap is

343 based on the superior epigastric artery. The DIEP flap has largely superseded

344 the TRAM flap, since no muscle is harvested in the DIEP flap. This reduces

345 the risk of complications such as abdominal bulge or hernia. The SIEA flap is

346 based on the superficial inferior epigastric artery, which is of smaller calibre

347 than the DIEP and often too small to perfuse the flap.

348

349 The traditional immediate two-stage Implant-based Reconstruction (IBR)

350 involves placement of a tissue expander in a sub-muscular pocket. This

351 provides a vascularised tissue layer in between the expander and the

352 mastectomy flap, protecting against mastectomy flap necrosis. While this is

353 still the mainstay of treatment, its limitations include lack of lower pole and

354 infero-lateral breast expansion, increased time for subsequent expansion,

355 increased pain associated with expansion, and, re-operative pocket

356 modification at a subsequent date (Hallberg et al., 2018, Smith et al., 2018a).

357 An acellular dermal matrix (ADM) was introduced to address some of these

358 limitations.

359

360 ADMs are utilised in immediate Direct-To-Implant (DTI) breast reconstruction,

361 as time-efficient and potentially less expensive alternative to tissue expanders

362 and free flap surgeries (Salzberg, 2006). In the sub-muscular approach, the

363 ADM is sutured between the inframammary fold and the inferior border of the

364 surgically released pectoralis major muscle to provide support and coverage

365 of the implant in the lower pole of the breast. The ADM properties also allow

366 for incorporation of the allograft to the native skin with minimal fibrosis or

367 contracture. From an aesthetic standpoint, a meta-analysis by (DeLong et al.,

368 2019) concluded that objective observers consider acellular dermal matrix-

369 assisted expander-to-implant breast reconstructions aesthetically superior to

370 reconstruction with only muscular coverage, but patients appear to be equally

371 satisfied with both reconstructive options. The BRIOS trial, comparing one-

372 stage versus two-stage IBR reconstruction, concluded that risks for adverse

373 outcomes were significantly higher in the one-stage ADM group, with no

374 differences in quality of life outcomes (Negenborn et al., 2018).

375

376 **1.3 Loco-regional post-mastectomy radiotherapy (PMRT)**

377

378 Adjuvant loco-regional post-mastectomy radiotherapy of the chest wall and

379 regional lymph nodes (regional lymph node irradiation, RNI: internal

380 mammary and supraclavicular) is historically indicated for locally advanced

381 disease (Yang and Ho, 2013, Macdonald et al., 2011). These indications

382 expanded based on level-one evidence by the Early Breast Cancer Trialists

383 Collaborative Group (EBCTCG) (McGale et al., 2014). The EBCTCG meta-

384 analysis showed significantly improved disease-free and overall survival after

385 PMRT and RNI in intermediate risk women with tumours $\leq$50 mm and 1-3

positive lymph nodes (Marks et al., 2017). Despite these findings, new USA guidelines (Recht et al., 2017) highlight that the EBCTCG review of 1133 patients was based on historical studies from the 1970s and 1980s without the benefits of contemporary systemic treatments, showing much lower risks for all cancer endpoints (Kunkler et al., 2017). The current PMRT recommendations for this intermediate risk group remains controversial and is awaiting the results of the SUPREMO (Selective Use of Postoperative Radiotherapy AftEr MastectOmy) trial, which is the only randomised trial of chest wall RT in which BRR and toxicity have been prospectively assessed (Russell et al., 2015, Donker et al., 2014).

Despite potential oncological advantages, PMRT may have deleterious effects on breast cosmetic outcomes and may increase surgery complications following immediate BRR (O'Halloran et al., 2017). Previous studies evaluating the impact of PMRT on types of immediate BRR showed its potential feasibility in this setting with lower morbidities compared to implant-based procedures (Bennett et al., 2018, Jagsi et al., 2018, Barry and Kell, 2011, Ho et al., 2017). Surprisingly, the rapid adoption of immediate implant-based reconstruction in about 70% of women compared to 34% of autologous procedures when PMRT is recommended may be influenced by surgeon and patient preferences, regardless of current evidence (Jagsi et al., 2018, Potter et al., 2019, O'Halloran et al., 2017). The MROC cohort comparing immediate versus delayed reconstruction evaluated complications requiring re-hospitalization or re-operation; these were designated as "major" complications (Yoon et al., 2018). Reconstructive failures, defined as

411 complications necessitating implant or flap removal, were also recorded.

412 Controlling for demographic and clinical covariates, delayed reconstruction

413 was associated with significantly lower odds of any complication and of major

414 complications, compared with immediate procedures. Delayed autologous

415 patients were at significantly lower risk of complications compared to

416 immediate autologous patients, but there was no difference for implant

417 patients (Yoon et al., 2018).

418

419 However, increasing recommendations for PMRT and growing numbers of

420 immediate BRR have prompted numerous questions about their optimal

421 combination (O'Halloran et al., 2018). The EBCTCG trials omitted patients

422 with immediate BRR and previous publications have not provided clarity

423 concerning the choice between immediate and delayed BRR (McGale et al.,

424 2014). Despite this, immediate autologous BRR is commonly recommended

425 in the setting of PMRT, given the potential long-term benefits on patient's

426 QOL and breast cosmetic satisfaction (Santosa et al., 2018, Velikova et al.,

427 2018). Currently, immediate autologous BRR and PMRT recommendations

428 are highly variable (Momoh et al., 2012, Kelley et al., 2014). The landmark

429 'Gap Analysis' publication in Lancet Oncology by the International Association

430 of Breast Surgery highlighted this as a key unanswered question in breast

431 cancer research (Cutress et al., 2018).

432

433 Potter and colleagues conducted a systematic review showing methodological

434 variations in the definitions of surgery complications, including their disparate

435 reporting, significantly precluding inter-study comparisons (Potter et al., 2011).

436 Complications of autologous breast reconstruction with PMRT include: flap-
437 related fat necrosis, partial/total flap loss, poor wound healing and
438 fibrosis/contracture that reduces breast volume (Ho et al., 2017). Surgical
439 complications contribute variably to decreases in patient satisfaction and
440 impaired cosmetic outcomes (Ho et al., 2017). Potter and colleagues
441 proposed a standardised BRR core outcomes set through expert consensus
442 using Delphi methodology (Potter et al., 2015). The range of complications,
443 including flap-related complications and unplanned surgery, were itemised.
444 The BRR core outcome set has yet to recommend a standardised
445 measurement tool for evaluating surgical complications. Currently, surgeons
446 are recommended to use the Clavien-Dindo classification (CDC) (Dindo et al.,
447 2004). Patient-reported QOL outcomes using validated BRR questionnaires
448 such as the BREAST-Q and the European Organisation for Research and
449 Treatment of Cancer (EORTC) (QLQ)-BRECON23 are recommended to
450 evaluate comparative effectiveness (Pusic et al., 2009, Winters et al., 2010,
451 Winters and Thomson, 2011, Cano et al., 2012, Klassen et al., 2009, Tevis et
452 al., 2018, Santosa et al., 2018, Winters et al., 2018).

453

454 **1.4 Quality of Life (QOL)**

455

456 Health-related quality of life (HRQoL) is a multi-dimensional concept with self-
457 reported domains related to physical, mental, emotional, and social
458 functioning. It goes beyond direct measures of population health, life
459 expectancy, and causes of mortality. Patient-Reported Outcomes (PROs)
460 pertain to measurement of any aspect of a patient's health status that comes

461   directly from the patient. Evaluation of clinical variables such as morbidity and

462   mortality are necessary but not sufficient for adequate outcome assessment,

463   as a mastectomy can have a profoundly negative impact on a woman's

464   physical, psychological and sexual wellbeing (Dean et al., 1983, Eltahir et al.,

465   2013, Winters et al., 2014, Cserni et al., 2018, Cohen et al., 2016, Chen et al.,

466   2010, Erdmann-Sager et al., 2018, Pusic et al., 2009, Klassen et al., 2009,

467   Winters et al., 2010, Cano et al., 2012, Pusic et al., 2012, Cano et al., 2014,

468   Efficace et al., 2015, Macadam et al., 2016, Dikmans et al., 2019). For many

469   procedures, the more discriminating outcome is the patient's own perception

470   of the surgical result and impact on quality of life.

471

472   QOL measurement is pertinent in comparative effectiveness research (CER).

473   Here, two active forms of treatment are compared or usual care in comparison

474   with usual care with an additional intervention element. Capturing the patient-

475   own perception is essential in a prospective clinical CER to examine real-

476   world outcomes related to treatment modalities. Patient-reported outcome

477   measures (PROMs) are standardised questionnaires that measure QOL

478   (Cano et al., 2009). The BREAST-Q and European Organisation for Research

479   and Treatment of Cancer (EORTC) Breast Cancer-Specific Quality of Life

480   Questionnaire (QLQ-BRECON23) are two psychometrically robust, validated

481   disease-specific PROMs to evaluate QoL for breast reconstruction (Pusic et

482   al., 2009, Winters et al., 2018). PROMs for assessing HRQoL in patients with

483   breast cancer comprise the 30-item EORTC QLQ-C30 (Aaronson et al., 1993)

484   and the disease-specific QLQ-BR23 (Sprangers et al., 1996). For breast

485   reconstruction, patient-reported QOL outcomes using these validated BRR

486 questionnaires are integral to comparative effectiveness studies (Pusic et al.,

487 2009, Winters et al., 2010, Winters and Thomson, 2011, Cano et al., 2012,

488 Klassen et al., 2009, Tevis et al., 2018, Santosa et al., 2018). Their

489 development and validation includes three phases (Pusic et al., 2009): 1)

490 Phase I: Item Generation and Conceptual Framework Formation - generates

491 a pool of items to ensure all important areas are considered for inclusion in

492 the final scale; encompasses literature review, patient interviews

493 (predominant component), and expert opinion. Item pool is then pre-tested on

494 small sample of patients – to check ambiguities; confirm appropriateness, and

495 determine acceptability and completion time; 2) Phase II: Item Reduction -

496 field testing using larger patient sample – to revise or eliminate items; and 3)

497 Phase III: Psychometric evaluation using Rasch measurement methods and

498 analyses to guide scale construction. Reliability, validity, and responsiveness

499 are confirmed.

500

501 The COSMIN (COnsensus-based Standards for the selection of health

502 Measurement INstruments) initiative has aimed to improve the selection of

503 outcome measurement instruments by developing tools for selecting the most

504 appropriate instrument. Selecting unsuitable/poor quality outcome

505 measurement instruments may generate bias, lead to waste of resources and

506 be potentially unethical as patients contribute little to knowledge but still suffer

507 from the burden/risks of the study.

508

509 When selecting an instrument, the outcome should be clearly defined. With

510 respect to HRQoL, it should be clarified which subdomains are relevant for the

511 target population. The COSMIN initiative has developed several tools to help

512 researchers choose the most appropriate initiative. These include: 1)

513 COSMIN taxonomy and definitions of measurement properties, clustered

514 within three domains, i.e. validity [construct, content and criterion validity],

515 reliability [internal consistency, reliability, measurement error] and

516 responsiveness; 2) COSMIN checklist to evaluate the methodological quality

517 of studies on measurement properties; 3) Search filter for finding studies on

518 measurement properties; 4) Protocol for systematic reviews of outcome

519 measurement instruments; 5) Database of systematic reviews of outcome

520 measurement instruments; and 6) Guideline for selecting outcome

521 measurement instruments for outcomes included in a Core Outcome Set.

522

523 A domain refers to the distinct area of experience that a given questionnaire is

524 designed to explore. As per the SPIRIT-PRO extension, PRO measures

525 maybe multidimensional (e.g. HRQOL) or unidimensional (e.g. specific

526 symptom such as pain). Defining the key objectives and hypothesis a priori

527 will encourage the key PRO domains to include in the study, reducing the risk

528 of multiple statistical testing, a Type I error and selective reporting of PROs

529 based on statistically significant results (Calvert et al., 2018). The BREAST-Q

530 breast reconstruction module has satisfaction domains (e.g. satisfaction with

531 back and abdomen). When evaluating patients undergoing abdominal-based

532 flaps, the 'satisfaction with abdomen' domain is pertinent. Conversely, if

533 evaluating the Latissimus Dorsi (LD) flap, the 'satisfaction with back' is more

534 pertinent. The chosen domains should be described in the study protocol a

535 priori.

536　Health utility measures include the EQ-5D instruments, with the 5-level EQ-5D

537　version (EQ-5D-5L) introduced by the EuroQol Group in 2009 to improve the

538　instrument's sensitivity (EuroQol, 2019). It consists of five dimensions:

539　mobility, self-care, usual activities, pain/discomfort and anxiety/depression.

540　Each dimension has 5 levels, from "no problems" through to "extreme

541　problems". The patient indicates his/her health state by choosing the most

542　appropriate statement in each of the five dimensions. This decision results in

543　a 1-digit number that identifies the level selected for that dimension (EuroQol,

544　2020). The digits for the five dimensions can be combined into a 5-digit

545　number that describes the patient's health state. This summary index score is

546　based on societal preference weights for the health state, with the weights

547　referred to as 'utilities'; these are often used to compute QALYs for economic

548　analyses (EuroQol, 2019). Health state index scores generally range from

549　less than 0 to 1 (the value of full health), with higher scores indicating higher

550　health utility.

551

552　The EORTC has developed the QLU-C10D, a multi-attribute utility instrument

553　derived from the cancer-specific quality of life questionnaire, EORTC QLQ-

554　C30 (King et al., 2016). U.K.-specific utility weights have been defined

555　(Norman et al., 2019) and will enable cost-utility analysis (CUA) for economic

556　evaluation of new oncology therapies and technologies in the UK, where cost

557　and resource allocation are fundamental. Nevertheless, there is a growing

558　view that measurement of health alone (for example through QALYs) in

559　economic evaluation is often insufficient (Ryan et al., 2006). This is especially

560　the case where there are significant spillover effects of intervention, for

561 example impacts on carers or family. Moreover, QALYs typically measure

562 one's health status, but do not measure what people are capable of doing, as

563 a sense of broader wellbeing. To address this, (Flynn et al., 2015) proposed

564 the use of an alternate cost utility index, the Investigating Choice Experiments

565 Capability Measure for Adults (ICECAP-A), which has 5 attributes, each of

566 which are scored between 1 to 4 ranging from full capability to no capability.

567 These include: 1) Stability (being able to feel settled and secure) 2)

568 Attachment (being able to have love, friendship and support) 3) Autonomy

569 (being able to be independent) 4) Achievement (being able to achieve and

570 progress) and 5) Enjoyment (being able to have enjoyment and pleasure).

571

572 An important concept when interpreting QOL outcomes is the minimally

573 important difference (MID). MID is defined as the smallest change in a HRQoL

574 domain, which is perceived as 'important' by the patient and clinician, which

575 may indicate a change in management (Cocks et al., 2012, King, 1996, Cano

576 et al., 2014). Small differences in QOL scores maybe statistically significant,

577 without clinical relevance. MID estimates can also facilitate clinical trial design

578 by informing the choice of sample size and specifying clinical trial endpoints.

579 MIDs are determined by anchor-based methods, which express a change in

580 HRQoL scores within specific domains of a patient and/or physician-derived

581 rating, and distribution-based methods, that utilise statistical distribution of

582 HRQoL scores (e.g. standard deviation), often considered as providing

583 supportive evidence to anchor-based methods (Musoro et al., 2019). For the

584 reconstruction module of the BREAST-Q, a minimal important difference

585 score of 4 points has been proposed to be clinically useful when assessing an

586   individual patient's outcome for the different domains (i.e. breast satisfaction;

587   psychosocial wellbeing; physical wellbeing and sexual wellbeing), based on

588   analysis of prospectively collected data from 3052 Mastectomy

589   Reconstruction Outcomes Consortium (MROC) patients (Voineskos et al.,

590   2020).

591

592   Longitudinal analysis of HRQOL is pertinent, as HRQOL maybe incorporated

593   by oncology trials as a major endpoint, in order to evaluate the clinical benefit

594   of new therapeutic strategies. Methods used to analyse longitudinal HRQOL

595   include: 1) general linear mixed model (GLMM) 2) Item Response Theory

596   (IRT) models and 3) time-to-event models such as the time-to-HRQoL score

597   deterioration (TTD). One challenge associated with longitudinal assessment

598   of HRQOL is the potential occurrence of a response shift (RS) effect. This is

599   defined as "a change in the meaning of one's self-evaluation of a target

600   construct". This maybe due to change in patients' internal standards of

601   measurement (i.e. scale recalibration); change in values (i.e. the domains

602   making up the target construct) or redefinition of the target construct

603   (reconceptualization). TTD has been recommended as the optimal method to

604   analyse longitudinal HRQOL, as it takes into account the occurrence of the

605   RS recalibration component by choosing different reference scores to qualify

606   the deterioration and it is reported using hazard ratios (HR), a format familiar

607   to clinicians. Anota et al. demonstrated that definition of TTD can influence

608   change in HRQOL results, precluding inter-study results comparison. In the

609   breast cancer study, the choice of the reference score impacted on the

610   median TTD. When the best previous score was used as the reference,

611   instead of the baseline score, the median TTD of cognitive functioning

612   decreased while that of the breast and arm symptoms increased. The TTD

613   approach has the advantage of taking recalibration into account without

614   additional questionnaires, by using changing scores as a reference.

615

616   To facilitate transparent and better reporting of QOL outcomes, a number of

617   tools have been designed. The CONSORT-PRO statement was designed to

618   promote transparent reporting of RCTs where PROs are primary or important

619   secondary outcomes (Calvert et al., 2013). The statement was based on the

620   methodological framework for guideline development proposed by the

621   EQUATOR Network (Moher et al., 2010), with the initial work led by ISOCOL.

622   The development process involved a systematic review of existing guidelines,

623   survey of key stakeholders, with final dissemination to ISOCOL members.

624   Subsequently, the final CONSORT PRO guidance was released. For trial

625   protocols, the SPIRIT PRO guidance has been published (Calvert et al.,

626   2018). Whilst guidance exists to facilitate transparent reporting, it is important

627   that PRO findings are obtained from robust methodological practices and are

628   analysed consistently. To address this, the Setting International Standards in

629   Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data

630   (SISAQOL) Consortium has published consensus recommendations for PRO

631   analysis in cancer RCTs (Coens et al., 2020).

632

633   For implant-based reconstruction, there is growing evidence that PMRT

634   increases the rate of serious adverse events, with a reduction in patient

635   satisfaction, quality of life (QoL) and objective cosmetic outcomes (Ricci et al.,

636 2017, Cordeiro et al., 2014, Pu et al., 2018). Conversely, for autologous flap

637 reconstruction, the evidence is more equivocal and the optimal sequence of

638 reconstruction and PMRT is unclear (Rogers and Allen, 2002, Chatterjee et

639 al., 2009, Cooke et al., 2017, Taghizadeh et al., 2015, O'Connell et al.,

640 2018a). Overall, the dogma is that patients who are expected to require

641 PMRT are advised to undergo delayed autologous breast reconstruction or

642 'delayed-immediate' reconstruction, which utilises a temporising implant,

643 facilitating preservation of native breast skin and a chest wall mound whilst

644 the patient awaits a planned exchange to autologous reconstruction

645 (Kronowitz et al., 2004). However, the patient must live without a breast for

646 substantial time and this may culminate in psychosocial morbidity associated

647 with mastectomy alone (Wilkins et al., 2000). Moreover, there is perceived

648 evidence to suggest potential satisfactory outcomes, comparable complication

649 rates and PROs, after immediate autologous breast reconstruction with

650 adjuvant radiotherapy (Chatterjee et al., 2009, Taghizadeh et al., 2015, Cooke

651 et al., 2017). Furthermore, some protagonists have reported higher baseline

652 QoL scores before intended immediate breast reconstruction with adjuvant

653 radiotherapy compared to intended delayed reconstruction (Billig et al., 2017).

654

655 Subsequent parts of the thesis will focus on key papers published by the

656 author on the aforementioned theme.

657

658 **1.5 Aim**

659 The aim of the author's work was to evaluate outcome measurement and

660 reporting quality in surgery, with a focus on plastic & reconstructive breast

surgery and autologous reconstruction with or without radiotherapy. The author has also assessed the quality of evidence, reporting of clinical complications as well as patient-reported complications, and evaluated compliance of RCTs against the CONSORT checklist for Non-Pharmacological Treatments (NPT).

2. Report Quality of Surgical Randomised Controlled Trials (RCTs)

RCTs represent the gold standard in evaluating intervention effectiveness and are classified as Level Ib by the Oxford Centre for Evidence Based Medicine. However, poor reporting can inhibit adequate critical appraisal by readers and lead to inconsistencies between results and conclusions. Adequate and accurate reporting is fundamental to facilitate critical appraisal and interpretation of the data by the readers.

The Consolidated Standards of Reporting Trials (CONSORT) statement was developed to provide a set of standards for transparent reporting of RCTs. Surgical RCTs have inherent challenges, with issues related to blinding, inconsistent care provider expertise and centres' volume all having an impact on the outcomes. The 2008 CONSORT extension for non-pharmacological treatment interventions (CONSORT NPT) (Boutron et al., 2008) is an extension on the 2001 CONSORT checklist, and takes into account the aforementioned challenges inherent to surgical RCTs.

The author conducted a systematic review and meta-analysis evaluating the compliance of RCTs in Surgery against the CONSORT NPT statement (**Yao et al., 2014)** [Appendix 1,2,10]. The aim of the project was to answer a specific research question, i.e. "in the Ophthalmic Surgery literature, is the reporting quality of RCTs against the CONSORT NPT statement optimal." A systematic review methodology was deemed most optimal to evaluate all the

710   literature published in the field, within a defined period, with pre-defined

711   inclusion and exclusion criteria.

712

713   The mean CONSORT score of the 65 RCTs was 8.9 out of 23 (39%, range

714   3.0–14.7, SD 2.49). The poorest-reported items were: title and abstract;

715   details of how adherence with protocol was assessed; and, interpretation of

716   results [Figure 2; Appendix 1]. No paper adequately reported all items in the

717   CONSORT checklist. There was no correlation between CONSORT score

718   versus the impact factor (Spearman rho = 0.14, P = 0.29, Cohen's d = 3.297),

719   or the number of authors (Spearman rho = 0.01, P = 0.93, Cohen's d = 1.533).

720   There was no statistically significant difference between the scores of single-

721   and multi-centre trials.

722

723   This work [Appendix 1,2,10] has been cited multiple times. The poor reporting

724   compliance identified corroborated the poor compliance in a number of other

725   surgical specialties (Camm et al., 2015). The work formed basis for change of

726   policy and the requirement for mandatory completion and uploading of a

727   CONSORT statement by authors when submitting articles to the peer-

728   reviewed journal, International Journal of Surgery (IJS), with significant

729   improvement in compliance (Agha et al., 2016).

730

731   The strengths of the review include the fact that subjectivity was minimised in

732   by predefining the scoring strategy among the reviewers. The item was only

733   scored if all elements were reported, on the basis that CONSORT items

734   represent absolutely fundamental information; 'the minimum criteria,' that

735   should be reported in a RCT. Furthermore, all items on the checklist were

736   given equal weighting. Whilst this may not reflect their relative importance, it

737   was nonetheless an objective approach to analyse deficits, patterns, as well

738   as overall compliance. Two independent authors performed the scoring to

739   reduce bias.

740

741   There were several limitations. The period studied was restricted to 2011, so

742   did not allow analysis of temporal trends in the CONSORT score. Many items

743   contain multiple elements. Whether reviewers score items in regard to the

744   multiple elements is a potential area of subjectivity, although this was

745   minimised by the aforementioned strategy. No correlation was identified

746   between CONSORT score and surrogate markers of papers quality; this

747   maybe attributed to inclusion of inadequately powered studies. Finally, the

748   search was restricted to the English language, with the potential of missing

749   articles for inclusion.

750

751   Given many journals have now made uploading a reporting guidance checklist

752   as a mandatory part of article submission, future work should analyse

753   temporal trends in compliance and reporting quality. Reviews should be

754   prospectively registered, with comprehensive search of the databases,

755   without language restriction.

756

757   I was involved in the conception, design, design of search strategy, database

758   searching, data extraction, statistics, interpretation, drafting and critical review

759   of manuscript.

3. Clinical outcomes, Patient-Reported Outcomes (PROs) and Cost of Deep Inferior Epigastric Perforator (DIEP) flap versus Implant-based Breast Reconstruction

Two of the commonest reconstructive modalities include autologous reconstruction using the deep inferior epigastric perforator (DIEP) flap and implant-based reconstruction (IBR). The treatment choice is determined by a number of patient and surgeon factors. Nevertheless, many plastic surgery units worldwide consider autologous reconstruction superior, replacing "like with like" (Sisco et al., 2012). There is growing evidence to suggest that autologous BRR may culminate in superior clinical and PROs (Matros et al., 2015, Santosa et al., 2018, Lagares-Borrego et al., 2016, Tonseth et al., 2008, Atherton et al., 2011). IBR is associated with complications, including infection, migration, exposure/extrusion, rupture, patient dissatisfaction due to edge visibility/implant animation and reduced/absent nipple sensation. Capsular contracture can result in pain, asymmetry, increased palpability and need for implant removal (Agha et al., 2015). Allergan's 10-year cumulative risk study found that 24.6% of patients who underwent IBR developed capsular contracture (Spear and Murphy, 2014). Conversely, DIEP flap is widely considered the "gold standard" for postmastectomy BRR.

The known surgical complications for DIEP and IBR are detailed below. The rates of complications are derived from the data from the Mastectomy Reconstruction Outcome Consortium (MROC) cohort (Wilkins et al., 2018).

Breast Complication	Implant (DTI and TE) (%)	Pedicled TRAM (%)	Free TRAM (%)	DIEP (%)	LD (%)
Haematoma	3.5	3.6	4.1	6.0	4.1
Wound dehiscence	1.6	1.2	1.0	3.6	1.4
Wound infection	10.0	6.0	4.1	3.8	8.2
Mastectomy skin flap necrosis	6.6	6.0	6.2	7.7	5.5
Seroma	2.9	2.4	0.0	0.8	2.7
Capsular contracture	0.8	-	-	-	1.4
Implant malposition	0.5	-	-	-	1.4
Implant leakage, rupture, and/or deflation	1.1	-	-	-	0.0
Acute partial flap necrosis	-	11.9	5.2	2.5	1.4
Total flap loss	-	1.2	2.1	1.4	0.0
Fat necrosis	-	7.1	5.2	9.0	0.0
Seroma	-	0.0	2.1	5.2	19.2
Abdominal wall bulge, laxity or hernia	-	4.8	3.1	1.6	0.0

785

786 From a cost standpoint, some authors have argued that DIEP reconstruction

787 is more cost-effective, resulting in lower overall complications and superior

788 PROs, compared with IBR (Matros et al., 2015, Atherton et al., 2011,

789 Lagares-Borrego et al., 2016). Whilst some North American and European

790 centres have published cost-analyses on DIEP and IBR, the data are sparse,

791 especially from public and free universal health care system settings.

792

793 The author conducted and published a meta-analysis evaluating clinical

794 outcomes, Patient-Reported Outcomes (PROs) and Cost of DIEP flap versus

795 Implant-based Breast Reconstruction (**Khajuria** et al., 2019) [Appendix 3-5,

796 11]. The aim of the project was to answer a specific research question, i.e. in

797 patients aged 18 or over with breast malignancy undergoing mastectomy and

798 breast reconstruction, does DIEP reconstruction lead to superior clinical,

799 patient-reported outcomes and cost compared with Implant-based

800 reconstruction. A systematic review methodology was deemed the most

801 optimal to answer the research question, as there had been a number of

802 studies published in the literature evaluating DIEP and IBR, and the SR was

803 performed to obtain overall summary measures for outcomes, to help facilitate

804 informed consent and the shared decision making process with the patient.

805

806 Robust Cochrane methodology was followed and comprehensive screening

807 of 6381 articles was undertaken. Cochrane's Review Manager 5.3 software

808 was used to perform the meta-analysis. Odds ratios [95% confidence intervals

809 (CI)] were used to evaluate dichotomous outcomes (surgical complications).

810 Standard mean differences (95% CI) were used for continuous outcomes

811 between treatment groups. Heterogeneity between studies was assessed in

812 Review Manager 5.3 (Liu et al., 2013) using the Higgins and Thompson's I^2

813 statistic.(Higgins and Thompson, 2002) Levels of heterogeneity were defined

814 as: low (I^2 <50%), moderate (I^2 50% - 80%), and high (I^2 >80%). A random-

815 effects model was used for cohorts with heterogeneity (I^2>50%) (DerSimonian

816 and Laird, 2015). As heterogeneity was generally moderate or high, and

817 outcome measures differed between studies, these were combined using the

818 DerSimian and Laird random-effects model.

819

820 Out of 6,381 articles screened, 16 were included [unilateral 782 DIEPs, 376

821 implants; mean age 49 years, follow-up (months): DIEP 29.9; IBR 35.5]. Mean

822 flap loss and fat necrosis rates were 3.97% (SD 4.90) and 9.67% (SD 17.0),

823 respectively. There was no difference in mean length of stay (MD 0.63

824 [confidence interval (CI) −9.17 to 10.43]; P =0.90) [Figure 5; Appendix 4]. The

825 number of reoperations for complications was significantly lower in DIEP

826 versus IBR [MD −0.29 (CI −0.48 to −0.09); P <0.01] [Figure 6; Appendix 5].

827 The mean difference (MD) is the difference in means of the intervention and

828 control groups, whereas the standardised mean difference (SMD) is the MD

829 divided by the standard deviation (SD), from either or both groups (Faraone,

830 2008). MD was employed as the studies included in the meta-analysis used

831 the same, continuous outcome and unit of measure. Conversely, a SMD

832 would be used when studies assess the same outcome but measure it in

833 different ways, e.g. measuring a clinical outcome where studies have used

834 different psychometric scales (Paramanandam and Roberts, 2014). It would

835 be necessary to standardise to a uniform scale before they can be combined.

836 The SMD is also easier to interpret compared to MD, as SMD can be

837 interpreted using Cohen's guidelines, where SMD of 0.2, 0.5 and 0.8 equates

838 to a small, medium and large effect respectively (Cohen, 1988). It can also be

839 easily converted to a number needed to treat (NNT) (da Costa et al., 2012);

840 NNTs are more intuitive and easier to interpret for clinicians.

841

842   There were no randomized controlled trials. Study quality was low with high

843   risk of bias. One study reported $11,941/Quality-adjusted Life Year

844   incremental cost effectiveness ratio for DIEP, with higher breast Quality-

845   adjusted Life Year (DIEP 19.5; IBR 17.7) using Breast Questionnaire; Two

846   comparative studies evaluating PROs favoured DIEP. Three studies

847   evaluating cost, favoured DIEP.

848

849   Study quality was assessed using the Grading of Recommendations

850   Assessment, Development and Evaluation (GRADE) tool (Atkins et al., 2004).

851   The quality of evidence was applied to each outcome. An overall GRADE

852   quality rating was then applied to the body of evidence across outcomes. Key

853   elements were considered: study design; study quality pertaining to study

854   methods/execution, consistency (i.e. how similar are the estimates of effect

855   across studies); directness (extent to which patient population, interventions

856   and outcome measures are similar to those of interest); and precision (width

857   of the confidence intervals) (Atkins et al., 2004). The highest quality rating is

858   for randomized trial evidence. This is downgraded to moderate, low, or very

859   low quality evidence, based on: limitations in design and implementation of

860   studies,   suggesting   high   likelihood   of   bias;   unexplained

861   heterogeneity/inconsistency in results; and imprecision of results (wide

862   confidence intervals).

863

864   The Risk of Bias in Nonrandomized Studies of Interventions (ROBINS-I) tool

865   was utilised to assess risk of bias (RoB). It covers 7 domains from which bias

866   may be introduced, with "signalling questions" facilitating judgments about

867   RoB for each outcome. The judgments within each domain were carried

868   forward for an overall RoB judgment across bias domains. For the breast

869   reconstruction reviews, the domains responsible for moderate/serious risk of

870   bias, included: 1) bias due to presence of confounders between groups, which

871   were unaccounted for. Majority of the studies also did not define the patient

872   baseline characteristics, while some studies, due to their retrospective nature,

873   were not able to report patient characteristics, as the data were unavailable.

874   No data were available on the level of care provider expertise and centre's

875   volume; 2) bias in the selection of participants into the study, with all studies

876   being non-randomised studies of the effects of interventions (NRSIs); 3) bias

877   due to deviations from intended interventions, with lack of pre-published

878   protocols or information on how adherence to protocol was assessed; 4)

879   missing data, insufficient follow-up and attrition bias; and 5) bias in outcome

880   measurement – heterogeneity, with outcomes not defined a priori or graded,

881   lack of blinding with ascertainment and response biases.

882

883   In this systematic review, the mean total flap loss rate was 3.97%. This is

884   greater than the 1.4% rate from the MROC cohort (Wilkins et al., 2018). The

885   mean fat necrosis rate in this systematic review was 9.67%; this is similar to

886   the 9.0% rate in the MROC cohort (Wilkins et al., 2018). The capsular

887   contracture rate reported in the systematic review was 3.33%, greater than

888   0.8% reported in the MROC cohort. The differences maybe explained by the

889   small sample sizes and significant heterogeneity in outcome reporting in the

890   studies included in the systematic review. Majority of the studies did not

891 define the complications a priori, precluding adequate interpretation of the

892 results.

893

894 This was the first published systematic review and meta-analysis in available

895 literature to evaluate clinical outcomes, PROs, and cost of DIEP versus IBR. It

896 was published in a peer-reviewed journal (**Khajuria** et al., 2019) [Appendix

897 11]. One of the most pertinent findings from this review was the poor reporting

898 of outcomes in the Plastic & Breast Reconstructive Surgery literature. Clinical

899 complications are poorly reported as percentages without standardised

900 reporting using classifications such as Clavien-Dindo (Dindo et al., 2004) and

901 complication-specific classifications, e.g. Baker's classification (for capsular

902 contracture) or flap necrosis classification (Lie et al., 2013). Reporting

903 complications as percentages is suboptimal, as there is no stratification

904 according to management (nonoperative/conservative or operative

905 management) (Khajuria and Mosahebi, 2019, Khajuria and Farhadi, 2020).

906

907

908 Moreover, there was inconsistency and heterogeneity in clinical outcome

909 reporting. Only 8/14 (57.1%) studies evaluating DIEP reported flap loss rates.

910 Only 3/6 (50.0%) studies evaluating IBR reported implant-specific

911 complications, including capsular contracture. Only 1 of these studies

912 classified capsular contracture according to the Baker's classification. Since

913 classification/grades help guide management strategies, inaccurate

914 classification, and grading of these complications, risks biased comparisons of

915 clinical outcomes between studies, rendering it difficult to interpret the study

916 findings. This corroborates the results from the systematic review by (Potter et

917 al., 2011), on reporting quality of BRR clinical outcomes, that identified poor

918 reporting quality and need for a core outcome set to facilitate outcome

919 assessment in effectiveness studies. Furthermore, no studies reported

920 outcomes using the validated CDC.

921

922 Out of 6 IBR studies, 3 reported implant-specific complications; 2 out of 6

923 studies did not categorize type of IBR and reported as "implant

924 reconstruction". Three out of 6 studies reported EP reconstruction, and 1

925 reported DTI. Due to the scarcity of IBR data, further subgroup analysis was

926 not possible. Future studies should clearly specify the type of reconstruction –

927 DTI/EP; subpectoral or prepectoral and whether acellular dermal matrix was

928 utilized. Adequate reporting as part of a core outcome set will facilitate inter-

929 study comparisons and meta-analyses.

930

931 Out of 16 studies, only 2 comparative studies (12.5%) reported PROs. A

932 major paradigm shift is needed to incorporate PROs in all studies evaluating

933 BRR, as also supported by the recent publication of the "Gap analysis" in

934 BRR (Cutress et al., 2018). Evaluating clinical outcomes without PROs is a

935 major drawback in evaluating outcomes in BRR, as the reconstruction must

936 satisfy the patient with regard to physical, psychological, and sexual well-

937 being (Pusic et al., 2009). Disregard of these domains renders outcome

938 assessment incomplete and suboptimal. Two comparative studies that

939 evaluated PROs in our review favored DIEP reconstruction. Matros et al.

940 utilized a robust, validated, disease-specific questionnaire, BREAST-Q.

941 BREAST-Q scores were reported as consistently higher for DIEP compared

942 with IBR in postoperative years 1–8, with a higher breast Quality-adjusted Life

943 Year for DIEP. Conversely, Tønseth et al. used generic PRO tools, SF-36,

944 which revealed no difference in QoL between DIEP and IBR, and Visual

945 Analog Scale, with superior cosmetic outcome with DIEP. The study also

946 used a non-validated study-specific questionnaire that demonstrated higher

947 breast satisfaction, improved social relationship, and body image satisfaction

948 for DIEP. The results from the author's review corroborated results from

949 (Santosa et al., 2018) who evaluated PROs for 2,013 patients (523

950 autologous reconstructions; 1,490 IBR) from the MROC cohort, pre and 2

951 years post BRR, using the BREAST-Q.

952 The 4 domains evaluated were as follows: satisfaction with breasts,

953 psychosocial well-being, physical well-being, and sexual well-being. At 2

954 years, patients who underwent autologous reconstruction had higher breast

955 satisfaction, higher psychosocial well-being, and sexual well-being than did

956 those who underwent IBR (Santosa et al., 2018). Lack of a significant

957 difference in QoL between DIEP and IBR reported by Tønseth et al in the

958 author's study may be due to the small sample size in the study (n = 50) and

959 use of a nonspecific, generic QoL tool, SF-36, which may not be sensitive

960 enough to measure changes as a result of BRR intervention or to capture all

961 aspects of outcome specific to breast surgery. Moreover, as purported by the

962 author, QoL domains should be defined a priori, facilitating estimations of

963 potential effect size (Winters and Khajuria, 2018b). Three comparative studies

964 evaluated cost, all favoring DIEP (Matros et al., 2015, Atherton et al., 2011,

965 Lagares-Borrego et al., 2016). Two studies were conducted in a universal

966 health care system (UK and Spain) and 1 was conducted in a health

967 insurance-based model (USA), making direct comparisons on cost difficult.

968 This is exacerbated by only 1 study performing robust cost effectiveness

969 analysis, calculating an ICER of $11,491 for DIEP (Matros et al., 2015). An

970 ICER is the additional cost for DIEP to obtain 1 year of perfect breast-related

971 QoL compared with IBR; a threshold of $50,000–$100,000 for a year in

972 perfect overall health has been deemed as acceptable for the adoption of new

973 technologies or techniques in developed countries (Laupacis et al., 1992).

974 Heterogeneity in cost-evaluation methods and reporting prevented the

975 calculation of an overall cost-effectiveness summary measure in the author's

976 systematic review.

977

978 Adequate reporting of core outcome measures is required to minimize

979 reporting bias and facilitate evidence synthesis. Prospective, multicentre,

980 cohort studies using robust PROMs tools, evaluating cost-effectiveness and

981 contributing to national/international registries, will facilitate national-level

982 policy and shared decision-making.

983

984 In the DIEP versus IBR review, I was involved in the conception, design,

985 PROSPERO registration, design of search strategy, database searching, data

986 extraction, performing all the analysis, interpretation of the data, drafting and

987 critical review of the manuscript.

988

989

990

4. Clinical outcomes and PROs of Immediate versus Delayed autologous breast reconstruction, in context of adjuvant and neoadjuvant Radiotherapy (RT)

The author conducted and published a meta-analysis to evaluate the quality and strengths of the evidence regarding surgical complications in autologous abdominal flaps in the context of RT receipt and timing (Khajuria et al., 2020) [Appendix 6-9, 12]. The aim of the project was to answer a specific research question, i.e. in patients aged 18 or over with breast malignancy undergoing mastectomy and breast reconstruction, does immediate reconstruction yield superior clinical and patient-reported outcomes compared with delayed reconstruction. A systematic review methodology was deemed the most optimal to answer the research question, as there had been a number of studies published in the literature evaluating immediate and delayed reconstruction, and the SR was performed to obtain overall summary measures for outcomes, to help facilitate informed consent and the shared decision making process with the patient. The recent Breast Cancer Campaign gap analysis publication in Lancet Oncology identified this a key clinical and translational research gap in breast cancer research (Cutress et al., 2018), with radiotherapy timing being a contentious issue.

The UK National Flap Registry (UKNFP) Report 2019 identified that the majority of breast reconstructions were, in fact, delayed reconstructions (49.0%), compared with immediate reconstructions (45.2%). There is inter-unit and regional variation in terms of timing of radiotherapy and breast

1016   reconstruction, which is a contentious issue, as also highlighted in the 'Gap

1017   Analysis' publication. So, the current question in the field is that in patients

1018   with breast cancer undergoing mastectomy, who need radiotherapy, should

1019   the flap be irradiated or should the reconstruction be delayed, with evaluation

1020   of clinical and quality of life outcomes.

1021

1022   Radiotherapy timing evaluated commonly used adjuvant and less commonly

1023   pre-operative RT (Neo RT), administered prior to skin-sparing mastectomy

1024   and immediate breast reconstruction (Zinzindohoue et al., 2016). QOL studies

1025   were evaluated for their methodological rigour. This was the first meta-

1026   analysis published in available literature to compare clinical and patient-

1027   reported outcomes for abdominal-based breast reconstructions in the context

1028   of adjuvant, neoadjuvant and no RT groups (Khajuria et al., 2020) [Appendix

1029   12].

1030

1031   Robust Cochrane methodology was employed. In this review, if CDC grades

1032   were not defined, the complications reported by the included studies were

1033   retrospectively graded by two independent authors according to CDC; any

1034   discrepancies were discussed and agreed by the senior author. Cochrane's

1035   Review Manager 5.3 software was used to perform the meta-analysis. Odds

1036   ratios [95% confidence intervals (CI)] were used to evaluate dichotomous

1037   outcomes (surgical complications). Standard mean differences (95% CI) were

1038   used for continuous outcomes between treatment groups.   Heterogeneity

1039   between studies was assessed in Review Manager 5.3 (Liu et al., 2013) using

1040   the Higgins and Thompson's I^2 statistic (Higgins and Thompson, 2002).

1041 Levels of heterogeneity were defined as: low (I^2 <50%), moderate (I^2 50% -

1042 80%), and high (I^2 >80%). A random-effects model was used for cohorts with

1043 heterogeneity (I^2>50%) (DerSimonian and Laird, 2015). As heterogeneity was

1044 generally moderate or high, and outcome measures differed between studies,

1045 these were combined using the DerSimian and Laird random-effects model.

1046

1047 No eligible studies prospectively graded surgical complications according to

1048 an accepted classification such as CDC [fat necrosis; partial or total flap loss;

1049 infection and wound complications (dehiscence, delayed wound healing)].

1050 One study graded partial flap loss using a novel flap necrosis classification

1051 system (Modarressi et al., 2017), adapted from Kwok et al (Lie et al., 2013).

1052 Only 30.30% (30/99) of all surgical complications reported across the 12

1053 included studies were defined a priori and none were classified as per CDC.

1054

1055 Meta-analyses comparing adjuvant versus no RT showed no inter-study

1056 differences in rates of: overall complications; CDC grade 3; CDC grade 2;

1057 surgical complications; fat necrosis; unplanned emergency re-operations for

1058 complications or infection [Figure 8, Appendix 7]. There were no total flap

1059 losses. Likewise, comparing neoadjuvant versus no RT showed no

1060 differences in overall complications, CDC grade 3, fat necrosis or total flap

1061 loss rates. Rates of partial flap loss were higher in the Neo RT versus no RT

1062 groups [Figure 9, Appendix 8].

1063

1064 Data were also pooled to provide an overall summary measure of combined

1065 RT (adjuvant and neoadjuvant) compared to no RT. The merit of this

1066 approach was discussed with a senior clinical oncologist and the rationale

1067 was to explore the impacts of radiotherapy in general and as an expanded

1068 patient group, that is potentially hypothesis-generating. This showed

1069 significantly higher overall complications in the combined RT groups

1070 compared with no RT, with no inter-study differences in: CDC grade 3; grade

1071 2 complications; rates of fat necrosis or emergency re-operations for

1072 complications [Figure 10; Appendix 9]. Rates of partial flap loss were also

1073 higher in the combined RT compared to no RT, with no differences in rates of

1074 total flap loss, infection or wound complications.

1075

1076 There was limited reporting of patient-reported QOL outcomes. Study designs

1077 comprised two prospective studies (Cooke et al., 2017, Billig et al., 2017) and

1078 one retrospective study (O'Connell et al., 2018a), limited by small patient

1079 numbers and short follow-up for the adjuvant groups. There was no

1080 standardized evaluation of cosmetic outcomes, precluding meta-analyses.

1081 Studies lacked robust methodology and quality and were based on

1082 independent panel assessments of medical photographs, with more recent

1083 use of Vectra XT 3-D (O'Connell et al., 2018b).

1084

1085 There were no significant intergroup differences in surgical complications

1086 following PMRT or Neo RT versus no RT. Reported meta-analyses of surgical

1087 complications in pooled RT groups (PMRT and neo-adjuvant) in this review

1088 however showed significantly higher rates of overall complications and partial

1089 flap loss following RT compared to no RT groups. Combined analyses of RT

1090 patients reflect the value of adequately large patient groups, where cohorts of

1091   at least 1000 women are recommended for the studies to be adequately

1092   powered to detect significant differences. It illustrates that a larger sample

1093   size with more events serves as the proof of principle that the individual

1094   studies are underpowered to detect statistical differences based on fewer

1095   event rates.

1096

1097   Current evidence for irradiating autologous abdominal flaps remains indirect

1098   and largely of poor quality within only two moderate quality studies out of

1099   twelve in this report. Future cohort studies should be designed and powered

1100   akin to quasi randomised trials and take advantage of newly evolving study

1101   designs including multiple cohort randomised controlled trials or trials within

1102   cohorts (Young-Afat et al., 2017). These designs permit collection of big data

1103   within registry or cohort platforms and allow multiple synchronous randomised

1104   trials to be conducted in a cost-effective manner (Young-Afat et al., 2017).

1105

1106   In the radiotherapy review, I was involved in the conception, design,

1107   PROSPERO registration, design of search strategy, database searching, data

1108   extraction, performing all the analysis, interpretation of the data, drafting and

1109   critical review of the manuscript.

1110

1111

1112

1113

1114

1115

5. The impact of mobile technology on teamwork and communication in surgical and medical settings

A PRISMA-compliant systematic review was conducted to evaluate the impact of mobile technology on teamwork and communication in surgical and medical settings (Martin et al., 2019). I was involved in the conception, design, development of search strategy, screening of articles, data extraction, analysis and drafting. The review highlights the potential benefits of mobile technology, which is ubiquitous among healthcare professionals. However, the paucity of high-quality evidence for its effectiveness and other common barriers limit widespread uptake.

The aim of the review was to assess the quality and breadth of evidence for the impact of mobile technologies on communication and teamwork within surgical and medical hospital settings. The systematic review methodology was deemed most appropriate to carry out a robust assessment of the evidence-base on which to base recommendations and identify areas for future research. The review was prospectively registered on PROSPERO (CRD42017064128) and conducted in accordance with the PRISMA Statement.

A robust search strategy was developed and comprehensive search was undertaken of MEDLINE, PsycINFO, EMBASE, CINAHL Plus, HMIC, the Cochrane Library, and National Institute of Health Research Health Technology Assessment Database. Mobile technology was defined as hand-

held devices (mobiles, smartphones, tablets, or bespoke mobile devices) that facilitate 2-way communication or data transfer and which directly impact patient care. Screening of the articles was conducted independently be me (AK) and another author, and Cohen's kappa was calculated to determine inter-rater reliability. Data on the study design, population, intervention, comparators, setting and quantitative and qualitative outcomes were extracted. The outcomes included: 1) workflow, efficiency and quality of communication; 2) accessibility and inter-team relationships and 3) professionals' views of mobile technology. In addition to the National Institutes of Health Quality Assessment Tools, the mobile health (mHealth) evidence reporting and assessment (mERA) checklist was used to assess quality and risk of bias (Agarwal et al., 2016). The mERA is a reporting guideline for mHealth, similar to CONSORT (Schulz et al., 2010) and PRISMA (Liberati et al., 2009).

Out of 8072 articles screened, 38 were included in the final analysis. The data demonstrated that overall there is a lack of high-quality evidence evaluating the impact of mobile technologies on communication and teamwork in hospital settings. Fourteen studies reported quantitative outcomes, all but 2 using questionnaires, and 7 used content analysis of mobile phone data. Two studies used direct observational data. One assessed time taken to complete handover while the other assessed the speed and latency of communication. Two further studies reported qualitative outcomes, with one using semi-structured interviews and focus groups and the other using an exploratory case study approach. Finally, a mixed-methods approach was adopted by 6

studies, with all including content analysis of messages sent or received; 4 included additional structured interviews, 2 included questionnaires, and 2 included more direct observation. Meta-analysis was not performed due to heterogeneity in study designs and outcomes. Instead, a narrative synthesis was conducted. Broadly speaking, the studies reported that introduction of mobile devices led to enhanced workflow, efficiency, and communication quality. They also reported improvement in clinical handover, faster response times to clinical messages, and facilitation in easy delivery of non-urgent information while also supporting the triage, prioritization, and timeliness of communication. Moreover, mobile devices improved accessibility, inter-professional interactions, and senior decision maker involvement in clinical care. The technology was valued by healthcare staff for being more convenient and was preferred to existing modes of communication such as traditional pagers/bleeps. Nevertheless, few studies reported that doctors felt frequently interrupted by low-value and unnecessary information, often inappropriate given the content and context. Other issues identified with mobile devices included cost, lack of institutional integration and support, poor battery life, reliability, small screen size and potential risk to security and confidentiality of patient information.

The strengths of the review include the fact that this is the first systematic review in available literature to evaluate the quality and breadth of evidence on the impact of mobile technology on teamwork and communication in surgical and medical hospital settings. The protocol was prospectively registered, and the review was conducted in accordance with PRISMA

guidance. Robust methodology was implemented, with comprehensive database searching, independent screening of articles, assessment of methodological quality and reporting quality. Due to the heterogeneity in study designs and outcome measures, a meta-analysis was not performed, and instead a narrative synthesis was performed. The conclusion has been put in context, in light of the quality of the evidence.

The limitations of the review include the paucity of high quality evidence. The reporting of studies as measured by the mERA checklist was also suboptimal, with a mean score of 6.1 of 16 (range, 3-11), and no study was fully compliant. The poorest reported items were: Item 6 (usability/content testing); Item 11 (limitations for delivery at scale); Item 12 (contextual adaptability) and Item 13 (replicability) [detailed descriptions of the items are documented by (Agarwal et al., 2016)]. Most of the studies were single-centre studies and examined small populations in restricted environments that do not truly represent complex real-world settings, limiting generalizability. It is difficult to draw clear conclusions due to methodological inadequacies including the lack of prospective randomization or assessment of matched comparator groups, the limited number of participants and truncated study lengths, and, due to significant variability in methodologies and outcomes employed, an inability to effectively pool results from multiple studies. Twenty-six studies included questionnaire-based data collection, yet only 6 discussed validity testing of the questionnaires used. While some of these methodological flaws may be attributed to the inherent difficulty of assessing such interventions in complex surgical and medical hospital settings, few studies clearly set out to try and

1216 overcome these challenges in a meaningful way. Of the 22 interventional

1217 studies reviewed, only 2 had any form or randomization or prospective

1218 assessment of matched comparator groups, and in the remainder only 5

1219 made reference to pre-intervention baseline data against which the mobile

1220 intervention was compared.

1221

1222 In summary, an evidence-based approach to the development, deployment

1223 and evaluation of new mobile communication devices is needed. Future,

1224 prospective randomized studies are required with a priori defined outcomes to

1225 evaluate comparative effectiveness. Studies should have larger sample sizes

1226 to ensure they are adequately powered. In addition, for questionnaire-based

1227 data, tools used must be validated. Studies should be adequately reported in

1228 line with the mERA checklist, and journal editors and key stakeholders should

1229 consider incorporating reporting guidelines into their 'instructions to authors',

1230 i.e. making mERA checklist submission as a mandatory part of manuscript

1231 submission. Mandatory requirement to complete reporting checklists has been

1232 shown to enhance compliance in other areas of surgery (Agha et al., 2016).

1233

1234

1235

1236

1237

1238

1239

1240

6. Discussion

The author has a large body of published work in high impact journals in evidence synthesis, outcome measurement and reporting quality in surgery, with a focus in Plastic & Reconstructive Surgery, forming the basis of an application for a PhD by Publication at the University of Portsmouth (Khajuria and Agha, 2013, Khajuria and Ahmed Agha, 2015, Khajuria and Mosahebi, 2019, Khajuria et al., 2019, Khajuria et al., 2017a, Smith et al., 2018b, Winters and Khajuria, 2018a, Yao et al., 2014, Khajuria et al., 2018).

The author has also been supervising students, with resultant publications (Ishak et al., 2019, Reddy et al., 2020, Mantelakis and Khajuria, 2020) [Appendix 16-17]. The PhD award will further facilitate the author in establishing a research group and to undertake postdoctoral research. This will include conducting a national stream funded large prospective cohort study (IDEAL framework Stage 2b)/RCT (IDEAL framework Stage 3) (McCulloch et al., 2009) in evaluating immediate versus delayed autologous breast reconstruction, in context of radiotherapy, with robust reporting of breast reconstruction core outcome set (Potter et al., 2015) and incorporation and measurement of disease-specific PROs and cost-effectiveness. The first step will include conducting a national clinician survey to establish if clinical equipoise (McCulloch et al., 2002) truly exists and a national patient survey to understand patient preferences and willingness to be recruited and/or randomised. The next step would be to set up a multicentre prospective cohort study (IDEAL framework stage 2B) or RCT (IDEAL framework stage 3) with follow-up of at least 1 year. Steering committee will consist of plastic and

1266 breast surgeons, patient advocates and oncologists, in collaboration with the

1267 Royal College of Surgeons (RCS)- affiliated UK Reconstructive Surgery Trials

1268 Network (RSTN) and the Association of Breast Surgery (ABS)-affiliated iBRA

1269 Net. Core outcome sets, with complications as per Clavien Dindo

1270 classification (Dindo et al., 2004) and PROs using BREAST-Q and EORTC

1271 QLQ-BRECON23 will be reported. This cohort may allow a novel RCT trial

1272 design called Trials within Cohorts (TWiCs) (Young-Afat et al., 2017) to be

1273 established, facilitating recruitment, with a priori quality of life domains

1274 (Winters and Khajuria, 2018). Moreover, ICERs (Matros et al., 2015) will be

1275 calculated, to determine a breast QALY, i.e. the increased cost for obtaining 1

1276 year of perfect breast health related quality of life between two groups.

1277

1278 The collaborative model, employed by bodies such as RSTN and the iBRA

1279 Net, has facilitated rapid, multicentre data collection with resultant high impact

1280 publications (Potter et al., 2019). Surgical research in the future will likely rely

1281 on these models to cultivate collaboration. This will encourage multicentre

1282 studies that will enhance power of the studies as well as their generalizability,

1283 whilst ensuring the rigorous evaluation of novel surgical techniques, products

1284 and implants within the field of oncoplastic breast surgery to optimise the

1285 interval validity.

1286

1287

1288

1289

1290

6.1 Strengths of this work

The author has addressed fundamental questions in the overarching theme of evidence-based healthcare as well as focussing on plastic and reconstructive breast surgery, with key research questions. All original articles published by the author are the first in available literature in their respective themes. The work was planned meticulously by the author as well as through collaborations with a multidisciplinary group of plastic surgeons, breast surgeons, clinical oncologists, methodologists who have experience in systematic reviews and reporting quality (Khajuria and Agha, 2013, Khajuria and Ahmed Agha, 2015, Khajuria et al., 2017b, Khajuria et al., 2017a, Winters and Khajuria, 2018a, Winters et al., 2010, McCulloch et al., 2009, Smith et al., 2018b, Martin et al., 2019, Mantelakis and Khajuria, 2020, Venkatesh et al., 2020). The review protocols were registered a priori on PROSPERO with pre-defined patient population, intervention, comparator and outcomes. The reviews were PRISMA-compliant and Cochrane methodology was followed. For the radiotherapy review, complications were graded retrospectively by two independent authors, as per the Clavien-Dindo classification. A random effects model was utilised, along with robust methods to assess the quality and risk of bias.

6.2 Limitations

The studies are limited by the quantity and quality of the available evidence. No Level-I evidence was included in the results of the breast reconstruction

51

studies, which were based on primarily low-quality studies. No reporting

guidelines were used for observational studies (STROBE checklist) (von Elm

et al., 2007) or for case series (PROCESS statement) (Agha et al., 2018). For

observational studies, key elements that were poorly reported included: the

eligibility criteria; the sources and methods of selection of participants; clearly

defined outcomes; participant characteristics; potential confounders; effect

modifiers; explanation of how sample size was calculated; statistical methods

used to adjust for confounding; explanation on how missing data and any loss

to follow-up was addressed; and comment of limitations and external validity.

It was particularly challenging to retrospectively grade surgical interventions

using CDC in the radiotherapy meta-analysis; in three studies, reported

complications were not amenable to retrospective grading, compromising the

interpretability and reliability of the results. Only 30.3% of all surgery

complications reported (30/99) across the 12 included studies in the

radiotherapy review were defined a priori, with potential for selective outcome

reporting. Panel assessments of cosmetic outcome also potentiate risk of

observer bias.

In the DIEP versus IBR review (Khajuria et al., 2019), there is imprecision with

wide confidence intervals and high heterogeneity, when evaluating length of

stay. Whilst the published paper states there is no effect, this should be

interpreted with caution in light of the wide confidence intervals and the

imprecision of the estimate, which may preclude any meaningful interpretation

of the summary effect, apart from highlighting the need for further research

with higher quality studies, larger sample sizes and low heterogeneity and

1341 lower variability in outcome measurement. The same conclusion may be

1342 drawn for the number of reoperations for complications. Whilst the paper

1343 states a lower mean number of reoperations for complications for DIEP

1344 reconstruction, the estimate is imprecise, with wide confidence intervals,

1345 necessitating further research with higher quality studies.

1346

1347 The radiotherapy review (Khajuria et al., 2020) further highlights the issues

1348 related to interpreting summary effect measures, based on imprecision of the

1349 individual study estimates, and wide confidence intervals. In addition, none of

1350 the studies considered were RCTs, so the evidence relates to association,

1351 rather than causality. Whilst the forest plots comparing adjuvant RT versus no

1352 RT and well as neoadjuvant RT versus no RT, state no differences in overall

1353 complications, examining combined RT versus no RT showed greater overall

1354 complications in RT group. The greater overall sample size gave greater

1355 precision and narrower confidence intervals; but the evidence remained very

1356 weak.

1357

1358 The best available evidence was used (RCTs if available, or prospective

1359 cohort studies). When these were not available, poorer quality evidence was

1360 used, with suitable warnings. It is well established that conducting a meta-

1361 analysis does not overcome limitations in the design and execution of the

1362 primary studies. Combining studies of poor quality with those more rigorously

1363 conducted may yield a false sense of precision of the true effect and in some

1364 cases may be misleading. Indeed concerns have been raised regarding

1365 interpreting meta-analyses in Plastic Surgery, given majority of included

1366   primary studies are of poor quality, with heterogeneity within and between

1367   primary studies (McGuire et al., 2019). For the DIEP versus IBR review

1368   (Khajuria et al., 2019), in the protocol, the use of sensitivity analysis was

1369   purported, with exclusion of poor quality studies to determine the impact on

1370   the effect summary. However, due to the paucity of studies (e.g. only two

1371   studies reported comparative data on length of stay and number of

1372   reoperations for complications), this approach was not possible. It may be

1373   argued that given the heterogeneity and poor quality of the studies, a

1374   narrative synthesis should have been performed for all outcomes, not just for

1375   PROs and cost, and that a meta-analysis should not have been performed.

1376

1377   For the radiotherapy review, there were 4 moderate quality studies, with the

1378   remaining studies of poor quality. Out of the 4 moderate quality studies, 3 did

1379   not have a control group, so were not included in the meta-analysis. Similarly,

1380   it may be argued that a narrative synthesis, as opposed to meta-analysis,

1381   would be the more suitable, in light of the poor quality and serious risk of bias,

1382   for studies included in the review. Meta-analysis using published means and

1383   percentages does not permit the adjustments that are possible with individual

1384   patient data. There is no way of adjusting the results. In the absence of

1385   randomised controlled trials, only associations can be shown; not cause and

1386   effect. Nevertheless, it is important to provide the best available evidence

1387   even when that evidence is very poor. The absence of high quality of

1388   evidence can be used for directing future research."

1389

1390 Generic PRO tools used in the studies in the reviews may not be sensitive

1391 enough to pick up clinically meaningful differences. There is also no evidence

1392 contrasting the psychometric robustness of the disease-specific BREAST-Q

1393 versus EORTC QLQ-BRECON23 questionnaires. As previously purported by

1394 the author in Lancet Oncology, further validation work of the BREAST-Q

1395 scales may be required, since in the large, multi-centre BRIOS trial evaluating

1396 one-stage ADM-implant versus 2-stage BRR found no correlation between

1397 complications and PRO scores, which was somewhat surprising (Winters and

1398 Khajuria, 2018a). Finally, the external validity of the reviews may be

1399 compromised, as majority of the included studies are single-centre studies,

1400 primarily from middle-high income countries, implicating the results may not

1401 be generalizable to practice and outcomes in low-income countries.

1402

1403 **6.3 Future work**

1404

1405 After the PhD award, the author will build on the work in this report as a post-

1406 doctoral research fellow, with the aim to address the aforementioned

1407 limitations and set up and lead a multi-centre IDEAL-2b/ IDEAL 3 study

1408 (McCulloch et al., 2009) to establish the optimal sequence of RT in

1409 abdominal-based autologous breast reconstruction. The IDEAL framework

1410 describes an evidence-based and step-wise approach towards conducting a

1411 RCT. Specific 2b features (prospective cohort) will include establishment of

1412 prospective databases, relevant outcome measures and learning curve

1413 evaluation using the cumulative sum (CUSUM) method (Maguire et al., 2013).

1414 First stage will involve conducting a national clinician survey to establish if

clinical equipoise (McCulloch et al., 2002) truly exists and to ascertain the feasibility and willingness of surgeons to recruit. Patient and public (PPI) engagement will help to understand patient preferences and willingness to be recruited and/or randomized. Core outcome sets, with a priori defined complications as per CDC (Dindo et al., 2004) and PROs using BREAST-Q and EORTC QLQ-BRECON23 will be reported. Moreover, Incremental Cost Effectiveness Ratios (ICERs) (Matros et al., 2015) will be calculated, to determine a breast Quality Adjusted Life Year (QALY), i.e. the increased cost for obtaining 1 year of perfect breast health-related quality of life between immediate and delayed groups. The results will be reported using established tools e.g. STROBE/CONSORT, with focus on the individual components both at design stage and at reporting stage. To improve the systematic reviews, I would ensure that if there were significant heterogeneity, wide confidence intervals and moderate-serious risk of bias, that a narrative synthesis is performed as opposed to a meta-analysis. The RoB issues, as previously discussed, could be addressed in future observational studies. Studies should use the best possible methods of adjusting for confounding, for example individual patient data meta-analysis and propensity score matching (Austin, 2011). Confounding can be controlled at the design stage, by restriction (i.e. using inclusion and exclusion criteria) or matching (where confounders are allocated equally in the different arms of the study); at the analysis stage, standardisation may be employed, where confounders such as BMI can be adjusted for or via statistical adjustment using multivariate regression. Methods to quantify selection bias have been proposed, including the 'relative odds ratio', which is the ratio of the odds among participants to the

corresponding estimate in the source population (Nohr and Liew, 2018). If both populations have the same ORs, ROR is 1, indicating no bias; a ROR >1 indicates overestimation and a ROR <1 indicates underestimation. Prospective, longitudinal studies with sufficient follow-ups are needed. Protocols should be published and outcomes should be defined a priori. Outcome assessment should be blinded. Outcomes should be adequately reported using classifications such as CDC and the core outcome set.

With regards to neoadjuvant RT, the Royal Marsden and Imperial College Healthcare NHS Trusts are conducting the Primary Radiotherapy And DIEP flAp Reconstruction Trial (PRADA) trial, investigating the feasibility of using neoadjuvant radiotherapy prior to mastectomy and reconstruction (Trust, 2016). The investigators set out to formally evaluate the safety of reversing the order of mastectomy plus immediate DIEP flap reconstruction and adjuvant radiotherapy in a phase II study, with a view to a subsequent randomised controlled trial testing local control, complication rates and quality of life outcomes, including patient satisfaction (BREAST-Q reconstruction module) as well as volume and symmetry using 3D-surface imaging. The trial results will further facilitate the informed consent process and contribute to development of national, evidence-based guidance in breast reconstruction. Other ongoing studies include the DBCG RT Recon Trial, a multicentre RCT, evaluating delayed-immediate versus delayed breast reconstruction, with estimated completion date in November 2023 (DBCG, 2018). Delayed-immediate reconstruction includes reconstruction with silicone implant or expander covered by pectoral muscle and mesh or matrix. Conversely, Delayed reconstruction

1465 includes autologous or implant-based (one- or two-stage, +/- acellular dermal

1466 matrix (ADM)) and is performed 6-12 months after completion of chemotherapy

1467 and PMRT. Whilst immediate reconstruction may be the preferred choice for

1468 many surgeons and patients, the MROC cohort demonstrated no difference in

1469 patient satisfaction or in psychological, sexual, or physical well-being. Thus, the

1470 DBCG RT Recon Trial may also provide valuable insights into the optimal timing

1471 of breast reconstruction and RT, especially in the context of UK breast

1472 reconstruction where majority of the breast reconstructions performed in the

1473 UKNFR were delayed.

1474

1475 Data from registries (such as the UKNFR) and high-quality cohort studies

1476 (e.g. MROC) will identify trends in practice and form basis for identifying novel

1477 research questions and hypotheses. Systematic reviews and meta-analyses

1478 will also be important to determine overall effect estimates, on the pre-

1479 requisite that good quality studies and data are available to meta-analyse.

1480 With regards to funding for a cohort study, funding streams will include

1481 national bodies, such as BAPRAS, ABS and RCS. Several grants are

1482 available, and an additional benefit and feature of the collaborative model has

1483 been the need for limited funding to conduct the studies.

1484

1485

1486

1487

1488

1489

7. Conclusion

The thesis evaluates the reporting quality and outcome reporting in surgery, with a focus on plastic and reconstructive breast surgery. The DIEP vs. IBR systematic review provides a weak recommendation that DIEP reconstruction maybe more cost-effective and yield higher PRO scores, with the major limitation and caveat that the results are based on poor quality studies with serious risk of bias. The results however, do corroborate the results from the MROC cohort, the most robust cohort in the BRR literature, which was excluded from the systematic review, as it did not fulfil the inclusion criteria. The key finding is the presence of significant heterogeneity in outcome measurement, suboptimal reporting of core outcome set and no grading of complications, with no inference that can be derived about clinical management in the breast reconstruction literature. The review has identified that there is insufficient high quality evidence and we need future studies to be more robust and of higher quality. It has provided key insights on areas to focus and improve in future studies to enhance the evidence-base of breast reconstruction.

The radiotherapy review identified no statistically significant difference in outcomes between immediate versus delayed breast reconstruction in context of radiotherapy, based, once again, on poor quality evidence. The result contributes to the knowledge and challenges the current dogma in the UK where majority of the patients who need radiotherapy, undergo delayed reconstruction, as opposed to immediate reconstruction (as per the UK

1515   National Flap Registry Data). The review has also demonstrated the paucity

1516   of high quality evidence and the need for future high quality studies.

1517

1518   Majority of the individual studies are non-randomised, with small sizes,

1519   inadequate follow-ups, single-centre and retrospective. Reporting guidelines

1520   are available but not being adhered to. CONSORT and SPIRIT PRO

1521   guidelines exist for trial reporting and protocol reporting respectively, but

1522   majority of the studies are observational studies and currently no STROBE

1523   PRO extension exists. These insights from the reviews highlight the need for

1524   future multicentre, prospective studies with sufficient follow-ups, with

1525   adequate outcome measurement and reporting.

1526

1527   There is a perennial need to enhance reporting quality of clinical

1528   complications as per core outcome set, using robust tools/classification

1529   systems such as CDC. Complications reported as percentages without

1530   standardised reporting using classifications such as CDC and complication-

1531   specific classifications, e.g. Baker's classification (for capsular contracture),

1532   should be avoided, as there is no stratification according to management

1533   (nonoperative/conservative or operative management) (Khajuria and

1534   Mosahebi, 2019). Mandatory inclusion of adequate reporting in papers

1535   enforced by journals will facilitate improved reporting, evidence

1536   synthesis/meta-analysis and enhanced interpretability of research findings.

1537

1538

1539

8. References

Aaronson, N. K., Ahmedzai, S., Bergman, B., et al. 1993. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst, 85, 365-76.

Agarwal, S., Lefevre, A. E., Lee, J., et al. 2016. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. Bmj, 352, i1174.

Agha, R. A., Borrelli, M. R., Farwana, R., et al. 2018. The PROCESS 2018 statement: Updating Consensus Preferred Reporting Of CasE Series in Surgery (PROCESS) guidelines. Int J Surg, 60, 279-282.

Agha, R. A., Fowler, A. J., Herlin, C., Goodacre, T. E. & Orgill, D. P. 2015. Use of autologous fat grafting for breast reconstruction: a systematic review with meta-analysis of oncological outcomes. J Plast Reconstr Aesthet Surg, 68, 143-61.

Agha, R. A., Fowler, A. J., Limb, C., et al. 2016. Impact of the mandatory implementation of reporting guidelines on reporting quality in a surgical journal: A before and after study. Int J Surg, 30, 169-72.

Allen, R. J. & Treece, P. 1994. Deep inferior epigastric perforator flap for breast reconstruction. Ann Plast Surg, 32, 32-8.

Atherton, D. D., Hills, A. J., Moradi, P., Muirhead, N. & Wood, S. H. 2011. The economic viability of breast reconstruction in the UK: comparison of a single surgeon's experience of implant; LD; TRAM and DIEP based reconstructions in 274 patients. J Plast Reconstr Aesthet Surg, 64, 710-5.

Atkins, D., Best, D., Briss, P. A., et al. 2004. Grading quality of evidence and strength of recommendations. Bmj, 328, 1490.

Austin, P. C. 2011. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res, 46, 399-424.

Barry, M. & Kell, M. R. 2011. Radiotherapy and breast reconstruction: a meta-analysis. Breast Cancer Res Treat, 127, 15-22.

Bennett, K. G., Qi, J., Kim, H. M., et al. 2018. Comparison of 2-Year Complication Rates Among Common Techniques for Postmastectomy Breast Reconstruction. JAMA Surg, 153, 901-908.

Billig, J., Jagsi, R., Qi, J., et al. 2017. Should Immediate Autologous Breast Reconstruction Be Considered in Women Who Require Postmastectomy Radiation Therapy? A Prospective Analysis of Outcomes. Plast Reconstr Surg, 139, 1279-1288.

Borenstein, M., Hedges, L. V., Higgins, J. P. & Rothstein, H. R. 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods, 1, 97-111.

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F. & Ravaud, P. 2008. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med, 148, 295-309.

Calvert, M., Blazeby, J., Altman, D. G., et al. 2013. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. Jama, 309, 814-22.

Calvert, M., Kyte, D., Mercieca-Bebber, R., et al. 2018. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. Jama, 319, 483-494.

Camm, C. F., Agha, R. A. & Edison, E. 2015. CONSORT adherence in journals is still far from perfect. Ann Surg, 261, e38.

Cano, S. J., Klassen, A. & Pusic, A. L. 2009. The science behind quality-of-life measurement: a primer for plastic surgeons. Plast Reconstr Surg, 123, 98e-106e.

Cano, S. J., Klassen, A. F., Scott, A., Alderman, A. & Pusic, A. L. 2014. Interpreting clinical differences in BREAST-Q scores: minimal important difference. Plast Reconstr Surg, 134, 173e-175e.

Cano, S. J., Klassen, A. F., Scott, A. M., Cordeiro, P. G. & Pusic, A. L. 2012. The BREAST-Q: further validation in independent clinical samples. Plast Reconstr Surg, 129, 293-302.

Chatterjee, J. S., Lee, A., Anderson, W., et al. 2009. Effect of postoperative radiotherapy on autologous deep inferior epigastric perforator flap volume after immediate breast reconstruction. Br J Surg, 96, 1135-40.

Chen, C. M., Cano, S. J., Klassen, A. F., et al. 2010. Measuring quality of life in oncologic breast surgery: a systematic review of patient-reported outcome measures. Breast J, 16, 587-97.

Chovan, J. 2019. Repair of Post-TRAM Bulge/Hernia [Online]. Available: https://www.medillsb.com/illustration_image_details.aspx?AID=9217&IID=186633 [Accessed 19.07.19 2019].

Cocks, K., King, M. T., Velikova, G., et al. 2012. Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. Eur J Cancer, 48, 1713-21.

Coens, C., Pe, M., Dueck, A. C., et al. 2020. International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. Lancet Oncol, 21, e83-e96.

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences. , Hillsdale, NJ Erlbaum.

Cohen, W. A., Mundy, L. R., Ballard, T. N., et al. 2016. The BREAST-Q in surgical research: A review of the literature 2009-2015. J Plast Reconstr Aesthet Surg, 69, 149-62.

Cooke, A. L., Diaz-Abele, J., Hayakawa, T., et al. 2017. Radiation Therapy Versus No Radiation Therapy to the Neo-breast Following Skin-Sparing Mastectomy and Immediate Autologous Free Flap Reconstruction for Breast Cancer: Patient-Reported and Surgical Outcomes at 1 Year-A Mastectomy Reconstruction Outcomes Consortium (MROC) Substudy. Int J Radiat Oncol Biol Phys, 99, 165-172.

Cordeiro, P. G., Albornoz, C. R., Mccormick, B., Hu, Q. & Van Zee, K. 2014. The impact of postmastectomy radiotherapy on two-stage implant breast reconstruction: an analysis of long-term surgical outcomes, aesthetic results, and satisfaction over 13 years. Plast Reconstr Surg, 134, 588-95.

Cserni, G., Chmielik, E., Cserni, B. & Tot, T. 2018. The new TNM-based staging of breast cancer. Virchows Arch, 472, 697-703.

1636 Cutress, R. I., Mcintosh, S. A., Potter, S., et al. 2018. Opportunities and priorities
1637　　　　for breast surgical research. Lancet Oncol, 19, e521-e533.
1638 Da Costa, B. R., Rutjes, A. W., Johnston, B. C., et al. 2012. Methods to convert
1639　　　　continuous outcomes into odds ratios of treatment response and numbers
1640　　　　needed to treat: meta-epidemiological study. Int J Epidemiol, 41, 1445-59.
1641 Davies, G., Mills, N., Holcombe, C. & Potter, S. 2020. Perceived barriers to
1642　　　　randomised controlled trials in breast reconstruction: obstacle to trial
1643　　　　initiation or opportunity to resolve? A qualitative study. Trials, 21, 316.
1644 Dbcg. 2018. Delayed-immediate Versus Delayed Breast Reconstruction in Breast
1645　　　　Cancer Patients With Mastectomy and Radiation Therapy [Online].
1646　　　　Available: https://clinicaltrials.gov/ct2/show/NCT03730922 [Accessed
1647　　　　08.08.2020].
1648 Dean, C., Chetty, U. & Forrest, A. P. 1983. Effects of immediate breast
1649　　　　reconstruction on psychosocial morbidity after mastectomy. Lancet, 1,
1650　　　　459-62.
1651 Delong, M. R., Tandon, V. J., Farajzadeh, M., et al. 2019. Systematic Review of the
1652　　　　Impact of Acellular Dermal Matrix on Aesthetics and Patient Satisfaction
1653　　　　in Tissue Expander-to-Implant Breast Reconstructions. Plast Reconstr
1654　　　　Surg, 144, 967e-974e.
1655 Dersimonian, R. & Laird, N. 2015. Meta-analysis in clinical trials revisited.
1656　　　　Contemp Clin Trials, 45, 139-45.
1657 Dikmans, R. E. G., Van De Grift, T. C., Bouman, M. B., Pusic, A. L. & Mullender, M. G.
1658　　　　2019. Sexuality, a topic that surgeons should discuss with women before
1659　　　　risk-reducing mastectomy and breast reconstruction. Breast, 43, 120-122.
1660 Dindo, D., Demartines, N. & Clavien, P. A. 2004. Classification of surgical
1661　　　　complications: a new proposal with evaluation in a cohort of 6336
1662　　　　patients and results of a survey. Ann Surg, 240, 205-13.
1663 Donker, M., Van Tienhoven, G., Straver, M. E., et al. 2014. Radiotherapy or surgery
1664　　　　of the axilla after a positive sentinel node in breast cancer (EORTC 10981-
1665　　　　22023 AMAROS): a randomised, multicentre, open-label, phase 3 non-
1666　　　　inferiority trial. Lancet Oncol, 15, 1303-10.
1667 Efficace, F., Fayers, P., Pusic, A., et al. 2015. Quality of patient-reported outcome
1668　　　　reporting across cancer randomized controlled trials according to the
1669　　　　CONSORT patient-reported outcome extension: A pooled analysis of 557
1670　　　　trials. Cancer, 121, 3335-42.
1671 Eltahir, Y., Werners, L. L., Dreise, M. M., et al. 2013. Quality-of-life outcomes
1672　　　　between mastectomy alone and breast reconstruction: comparison of
1673　　　　patient-reported BREAST-Q and other health-related quality-of-life
1674　　　　measures. Plast Reconstr Surg, 132, 201e-209e.
1675 Erdmann-Sager, J., Wilkins, E. G., Pusic, A. L., et al. 2018. Complications and
1676　　　　Patient-Reported Outcomes after Abdominally Based Breast
1677　　　　Reconstruction: Results of the Mastectomy Reconstruction Outcomes
1678　　　　Consortium Study. Plast Reconstr Surg, 141, 271-281.
1679 Euroqol. 2019. EQ-5D-5L User Guide [Online]. Available:
1680　　　　https://euroqol.org/wp-content/uploads/2019/09/EQ-5D-5L-English-
1681　　　　User-Guide_version-3.0-Sept-2019-secured.pdf [Accessed 10.08.2020].
1682 Euroqol. 2020. EQ-5D [Online]. Available: https://euroqol.org/eq-5d-
1683　　　　instruments/eq-5d-5l-about/ [Accessed 10.08.2020].

1684 Faraone, S. V. 2008. Interpreting estimates of treatment effects: implications for
1685   managed care. P t, 33, 700-11.
1686 Flynn, T. N., Huynh, E., Peters, T. J., et al. 2015. Scoring the Icecap-a capability
1687   instrument. Estimation of a UK general population tariff. Health Econ, 24,
1688   258-69.
1689 Ginsburg, O., Bray, F., Coleman, M. P., et al. 2017. The global burden of women's
1690   cancers: a grand challenge in global health. Lancet, 389, 847-860.
1691 Hallberg, H., Rafnsdottir, S., Selvaggi, G., et al. 2018. Benefits and risks with
1692   acellular dermal matrix (ADM) and mesh support in immediate breast
1693   reconstruction: a systematic review and meta-analysis. J Plast Surg Hand
1694   Surg, 52, 130-147.
1695 Higgins, J. P. & Thompson, S. G. 2002. Quantifying heterogeneity in a meta-
1696   analysis. Stat Med, 21, 1539-58.
1697 Ho, A. Y., Hu, Z. I., Mehrara, B. J. & Wilkins, E. G. 2017. Radiotherapy in the setting
1698   of breast reconstruction: types, techniques, and timing. Lancet Oncol, 18,
1699   e742-e753.
1700 Horton, R. 1996. Surgical research or comic opera: questions, but few answers.
1701   Lancet, 347, 984-5.
1702 Howell, A., Anderson, A. S., Clarke, R. B., et al. 2014. Risk determination and
1703   prevention of breast cancer. Breast Cancer Res, 16, 446.
1704 Ishak, A., Rajangam, A. & Khajuria, A. 2019. The evidence-base for the
1705   management of flexor tendon injuries of the hand: Review. Ann Med Surg
1706   (Lond), 48, 1-6.
1707 Jagsi, R., Momoh, A. O., Qi, J., et al. 2018. Impact of Radiotherapy on Complications
1708   and Patient-Reported Outcomes After Breast Reconstruction. J Natl
1709   Cancer Inst, 110.
1710 Kang, H. 2016. How to understand and conduct evidence-based medicine. Korean
1711   J Anesthesiol, 69, 435-445.
1712 Kelley, B. P., Ahmed, R., Kidwell, K. M., et al. 2014. A systematic review of
1713   morbidity associated with autologous breast reconstruction before and
1714   after exposure to radiotherapy: are current practices ideal? Ann Surg
1715   Oncol, 21, 1732-8.
1716 Khajuria, A. & Agha, R. A. 2013. Surgical clinical trials--need for quantity and
1717   quality. Lancet, 382, 1876.
1718 Khajuria, A. & Ahmed Agha, R. 2015. CONSORT compliance in surgical
1719   randomized trials: possible solutions. Ann Surg, 261, e135.
1720 Khajuria, A., Charles, W. N., Prokopenko, M., et al. 2020. Immediate and delayed
1721   autologous abdominal microvascular flap breast reconstruction in
1722   patients receiving adjuvant, neoadjuvant or no radiotherapy: a meta-
1723   analysis of clinical and quality-of-life outcomes. BJS Open, 4, 182-196.
1724 Khajuria, A. & Farhadi, J. 2020. Immediate versus delayed autologous breast
1725   reconstruction. J Plast Reconstr Aesthet Surg.
1726 Khajuria, A., Geoghegan, L., Solberg, Y., et al. 2018. Selective non-operative
1727   management for penetrating extremity trauma: A paradigm shift in
1728   management? J Plast Reconstr Aesthet Surg, 71, 1239-1244.
1729 Khajuria, A. & Mosahebi, A. 2019. Outcome reporting in breast reconstruction. J
1730   Plast Reconstr Aesthet Surg.

Khajuria, A., Prokopenko, M., Greenfield, M., et al. 2019. A Meta-analysis of Clinical, Patient-Reported Outcomes and Cost of DIEP versus Implant-based Breast Reconstruction. Plast Reconstr Surg Glob Open, 7**,** e2486.

Khajuria, A., Smith, O. J., Prokopenko, M., Greenfield, M. & Mosahebi, A. 2017a. Protocol for a systematic review and meta-analysis on the clinical outcomes and cost of deep inferior epigastric perforator (DIEP) flap versus implants for breast reconstruction. Syst Rev, 6**,** 232.

Khajuria, A., Winters, Z. & Mosahebi, A. 2017b. A systematic review and meta-analysis of clinical and patient-reported outcomes (PROs) of immediate versus delayed autologous abdominal-based flap breast reconstruction in the context of post-mastectomy radiotherapy [PROSPERO CRD42017077945] [Online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD4201 7077945 [Accessed 12.02.2020].

King, M. C., Marks, J. H. & Mandell, J. B. 2003. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science, 302**,** 643-6.

King, M. T. 1996. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. Qual Life Res, 5**,** 555-67.

King, M. T., Costa, D. S., Aaronson, N. K., et al. 2016. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. Qual Life Res, 25**,** 625-36.

Klassen, A. F., Pusic, A. L., Scott, A., Klok, J. & Cano, S. J. 2009. Satisfaction and quality of life in women who undergo breast surgery: a qualitative study. BMC Womens Health, 9**,** 11.

Koshima, I. & Soeda, S. 1989. Inferior epigastric artery skin flaps without rectus abdominis muscle. Br J Plast Surg, 42**,** 645-8.

Kronowitz, S. J., Hunt, K. K., Kuerer, H. M., et al. 2004. Delayed-immediate breast reconstruction. Plast Reconstr Surg, 113**,** 1617-28.

Kunkler, I. H., Dixon, J. M., Maclennan, M. & Russell, N. S. 2017. European interpretation of North American post mastectomy radiotherapy guideline update. Eur J Surg Oncol, 43**,** 1805-1807.

Lagares-Borrego, A., Gacto-Sanchez, P., Infante-Cossio, P., et al. 2016. A comparison of long-term cost and clinical outcomes between the two-stage sequence expander/prosthesis and autologous deep inferior epigastric flap methods for breast reconstruction in a public hospital. J Plast Reconstr Aesthet Surg, 69**,** 196-205.

Laupacis, A., Feeny, D., Detsky, A. S. & Tugwell, P. X. 1992. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. Cmaj, 146**,** 473-81.

Liberati, A., Altman, D. G., Tetzlaff, J., et al. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Bmj, 339**,** b2700.

Lie, K. H., Barker, A. S. & Ashton, M. W. 2013. A classification system for partial and complete DIEP flap necrosis based on a review of 17,096 DIEP flaps in 693 articles including analysis of 152 total flap failures. Plast Reconstr Surg, 132**,** 1401-8.

Liu, Z., Yao, Z., Li, C., et al. 2013. A step-by-step guide to the systematic review and meta-analysis of diagnostic and prognostic test accuracy evaluations. Br J Cancer, 108**,** 2299-303.

1780   Macadam, S. A., Zhong, T., Weichman, K., et al. 2016. Quality of Life and Patient-
1781   Reported Outcomes in Breast Cancer Survivors: A Multicenter
1782   Comparison of Four Abdominally Based Autologous Reconstruction
1783   Methods. Plast Reconstr Surg, 137**,** 758-71.
1784   Macdonald, S. M., Harris, E. E., Arthur, D. W., et al. 2011. ACR appropriateness
1785   criteria(R) locally advanced breast cancer. Breast J, 17**,** 579-85.
1786   Maguire, T., Mayne, C. J., Terry, T. & Tincello, D. G. 2013. Analysis of the surgical
1787   learning curve using the cumulative sum (CUSUM) method. Neurourol
1788   Urodyn, 32**,** 964-7.
1789   Mantelakis, A. & Khajuria, A. 2020. The applications of machine learning in
1790   plastic and reconstructive surgery: protocol of a systematic review. Syst
1791   Rev, 9**,** 44.
1792   Marks, L. B., Kaidar-Person, O. & Poortmans, P. 2017. Regarding Current
1793   Recommendations for Postmastectomy Radiation Therapy in Patients
1794   With One to Three Positive Axillary Lymph Nodes. J Clin Oncol, 35**,** 1256-
1795   1258.
1796   Martin, G., Khajuria, A., Arora, S., et al. 2019. The impact of mobile technology on
1797   teamwork and communication in hospitals: a systematic review. J Am
1798   Med Inform Assoc, 26**,** 339-355.
1799   Matros, E., Albornoz, C. R., Razdan, S. N., et al. 2015. Cost-effectiveness analysis of
1800   implants versus autologous perforator flaps using the BREAST-Q. Plast
1801   Reconstr Surg, 135**,** 937-46.
1802   Mcculloch, P., Altman, D. G., Campbell, W. B., et al. 2009. No surgical innovation
1803   without evaluation: the IDEAL recommendations. Lancet, 374**,** 1105-12.
1804   Mcculloch, P., Taylor, I., Sasako, M., Lovett, B. & Griffin, D. 2002. Randomised
1805   trials in surgery: problems and possible solutions. Bmj, 324**,** 1448-51.
1806   Mcgale, P., Taylor, C., Correa, C., et al. 2014. Effect of radiotherapy after
1807   mastectomy and axillary surgery on 10-year recurrence and 20-year
1808   breast cancer mortality: meta-analysis of individual patient data for 8135
1809   women in 22 randomised trials. Lancet, 383**,** 2127-35.
1810   Mcguire, C., Samargandi, O. A., Corkum, J., Retrouvey, H. & Bezuhly, M. 2019.
1811   Meta-Analyses in Plastic Surgery: Can We Trust Their Results? Plast
1812   Reconstr Surg, 144**,** 519-530.
1813   Modarressi, A., Muller, C. T., Montet, X., Ruegg, E. M. & Pittet-Cuenod, B. 2017.
1814   DIEP flap for breast reconstruction: Is abdominal fat thickness associated
1815   with post-operative complications? J Plast Reconstr Aesthet Surg, 70**,**
1816   1068-1075.
1817   Moher, D., Schulz, K. F., Simera, I. & Altman, D. G. 2010. Guidance for developers
1818   of health research reporting guidelines. PLoS Med, 7**,** e1000217.
1819   Momoh, A. O., Colakoglu, S., De Blacam, C., et al. 2012. Delayed autologous breast
1820   reconstruction after postmastectomy radiation therapy: is there an
1821   optimal time? Ann Plast Surg, 69**,** 14-8.
1822   Musoro, J. Z., Coens, C., Fiteni, F., et al. 2019. Minimally Important Differences for
1823   Interpreting EORTC QLQ-C30 Scores in Patients With Advanced Breast
1824   Cancer. JNCI Cancer Spectr, 3**,** pkz037.
1825   Negenborn, V. L., Young-Afat, D. A., Dikmans, R. E. G., et al. 2018. Quality of life
1826   and patient satisfaction after one-stage implant-based breast
1827   reconstruction with an acellular dermal matrix versus two-stage breast

reconstruction (BRIOS): primary outcome of a randomised, controlled trial. Lancet Oncol, 19, 1205-1214.

Nohr, E. A. & Liew, Z. 2018. How to investigate and adjust for selection bias in cohort studies. Acta Obstet Gynecol Scand, 97, 407-416.

Norman, R., Mercieca-Bebber, R., Rowen, D., et al. 2019. U.K. utility weights for the EORTC QLU-C10D. Health Econ, 28, 1385-1401.

O'connell, R. L., Di Micco, R., Khabra, K., et al. 2018a. Comparison of immediate versus delayed deep inferior epigastric artery perforator (DIEP) flap reconstruction in women who require post mastectomy radiotherapy. Plast Reconstr Surg.

O'connell, R. L., Khabra, K., Bamber, J. C., et al. 2018b. Validation of the Vectra XT three-dimensional imaging system for measuring breast volume and symmetry following oncological reconstruction. Breast Cancer Res Treat, 171, 391-398.

O'halloran, N., Lowery, A., Kalinina, O., et al. 2017. Trends in breast reconstruction practices in a specialized breast tertiary referral centre. BJS Open, 1, 148-157.

O'halloran, N., Potter, S., Kerin, M. & Lowery, A. 2018. Recent Advances and Future Directions in Postmastectomy Breast Reconstruction. Clin Breast Cancer, 18, e571-e585.

Paramanandam, V. S. & Roberts, D. 2014. Weight training is not harmful for women with breast cancer-related lymphoedema: a systematic review. J Physiother, 60, 136-43.

Potter, S., Brigic, A., Whiting, P. F., et al. 2011. Reporting clinical outcomes of breast reconstruction: a systematic review. J Natl Cancer Inst, 103, 31-46.

Potter, S., Conroy, E. J., Cutress, R. I., et al. 2019. Short-term safety outcomes of mastectomy and immediate implant-based breast reconstruction with and without mesh (iBRA): a multicentre, prospective cohort study. Lancet Oncol, 20, 254-266.

Potter, S., Conroy, E. J., Williamson, P. R., et al. 2016. The iBRA (implant breast reconstruction evaluation) study: protocol for a prospective multi-centre cohort study to inform the feasibility, design and conduct of a pragmatic randomised clinical trial comparing new techniques of implant-based breast reconstruction. Pilot Feasibility Stud, 2, 41.

Potter, S., Holcombe, C., Ward, J. A. & Blazeby, J. M. 2015. Development of a core outcome set for research and audit studies in reconstructive breast surgery. Br J Surg, 102, 1360-71.

Pu, Y., Mao, T. C., Zhang, Y. M., Wang, S. L. & Fan, D. L. 2018. The role of postmastectomy radiation therapy in patients with immediate prosthetic breast reconstruction: A meta-analysis. Medicine (Baltimore), 97, e9548.

Pusic, A. L., Klassen, A. F., Scott, A. M., et al. 2009. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. Plast Reconstr Surg, 124, 345-53.

Pusic, A. L., Klassen, A. F., Snell, L., et al. 2012. Measuring and managing patient expectations for breast reconstruction: impact on quality of life and patient satisfaction. Expert Rev Pharmacoecon Outcomes Res, 12, 149-58.

Recht, A., Comen, E. A., Fine, R. E., et al. 2017. Postmastectomy Radiotherapy: An American Society of Clinical Oncology, American Society for Radiation

1876   Oncology, and Society of Surgical Oncology Focused Guideline Update.
1877   Ann Surg Oncol, 24, 38-51.
1878 Reddy, R. K., Charles, W. N., Sklavounos, A., et al. 2020. The effect of smoking on
1879   COVID-19 severity: a systematic review and meta-analysis. J Med Virol.
1880 Ricci, J. A., Epstein, S., Momoh, A. O., et al. 2017. A meta-analysis of implant-based
1881   breast reconstruction and timing of adjuvant radiation therapy. J Surg
1882   Res, 218, 108-116.
1883 Rogers, N. E. & Allen, R. J. 2002. Radiation effects on breast reconstruction with
1884   the deep inferior epigastric perforator flap. Plast Reconstr Surg, 109,
1885   1919-24; discussion 1925-6.
1886 Russell, N. S., Kunkler, I. H. & Van Tienhoven, G. 2015. Determining the
1887   indications for post mastectomy radiotherapy: moving from 20th century
1888   clinical staging to 21st century biological criteria. Ann Oncol, 26, 1043-4.
1889 Ryan, M., Netten, A., Skåtun, D. & Smith, P. 2006. Using discrete choice
1890   experiments to estimate a preference-based measure of outcome--an
1891   application to social care for older people. J Health Econ, 25, 927-44.
1892 Salzberg, C. A. 2006. Nonexpansive immediate breast reconstruction using
1893   human acellular tissue matrix graft (AlloDerm). Ann Plast Surg, 57, 1-5.
1894 Santosa, K. B., Qi, J., Kim, H. M., et al. 2018. Long-term Patient-Reported Outcomes
1895   in Postmastectomy Breast Reconstruction. JAMA Surg.
1896 Schmidt, F. L., Oh, I. S. & Hayes, T. L. 2009. Fixed- versus random-effects models
1897   in meta-analysis: model properties and an empirical comparison of
1898   differences in results. Br J Math Stat Psychol, 62, 97-128.
1899 Schulz, K. F., Altman, D. G. & Moher, D. 2010. CONSORT 2010 Statement: updated
1900   guidelines for reporting parallel group randomised trials. Trials, 11, 32.
1901 Singletary, S. E. 2003. Rating the risk factors for breast cancer. Ann Surg, 237,
1902   474-82.
1903 Sisco, M., Du, H., Warner, J. P., et al. 2012. Have we expanded the equitable
1904   delivery of postmastectomy breast reconstruction in the new millennium?
1905   Evidence from the national cancer data base. J Am Coll Surg, 215, 658-66;
1906   discussion 666.
1907 Smith, J. M., Broyles, J. M., Guo, Y., et al. 2018a. Human acellular dermis increases
1908   surgical site infection and overall complication profile when compared
1909   with submuscular breast reconstruction: An updated meta-analysis
1910   incorporating new products(*). J Plast Reconstr Aesthet Surg, 71, 1547-
1911   1556.
1912 Smith, O. J., Kanapathy, M., Khajuria, A., et al. 2018b. Systematic review of the
1913   efficacy of fat grafting and platelet-rich plasma for wound healing. Int
1914   Wound J, 15, 519-526.
1915 Solomon, M. J., Laxamana, A., Devore, L. & Mcleod, R. S. 1994. Randomized
1916   controlled trials in surgery. Surgery, 115, 707-12.
1917 Spear, S. L. & Murphy, D. K. 2014. Natrelle round silicone breast implants: Core
1918   Study results at 10 years. Plast Reconstr Surg, 133, 1354-61.
1919 Sprangers, M. A., Groenvold, M., Arraras, J. I., et al. 1996. The European
1920   Organization for Research and Treatment of Cancer breast cancer-specific
1921   quality-of-life questionnaire module: first results from a three-country
1922   field study. J Clin Oncol, 14, 2756-68.
1923 Taghizadeh, R., Moustaki, M., Harris, S., Roblin, P. & Farhadi, J. 2015. Does post-
1924   mastectomy radiotherapy affect the outcome and prevalence of

complications in immediate DIEP breast reconstruction? A prospective cohort study. J Plast Reconstr Aesthet Surg, 68, 1379-85.

Tevis, S. E., James, T. A., Kuerer, H. M., et al. 2018. Patient-Reported Outcomes for Breast Cancer. Ann Surg Oncol, 25, 2839-2845.

Tonseth, K. A., Hokland, B. M., Tindholdt, T. T., Abyholm, F. E. & Stavem, K. 2008. Quality of life, patient satisfaction and cosmetic outcome after breast reconstruction using DIEP flap or expandable breast implant. J Plast Reconstr Aesthet Surg, 61, 1188-94.

Trust, R. M. N. F. 2016. Primary Radiotherapy And DIEP flAp Reconstruction Trial (PRADA) [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT02771938 [Accessed 08.08.2020].

Van Maaren, M. C., Le Cessie, S., Strobbe, L. J. A., et al. 2019. Different statistical techniques dealing with confounding in observational research: measuring the effect of breast-conserving therapy and mastectomy on survival. J Cancer Res Clin Oncol, 145, 1485-1493.

Velikova, G., Williams, L. J., Willis, S., et al. 2018. Quality of life after postmastectomy radiotherapy in patients with intermediate-risk breast cancer (SUPREMO): 2-year follow-up results of a randomised controlled trial. Lancet Oncol, 19, 1516-1529.

Venkatesh, A., Khajuria, A. & Greig, A. 2020. Management of Pediatric Distal Fingertip Injuries: A Systematic Literature Review. Plast Reconstr Surg Glob Open, 8, e2595.

Veronesi, U., Cascinelli, N., Mariani, L., et al. 2002. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med, 347, 1227-32.

Voineskos, S. H., Klassen, A. F., Cano, S. J., Pusic, A. L. & Gibbons, C. J. 2020. Giving Meaning to Differences in BREAST-Q Scores: Minimal Important Difference for Breast Reconstruction Patients. Plast Reconstr Surg, 145, 11e-20e.

Von Elm, E., Altman, D. G., Egger, M., et al. 2007. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Bmj, 335, 806-8.

Wilkins, E. G., Cederna, P. S., Lowery, J. C., et al. 2000. Prospective analysis of psychosocial outcomes in breast reconstruction: one-year postoperative results from the Michigan Breast Reconstruction Outcome Study. Plast Reconstr Surg, 106, 1014-25; discussion 1026-7.

Wilkins, E. G., Hamill, J. B., Kim, H. M., et al. 2018. Complications in Postmastectomy Breast Reconstruction: One-year Outcomes of the Mastectomy Reconstruction Outcomes Consortium (MROC) Study. Ann Surg, 267, 164-170.

Winters, S., Martin, C., Murphy, D. & Shokar, N. K. 2017. Breast Cancer Epidemiology, Prevention, and Screening. Prog Mol Biol Transl Sci, 151, 1-32.

Winters, Z. & Khajuria, A. 2018a. Quality of life after breast reconstruction - the BRIOS study. Lancet Oncology, 19, PE579.

Winters, Z. E., Afzal, M., Rutherford, C., et al. 2018. International validation of the European Organisation for Research and Treatment of Cancer QLQ-

BRECON23 quality-of-life questionnaire for women undergoing breast reconstruction. Br J Surg, 105, 209-222.

Winters, Z. E., Balta, V., Thomson, H. J., et al. 2014. Phase III development of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire module for women undergoing breast reconstruction. Br J Surg, 101, 371-82.

Winters, Z. E., Benson, J. R. & Pusic, A. L. 2010. A systematic review of the clinical evidence to guide treatment recommendations in breast reconstruction based on patient- reported outcome measures and health-related quality of life. Ann Surg, 252, 929-42.

Winters, Z. E., Emson, M., Griffin, C., et al. 2015. Learning from the QUEST multicentre feasibility randomization trials in breast reconstruction after mastectomy. Br J Surg, 102, 45-56.

Winters, Z. E. & Khajuria, A. 2018b. Quality of life after breast reconstruction-the BRIOS study. Lancet Oncol, 19, e579.

Winters, Z. E. & Thomson, H. J. 2011. Assessing the clinical effectiveness of breast reconstruction through patient-reported outcome measures. Br J Surg, 98, 323-5.

Yang, T. J. & Ho, A. Y. 2013. Radiation therapy in the management of breast cancer. Surg Clin North Am, 93, 455-71.

Yao, A. C., Khajuria, A., Camm, C. F., Edison, E. & Agha, R. 2014. The reporting quality of parallel randomised controlled trials in ophthalmic surgery in 2011: a systematic review. Eye (Lond), 28, 1341-9.

Yoon, A. P., Qi, J., Brown, D. L., et al. 2018. Outcomes of immediate versus delayed breast reconstruction: Results of a multicenter prospective study. Breast, 37, 72-79.

Young-Afat, D. A., Van Gils, C. H., Van Den Bongard, H. & Verkooijen, H. M. 2017. The Utrecht cohort for Multiple BREast cancer intervention studies and Long-term evaLuAtion (UMBRELLA): objectives, design, and baseline results. Breast Cancer Res Treat, 164, 445-450.

Zinzindohoue, C., Bertrand, P., Michel, A., et al. 2016. A Prospective Study on Skin-Sparing Mastectomy for Immediate Breast Reconstruction with Latissimus Dorsi Flap After Neoadjuvant Chemotherapy and Radiotherapy in Invasive Breast Carcinoma. Ann Surg Oncol, 23, 2350-6.

2019 **CHAPTER 8. APPENDICES**
2020
2021
2022

2023
2024
2025 Appendix 1: Figure 2. Adherence (%) of RCTs to individual items of the

2026 CONSORT extension for the NPT checklist. Overall, the mean adherence to

2027 any given item, including those subdivided, was 36.9%. Adherence ranged

2028 from 1 RCT (1.6%), in item 1, to 64 RCTs (98.4%), in item 2. (New...=New

2029 Item).

2030

2031 **The reporting quality of parallel randomised controlled trials in ophthalmic surgery in 2011: a systematic review.**

Yao AC[1], Khajuria A[1], Camm CF[2], Edison E[3], Agha R[4].

⊖ Author information

1   School of Medicine, Imperial College London, London, UK.
2   New College, University of Oxford, Oxford, UK.
3   University College London, London, UK.
4   Stoke Mandeville Hospital, Buckinghamshire, UK.

Abstract

PURPOSE: Randomised controlled trials (RCTs) represent a gold standard for evaluating therapeutic interventions. However, poor reporting clarity can prevent readers from assessing potential bias that can arise from a lack of methodological rigour. The Consolidated Standards of Reporting Trials statement for non-pharmacological interventions 2008 (CONSORT NPT) was developed to aid reporting. RCTs in ophthalmic surgery pose particular challenges in study design and implementation. We aim to provide the first assessment of the compliance of RCTs in ophthalmic surgery to the CONSORT NPT statement.

METHOD: In August 2012, the Medline database was searched for RCTs in ophthalmic surgery reported between 1 January 2011 and 31 December 2011. Results were searched by two authors and relevant papers selected. Papers were scored against the 23-item CONSORT NPT checklist and compared against surrogate markers of paper quality. The CONSORT score was also compared between different RCT designs.

RESULTS: In all, 186 papers were retrieved. Sixty-five RCTs, involving 5803 patients, met the inclusion criteria. The mean CONSORT score was 8.9 out of 23 (39%, range 3.0-14.7, SD 2.49). The least reported items related to the title and abstract (1.6%), reporting intervention adherence (3.1%), and interpretation of results (4.7%). No significant correlation was found between CONSORT score and journal impact factor (R=0.14, P=0.29), number of authors (R=0.01, P=0.93), or whether the RCT used paired-eye, one-eye, or two-eye designs in their randomisation (P=0.97).

CONCLUSIONS: The reporting of RCTs in ophthalmic surgery is suboptimal. Further work is needed by trial groups, funding agencies, authors, and journals to improve reporting clarity.

2039

2040 Appendix 2: Figure 3. Yao, A. C., **Khajuria, A**., Camm, C. F., Edison, E. &

2041 Agha, R. 2014. The reporting quality of parallel randomised controlled trials in

2042 ophthalmic surgery in 2011: a systematic review. Eye (Lond), 28, 1341-9.

A Meta-analysis of Clinical, Patient-Reported Outcomes and Cost of DIEP versus Implant-based Breast Reconstruction.

Khajuria A[1,2], Prokopenko M[3], Greenfield M[3], Smith O[3], Pusic AL[4], Mosahebi A[3].

⊖ Author information

1   Department of Continuing Education, Kellogg College, University of Oxford, UK.
2   Department of Surgery and Cancer, Imperial College London, UK.
3   Department of Plastic Surgery, Royal Free Hospital, London, UK.
4   Patient-Reported Outcomes, Value and Experience (PROVE) Centre, Brigham and Women's Hospital, Harvard Medical School.

Abstract

INTRODUCTION: Comparative data on clinical outcomes and cost of deep inferior epigastric perforator (DIEP) and implant-based reconstruction (IBR) are limited. We conducted a Preferred Reporting Items for Systematic Review and Meta-analysis-compliant systematic review and meta-analysis to compare clinical, patient-reported outcomes (PROs) and cost.

METHODS: The protocol was published a priori on PROSPERO (CRD42017072557). EMBASE, MEDLINE, Google Scholar, Cochrane Controlled Register of Trials, Science Citation Index, and ClinicalTrials.gov were searched from January 1994 to August 2018. Two independent reviewers evaluated the articles for inclusion. Study quality was assessed using Grading of Recommendations Assessment, Development, and Evaluation, and risk of bias (RoB) was assessed using Cochrane's RoB in Nonrandomized Studies of Interventions tool.

RESULTS: Out of 6,381 articles screened, 16 were included [unilateral 782 DIEPs, 376 implants; mean age 49 years, follow-up (months): DIEP 29.9; IBR 35.5]. Mean flap loss and fat necrosis rates were 3.97% (SD 4.90) and 9.67% (SD 17.0), respectively. There was no difference in mean length of stay {standard mean difference 0.63 [confidence interval (CI) -9.17 to 10.43]; $P =0.90$}. The number of reoperations for complications was significantly lower in DIEP versus IBR [SMD -0.29 (CI -0.48 to -0.09); $P < 0.01$]. There were no randomized controlled trials. Study quality was low with high RoB. One study reported $11,941/Quality-adjusted Life Year incremental cost-effectiveness ratio for DIEP, with higher breast Quality-adjusted Life Year (DIEP 19.5; IBR 17.7) using Breast Questionnaire; 3 studies evaluated cost, favoring DIEP. Two comparative studies evaluating PROs favored DIEP.

CONCLUSIONS: DIEP reconstruction maybe more cost-effective and yield superior PROs. However, poor-quality, bias-ridden studies limit the findings. Adequate reporting of core outcome measures is required to minimize reporting bias and facilitate evidence synthesis. Prospective, multicenter, cohort studies using robust patient-reported outcome measures (PROMs) tools, evaluating cost-effectiveness and contributing to national/international registries, will facilitate national-level policy and shared decision-making.

2043

2044   Appendix 3: Figure 4. **Khajuria, A**., Prokopenko, M., Greenfield, M., et al.

2045   2019. A Meta-analysis of Clinical, Patient-Reported Outcomes and Cost of

2046   DIEP versus Implant-based Breast Reconstruction. Plast Reconstr Surg Glob

2047   Open, 7**,** e2486.
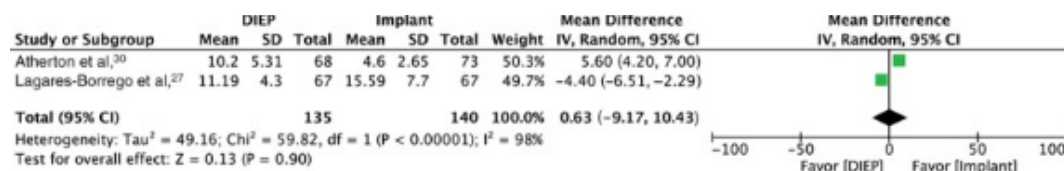
2048

2049   Appendix 4: Figure 5. Forest plot for 2 comparative studies, evaluating mean
2050   length of stay (days)
2051
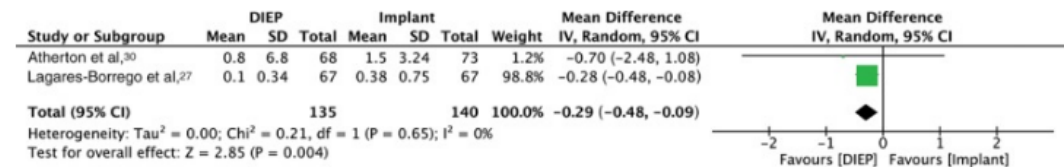
2052

2053   Appendix 5: Figure 6. Forest plot for 2 comparative studies, evaluating mean

2054   number of reoperations for complications.

Immediate and delayed autologous abdominal microvascular flap breast reconstruction in patients receiving adjuvant, neoadjuvant or no radiotherapy: a meta-analysis of clinical and quality-of-life outcomes.

Khajuria A[1,2], Charles WN[2], Prokopenko M[3], Beswick A[4], Pusic AL[5], Mosahebi A[3], Dodwell DJ[6], Winters ZE[7].

⊟ Author information

1   Kellogg College, Nuffield Department of Surgery, University of Oxford, Oxford, UK.
2   Department of Surgery and Cancer, Imperial College London, London, UK.
3   Department of Plastic Surgery, Royal Free Hospital, London, UK.
4   School of Clinical Sciences, University of Bristol, Bristol, UK.
5   Patient-Reported Outcomes, Value and Experience Centre, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.
6   Nuffield Department of Population Health, University of Oxford, Oxford, UK.
7   Surgical Intervention Trials Unit, Division of Surgery and Interventional Science, University College London, London, UK.

Abstract in English, Spanish

BACKGROUND: Effects of postmastectomy radiotherapy (PMRT) on autologous breast reconstruction (BRR) are controversial regarding surgical complications, cosmetic appearance and quality of life (QOL). This systematic review evaluated these outcomes after abdominal free flap reconstruction in patients undergoing postoperative adjuvant radiotherapy (PMRT), preoperative radiotherapy (neoadjuvant radiotherapy) and no radiotherapy, aiming to establish evidence-based optimal timings for radiotherapy and BRR to guide contemporary management.

METHODS: The study was registered on PROSPERO (CRD42017077945). Embase, MEDLINE, Google Scholar, CENTRAL, Science Citation Index and ClinicalTrials.gov were searched (January 2000 to August 2018). Study quality and risk of bias were assessed using GRADE and Cochrane's ROBINS-I respectively.

RESULTS: Some 12 studies were identified, involving 1756 patients (350 PMRT, 683 no radiotherapy and 723 neoadjuvant radiotherapy), with a mean follow-up of 27·1 (range 12·0-54·0) months for those having PMRT, 16·8 (1·0-50·3) months for neoadjuvant radiotherapy, and 18·3 (1·0-48·7) months for no radiotherapy. Three prospective and nine retrospective cohorts were included. There were no randomized studies. Five comparative radiotherapy studies evaluated PMRT and four assessed neoadjuvant radiotherapy. Studies were of low quality, with moderate to serious risk of bias. Severe complications were similar between the groups: PMRT versus no radiotherapy (92 versus 141 patients respectively; odds ratio (OR) 2·35, 95 per cent c.i. 0·63 to 8·81, P = 0·200); neoadjuvant radiotherapy versus no radiotherapy (180 versus 392 patients; OR 1·24, 0·76 to 2·04, P = 0·390); and combined PMRT plus neoadjuvant radiotherapy versus no radiotherapy (272 versus 453 patients; OR 1·38, 0·83 to 2·32, P = 0·220). QOL and cosmetic studies used inconsistent methodologies.

CONCLUSION: Evidence is conflicting and study quality was poor, limiting recommendations for the timing of autologous BRR and radiotherapy. The impact of PMRT and neoadjuvant radiotherapy appeared to be similar.

2055

2056   Appendix 6: Figure 7. **Khajuria, A**., Charles, W. N., Prokopenko, M., et al.

2057   2020. Immediate and delayed autologous abdominal microvascular flap

2058   breast reconstruction in patients receiving adjuvant, neoadjuvant or no

2059   radiotherapy: a meta-analysis of clinical and quality-of-life outcomes. BJS

2060   Open, 4**,** 182-196.

2061

a Overall complications

Reference	RT	No RT	Weight (%)	Odds ratio	Odds ratio
Cooke et al.[60]	20 of 64	16 of 61	57·9	1·28 (0·59, 2·78)	
O'Connell et al.[58]	11 of 28	20 of 80	42·1	1·94 (0·78, 4·83)	
Total	31 of 92	36 of 141	100·0	1·52 (0·84, 2·75)	

Heterogeneity: $\tau^2 = 0·00$; $\chi^2 = 0·47$, 1 d.f., $P = 0·49$; $I^2 = 0\%$
Test for overall effect: $Z = 1·40$, $P = 0·16$

Favours RT   Favours no RT

b CDC grade III complications

Reference	RT	No RT	Weight (%)	Odds ratio	Odds ratio
Cooke et al.[60]	6 of 64	1 of 61	31·6	6·21 (0·72, 53·15)	
O'Connell et al.[58]	4 of 28	8 of 80	68·4	1·50 (0·41, 5·43)	
Total	10 of 92	9 of 141	100·0	2·35 (0·63, 8·81)	

Heterogeneity: $\tau^2 = 0·23$; $\chi^2 = 1·29$, 1 d.f., $P = 0·26$; $I^2 = 22\%$
Test for overall effect: $Z = 1·27$, $P = 0·20$

Favours RT   Favours no RT

c CDC grade II complications

Reference	RT	No RT	Weight (%)	Odds ratio	Odds ratio
Cooke et al.[60]	2 of 64	4 of 61	28·0	0·46 (0·08, 2·61)	
O'Connell et al.[58]	3 of 28	3 of 80	29·8	3·08 (0·58, 16·24)	
Rogers and Allen [61]	5 of 30	7 of 30	42·1	0·66 (0·18, 2·36)	
Total	10 of 122	14 of 171	100·0	0·94 (0·32, 2·76)	

Heterogeneity: $\tau^2 = 0·29$; $\chi^2 = 2·92$, 2 d.f., $P = 0·23$; $I^2 = 31\%$
Test for overall effect: $Z = 0·11$, $P = 0·91$

Favours RT   Favours no RT

2062

d Fat necrosis

Reference	RT	No RT	Weight (%)	Odds ratio	Odds ratio
Cooke et al.[60]	2 of 64	1 of 61	15·7	1·94 (0·17, 21·91)	
O'Connell et al.[58]	1 of 28	2 of 80	15·5	1·44 (0·13, 16·57)	
Peeters et al.[56]	6 of 16	36 of 109	57·7	1·22 (0·41, 3·61)	
Rogers and Allen[61]	7 of 30	0 of 30	11·2	19·47 (1·06, 358·38)	
Total	16 of 138	39 of 280	100·0	1·83 (0·67, 5·00)	

Heterogeneity: $\tau^2 = 0·15$; $\chi^2 = 3·40$, 3 d.f., $P = 0·33$; $I^2 = 12\%$
Test for overall effect: $Z = 1·18$, $P = 0·24$

Favours RT   Favours no RT

2063

2064

2065 Appendix 7: Figure 8. Forest plots comparing adjuvant radiotherapy with no

2066 radiotherapy; a. Overall complications, b. Clavien–Dindo classification (CDC)

2067 grade III complications, c. CDC grade II complications, d fat necrosis. RT,

2068 radiotherapy.

2069

a Overall complications

Reference	Neoadjuvant RT	No RT	Weight (%)	Odds ratio	Odds ratio
Modarressi et al.[64]	20 of 60	9 of 45	19·8	2·00 (0·81, 4·95)	
Mull et al.[65]	26 of 142	45 of 312	57·8	1·33 (0·78, 2·26)	
O'Connell et al.[58]	12 of 38	20 of 80	22·4	1·38 (0·59, 3·24)	
Total	58 of 240	74 of 437	100·0	1·45 (0·97, 2·18)	

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.60$, 2 d.f., $P = 0.74$; $I^2 = 0\%$
Test for overall effect: $Z = 1.82$, $P = 0.07$

Favours neoadjuvant RT   Favours no RT

b CDC grade III complications

Reference	Neoadjuvant RT	No RT	Weight (%)	Odds ratio	Odds ratio
Mull et al.[65]	26 of 142	45 of 312	87·3	1·33 (0·78, 2·26)	
O'Connell et al.[58]	3 of 38	8 of 80	12·7	0·77 (0·19, 3·09)	
Total	29 of 180	53 of 392	100·0	1·24 (0·76, 2·04)	

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.52$, 1 d.f., $P = 0.47$; $I^2 = 0\%$
Test for overall effect: $Z = 0.85$, $P = 0.39$

Favours neoadjuvant RT   Favours no RT

c Fat necrosis

Reference	Neoadjuvant RT	No RT	Weight (%)	Odds ratio	Odds ratio
O'Connell et al.[58]	2 of 38	2 of 80	8·5	2·17 (0·29, 16·00)	
Peeters et al.[66]	29 of 77	36 of 109	91·5	1·23 (0·67, 2·25)	
Total	31 of 115	38 of 189	100·0	1·29 (0·72, 2·30)	

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.29$, 1 d.f., $P = 0.59$; $I^2 = 0\%$
Test for overall effect: $Z = 0.85$, $P = 0.40$

Favours neoadjuvant RT   Favours no RT

2070

2071   d Partial flap loss

Study or Subgroup	Neoadjuvant RT		No RT		Weight	Odds Ratio M-H, Random, 95% CI	Odds Ratio M-H, Random, 95% CI
	Events	Total	Events	Total			
Modarressi 2017	12	60	2	45	36.0%	5.38 [1.14, 25.39]	
Mull 2017	7	142	5	312	64.0%	3.18 [0.99, 10.21]	
Total (95% CI)		202		357	100.0%	3.85 [1.51, 9.76]	
Total events	19		7				

Heterogeneity: Tau² = 0.00; Chi² = 0.29, df = 1 (P = 0.59); I² = 0%
Test for overall effect: Z = 2.83 (P = 0.005)

Favours [Neoadjuvant RT]   Favours [No RT]

2072

2073   Appendix 8: Figure 9. Forest plot comparing neoadjuvant radiotherapy with no

2074   radiotherapy; a. Overall complications, b. Clavien–Dindo classification (CDC)

2075   grade III complications, c. fat necrosis, d. partial flap loss. RT, radiotherapy.

2076

a Overall complications

Reference	RT	No RT	Weight (%)	Odds ratio	
Cooke et al.[60]	20 of 64	16 of 61	19·7	1·28 (0·59, 2·78)	
Modarressi et al.[64]	20 of 60	9 of 45	14·5	2·00 (0·81, 4·95)	
Mull et al.[65]	26 of 142	45 of 312	42·5	1·33 (0·78, 2·26)	
O'Connell et al.[58]	23 of 66	20 of 80	23·3	1·60 (0·78, 3·28)	
Total	89 of 332	90 of 498	100·0	1·46 (1·04, 2·07)	

Heterogeneity: $\tau^2=0·00$; $\chi^2=0·76$, 3 d.f., $P=0·86$; $I^2=0\%$
Test for overall effect: $Z=2·16$, $P=0·03$

Odds ratio scale: 0·01 0·1 1 10 100
Favours RT Favours no RT

b CDC grade III complications

Reference	RT	No RT	Weight (%)	Odds ratio	
Cooke et al.[60]	6 of 64	1 of 61	5·6	6·21 (0·72, 53·15)	
Mull et al.[65]	26 of 142	45 of 312	72·8	1·33 (0·78, 2·26)	
O'Connell et al.[58]	7 of 66	8 of 80	21·5	1·07 (0·37, 3·12)	
Total	39 of 272	54 of 453	100·0	1·38 (0·83, 2·32)	

Heterogeneity: $\tau^2=0·02$; $\chi^2=2·15$, 2 d.f., $P=0·34$; $I^2=7\%$
Test for overall effect: $Z=1·24$, $P=0·22$

Odds ratio scale: 0·01 0·1 1 10 100
Favours RT Favours no RT

c Fat necrosis

Reference	RT	No RT	Weight (%)	Odds ratio	
Cooke et al.[60]	2 of 64	1 of 61	4·2	1·94 (0·17, 21·91)	
O'Connell et al.[58]	3 of 66	2 of 80	7·3	1·86 (0·30, 11·46)	
Peeters et al.[66]	35 of 93	36 of 109	87·0	1·22 (0·69, 2·18)	
Rogers and Allen[61]	7 of 30	0 of 30	1·6	19·47 (1·06, 358·38)	
Total	47 of 253	39 of 280	100·0	1·59 (0·96, 2·64)	

Heterogeneity: $\chi^2=3·68$, 3 d.f., $P=0·30$; $I^2=18\%$
Test for overall effect: $Z=1·79$, $P=0·07$

Odds ratio scale: 0·01 0·1 1 10 100
Favours RT Favours no RT

d Partial flap loss
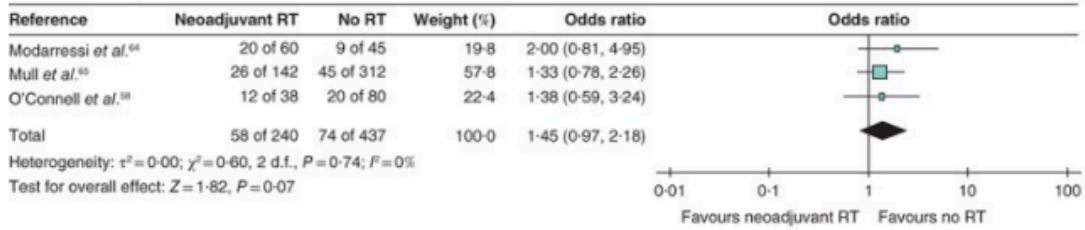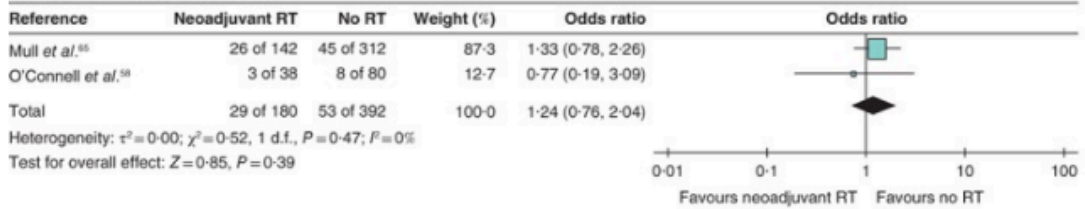
Study or Subgroup	RT Events	RT Total	No RT Events	No RT Total	Weight	Odds Ratio M-H, Random, 95% CI	
Cooke 2017	9	64	6	61	41·9%	1·50 [0·50, 4·50]	
Modarressi 2017	12	60	2	45	21·0%	5·38 [1·14, 25·39]	
Mull 2017	7	142	5	312	37·2%	3·18 [0·99, 10·21]	
Total (95% CI)		266		418	100·0%	2·59 [1·27, 5·28]	
Total events	28		13				

Heterogeneity: Tau² = 0·00; Chi² = 1·93, df = 2 (P = 0·38); I² = 0%
Test for overall effect: Z = 2·63 (P = 0·009)

Odds Ratio M-H, Random, 95% CI scale: 0·01 0·1 1 10 100
Favours [RT] Favours [No RT]

Appendix 9: Figure 10. Forest plot comparing combined adjuvant and neoadjuvant radiotherapy with no radiotherapy; a. Overall complications, b. CDC grade III complications, c. fat necrosis; d. partial flap loss. RT, radiotherapy.

npg

The reporting quality of parallel randomised controlled trials in ophthalmic surgery in 2011: a systematic review

AC Yao[1], A Khajuria[1], CF Camm[2], E Edison[3] and R Agha[4]

Abstract

Purpose   Randomised controlled trials (RCTs) represent a gold standard for evaluating therapeutic interventions. However, poor reporting clarity can prevent readers from assessing potential bias that can arise from a lack of methodological rigour. The Consolidated Standards of Reporting Trials statement for non-pharmacological interventions 2008 (CONSORT NPT) was developed to aid reporting. RCTs in ophthalmic surgery pose particular challenges in study design and implementation. We aim to provide the first assessment of the compliance of RCTs in ophthalmic surgery to the CONSORT NPT statement.

Method   In August 2012, the Medline database was searched for RCTs in ophthalmic surgery reported between 1 January 2011 and 31 December 2011. Results were searched by two authors and relevant papers selected. Papers were scored against the 23-item CONSORT NPT checklist and compared against surrogate markers of paper quality. The CONSORT score was also compared between different RCT designs.

Results   In all, 186 papers were retrieved. Sixty-five RCTs, involving 5803 patients, met the inclusion criteria. The mean CONSORT score was 8.9 out of 23 (39%, range 3.0–14.7, SD 2.49). The least reported items related to the title and abstract (1.6%), reporting intervention adherence (3.1%), and interpretation of results (4.7%). No significant correlation was found between CONSORT score and journal impact factor ($R = 0.14$, $P = 0.29$), number of authors ($R = 0.01$, $P = 0.93$), or whether the RCT used paired-eye, one-eye, or two-eye designs in their randomisation ($P = 0.97$).

Conclusions   The reporting of RCTs in ophthalmic surgery is suboptimal. Further work is needed by trial groups, funding agencies, authors, and journals to improve reporting clarity.

Eye advance online publication, 12 September 2014; doi:10.1038/eye.2014.206

Introduction

The randomised controlled trial (RCT) is a cornerstone of medical research and evidence-based medicine. RCTs are widely regarded as the 'criterion standard' for evaluating the effectiveness of an intervention. They are classed in the Levels of Evidence as level 1b by the Oxford Centre for Evidence-based Medicine.[1] However, poorly reported RCTs are associated with bias in estimating the effectiveness of interventions,[2,3] and inconsistencies between the conclusions and results.[4] Adequate and accurate reporting is vital to facilitate critical appraisal and interpretation of the data by the readers.

The Consolidated Standards of Reporting Trials (CONSORT) statement was developed to provide a minimum set of standards for transparent reporting of RCTs. The original CONSORT statement, published in 1996,[5] has since been revised in 2001,[6,7] and updated most recently in 2010.[8] Additionally, an extension to the statement was developed to address specific issues surrounding the reporting of RCTs evaluating surgical interventions.[9] The 2008 CONSORT extension for non-pharmacological treatment interventions (CONSORT NPT) is an extension on the 2001 CONSORT checklist that incorporates additional issues relating to masking difficulty, intervention complexity, and

[1]School of Medicine, Imperial College London, London, UK

[2]New College, University of Oxford, Oxford, UK

[3]University College London, London, UK

[4]Stoke Mandeville Hospital, Buckinghamshire, UK

Correspondence:
AC Yao, School of Medicine, Imperial College London, Sir Alexander Fleming Building, Exhibition Road, London SW7 2AZ, UK
Tel: +44 (0)7807096773;
E-mail: alexander.c.yao@gmail.com

inconsistent care providers' expertise that commonly affect surgical RCTs.[10,11]

RCTs in ophthalmology represent further challenges for researchers;[12] for example, each patient has the potential to contribute two data points. Studies in ophthalmology may require alternative designs and hence alternative methods of analysis to accommodate this.[13,14] Previously, reporting of RCT abstracts in ophthalmology has been suboptimal.[15] A review of 24 ophthalmology RCTs published in 1999 found that only an average of 33.4 out of 57 descriptors were adequately reported to the standard described in the 1996 CONSORT statement.[16] We are unaware of previous assessments regarding the compliance of RCTs in ophthalmic surgery to the CONSORT NPT, and could find no reference in a computerised search of the PubMed database.

The primary objective of this study was to assess the compliance of recent RCTs in ophthalmic surgery to the 2008 CONSORT NPT extension of the CONSORT 2001 statement. The secondary objectives included identifying any associations between CONSORT NPT compliance and surrogate markers of article quality, including ISI 2011 impact factor of the publishing journal, number of authors, number of patients in the trial, and whether the study was a single- or multi-centre study. The association between CONSORT score and different designs in randomisation of ophthalmology RCTs was also analysed.

Materials and methods

Search method

The Medline database was searched during August 2012 for RCTs from the period 1 January 2011 to 31 December 2011 for the Medical Subject Headings 'Ophthalmic Surgical Procedures' NOT 'Pharmacology', with the 'explode' function activated. Limitations were set for English language and trials on human subjects. Results were then manually searched independently by two authors (ACY and AK) for RCTs that satisfied the inclusion criteria. The RCTs were identified by reviewing the titles and abstracts of the results. Where there was insufficient information in the title and abstract for determining inclusion, the full article was obtained and reviewed. The two authors then resolved any conflicts in article selection by consensus. Where differences remained, a third author (CFC) was consulted to make the final decision. After the final selection was confirmed, all full articles were obtained. The search protocol is summarised in the PRISMA (Preferred Reporting for Systematic Reviews and Meta-Analyses) flow diagram (Figure 1).

Studies were only included if they were randomised, parallel-group RCTs in humans, involving a surgical procedure as at least one intervention arm. Excluded were studies involving purely pharmacological interventions, cost-effectiveness or economic analyses, interim analyses, short communications, simulation studies, and studies involving only cadaveric eyes.

Scoring

The papers were then scored independently by two authors (ACY and AK) against the 23 items on the 2008 CONSORT NPT extension of the 2001 CONSORT checklist. Each item was given an equal weighting, scoring 1 each, for a total of 23. Articles were scored 1 for an item if all information detailed in the respective item was reported, an approach reflective of the latest CONSORT 2010 guidelines.[8] Otherwise the item was scored 0. Two items were subdivided in the CONSORT NPT statement: item 4 included three parts (4A, 4B, and 4C), and item 11 had two parts (11A, 11B). For these items each had its parts scored independently, with each worth a third and one-half, respectively. The resulting mark out of 23 was termed the 'CONSORT score'. After initial scoring, any discrepancies in scores between the two authors were settled by consensus. If agreement could not be reached, the third author (CFC) was consulted for the final decision.

Secondary analyses

The relationship between the CONSORT score and several surrogate markers of article quality were also analysed (all prespecified). These included the number of authors;[17,18] number of patients; ISI 2011 impact factor of publishing journal;[19] and whether the study was a single or multi-centre study. The relationship between the CONSORT score and different designs in randomisation of ophthalmology RCTs, as defined by Lee *et al*,[12] was analysed: paired-eye design, one-eye design, and two-eye design.

Statistical analyses

Inter-rater reliability was assessed using the Cohen's kappa score calculation. Spearman Rank correlation coefficient was used to assess the relationship between CONSORT score and surrogate markers of article quality. The Mann–Whitney U test was used to measure inter-group differences between single- and multi-centre trials. The Kruskal–Wallis test was used to analyse the CONSORT scores between different study designs: paired-eye, one-eye, and two-eye designs. Differences in CONSORT score between same-group, different-group, and mixed two-eye designs were also analysed using the Kruskal–Wallis test. All statistical analyses were carried out using SPSS (version 22.0; SPSS Inc., Chicago, IL, USA).
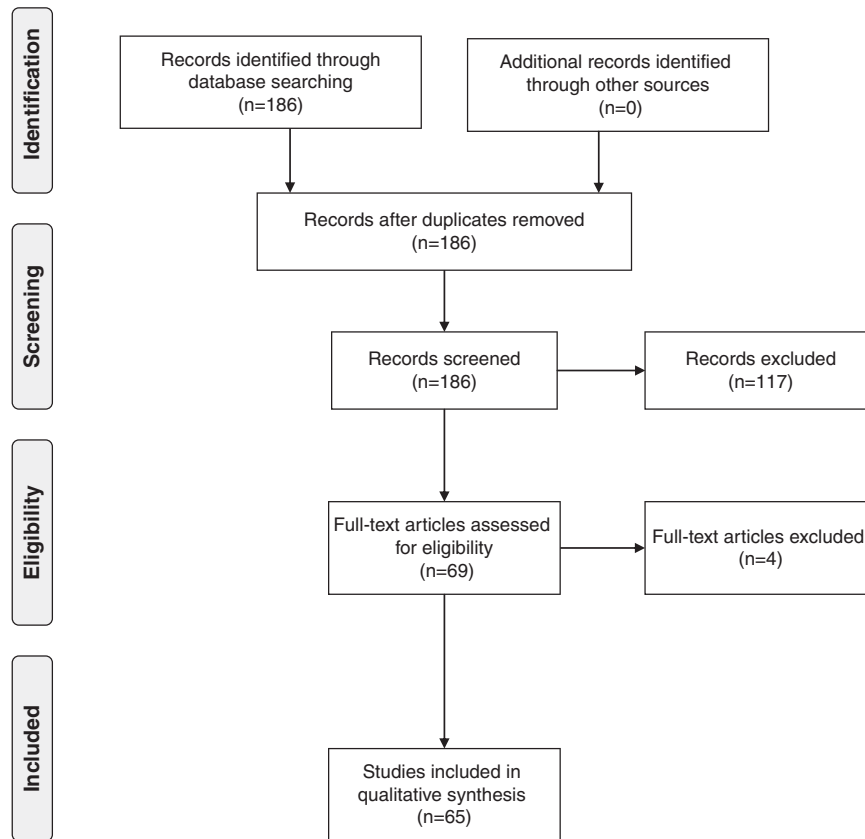
Figure 1 PRISMA flow diagram of article selection for scoring.

Results

In all, 186 articles were retrieved from the search of the Medline database (Figure 1). Of these, 69 articles were selected. Following review of the full articles, four articles were excluded: two for not being RCTs, and two for being unrelated to ophthalmology. The remaining 65 RCTs, involving 5803 patients, met the inclusion criteria. Inter-observer concordance for article selection had a kappa score of 0.91. In total 1495 items were scored. Following the initial round of scoring, the authors' scores were disputed on 50 items (2.8%). All 50 disputed items were resolved following discussion. The kappa score for the initial round of scoring was 0.94.

The mean CONSORT score of the 65 RCTs was 8.9 out of 23 (39%, range 3.0–14.7, SD 2.49). The compliance for individual items is shown in Table 1 and Figure 2. The poorest-reported items were item 1: title and abstract (one paper, 1.6%), item 4c: details of how adherence with protocol was assessed (two papers, 3.1%), and item 20: interpretation of results (three papers, 4.7%). No paper adequately reported all items in the CONSORT checklist.

Six journals' impact factors were not listed in ThompsonReuters' Journal Citation Reports,[19] which included 7 of the 65 RCTs. For the 58 remaining papers,

there was no correlation between CONSORT score and the impact factor (Spearman rho $= 0.14$, $P = 0.29$, Cohen's $d = 3.297$), Figure 3. There was no correlation between CONSORT score and the number of authors (Spearman rho $= 0.01$, $P = 0.93$, Cohen's $d = 1.533$). There was no statistically significant difference between the scores of single- and multi-centre trials ($P = 0.58$, Cohen's $d = 0.226$), or between paired-eye, one-eye, or two-eye RCT designs ($P = 0.98$, partial $\eta^2 = 0.001$). In addition, there was no statistical difference in CONSORT score between same-group, different-group, and mixed two-eye RCT designs ($P = 0.97$, partial $\eta^2 = 0.005$).

Discussion

RCT adherence to the CONSORT NPT checklist varied considerably. The CONSORT score ranged widely from 3 to 14.7 out of 23 items in this study. Several items integral to trial reporting, such as the background, rationale, objectives, and hypotheses, were well reported. Notably, adherence was over 95% to item 2: background, item 5: specifying objectives/hypotheses, and item 22: general interpretation of results in the context of current evidence. Despite this, the mean score was only 8.9 out of

Table 1 Adherence of RCTs to individual items of the CONSORT NPT checklist

Item	Descriptor	Adherence (number of articles (%))
Title and abstract		
1	Title and abstract	1 (1.6)
Introduction		
2	Scientific background	63 (98.4)
Methods		
3	Participant's eligibility, settings and locations	34 (53.1)
4a	Intervention details	53 (82.8)
4b	Intervention standardisation	46 (71.9)
4c	Assessment or enhancement of protocol adherence	2 (3.1)
5	Objectives and hypotheses	62 (96.9)
6	Primary and secondary outcome measures	24 (37.5)
7	Sample size, interim analyses, stopping rules	7 (10.9)
8	Random allocation sequence generation	28 (43.8)
9	Allocation concealment	13 (20.3)
10	Implementing allocation sequence	6 (9.4)
11a	Blinding (masking) status	18 (28.1)
11b	Method of blinding	6 (9.4)
12	Statistical methods	32 (50.0)
Results		
13	Participant flow	9 (14.1)
New item	Details of treatment as they were implemented	11 (17.2)
14	Recruitment and follow-up dates	33 (51.6)
15	Baseline demographic and clinical characteristics	16 (25.0)
16	Numbers analysed	6 (9.4)
17	Outcomes and estimation	4 (6.3)
18	Ancillary analyses	31 (48.4)
19	Adverse events	37 (57.8)
Discussion		
20	Interpretation of results taking into account potential bias	3 (4.7)
21	Generalisability	8 (12.5)
22	General interpretation in the context of current evidence	61 (95.3)

23 items (39%) on the CONSORT NPT. No RCTs obtained a full score.

Suboptimal compliance of RCT reporting to CONSORT is also found across many other surgical specialties including urological surgery,[20] general surgery,[21] neurosurgery,[22] orthopaedic surgery,[23] plastic surgery,[24] and vascular surgery,[20] as well as medical specialties such as cardiology.[25] The deficiencies identified in previous studies include particularly poor reporting of randomisation implementation, masking status, and healthcare providers.[26,27] Similar deficiencies in reporting quality were found in our study. A review of 164 RCTs by Agha *et al*[20] in six surgical specialties reported an average CONSORT score of only 11.2 out of the 22 items (51%) using the 2001 CONSORT statement. In our study, the same statement was used with the additional CONSORT NPT extension. The slightly lower CONSORT scores in our study is likely accounted for by the additional criteria within the extension.

The compliance to individual items was similarly varied. Inter-item variability appears globally consistent across other specialties.[20–25,28] In our study, over 90% of RCTs adequately reported scientific background and explaining rationale (item 2), reporting objectives or hypotheses (item 5), and interpreting results in the context of current evidence (item 22). This might be considered unsurprising, as these items represent the better recognised and readily achievable standards in the reporting of RCTs. High levels of reporting to item 2,[20,25] item 5,[21,24,25] and item 22[20,25] have also been reported in other specialties. Despite this, 15 of the 23 items were reported in less than 50% of the RCTs. Of these items, nine items were reported in less than 25% of the RCTs. Similar findings have been found in a wide range of surgical specialties.[20–24,28] Although most RCTs reported at least one aspect described by the item, a common reason for failure to score on an item was a failure to report all aspects highlighted by that item.
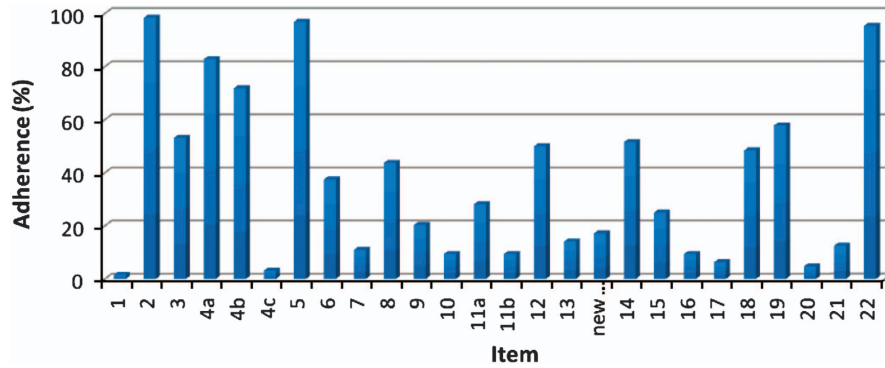
Figure 2 Adherence (%) of RCTs to individual items of the CONSORT extension for the NPT checklist. Overall, the mean adherence to any given item, including those subdivided, was 36.9%. Adherence ranged from 1 RCT (1.6%), in item 1, to 64 RCTs (98.4%), in item 2. (New... = New Item).
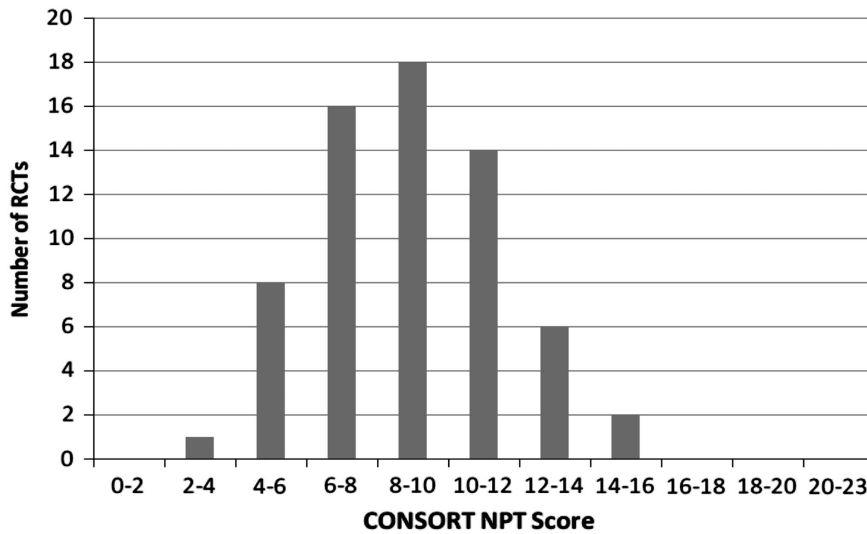
Figure 3 Histogram illustrating the distribution of RCTs obtaining particular CONSORT NPT scores.

The least reported item was related to the title and abstract (item 1). This was adequately reported in only 1 RCT (1.6%). Previous studies have shown this item to be well reported in other specialties.[20,22,24] However, these studies assess compliance against the CONSORT 2001 statement. In our study, RCTs generally mentioned 'randomisation' in the abstract or title fulfilling one aspect of the item. However, RCTs often failed to describe additional aspects of items as defined by the CONSORT NPT extension: the experimental treatment, care provider, centres involved, and masking status. Our pre-determined scoring strategy required all aspects of the item to be described to award the score, reflective of the CONSORT 2010 guidelines.[8] Indeed, these findings are consistent with Camm *et al*[25] assessing reporting of items to the CONSORT 2010 statement. Sufficiently detailed abstracts are essential as the readers often base their assessment of trials on the abstract information.

The value of complete abstract reporting is highlighted by the CONSORT Extension for Abstracts checklist.[29] Despite the publication of the checklist, Knobloch and Vogt[30] identified a mean compliance of only 9.46 out of the 17 items in the abstract extension checklist in 39 abstracts from the *Annals of Surgery*. Similarly, Berwanger *et al*[31] reviewed 227 abstracts from the *NEJM*, *JAMA*, *BMJ*, and *The Lancet*, finding that only 21 abstracts (9.3%) specified masking status.

There was no correlation between CONSORT score and surrogate markers of article quality. This is perhaps an unsurprising reflection that the CONSORT statement is more an assessment tool for the quality of RCT reporting rather than an assessment tool for the quality of RCT design itself. Neither the higher number of authors nor the higher journal impact factor was associated with improved CONSORT compliance, contrary to the popular belief that such markers help identify superior

articles.[17,18] Indeed, the evidence for association between surrogate markers of quality and CONSORT score is inconsistent. Camm et al[25] highlighted a significant association between impact factor and CONSORT 2010 score in RCTs concerning anti-arrhythmic agents. Balasubramanian et al[21] found that CONSORT score was significantly associated with higher author number, multi-centre studies, and impact factor in general surgery. However, Agha et al[20] reported no significant difference between CONSORT score and the same surrogate markers. Additionally, previous studies have also shown no link between higher impact factor and improved trial methodology.[32] Rigorous adoption of CONSORT by journals, however, has been shown to correlate with improved reporting quality.[33–37]

Fulfilment of the CONSORT checklist items was suboptimal across different types of RCT design. There was no significant difference in CONSORT score between single- and multi-centre trials ($P = 0.16$). In addition, there was no significant difference ($P = 0.46$) in trials randomising two eyes to the same group, different group, or a combination of same group and different group (mixed). This indicates that the need for improvement in reporting quality is not confined to specific types of study, but is applicable globally.

Healthcare providers face particular challenges in conducting surgical RCTs compared to pharmaceutical trials.[12,38–41] Notable difficulties include achieving and implementing masking, addressing varying expertise levels of care providers, and varying patient volumes of centres. Furthermore, inadequate funding and difficulty in securing consent may contribute to the lack of sufficient patient numbers, leading to low sample size and inadequate study power.[42,43] These factors may affect the accuracy in evaluating the effectiveness of interventions.[44] The CONSORT NPT extension provides a specific checklist to highlight the standards of reporting of these factors, which are not necessarily relevant to pharmaceutical trials.

Accurate and complete reporting of RCTs in ophthalmic surgery is especially important due to the potential added level of complexity of study design. The presence of two potential data points (ie two eyes) may lead to considerable heterogeneity in design, randomisation method, and statistical analysis.[12,45] Although there is a need to accurately inform readers of alternative statistical methodology, statistical consideration with respect to study design is often under-reported in many RCTs in ophthalmology.[12] In our study, 32 of the 64 RCTs (50%) adequately satisfied item 12 (regarding statistical methods). Poor reporting quality can prevent readers from assessing the potential bias that can arise from a lack of methodological rigour.[46]

Inadequate adherence to the CONSORT NPT may arise from failure at any of the four stages of the awareness-to-adherence model of compliance to guidelines (awareness, agreement, adoption, and adherence) defined by Pathman et al.[47] Given the heterogeneity of study designs in ophthalmic surgery, authors may be reluctant to consider using a checklist tool that was not developed for such a design. In addition, the adoption of the CONSORT statement and its extensions into journals' 'Instructions to Authors' has been suboptimal.[48–51] Despite a 73% increase since 2003, Hopewell et al[49] found that only 62 of 165 (38%) high-impact journals mentioned the CONSORT statement in their 'Instructions to Authors.' Although 50 of 57 responding editors (88%) stated that their journal recommended CONSORT, only 35 of 56 respondents (62%) stated that this was a requirement. Endorsement of the CONSORT extensions was noted to be especially lacking. The possibility should be considered that other factors such as journal word counts may encourage authors to include CONSORT items only selectively.

There are various limitations to this study. The search was restricted to articles in the English language and from the Medline database. The period studied was restricted to 2011, preventing any analysis of the temporal trends in CONSORT score. The number of RCTs including in this period was relatively small, limiting the power to examine the relationship between CONSORT scores and surrogate markers of RCT quality. Some CONSORT items may be included in associated RCT protocols in the public domain that were not analysed. Pragmatic difficulties arise in the scoring of RCT compliance to the CONSORT NPT. Many items contain multiple elements. Whether reviewers score items in regard to the multiple elements is a potential area of subjectivity. Subjectivity was minimised in this study by predefining the scoring strategy among the reviewers. The item was only scored if all elements were reported. This is on the basis that CONSORT items represent absolutely fundamental information; 'the minimum criteria,' that should be reported in a RCT.[8] Furthermore, all items on the checklist were given equal weighting to minimise subjectivity. Although this may not reflect their relative importance, it is nonetheless an objective approach to analyse deficits, patterns, as well as overall compliance.

The 2008 CONSORT NPT extension will benefit from updating to be brought in line with the CONSORT 2010 checklist. Key updates would include addition of the three new items regarding trial registration, availability of the trial protocol, and the declaration of funding. General changes might focus on reducing obfuscation by alterations in wording: replacing, simplifying, or removing misused words or phrases. In addition, greater

specificity and subdivisions of items would help to address the additional requirements for NPTs.

There is a need to improve the quality of reporting of RCTs in ophthalmic surgery. The adoption of CONSORT by journals is associated with improved reporting quality,[33–37,52] and therefore we recommend journals are explicit towards authors regarding CONSORT before submission and peer review. Editors, peer reviewers, authors, and developers of reporting guidelines will benefit from working closely with groups such as the Enhancing the Quality and Transparency of Health Research Network to support development and dissemination of reporting guidelines.[53] Further development of the CONSORT Statement may help to improve compatibility to RCTs with alternative methodologies including within-person randomised trials, common in ophthalmic surgery. Future extensions to the CONSORT Statement will hopefully start to address this.[27]

Conclusion

In conclusion, our findings suggest that the 2008 CONSORT NPT guidelines are not being met in 2011. It is recommended that the authors, funding agencies, peer-reviewers, and journal-editors in ophthalmology collaborate to enhance the integration of CONSORT into the RCT publication process. Evolution and further extension of CONSORT will hopefully help to incorporate studies with alternative methodologies such as are seen in ophthalmology.

Summary

What was known before
- Despite the importance in the levels of evidence, randomised controlled trials (RCTs) in many surgical specialties are often inadequately reported.
- Previous studies have suggested similar inadequacies as applying to RCTs in ophthalmic surgery, in the reporting of abstracts.

What this study adds
- This study formally analysed the reporting quality of RCTs in ophthalmic surgery by assessing compliance to the 2008 CONSORT extension for Non-Pharmacological Treatment interventions (CONSORT NPT) guidelines.
- Overall, there was suboptimal compliance of RCTs in ophthalmic surgery in 2011 to the 2008 CONSORT NPT guidelines.
- Similar levels of RCT reporting quality were found in ophthalmic surgery compared with other surgical specialties.

Conflict of interest

The authors declare no conflict of interest.

References

1 Oxford Centre for Evidence-based MedicineLevels of Evidence, 2009. http://www.cebm.net/index.aspx?o=1025 (accessed 9 March 2013).

2 Schulz KF, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**(5): 408–412.

3 Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M *et al*. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; **352**(9128): 609–613

4 Boutron I, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010; **303**(20): 2058–2064.

5 Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I *et al*. Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *JAMA* 1996; **276**(8): 637–639.

6 CONSORT group. CONSORT Statement 2001 Checklist, 2001. Available from http://www.consort-statement.org/mod_product/uploads/CONSORT%202001%20checklist.pdf (accessed 9 March 2013).

7 Moher D, Schulz KF, Altman D. CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001; **285**(15): 1987–1991.

8 Schulz KF, Altman DG, Moher D. CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; **340**: c332.

9 CONSORT group. Extensions of the CONSORT statement, 2012. Available from http://www.consort-statement.org/extensions/ (accessed 9 March 2013).

10 Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. CONSORT Group. Extending the CONSORT statement to randomized trials of non pharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008; **148**(4): 295–309.

11 Boutron I, Moher D, Altman DG, Schulz KF, Methods Ravaud P. Processes of the CONSORT Group. Example of an extension for trials assessing non pharmacologic treatments. *Ann Intern Med* 2008; **148**(4): W60–W66.

12 Lee CF, Cheng AC, Fong DY. Eyes or subjects: are ophthalmic randomized controlled trials properly designed and analyzed? *Ophthalmology* 2012; **119**(4): 869–872

13 Ray WA, O'Day DM. Statistical analysis of multi-eye data in ophthalmic research. *Invest Ophthalmol Vis Sci* 1985; **26**(8): 1186–1188.

14 Gauderman WJ, Barlow WE. Sample size calculations for ophthalmologic studies. *Arch Ophthalmol* 1992; **110**(5): 690–692.

15 Scherer RW, Crawley B. Reporting of randomized clinical trial descriptors and use of structured abstracts. *JAMA* 1998; **280**(3): 269–272.

16 Sánchez-Thorin JC, Cortés MC, Montenegro M, Villate N. The quality of reporting of randomized clinical trials published in Ophthalmology. *Ophthalmology* 2001; **108**(2): 410–415.

17 Figg WD, Dunn L, Liewehr DJ, Steinberg SM, Thurman PW, Barrett JC et al. Scientific collaboration results in higher citation rates of published articles. *Pharmacotherapy* 2006; **26**(6): 759–767.

18 Willis DL, Bahler CD, Neuberger MM, Dahm P. Predictors of citations in the urological literature. *BJU Int* 2011; **107**(12): 1876–1880.

19 Thomson ReutersJournal Citation Reports, 2011. Available from http://admin-apps.webofknowledge.com/JCR/JCR (accessed 9 March 2013).

20 Agha R, Cooper D, Muir G. The reporting quality of randomised controlled trials in surgery: a systematic review. *Int J Surg* 2007; **5**(6): 413–422.

21 Balasubramanian SP, Wiener M, Alshameeri Z, Tiruvoipati R, Elbourne D, Reed MW. Standards of reporting of randomized controlled trials in general surgery: can we do better? *Ann Surg* 2006; **244**(5): 663–667

22 Kiehna EN, Starke RM, Pouratian N, Dumont AS. Standards for reporting randomized controlled trials in neurosurgery. *J Neurosurg* 2011; **114**(2): 280–285.

23 Parsons NR, Hiskens R, Price CL, Achten J, Costa ML. A systematic survey of the quality of research reporting in general orthopaedic journals. *J Bone Joint Surg Br* 2011; **93**(9): 1154–1159.

24 Karri V. Randomised clinical trials in plastic surgery: survey of output and quality of reporting. *J Plast Reconstr Aesthet Surg* 2006; **59**(8): 787–796.

25 Camm CF, Chen Y, Sunderland N. An assessment of the Reporting Quality of Randomised Controlled Trials Relating to Anti-Arrhythmic Agents (2002-2011). *Int J Cardiol* 2013; **168**(2): 1393–1396.

26 Mills EJ, Wu P, Gagnier J, Devereaux PJ. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. *Contemp ClinTrials* 2005; **26**(4): 480–487.

27 Altman DG, Moher D, Schulz KF. Improving the reporting of randomised trials: the CONSORT Statement and beyond. *Stat Med* 2012; **31**(25): 2985–2997.

28 Chen HL, Liu K. Reporting in randomized trials published in International Journal of Cardiology in 2011 compared to the recommendations made in CONSORT 2010. *Int J Cardiol* 2012; **160**(3): 208–210.

29 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet* 2008; **371**(9609): 281–283.

30 Knobloch K, Vogt PM. Adherence to CONSORT abstract reporting suggestions in surgical randomized-controlled trials published in Annals of Surgery. *Ann Surg* 2011; **254**(3): 546(author reply 546–547).

31 Berwanger O, Ribeiro RA, Finkelsztejn A, Watanabe M, Suzumura EA, Duncan BB et al. The quality of reporting of trial abstracts is suboptimal: survey of major general medical journals. *J Clin Epidemiol* 2009; **62**(4): 387–392.

32 Agha RA, Camm CF, Edison E, Orgill DP. The methodological quality of randomized controlled trials in plastic surgery needs improvement: A systematic review. *J Plast Reconstr Aesthet Surg* 2012; **66**(4): 447–452.

33 Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist. *Control Clin Trials* 2002; **23**(4): 380–388.

34 Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001; **285**(15): 1992–1995; Available from http://www.ncbi.nlm.nih.gov/pubmed/11308436.

35 Alvarez F, Meyer N, Gourraud PA, Paul C. CONSORT adoption and quality of reporting of randomized controlled trials: a systematic analysis in two dermatology journals. *Br J Dermatol* 2009; **161**(5): 1159–1165.

36 Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006; **185**(5): 263–267.

37 Han C, Kwak K, Marks DM, Pae CU, Wu LT, Bhatia KS et al. The impact of the CONSORT statement on reporting of randomized clinical trials in psychiatry. *Contemp Clin Trials* 2009; **30**(2): 116–122.

38 McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ* 2002; **324**(7351): 1448–1451.

39 Gelijns AC, Ascheim DD, Parides MK, Kent KC, Moskowitz AJ. Randomized trials in surgery. *Surgery* 2009; **145**(6): 581–587.

40 Farrokhyar F, Karanicolas PJ, Thoma A, Simunovic M, Bhandari M, Devereaux PJ et al. Randomized controlled trials of surgical interventions. *Ann Surg* 2010; **251**(3): 409–416.

41 Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing non pharmacological treatments. *PLoS Med* 2007; **4**(2): e61.

42 Watson A, Frizelle F. The end of the one-eyed surgeon? Time for more randomised controlled trials of surgical procedures. *N Z Med J* 2004; **117**(1203): U1096.

43 Solomon M, McLeod R. Surgery and the randomised controlled trial: past, present and future. *Med J Aust* 1998; **169**(7): 380–383.

44 Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med* 2002; **137**(6): 511–520.

45 Evans J. Reliable and accessible reviews of the evidence for the effect of health care: the role of the Cochrane Collaboration and the CONSORT statement. *Eye (London)* 1998; **12**(Pt 1): 2–4.

46 Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001; **323**(7303): 42–46.

47 Pathman DE, Konrad TR, Freed GL, Freeman VA, Koch GG. The awareness-to-adherence model of the steps to clinical guideline compliance. The case of pediatric vaccine recommendations. *Med Care* 1996; **34**(9): 873–889.

48 Altman DG. Endorsement of the CONSORT statement by high impact medical journals: survey of instructions for authors. *BMJ* 2005; **330**(7499): 1056–1057.

49 Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT Statement by high impact factor medical journals: a survey of journal editors and journal 'Instructions to Authors'. *Trials* 2008; **9**: 20.

50 Schriger DL, Arora S, Altman DG. The content of medical journal Instructions for authors. *Ann Emerg Med* 2006; **48**(6): 743–749; 749.e1–4.

51 Tharyan P, Premkumar TS, Mathew V, Barnabas JP, Manuelraj. Editorial policy and the reporting of randomized controlled trials: a survey of instructions for authors and assessment of trial reports in Indian medical journals 2004-05; *Nat Med J India* **21**(2): 62–68.

52 Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J Clin Epidemiol* 2007; **60**(3): 241–249.

53 EQUATOR NetworkWelcome to the EQUATOR Network website. Available from http://www.equator-network.org/home/ (accessed 9 March 2013).

A Meta-analysis of Clinical, Patient-Reported Outcomes and Cost of DIEP versus Implant-based Breast Reconstruction

Ankur Khajuria, BSc (Hons), MD, FHEA, FRSPH, MRCS (Eng), MAcadMEd (Eng)*†
Max Prokopenko, BSc (Hons), MD‡
Max Greenfield, BSc (Hons), MD‡
Oliver Smith, BSc, MD, MRCS‡
Andrea L. Pusic, MD, MHS, FACS, FRCSC§
Afshin Mosahebi, MD, PhD, FRCS (Plast), MBA‡

Introduction: Comparative data on clinical outcomes and cost of deep inferior epigastric perforator (DIEP) and implant-based reconstruction (IBR) are limited. We conducted a Preferred Reporting Items for Systematic Review and Meta-analysis-compliant systematic review and meta-analysis to compare clinical, patient-reported outcomes (PROs) and cost.

Methods: The protocol was published a priori on PROSPERO (CRD42017072557). EMBASE, MEDLINE, Google Scholar, Cochrane Controlled Register of Trials, Science Citation Index, and ClinicalTrials.gov were searched from January 1994 to August 2018. Two independent reviewers evaluated the articles for inclusion. Study quality was assessed using Grading of Recommendations Assessment, Development, and Evaluation, and risk of bias (RoB) was assessed using Cochrane's RoB in Nonrandomized Studies of Interventions tool.

Results: Out of 6,381 articles screened, 16 were included [unilateral 782 DIEPs, 376 implants; mean age 49 years, follow-up (months): DIEP 29.9; IBR 35.5]. Mean flap loss and fat necrosis rates were 3.97% (SD 4.90) and 9.67% (SD 17.0), respectively. There was no difference in mean length of stay {standard mean difference 0.63 [confidence interval (CI) –9.17 to 10.43]; P =0.90}. The number of reoperations for complications was significantly lower in DIEP versus IBR [SMD –0.29 (CI –0.48 to –0.09); P< 0.01]. There were no randomized controlled trials. Study quality was low with high RoB. One study reported $11,941/Quality-adjusted Life Year incremental cost-effectiveness ratio for DIEP, with higher breast Quality-adjusted Life Year (DIEP 19.5; IBR 17.7) using Breast Questionnaire; 3 studies evaluated cost, favoring DIEP. Two comparative studies evaluating PROs favored DIEP.

Conclusions: DIEP reconstruction maybe more cost-effective and yield superior PROs. However, poor-quality, bias-ridden studies limit the findings. Adequate reporting of core outcome measures is required to minimize reporting bias and facilitate evidence synthesis. Prospective, multicenter, cohort studies using robust patient-reported outcome measures (PROMs) tools, evaluating cost-effectiveness and contributing to national/international registries, will facilitate national-level policy and shared decision-making. (*Plast Reconstr Surg Glob Open 2019;7:e2486; doi: 10.1097/GOX.0000000000002486; Published online 29 October 2019.*)

INTRODUCTION

Breast cancer is the most common malignancy and the principal cause of cancer-related mortality in women.[1,2] Breast-conserving surgery or mastectomy is normally offered as management strategies.[3] However, mastectomy has been associated with a profoundly negative impact on a woman's physical, psychological, and sexual well-being.[4] Assessment of quality of life (QoL) and patient-reported outcomes (PROs) is thus especially pertinent in breast reconstruction (BRR) surgery, and morbidity and mortality are necessary but not sufficient for adequate outcome

Disclosure: *AK is a Kellogg Scholar at the University of Oxford and receives funding equating to the scholarship amount. ALP is the codeveloper of the Breast Questionnaire and receives royalties when the Breast Questionnaire is used in industry-sponsored clinical trials. None Neither of the other authors has any financial disclosures.*

assessment.[5–17] The reconstruction must satisfy the patient with regard to physical, psychological, and sexual well-being. The exponential rise in QoL and PRO research highlights their importance.[12,18] Development and validation of psychometrically robust, validated disease-specific PRO tools such as the Breast Questionnaire (BREAST-Q) and the European Organization for Research and Treatment of Cancer (EORTC) Breast Cancer-specific Quality of Life Questionnaire-23 further exemplify this. Their development and validation have been described previously.[9,18–21]

Patient demands for BRR have significantly increased over the last 2 decades with the doubling of postmastectomy BRR rates from 13% to 26% between 1998 and 2007.[22] This is not only due to advances in oncological management but also due to the clearly demonstrable functional and psychological benefits.[23–26] Two of the commonest reconstructive modalities include autologous reconstruction using the deep inferior epigastric perforator (DIEP) flap and implant-based reconstruction (IBR).[27] The treatment choice is determined by patient factors (individual preference, body image) and surgeon factors (resource availability and experience).[28] Nevertheless, many plastic surgery units worldwide regard autologous reconstruction, compared with IBR, as the superior modality, replacing "like with like."[22] There is emerging evidence that autologous abdominal-based flap BRR may yield superior clinical and PROs.[15,27,29–31]

IBR is associated with complications, including implant rupture, infection, migration, exposure/extrusion, patient dissatisfaction with edge visibility/implant animation and reduced/absent sensation at the nipple.[32] Capsular contracture can culminate in pain, increased palpability, asymmetry, and implant removal requirement.[33] Allergan's 10-year cumulative risk study found that 24.6% of patients who underwent IBR developed capsular contracture.[34] Conversely, DIEP flap is widely considered the "gold standard" for postmastectomy BRR. It has largely superseded the traditional transverse rectus abdominus myocutaneous (TRAM) flap, by preserving the rectus abdominis muscle continuity and integrity, limiting donor site complications such as abdominal bulge/hernia.[35]

From an economic standpoint, some protagonists have argued that DIEP reconstruction is more cost-effective, yielding fewer overall complications and superior PROs, compared with IBR.[15,27,30] Although some European and North American centers have published cost-analyses on DIEP and IBR, the data are sparse with relative scarcity of data from public and free universal health care system settings. We systematically evaluated the quality of evidence and analyzed cost, clinical outcomes, and PROs of unilateral DIEP versus IBR in context of breast malignancy. The aim was to help evaluate which technique is superior in terms of clinical outcomes, PROs, and cost and thus inform worldwide clinical practice and facilitate informed consent and patient–clinician-shared decision-making.

METHODS

Our protocol was registered and published a priori on the National Institute of Health Research Prospective Register of Systematic Reviews PROSPERO (CRD42017072557) and *Systematic Reviews* peer-reviewed journal.[35–37] In the section below, we have detailed the search strategy used, the identification and selection of studies, and the design with inclusion/exclusion criteria. We have subsequently described the risk of bias (RoB) and quality assessment, outcomes, data extraction, collection and management, and the statistical methods utilized.

Search Strategies

We conducted a comprehensive search of the MEDLINE (OVID SP), EMBASE (OVID SP), Google Scholar, Cochrane Controlled Register of Trials, Science Citation Index databases, and ClinicalTrials.gov from January 1994 up to August 2018 to identify studies relevant to the review. A combination of Medical Subject Headings terms, free text, and Boolean logical operators were used to construct the search strategy, in consultation with a literature search expert. Explode function was utilized to capture narrower terms. No language restrictions were applied. The reference list of all included articles was also screened for relevance. A sample search strategy, for EMBASE (OVID SP), is shown below; a similar search strategy was adapted for the other databases:

1. exp Breast Neoplasms/OR ((breast adj6 cancer*) or (breast adj6 neoplasm*) or (breast adj6 carcinoma*) or (breast adj6 tumour*) or (breast adj6 tumor*) or (breast* adj4 reconstruct*))
2. exp deep inferior epigastric perforator flap/ OR DIEP flap* OR DIEAP flap* OR ((Deep and inferior and epigastric and perforator) adj2 flap*) OR Deep and inferior and epigastric and perforator and flap*)
3. exp breast implant/ OR breast adj3 implant* OR exp silicone prosthesis/147 – [(1) AND (2)] OR [(1) AND (3)]; publication date: January 1994 to August 2018

Identification and Selection of Studies

Studies were extracted following database searching and were populated into an Endnote X8 library (Clarivate Analytics, USA). Using prespecified screening criteria, the screening was carried out in 2 stages, by 2 independent reviewers.

Stage 1: Title and abstract screening carried out by 2 researchers independently (MP, MG). Any discrepancies were resolved by consensus. If any doubts remained, the article proceeded to full-text review.

Stage 2: The full texts of the studies included in stage 1 were downloaded and screened for eligibility by 2 researchers independently (MP, MG). Discrepancies were resolved by consensus. If this was not possible, the senior author (AM) was consulted for the final determination for inclusion/exclusion of the article.

Study Design

All primary human studies evaluating clinical outcomes, PROs, or cost for unilateral DIEP flap BRR or

IBR in context of breast malignancy were included. The intervention included unilateral DIEP BRR, and the comparator was IBR. The inclusion and exclusion criteria are highlighted below.

Inclusion Criteria

1. Studies involving adult patients aged ≥18 years old
2. Studies involving unilateral autologous DIEP flap BRR or IBR in context of breast malignancy
3. Clinical studies [randomized controlled trials (RCTs), prospective and retrospective cohort studies and case series with 10 or more patients]

Exclusion Criteria

Duplicates, case reports, conference abstracts, simulation studies, review articles, clinical studies in nonhuman subjects, patients with segmental or partial mastectomy, technical operative repair descriptions with no outcome measures, BRR unrelated to cancer, and autologous flap techniques other than DIEP were excluded. Studies of patients receiving adjuvant postmastectomy radiotherapy (PMRT) were also excluded, as adjuvant PMRT is associated with serious adverse events and reduced QoL in IBR, although the evidence is more equivocal for autologous reconstruction, and thus would introduce bias and preclude outcome analysis when comparing IBR and DIEP. Our group is currently conducting a separate systematic review and meta-analysis to investigate outcomes for immediate versus delayed autologous reconstruction in context of PMRT (PROSPERO CRD42017077945).[38]

RoB and Quality Assessment

For nonrandomized comparative studies, the RoB in Nonrandomized Studies of Interventions (ROBINS-I) by the Cochrane collaboration was used.[39] ROBINS-I covers 7 domains from which bias may be introduced, with "signaling questions" facilitating judgments about RoB. These domains include: (1) bias due to confounding; (2) bias in the selection of participants into the study; (3) bias in the classification of interventions; (4) bias due to deviations from intended interventions; (5) bias due to missing data; (6) bias in the measurement of outcomes; and (7) bias in the selection of the reported result. The judgments within each domain were carried forward for an overall RoB judgment across bias domains.[39] To assess individual study methodological quality, the Grading of Recommendations, Assessment, Development and Evaluation approach[40] was utilized.

Outcomes

The primary outcomes were as follows:

1. Clinical (complications: fat necrosis, partial/total flap loss, infection, number of reoperation procedures for complications and implant-specific complications, including capsular contracture, implant rupture, displacement, deflation and scarring), with grades of complications where reported
2. PRO measures (generic and disease-specific PROMs tools, eg, BREAST-Q and EORTC-QLQ-BR23)
3. Cost-analyses

Data Extraction, Collection, and Management

A standardized extraction form was used to extract data from the full-text articles by 2 independent authors (MP, MG). Any discrepancy was resolved by consensus or with referral to the senior author (AM). The following data were extracted:

- first author; year of publication; study design; participant demographics (sex, age, BMI and comorbidity, where reported); study setting; length of follow-up;
- primary outcomes, as above.

Statistical Methods

Using Review Manager 5.3,[41] provided by the Cochrane Collaboration, an assessment of heterogeneity was performed. The Higgins and Thompson's I^2 statistic was used to quantify statistical heterogeneity.[42] The DerSimonian and Laird random-effects model, which is well established for evaluating heterogeneous cohorts, was employed.[43] Odds ratios (ORs) with 95% confidence interval (CI) were used to determine dichotomous outcomes (complications). Continuous outcomes were evaluated by standardized mean differences with 95% CI.

RESULTS

A total of 6,381 records were identified. Out of those, 16 fulfilled the inclusion criteria and were considered for quantitative synthesis.[15,27,30,31,44–55] The Preferred Reporting Items for Systematic Review and Meta-analysis diagram (Fig. 1) depicts how studies were included and the reasons for exclusion. The 16 studies included 782 unilateral DIEPs and 376 implants; mean age 49 years, mean follow-up (months): DIEP 29.9; IBR 35.5. There were 6 prospective cohort studies,[27,31,44,47,50,52] 8 retrospective cohort studies,[15,30,45,48,51,53–55] 2 case series,[46,49] and no RCTs. There was 1 multicenter study[15]; the remaining 15 were single-center studies. The overall quality of the studies using the Grading of Recommendations, Assessment, Development and Evaluation criteria was low, with serious RoB using the ROBINS-I tool. Tables 1 and 2 summarize the baseline characteristics and results (clinical, PROs, and cost).

Clinical Outcomes

Two studies provided comparative data on mean length of stay (days),[27,30] with no difference between DIEP and IBR [SMD 0.63 (CI –9.17 to 10.43); $P = 0.90$] and significant statistical heterogeneity ($I^2 = 98\%$) (Fig. 2). Moreover, combining data from single-arm studies (7 studies),[44,45,49,50,52,53,55] further revealed no difference in mean length of stay in days [DIEP (8.32; SD 2.05) versus IBR (9.80; SD 8.20), $P = 0.89$]. Two studies provided comparative data on the mean number of reoperations for complications,[27,30] with a statistically significant lower number for DIEP versus IBR [SMD –0.29 (CI –0.48 to –0.09), $P < 0.01$] with $I^2 = 0\%$. The combined data from single-arm studies (7 studies)[44,51–53,55] showed lower mean number of revision procedures for DIEP (0.22; SD 0.27) versus IBR (0.50; SD 0.68), but without statistical significance ($P = 0.65$). There was no statistically significant difference in mean infection rates between DIEP (5 studies)[27,44,49,54,55]
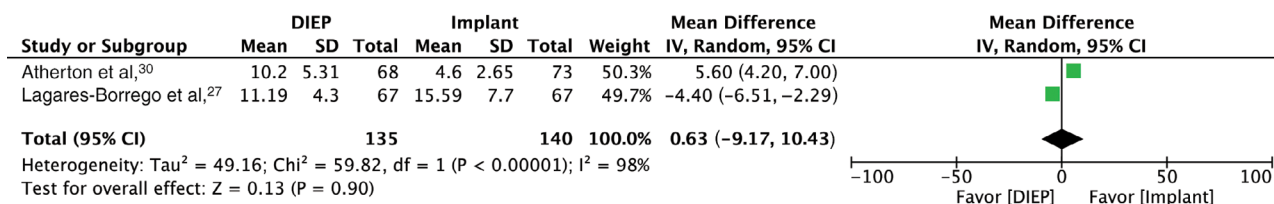
Fig. 1. Forest plot for 2 comparative studies, evaluating mean length of stay (days).

[1.67% (SD 2.29)] and IBR (2 studies)[27,52] [5.40% (SD 2.92)], P = 0.38. Three studies[27,46,52] reported mean implant-specific complication rates of 11.1% (SD 9.98); 1 classified capsular contracture grades as per the Baker's classification, with 1/30 (3.33%) grade IV contracture.[52] Out of all 6 IBR studies, 2 studies did not specify whether direct to implant (DTI) or expander–prosthesis (EP) reconstruction was employed.[15,30] One reported DTI,[52] and 3 reported EP reconstruction.[27,31,46] Mean flap loss and fat necrosis rates, reported by 9 studies, were 3.97% (SD 4.90) and 9.67% (SD 17.0), respectively.[27,31,44,47,49,51,53–55] No studies were reported as per the Clavien–Dindo classification (CDC).[56] Other than capsular contracture being graded as per the Baker's classification by 1 study,[52] none of the other complications were graded.

QoL

Two comparative studies evaluated QoL.[15,31] Tønseth et al[31] evaluated 29 patients with DIEP BRR and 21 patients with IBR. They utilized a generic PRO tool, Short-Form 36 (SF-36), which showed no difference in QoL, a nonvalidated study-specific questionnaire that showed higher breast satisfaction ($P < 0.001$), improved social relationship ($P = 0.02$) and body image satisfaction ($P = 0.01$) for DIEP, and a nonvalidated Visual Analog Scale, with superior cosmetic outcome with DIEP (Table 2). Matros et al[15] prospectively evaluated 103 patients with DIEP BRR and 172 patients with IBR and utilized the BREAST-Q. BREAST-Q scores were consistently higher for DIEP compared with implants in postoperative years 1–8, with a higher breast Quality-adjusted Life Year for DIEP (19.5) versus IBR (17.7).

Cost

Three comparative studies evaluated cost.[15,27,30] Matros et al (USA) calculated an incremental cost-effectiveness ratio (ICER) of $11,491 for DIEP, ie, the additional cost of DIEP BRR to obtain 1 year of perfect breast-related QoL compared with IBR. Lagares-Borrego et al, 2015 (Spain) reported no difference in overall cost between DIEP BRR (€18,857.77) versus IBR (€20,502.08); $P = 0.89$. However, when considering surgical complications, cost of DIEP (€2,859.90) was significantly lower than IBR (€5,837.9), $P < 0.001$. Cost of DIEP was also lower owing to length of hospital stay ($P < 0.001$), consultations ($P < 0.001$), and materials and tests used ($P < 0.001$), but higher owing to duration of procedure ($P < 0.001$). Atherton et al estimated cost at 3 years: DIEP £10,910 versus IBR £8,034. No statistics were performed; however, the authors reported that the cost "difference is small and patient will still require more revisions (with IBR), and if followed up

enough will lose this small financial benefit"; the cost difference maybe "justified by the increased patient satisfaction and cosmetic outcome (with DIEP)."

DISCUSSION

To our knowledge, this is the first systematic review and meta-analysis in available literature to evaluate clinical outcomes, PROs, and cost of DIEP versus IBR. Overall study quality is low with serious RoB, weakly supporting DIEP as a more cost-effective strategy that confers higher QoL compared with IBR. Factors limiting the quality of evidence include study designs, absence/heterogeneous reporting of clinical outcomes, study exclusion due to combined reporting of different flaps, and no breakdown of outcomes between unilateral and bilateral reconstruction. Majority of the studies have small sample sizes, were conducted in a retrospective manner with potentially biased patient recall after variable and delayed lengths of time post surgery, and failed to achieve adequate follow-up periods. Complications such as capsular contracture may occur well beyond this time frame. Fifteen or sixteen studies were single-center studies, negatively impacting generalizability.

Our systematic review demonstrates the inconsistency and heterogeneity in clinical outcome reporting, which presents a limitation. Only 8/14 (57.1%) studies evaluating DIEP reported flap loss rates. Likewise, only 3/6 (50.0%) studies evaluating IBR reported implant-specific complications, including capsular contracture. Only 1 of these studies classified capsular contracture according to the Baker's classification.[57] Because classification/grades help inform management strategies, inaccurate classification, and grading of these complications, risks biased comparisons of clinical outcomes between studies, rendering it difficult to interpret the study findings. This corroborates the results from the systematic review by Potter et al,[58] on reporting quality of BRR clinical outcomes, that identified poor reporting quality and need for a core outcome set to facilitate outcome assessment in effectiveness studies. Furthermore, no studies reported outcomes using the validated CDC.[56] Moreover, no studies reported grade of fat necrosis.

Standardization of outcome reporting, with uptake of validated tools such as CDC and incorporation into journal submission guidelines by editors, may promote higher quality, standardized reporting and facilitate homogeneity and meta-analysis.

Out of 6 IBR studies, 3 reported implant-specific complications; 2 out of 6 studies did not categorize type of IBR and reported as "implant reconstruction". Three out of 6

Table 1. Studies Evaluating DIEP or IBR Reconstruction; Comparative Studies' Data Presented as DIEP versus IBR

Reference Study, Location, Design	GRADE	ROBINS-I	No. Pts (DIEP)	No. Pts (IBR)	Mean F/u (mo) with SD/Range Where Reported	No. Overall Comps.	Fat Necrosis	Venous Congestion	Arterial Thrombosis	Flap Loss	Infection	Hematoma / Seroma	Mean LOS (days) with SD/Range Where Reported	Mean Number of Times Return to Theatre for Correction of Complications ± SD	Other Implant Comps.*	Cost-Analysis
Atherton et al,[30] UK, Cohort†	Low	Serious	68	73	36	NA	NA	NA	NA	NA	NA	NA	10.20 ± 5.31 cw 4.60 ± 2.65‡	0.8 ± 6.80 cw 1.5 ± 3.24§	NA	£10,910 cw £8,034
Cheng et al,[44] Taiwan, one-arm clinical trial‖	Moderate	Moderate	30	NA	NA	1	0	NA	NA	0	0	1	8.40	0.03	NA	(cost at 3 y)¶ £2,951
Kroll et al,[45] USA, Cohort‖	Low	Serious	21	NA	NA	NA	NA	NA	NA	NA	NA	NA	6.29	NA	NA	$18,941
Lagares-Borrego et al,[27] Spain, Cohort‖	Moderate	Moderate	67	67	45.31 ±15.65 cw 80.38 ±11.60	22 cw 26 (NS)	1 cw 0¶	1 cw 1¶	0 cw 0¶	2 cw 1¶	3 cw 5¶	2 cw 4¶	11.19 ± 4.3 [7–32] cw 15.59 ± 7.7 [6–51]‡	0.10±0.34 [0–1] cw 0.38±0.75§ [0–3]	15	€18,857.77 cw €20,502.08 (NS)
Matros et al,[15] USA, Cohort†	Moderate	Moderate	103	172	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	$75,184 cw $53,571¶ ICER: $11,941
McGeorge et al,[46] UK, Case series†	Low	Serious	NA	13	6	2	NA	NA	NA	NA	NA	1	NA	NA	1	NA
Moradi et al,[50] UK, Cohort‖	Moderate	Moderate	27	NA	NA	NA	1	NA	NA	NA	NA	1	9.1	NA	NA	NA
Nahabedian et al,[47] USA, Cohort‖	Moderate	Moderate	66	NA	NA	NA	6	1	NA	1	NA	NA	NA	NA	NA	NA
Niddam et al,[48] France, Cohort†	Low	Serious	50	NA	Median 18.3 [6–34] NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Paget et al,[49] UK, Case series‖	Low	Serious	10	NA	NA	NA	0	0	0	0	0	0	5.7 [5–7]	NA	NA	£7,628 ± £754 [£6,324.06– £8,332.68]
Paik et al,[51] South Korea, Cohort†	Moderate	Moderate	217	NA	NA	51	8	NA	NA	0	NA	17	NA	0.17	NA	NA
Schaverien et al,[55] UK, Cohort†	Low	Serious	26	NA	14 [6–20]	6	NA	NA	NA	2	0	NA	7.4 ± 3.7	0.1	NA	NA
Scheer et al,[54] Canada, Cohort†	Low	Serious	52	NA	NA	45	24	3	1	3	2	6	NA	NA	NA	NA
Tan et al,[53] Singapore, Cohort†	Low	Serious	16	NA	24	5	4	NA	NA	NA	NA	1	7.56 [5–10]	0.06	NA	$8,864.67
Tønseth et al,[31] Norway, Cohort‖	Moderate	Moderate	29	21	30±12 cw 33.6±12	9 cw 0	NA	NA	NA	4	NA	NA	NA	0.26 cw NA	NA	NA
Wang et al,[52] Taiwan, Cohort‖	Moderate	Moderate	NA	30	21.5 [6–40]	4	NA	NA	NA	NA	1	NA	4.0±NA	0.13	1**	NA

[], brackets for range. GRADE, tool for grading the quality of evidence; ICER, the additional cost of obtaining 1 year of perfect breast-related health for DIEP cw IBR; ROBINS-I, tool for assessing the risk of bias.
*Capsular contracture, scarring, implant deflation/rupture/displacement.
†Retrospective.
‡Statistically significant (P < 0.01).
§Statistically significant (P < 0.05).
¶Statistical significance not reported.
‖Prospective.
** Grade IV contracture (Baker's classification); ± SD.
F/u, follow up; Comps, complications; cw, compared with; GRADE, Grading of Recommendations, Assessment, Development and Evaluation; NA, not applicable/available; NS, no significance; pts, patients; UL, unilateral.

Table 2. Comparative Studies Evaluating PROs for DIEP versus IBR Reconstruction

Reference Study, Location, Design	No. Pts (DIEP)	No. Pts (IBR)	Mean F/u (mo) with SD Where Reported	PROs
Matros et al,[15] USA, cohort*	103	172	NA	BREAST-Q scores consistently higher for DIEP, 1–8 y postoperatively† Breast QALY: 19.5 cw 17.7†
Tønseth et al,[31] Norway, Cohort‡	29	21	30±12 cw 33.6±12	SF-36 scores: Physical functioning 85.0 cw 89.0 (NS); role physical 77.5 cw 78.7 (NS); bodily pain 72.9 cw 74.6 (NS); general health 78.0 cw 80.4 (NS); vitality 60.0 cw 63.8 (NS); social functioning 87.3 cw 90.0 (NS); role emotional 75.6 cw 69.8 (NS); mental health 79.6 cw 77.2 (NS) Study-specific questionnaire scores: Satisfied with appearance of breast: Yes: 24 cw 5; neither yes/no: 3 cw 8; no: 2 cw 8§ (*P* < 0.0005) Social relationship: Improved: 5 cw 0; unchanged: 24 cw 20; worse: o cw 1¶ (*P* = 0.02) Sad about body image: Yes: 3 cw 5; neither yes/no: 1 cw 6; no: 25 cw 10¶ (*P* = 0.01) Study-specific questions concerning self image (NS), social and intimate relationship (NS), general health (NS), and general satisfaction (NS) Visual Analog Scale: Breast shape: 7.9±2.2 cw 5.1±2.5§ (*P* < 0.0005) Breast symmetry: 7.6±2.1 cw 6.0±2.9¶ (*P* = 0.023) Breast volume: 7.7±2.1 cw 5.4±2.7§ (*P* = 0.006) Breast position: 8.8±1.3 cw 6.8±2.6§ (*P* = 0.003) Breast consistency: 5.6±2.9 cw 3.8±3.0§ (*P* = 0.008)

*Retrospective.
†Statistical significance not reported.
‡Prospective.
§Statistically significant (*P* < 0.01).
¶Statistically significant (*P* < 0.05).
Cw, compared with (assessing the RoB); QALY, Quality-Adjusted Life Year; NA, not available; NS, no significance.
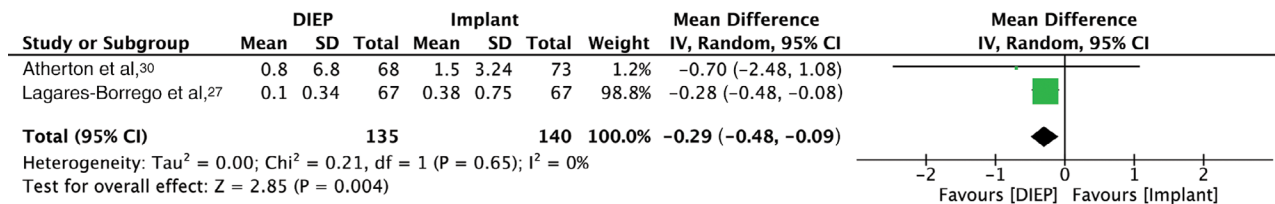
Fig. 2. Forest plot for 2 comparative studies, evaluating mean number of reoperations for complications.

studies reported EP reconstruction, and 1 reported DTI. Due to the scarcity of IBR data, further subgroup analysis was not possible. Future studies should clearly specify the type of reconstruction – DTI/EP; subpectoral or prepectoral and whether acellular dermal matrix was utilized. Adequate reporting as part of a core outcome set will facilitate inter-study comparisons and meta-analyses.

The Mastectomy Reconstruction Outcomes Consortium (MROC) is a large, multicenter prospective cohort study involving 9 academic and 2 private practices in the United States and Canada with high volumes of BRR.[59] This did not meet our systematic review's inclusion criteria, due to combined data reporting of unilateral and bilateral reconstructions, as well as reporting of clinical outcomes with combined results from a range of autologous reconstruction techniques, including DIEP, TRAM, free TRAM, and latissimus dorsi (LD) and superficial

inferior epigastric perforator flaps. Nevertheless, it is pertinent to discuss the results from this cohort.

Bennett et al[8] prospectively evaluated 2,343 patients undergoing postmastectomy autologous reconstruction (706), using DIEP, pedicled TRAM, free TRAM, superficial inferior epigastric perforator, latissimus dorsi or IBR (1,637), with comparison of 2-year complication rates. The authors found that DIEP had lower failure rates compared with IBR (1.3% versus 7.1%, *P* < 0.001) and lower odds of developing infection (OR 0.45; CI: 0.25–0.29; *P* = 0.006). This corroborates with the findings from our systematic review with lower rates of infection and revision procedures in DIEP compared with IBR. However, Bennett et al[8] reported higher odds of developing any complication with DIEP (OR 1.97; CI 1.41–2.76; *P* < 0.001), including reoperative complications (OR 2.76; CI 1.87–4.07; *P* < 0.001). This in part could be explained by outcomes following

adjuvant radiotherapy. Although the detrimental effect of PMRT on IBR is well established, the effect on autologous reconstruction is more equivocal. This is being evaluated in a separate systematic review and meta-analysis by our group (PROSPERO CRD42017077945).[38] Moreover, confounders such as the level-of-care provider expertise, nonstandardized operative technique, differences in centers' volume, and learning curves may further bias the results and their interpretability. Indeed, another MROC study evaluating hospital variations in clinical complications and PROs at 2 years post autologous BRR or IBR demonstrated that complications varied widely between hospitals.[10] It also highlighted the limitations of extrapolating single-institution level data and the challenges of evaluating hospital-based outcomes in BRR patients.

In our systematic review, out of 16 studies, only 2 comparative studies (12.5%)[15,31] reported PROs. A major paradigm shift is needed to incorporate PROs in all studies evaluating BRR, as also supported by the recent publication of the "Gap analysis" in BRR.[12] Evaluating clinical outcomes without PROs is a major drawback in evaluating outcomes in BRR, as the reconstruction must satisfy the patient with regard to physical, psychological, and sexual well-being.[59] Disregard of these domains renders outcome assessment incomplete and suboptimal.

Two comparative studies that evaluated PROs in our review favored DIEP reconstruction. Matros et al[15] utilized a robust, validated, disease-specific questionnaire, BREAST-Q. BREAST-Q scores were reported as consistently higher for DIEP compared with IBR in postoperative years 1–8, with a higher breast Quality-adjusted Life Year for DIEP. Conversely, Tønseth et al[31] used generic PRO tools, SF-36, which revealed no difference in QoL between DIEP and IBR, and Visual Analog Scale, with superior cosmetic outcome with DIEP. The study also used a nonvalidated study-specific questionnaire that demonstrated higher breast satisfaction, improved social relationship, and body image satisfaction for DIEP.

The results from our review corroborate results from Santosa et al[29] who evaluated PROs for 2,013 patients (523 autologous reconstructions; 1,490 IBR) from the MROC cohort, pre and 2 years post BRR, using the BREAST-Q. The 4 domains evaluated were as follows: satisfaction with breasts, psychosocial well-being, physical well-being, and sexual well-being. At 2 years, patients who underwent autologous reconstruction had higher breast satisfaction, higher psychosocial well-being, and sexual well-being than did those who underwent IBR.[29] Lack of a significant difference in QoL between DIEP and IBR reported by Tønseth et al in our review may be due to the small sample size in the study (n = 50) and use of a nonspecific, generic QoL tool, SF-36, which may not be sensitive enough to measure changes as a result of BRR intervention or to capture all aspects of outcome specific to breast surgery.[18] Moreover, as purported by our group, QoL domains should be defined a priori, facilitating estimations of potential effect size.[17]

Three comparative studies evaluated cost, all favoring DIEP.[15,27,30] Two studies were conducted in a universal health care system (UK and Spain)[27,30] and 1 was conducted in a health insurance-based model (USA),[15]

making direct comparisons on cost difficult. This is exacerbated by only 1 study performing robust cost-effectiveness analysis, calculating an ICER of $11,491 for DIEP.[15] An ICER is the additional cost for DIEP to obtain 1 year of perfect breast-related QoL compared with IBR; a threshold of $50,000–$100,000 for a year in perfect overall health has been deemed as acceptable for the adoption of new technologies or techniques in developed countries.[60] Heterogeneity in cost-evaluation methods and reporting prevented the calculation of an overall cost-effectiveness summary measure in our systematic review.

CONCLUSIONS

Limitations in study design and outcome reporting preclude firm consensus on best recommendations for postmastectomy BRR. However, the evidence supports a weak recommendation for DIEP reconstruction being more cost-effective and yielding higher QoL compared with IBR. There is a pressing need for level I and II data, in the form of RCTs and prospective, multicenter, longitudinal cohort studies, with long-term follow-up. These must incorporate validated, disease-specific PRO tools such as BREAST-Q. Evaluation of a priori core outcome set and cost-effectiveness is required for national guidelines, optimizing informed consent and facilitating clinician–patient-shared decision-making.

Ankur Khajuria, BSc (Hons), MD, FHEA, FRSPH,
MRCS (Eng), MAcadMEd
Department of Continuing Education
Kellogg College
University of Oxford
Oxford, 60-62
Banbury Road, OX1 2JA
E-mail: ankur.khajuria09@imperial.ac.uk

REFERENCES

1. Ginsburg O, Bray F, Coleman MP, et al. The global burden of women's cancers: a grand challenge in global health. *Lancet.* 2017;389:847–860.
2. Winters S, Martin C, Murphy D, et al. Breast cancer epidemiology, prevention, and screening. *Prog Mol Biol Transl Sci.* 2017;151:1–32.
3. Veronesi U, Cascinelli N, Mariani L, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *N Engl J Med.* 2002;347:1227–1232.
4. Helms RL, O'Hea EL, Corso M. Body image issues in women with breast cancer. *Psychol Health Med.* 2008;13:313–325.
5. Klassen AF, Pusic AL, Scott A, et al. Satisfaction and quality of life in women who undergo breast surgery: a qualitative study. *BMC Womens Health.* 2009;9:11.
6. Atisha DM, Tessiatore KM, Rushing CN, et al. A national snapshot of patient-reported outcomes comparing types of abdominal flaps for breast reconstruction. *Plast Reconstr Surg.* 2019;143:667–677.
7. Atisha D, Alderman AK, Lowery JC, et al. Prospective analysis of long-term psychosocial outcomes in breast reconstruction: two-year postoperative results from the Michigan breast reconstruction outcomes study. *Ann Surg.* 2008;247:1019–1028.
8. Bennett KG, Qi J, Kim HM, et al. Comparison of 2-year complication rates among common techniques for postmastectomy breast reconstruction. *JAMA Surg.* 2018;153:901–908.

9. Winters ZE, Balta V, Thomson HJ, et al; European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Group. Phase III development of the European organization for research and treatment of cancer quality of life questionnaire module for women undergoing breast reconstruction. *Br J Surg.* 2014;101:371–382.

10. Berlin NL, Tandon VJ, Qi J, et al. Hospital variations in clinical complications and patient-reported outcomes at 2 years after immediate breast reconstruction. *Ann Surg.* 2019;269:959–965.

11. Cohen WA, Mundy LR, Ballard TN, et al. The BREAST-Q in surgical research: a review of the literature 2009-2015. *J Plast Reconstr Aesthet Surg.* 2016;69:149–162.

12. Cutress RI, McIntosh SA, Potter S, et al; Association of Breast Surgery Surgical Gap Analysis Working Group. Opportunities and priorities for breast surgical research. *Lancet Oncol.* 2018;19:e521–e533.

13. Erdmann-Sager J, Wilkins EG, Pusic AL, et al. Complications and patient-reported outcomes after abdominally based breast reconstruction: results of the mastectomy reconstruction outcomes consortium study. *Plast Reconstr Surg.* 2018;141:271–281.

14. Howard MA, Sisco M, Yao K, et al. Patient satisfaction with nipple-sparing mastectomy: a prospective study of patient reported outcomes using the BREAST-Q. *J Surg Oncol.* 2016;114:416–422.

15. Matros E, Albornoz CR, Razdan SN, et al. Cost-effectiveness analysis of implants versus autologous perforator flaps using the BREAST-Q. *Plast Reconstr Surg.* 2015;135:937–946.

16. Macadam SA, Zhong T, Weichman K, et al. Quality of life and patient-reported outcomes in breast cancer survivors: a multicenter comparison of four abdominally based autologous reconstruction methods. *Plast Reconstr Surg.* 2016;137:758–771.

17. Winters ZE, Khajuria A. Quality of life after breast reconstruction-the BRIOS study. *Lancet Oncol.* 2018;19:e579.

18. Pusic AL, Klassen AF, Scott AM, et al. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg.* 2009;124:345–353.

19. Cano SJ, Klassen AF, Scott AM, et al. The BREAST-Q: further validation in independent clinical samples. *Plast Reconstr Surg.* 2012;129:293–302.

20. Winters ZE, Afzal M, Rutherford C, et al; European Organisation for Research and Treatment of Cancer Quality of Life Group. International validation of the European organisation for research and treatment of cancer QLQ-BRECON23 quality-of-life questionnaire for women undergoing breast reconstruction. *Br J Surg.* 2018;105:209–222.

21. Thomson HJ, Winters ZE, Brandberg Y, et al. The early development phases of a European Organisation for Research and Treatment of Cancer (EORTC) module to assess patient reported outcomes (pros) in women undergoing breast reconstruction. *Eur J Cancer.* 2013;49:1018–1026.

22. Sisco M, Du H, Warner JP, et al. Have we expanded the equitable delivery of postmastectomy breast reconstruction in the new millennium? Evidence from the national cancer data base. *J Am Coll Surg.* 2012;215:658–666; discussion 666.

23. Eltahir Y, Werners LL, Dreise MM, et al. Quality-of-life outcomes between mastectomy alone and breast reconstruction: comparison of patient-reported BREAST-Q and other health-related quality-of-life measures. *Plast Reconstr Surg.* 2013;132:201e–209e.

24. Guyomard V, Leinster S, Wilkinson M. Systematic review of studies of patients' satisfaction with breast reconstruction after mastectomy. *Breast.* 2007;16:547–567.

25. Dean C, Chetty U, Forrest AP. Effects of immediate breast reconstruction on psychosocial morbidity after mastectomy. *Lancet.* 1983;1:459–462.

26. Rowland JH, Holland JC, Chaglassian T, et al. Psychological response to breast reconstruction. Expectations for and impact on postmastectomy functioning. *Psychosomatics.* 1993;34:241–250.

27. Lagares-Borrego A, Gacto-Sanchez P, Infante-Cossio P, et al. A comparison of long-term cost and clinical outcomes between the two-stage sequence expander/prosthesis and autologous deep inferior epigastric flap methods for breast reconstruction in a public hospital. *J Plast Reconstr Aesthet Surg.* 2016;69:196–205.

28. Gu J, Groot G, Boden C, et al. Review of factors influencing women's choice of mastectomy versus breast conserving therapy in early stage breast cancer: a systematic review. *Clin Breast Cancer.* 2018;18:e539–e554.

29. Santosa KB, Qi J, Kim HM, et al. Long-term patient-reported outcomes in postmastectomy breast reconstruction. *JAMA Surg.* 2018;153:891–899.

30. Atherton DD, Hills AJ, Moradi P, et al. The economic viability of breast reconstruction in the UK: comparison of a single surgeon's experience of implant; LD; TRAM and DIEP based reconstructions in 274 patients. *J Plast Reconstr Aesthet Surg.* 2011;64:710–715.

31. Tønseth KA, Hokland BM, Tindholdt TT, et al. Quality of life, patient satisfaction and cosmetic outcome after breast reconstruction using DIEP flap or expandable breast implant. *J Plast Reconstr Aesthet Surg.* 2008;61:1188–1194.

32. Agha RA, Fowler AJ, Herlin C, et al. Use of autologous fat grafting for breast reconstruction: a systematic review with meta-analysis of oncological outcomes. *J Plast Reconstr Aesthet Surg.* 2015;68:143–161.

33. Steiert AE, Boyce M, Sorg H. Capsular contracture by silicone breast implants: possible causes, biocompatibility, and prophylactic strategies. *Med Devices (Auckl).* 2013;6:211–218.

34. Spear SL, Murphy DK; Allergan Silicone Breast Implant U.S. Core Clinical Study Group. Natrelle round silicone breast implants: core study results at 10 years. *Plast Reconstr Surg.* 2014;133:1354–1361.

35. Bajaj AK, Chevray PM, Chang DW. Comparison of donor-site complications and functional outcomes in free muscle-sparing TRAM flap and free DIEP flap breast reconstruction. *Plast Reconstr Surg.* 2006;117:737–746; discussion 747.

36. Khajuria A, Smith O, Mosahebi A. A systematic review and meta-analysis of the clinical outcomes and cost of using autologous deep inferior epigastric perforator (DIEP) flap versus implants for breast reconstruction. http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017072557. Accessed June 2, 2019.

37. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 2009;6:e1000100.

38. Khajuria A, Winters Z, Mosahebi A. A systematic review and meta-analysis of clinical and patient-reported outcomes (PROs) of immediate versus delayed autologous abdominal-based flap breast reconstruction in the context of post-mastectomy radiotherapy [PROSPERO CRD42017077945]. http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017077945.

39. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ.* 2016;355:i4919.

40. Atkins D, Best D, Briss PA, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ.* 2004;328:1490.

41. Collaboration TC. *Review Manager (RevMan) Version 5.2.* Copenhagen, Denmark: The Nordic Cochrane Centre; 2012. https://community.cochrane.org/help/tools-and-software/revman-5.

42. Higgins JP, Altman DG, Gøtzsche PC, et al; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 2011;343:d5928.

43. DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. *Contemp Clin Trials.* 2015;45(Pt A):139–145.

44. Cheng MH, Lin JY, Ulusal BG, et al. Comparisons of resource costs and success rates between immediate and delayed breast reconstruction using DIEP or SIEA flaps under a well-controlled clinical trial. *Plast Reconstr Surg.* 2006;117:2139–2142; discussion 2143.

45. Kroll SS, Reece GP, Miller MJ, et al. Comparison of cost for DIEP and free TRAM flap breast reconstructions. *Plast Reconstr Surg.* 2001;107:1413–1416; discussion 1417.

46. McGeorge DD, Mahdi S, Tsekouras A. Breast reconstruction with anatomical expanders and implants: our early experience. *Br J Plast Surg.* 1996;49:352–357.

47. Nahabedian MY, Tsangaris T, Momen B. Breast reconstruction with the DIEP flap or the muscle-sparing (MS-2) free TRAM flap: is there a difference? *Plast Reconstr Surg.* 2005;115:436–444; discussion 445–446.

48. Niddam J, Bosc R, Lange F, et al. DIEP flap for breast reconstruction: retrospective evaluation of patient satisfaction on abdominal results. *J Plast Reconstr Aesthet Surg.* 2014;67:789–796.

49. Paget JT, Young KC, Wilson SM. Accurately costing unilateral delayed DIEP flap breast reconstruction. *J Plast Reconstr Aesthet Surg.* 2013;66:926–930.

50. Moradi P, Durrant C, Glass GE, et al. SIEA flap leads to an increase in abdominal seroma rates compared to DIEP flap for breast reconstruction. *Eur J Plast Surg.* 2010;34:87–91.

51. Paik JM, Lee KT, Jeon BJ, et al. Donor site morbidity following DIEP flap for breast reconstruction in Asian patients: is it different? *Microsurgery.* 2015;35:596–602.

52. Wang HY, Ali RS, Chen SC, et al. One-stage immediate breast reconstruction with implant following skin-sparing mastectomy in Asian patients. *Ann Plast Surg.* 2008;60:362–366.

53. Tan S, Lim J, Yek J, et al. The deep inferior epigastric perforator and pedicled transverse rectus abdominis myocutaneous flap in breast reconstruction: a comparative study. *Arch Plast Surg.* 2013;40:187–191.

54. Scheer AS, Novak CB, Neligan PC, et al. Complications associated with breast reconstruction using a perforator flap compared with a free TRAM flap. *Ann Plast Surg.* 2006;56:355–358.

55. Schaverien MV, Perks AG, McCulley SJ. Comparison of outcomes and donor-site morbidity in unilateral free TRAM versus DIEP flap breast reconstruction. *J Plast Reconstr Aesthet Surg.* 2007;60:1219–1224.

56. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg.* 2004;240:205–213.

57. Spear SL, Baker JL Jr. Classification of capsular contracture after prosthetic breast reconstruction. *Plast Reconstr Surg.* 1995;96:1119–1123; discussion 1124.

58. Potter S, Brigic A, Whiting PF, et al. Reporting clinical outcomes of breast reconstruction: a systematic review. *J Natl Cancer Inst.* 2011;103:31–46.

59. Pusic AL, Matros E, Fine N, et al. Patient-reported outcomes 1 year after immediate breast reconstruction: results of the mastectomy reconstruction outcomes consortium study. *J Clin Oncol.* 2017;35:2499–2506.

60. Laupacis A, Feeny D, Detsky AS, et al. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *CMAJ.* 1992;146:473–481.

Immediate and delayed autologous abdominal microvascular flap breast reconstruction in patients receiving adjuvant, neoadjuvant or no radiotherapy: a meta-analysis of clinical and quality-of-life outcomes

A. Khajuria[1,3] , **W. N. Charles[3]**, **M. Prokopenko[4]**, **A. Beswick[6]**, **A. L. Pusic[7]**, **A. Mosahebi[4]**,
D. J. Dodwell[2] and **Z. E. Winters[5]**

[1]Kellogg College, Nuffield Department of Surgery, and [2]Nuffield Department of Population Health, University of Oxford, Oxford, [3]Department of Surgery and Cancer, Imperial College London, [4]Department of Plastic Surgery, Royal Free Hospital, [5]Surgical Intervention Trials Unit, Division of Surgery and Interventional Science, University College London, London, and [6]School of Clinical Sciences, University of Bristol, Bristol, UK, and [7]Patient-Reported Outcomes, Value and Experience Centre, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
Correspondence to: Mr A. Khajuria, Department of Surgery and Cancer, Imperial College London, Praed Street, London W2 1NY, UK (e-mail: ak8609@imperial.ac.uk)

Background: Effects of postmastectomy radiotherapy (PMRT) on autologous breast reconstruction (BRR) are controversial regarding surgical complications, cosmetic appearance and quality of life (QOL). This systematic review evaluated these outcomes after abdominal free flap reconstruction in patients undergoing postoperative adjuvant radiotherapy (PMRT), preoperative radiotherapy (neoadjuvant radiotherapy) and no radiotherapy, aiming to establish evidence-based optimal timings for radiotherapy and BRR to guide contemporary management.
Methods: The study was registered on PROSPERO (CRD42017077945). Embase, MEDLINE, Google Scholar, CENTRAL, Science Citation Index and ClinicalTrials.gov were searched (January 2000 to August 2018). Study quality and risk of bias were assessed using GRADE and Cochrane's ROBINS-I respectively.
Results: Some 12 studies were identified, involving 1756 patients (350 PMRT, 683 no radiotherapy and 723 neoadjuvant radiotherapy), with a mean follow-up of 27·1 (range 12·0–54·0) months for those having PMRT, 16·8 (1·0–50·3) months for neoadjuvant radiotherapy, and 18·3 (1·0–48·7) months for no radiotherapy. Three prospective and nine retrospective cohorts were included. There were no randomized studies. Five comparative radiotherapy studies evaluated PMRT and four assessed neoadjuvant radiotherapy. Studies were of low quality, with moderate to serious risk of bias. Severe complications were similar between the groups: PMRT *versus* no radiotherapy (92 *versus* 141 patients respectively; odds ratio (OR) 2·35, 95 per cent c.i. 0·63 to 8·81, $P = 0·200$); neoadjuvant radiotherapy *versus* no radiotherapy (180 *versus* 392 patients; OR 1·24, 0·76 to 2·04, $P = 0·390$); and combined PMRT plus neoadjuvant radiotherapy *versus* no radiotherapy (272 *versus* 453 patients; OR 1·38, 0·83 to 2·32, $P = 0·220$). QOL and cosmetic studies used inconsistent methodologies.
Conclusion: Evidence is conflicting and study quality was poor, limiting recommendations for the timing of autologous BRR and radiotherapy. The impact of PMRT and neoadjuvant radiotherapy appeared to be similar.

Introduction

Breast cancer is the commonest malignancy and leading cause of cancer-related mortality in women[1,2].

Breast-conserving surgery (BCS) with radiotherapy or mastectomy are recommended treatments, with comparable oncological outcomes[3,4]. Autologous abdominal-based

free flap and implant-based procedures are the approaches used most frequently in immediate breast reconstruction (BRR)[5]. Autologous BRR has the inherent advantage of using the patient's own tissues, taken from a different part of the body where there is excess fat and skin, to restore breast volume and appearance after mastectomy. Various donor sites can be used, most commonly the abdomen[6].

Adjuvant locoregional postmastectomy radiotherapy (PMRT) of the chest wall, and potentially of the regional lymph nodes, has been indicated historically for locally advanced disease[7,8]. These indications increased following the Early Breast Cancer Trialists' Collaborative Group[9] meta-analyses, which showed significantly improved disease-free and overall survival after PMRT and regional node irradiation in women at intermediate risk (tumour size 50 mm or less and 1–3 positive lymph nodes)[10]. Newly proposed US guidelines[11] emphasize the need to consider the lower recurrence rates associated with contemporary practice and the benefits of systemic therapy[12]. Current recommendations for PMRT in the intermediate-risk group remain controversial, pending the results of the SUPREMO (Selective Use of Postoperative Radiotherapy aftEr MastectOmy) trial, evaluating chest wall and/or axillary radiotherapy[13,14].

Adjuvant radiotherapy (PMRT) may have deleterious effects on breast cosmetic outcomes, quality of life (QOL) and surgical complications after immediate BRR[15]. Previous studies evaluating the impact of PMRT on types of immediate BRR showed its potential feasibility in this setting, with lower morbidity rates compared with those of implant-based procedures[5,16–18]. Surprisingly, the rapid adoption of immediate implant-based reconstruction in about 70 per cent of women, compared with 34 per cent of autologous procedures when PMRT is recommended, may be influenced by surgeon and patient preferences, regardless of current evidence[15,17,19].

Increasing recommendations for PMRT and immediate BRR have prompted a need to consider their optimal sequence. Previous systematic reviews have not provided clarity concerning the choice between immediate and delayed BRR[9]. Despite this, immediate autologous BRR is commonly recommended in the setting of PMRT, given the potential long-term benefits on patients' QOL and breast cosmetic satisfaction[20,21]. Currently, immediate autologous BRR and PMRT recommendations are variable[22,23]. A systematic review[24] in 2011 showed methodological variations in the definitions of surgical complications, precluding interstudy comparisons.

Complications of autologous breast reconstruction with PMRT include: poor wound-healing, flap-related fat necrosis, fibrosis and contracture, which reduce breast volume[5]. Surgical complications contribute variably to decreased patient satisfaction and impaired cosmetic outcomes[5]. A standardized core set of outcomes for BRR has been proposed[25] involving a range of complications, including flap-related complications and the need for further unplanned surgery. The BRR core outcome set has yet to recommend a standardized measurement tool for evaluating surgical complications. Most surgeons use the Clavien–Dindo classification (CDC)[26]. Patient-reported QOL outcomes using validated BRR questionnaires, such as the BREAST-Q and the European Organisation for Research and Treatment of Cancer (EORTC) Quality-of-Life Questionnaire (QLQ)-BRECON23, are recommended to evaluate comparative effectiveness[20,27–32].

This systematic review aimed to evaluate the quality and strengths of the current evidence regarding surgical complications in autologous abdominal flaps in the context of the receipt and timing of radiotherapy related to PMRT[5,6] and, less commonly, neoadjuvant radiotherapy, generally administered before skin-sparing mastectomy and immediate breast reconstruction[33], including assessment of QOL[34].

Methods

The protocol was registered and published on the Prospective Register of Systematic Reviews PROSPERO (CRD42017077945)[35]. The authors adhered to the PRISMA statement[36].

Search strategies

A comprehensive search of the MEDLINE (Ovid SP), Embase (Ovid SP), Google Scholar, Cochrane Controlled Register of Trials (CENTRAL), Science citation index databases and ClinicalTrials.gov (January 2000 to August 2018) was conducted, identifying the relevant studies. Combinations of Medical Subject Headings (MeSH) terms and free text were used, including Boolean logical operators for the search strategy. References of included articles were also screened for their relevance. The example of an Embase (Ovid SP) search strategy was adopted for other databases (*Appendix S1*, supporting information).

Identification and selection of studies

Database-related searches were entered into an EndNote™ X8 library (Clarivate Analytics, Philadelphia, Pennsylvania, USA). Study screening was performed independently in two stages by two investigators using prespecified screening criteria.

In stage 1, two authors independently screened titles and abstracts. Discrepancies were resolved by consensus with the senior author. Remaining doubts regarding an article resulted in a review of the complete publication.

In stage 2, full-text studies from stage 1 were screened independently for their eligibility by two reviewers. Discrepancies were resolved by consensus with a third reviewer. Authors of eligible studies were contacted (via e-mail) to reconcile any methodological issues or to provide more detailed information on data for individual types of autologous flap.

Study design

All primary human studies evaluating surgical complications for autologous free flap (microvascular) abdominal BRR in breast cancer and types of radiotherapy (PMRT, neoadjuvant and no radiotherapy) were included. Outcomes also included patient-reported QOL and cosmetic assessments. Radiotherapy groups were compared with a control or no radiotherapy group in comparative studies, compatible with immediate and delayed BRR. Commonly performed autologous abdominal flaps included: deep inferior epigastric perforator (DIEP), transverse rectus abdominis myocutaneous (TRAM) and the superficial inferior epigastric artery perforator (SIEA)[6].

Inclusion criteria

Inclusion criteria were: women aged at least 18 years with a diagnosis of invasive breast cancer (TNM categories: T0–3, N1–3, Mx, M0), undergoing immediate or delayed abdominal autologous BRR using free flaps (DIEP, TRAM or SIEA) who received adjuvant radiotherapy (PMRT), neoadjuvant radiotherapy or no radiotherapy.

Clinical studies that involved at least 50 patients were included (RCTs, prospective and retrospective comparative observational studies, and case series).

Exclusion criteria

Review articles, conference abstracts, simulation studies and clinical studies in non-human subjects were not included, along with studies involving patients who received segmental or partial mastectomy, technical descriptions of operative repair with no outcome measures, BRR unrelated to breast cancer, implant-based reconstructions and other non-abdominal autologous flaps.

Risk of bias and quality of studies

Cochrane's ROBINS-I (Risk Of Bias In Non-randomised Studies – of Interventions) tool was used for comparative studies[37]. This comprises seven domains from which the risk of bias may be ascertained to produce an overall risk-of-bias score[37]. The Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) tool[38] was used to evaluate the methodological quality of individual studies.

Study outcomes

Primary outcomes were surgical complications including: Clavien–Dindo classification (CDC) grades II and III[26]; partial flap loss; total flap loss; fat necrosis (CDC grades, when reported)[39]; number(s) of unplanned reoperations for surgical complications (excluding cosmetic revisions); and number(s) of total complications. A surgical complication was defined as an adverse, postoperative, surgery-related event that required additional treatment[16]. If CDC grades were not defined, the complications reported by the included studies were graded retrospectively according to the CDC by two independent authors; any discrepancy was discussed and agreed with the senior author.

Secondary outcomes were assessed using patient-reported QOL-validated questionnaires (COnsensus-based Standards for the Selection of health Measurement INstruments (COSMIN)[40,41], Breast Questionnaire (BREAST-Q), the EORTC Quality-of-Life Questionnaire (QLQ) – Breast Cancer 23[42], the Quality-of-Life Cancer Generic Questionnaire (QLQ-C30)[43], the Numerical Pain Rating Scale (NPRS)[44,45], the Patient-Reported Outcomes Measurement Information System – Profile 29 (PROMIS-29)[46], the McGill Pain Questionnaire (MPQ)[47], the Generalized Anxiety Disorder Scale (GAD-7)[48] and the Patient Health Questionnaire (PHQ-9)[49]), as well as assessment of cosmetic outcomes using independent panel or self assessments of medical photographs, and surface imaging using the Vectra® XT three-dimensional system[50] (Canfield Scientific, Parsippany, New Jersey, USA).

Data extraction, collection and management

Two authors independently extracted data from full-text articles using a standard data form. Any discrepancies were resolved by consensus with a third reviewer. Reporting authors of original articles were contacted on up to two occasions relating to missing data or where additional information was required.

Data extraction included: first author, year of publication, study design, study setting, number of centres, duration of follow-up, study population and participant demographics (mean age, BMI, smoking, co-morbidities).

Fig. 1 PRISMA diagram for the review
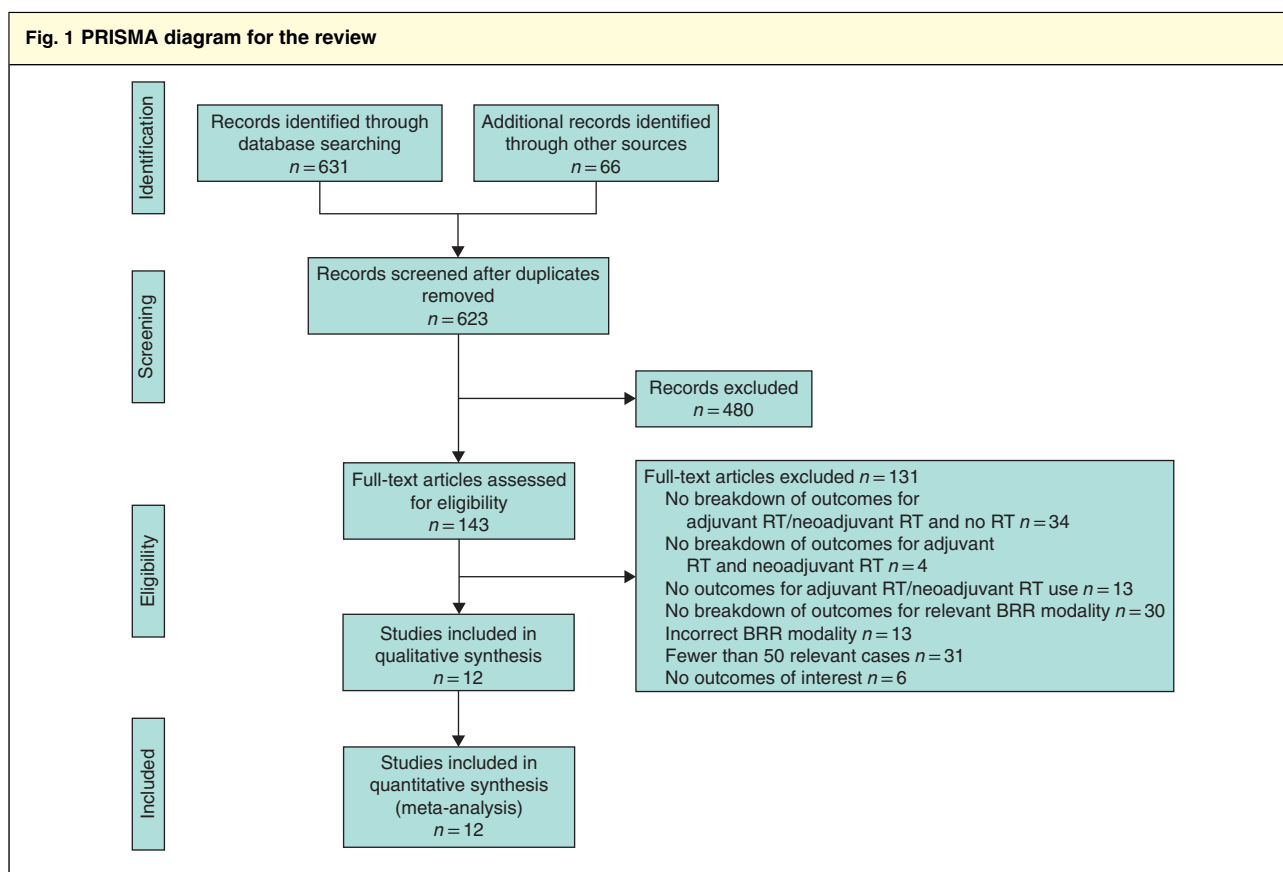
RT, radiotherapy; BRR, breast reconstruction.

Surgical complications were recorded using CDC: grades II–III[26]. Two authors reviewed eligible studies and classified each complication according to the CDC[26] if unreported.

QOL and cosmetic outcomes were listed.

Statistical analysis

When two or more studies reported outcome data, these were pooled using Review Manager 5.3 software (The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark). Odds ratios with 95 per cent confidence intervals were used to evaluate dichotomous outcomes (surgical complications). Standard mean differences (with 95 per cent c.i.) were used for continuous outcomes between treatment groups. Rates of each complication (fat necrosis, partial and total flap loss, infection and wound complications (dehiscence and delayed wound healing)) were compared for PMRT (*versus* no radiotherapy) and neoadjuvant radiotherapy (*versus* no radiotherapy). Data were also pooled to provide an overall summary measure

of combined radiotherapy (adjuvant and neoadjuvant) compared with no radiotherapy.

Heterogeneity between studies[51] was assessed in Review Manager 5.3 using the Higgins and Thompson I^2 statistic[52]. Levels of heterogeneity were defined as: low (I^2 less than 50 per cent), moderate ($I^2 = 50$–80 per cent) and high (I^2 above 80 per cent). A random-effects model was used for cohorts with heterogeneity (I^2 above 50 per cent)[53]. As heterogeneity was generally moderate or high, and outcome measures differed between studies, these were combined using the DerSimonian and Laird random-effects model. Results of meta-analyses are shown as forest plots. A sensitivity analysis was performed where possible, to evaluate whether outcomes differed when restricting the analysis exclusively to high-quality studies.

Clinically meaningful differences in QOL items/ questions or domain scores may vary depending on response shift, that is a change in the meaning of QOL scores over time[54]. This is relevant in longitudinal studies and may influence clinical significance, defined as greater than 5-point score differences for EORTC QLQ-C30

Table 1 Study summaries: comparative adjuvant or neoadjuvant radiotherapy in autologous breast reconstruction, and non-comparative studies (adjuvant radiotherapy or neoadjuvant radiotherapy only)

Reference	Years	Country	No. of centres	Type of BRR flap	Overall follow-up (months)	Group differences in baseline characteristics¶	RT dose and regimen
Baumann *et al.*[69]‡	2005–2009	USA	1	msTRAM; DIEP; SIEA	11*	n.a.	Total 60 Gy; missing details
Billig *et al.*[62]§	2012–2017	USA and Canada	11	TRAM; DIEP; SIEA	24	Adjuvant RT: more non-Hispanic patients ($P = 0.001$), bilateral BRR ($P = 0.002$), DIEP/SIEA ($P < 0.001$), adjuvant chemotherapy ($P < 0.001$); less TRAM ($P < 0.001$)#	Total 50.4 Gy over 4 weeks, daily (28 fractions of 1.8 Gy)
Chatterjee *et al.*[59]§	1995–2005	UK	1	DIEP	42 (12–120)†	Adjuvant RT: more IDC ($P = 0.02$), LVI ($P = 0.044$), positive axillary LN ($P < 0.001$)	Total 45 Gy over 4 weeks (20 fractions)
Cooke *et al.*[60]§	2012–2015	Canada	1	DIEP; SIEA	12	Adjuvant RT: higher TNM staging, positive LN, more chemotherapy (P values not provided)	Total 50/50.4 Gy over 4 weeks, daily (25 fractions of 2 Gy/28 fractions of 1.8 Gy)
Huang *et al.*[63]‡	1997–2001	Taiwan	1	TRAM	40 (24–74)†	n.a.	Total 50 Gy; missing details
Levine *et al.*[67]‡	1999–2011	USA	1	msTRAM; DIEP; SIEA	22.7*	n.a.	Missing details
Modarressi *et al.*[64]‡	2007–2013	Switzerland	1	DIEP	1	n.a.	Missing details
Mull *et al.*[65]‡	2003–2014	USA	1	msTRAM; TRAM; DIEP	1	Neoadjuvant RT: more chemotherapy ($P < 0.01$), higher TNM staging ($P < 0.01$); less hypertension/CAD ($P = 0.03$)	Missing details
O'Connell *et al.*[58]‡	2009–2014	UK	1	DIEP	44.3 (i.q.r. 31.1–56.4)†	Adjuvant and neoadjuvant RT: more chemotherapy and endocrine therapy as less DCIS/less advanced invasive disease (P values not provided)	Total 40 Gy over 3 weeks (15 fractions)
Peeters *et al.*[66]‡	1997–2003	Belgium	2	DIEP	≥ 12	n.a.	Total 50 Gy; missing details
Rogers and Allen[61]‡	1994–1999	USA	1	DIEP	18.7*	n.a.	Total 50.5 Gy over 6.5 weeks (missing details)
Temple *et al.*[68]‡	1990–2001	USA	1	TRAM	≥ 12	n.a.	Total 58 Gy; missing details

Values are *mean and †median (range), unless indicated otherwise. ‡Retrospective study; §prospective study. ¶Radiotherapy (RT) *versus* no RT, except #group difference values are for adjuvant RT *versus* neoadjuvant RT. BRR, breast reconstruction; (ms)TRAM, (muscle-sparing) transverse rectus abdominis myocutaneous; DIEP, deep inferior epigastric artery perforator; SIEA, superficial inferior epigastric artery perforator; IDC, invasive ductal carcinoma; LVI, lymphovascular invasion; LN, lymph node; n.a., not applicable/available; CAD, coronary artery disease; DCIS, ductal carcinoma *in situ*.

and QLQ-BR23[42,43,54]. Clinically meaningful differences are currently being evaluated using a number of methods such as qualitative interviews and using predefined clinical anchors[55]. Clinically meaningful differences in QOL should be differentiated from statistical significance[55]. BREAST-Q findings have been compared with large population-derived normative data, facilitating clinically meaningful interpretation of data[56,57].

Table 2 Surgical complications: immediate autologous breast reconstruction and adjuvant radiotherapy including non-comparative studies (adjuvant radiotherapy only)

Reference	GRADE	ROBINS-I	No. of patients		Follow-up (months)		Total no. of complications		No. of reoperations for complications	
			Adjuvant RT	No adjuvant RT	Adjuvant RT	No adjuvant RT	Adjuvant RT	No adjuvant RT	Adjuvant RT	No adjuvant RT
Chatterjee et al.[59]	Low	Serious	22	46	54*	36*	n.a.	n.a.	n.a.	n.a.
Cooke et al.[60]	Moderate	Moderate	64	61	12	12	20	16	6	1
O'Connell et al.[58]	Low	Serious	28	80	27·5*	48·7*	11	20	4	8
Peeters et al.[66]	Low	Serious	16	109	≥ 12	≥ 12	n.a.	n.a.	n.a.	n.a.
Rogers and Allen[61]	Low	Serious	30	30	19·9	17·4	65	41	32	26
Billig et al.[62]	Moderate	Moderate	108	n.a.	24	n.a.	81	n.a.	5	n.a.
Huang et al.[63]	Low	Serious	82	n.a.	40*	n.a.	131	n.a.	5	n.a.

*Values are median. GRADE, Grading of Recommendation, Assessment, Development, and Evaluation (tool for grading the quality of evidence); ROBINS-I, Risk Of Bias In Non-randomised Studies – of Interventions (tool for assessing risk of bias); RT, radiotherapy; n.a., not applicable/available.

Table 3 Clavien–Dindo classification of surgical complications: immediate autologous breast reconstruction and adjuvant radiotherapy including non-comparative studies (adjuvant radiotherapy only)

Reference	Adjuvant RT versus no adjuvant RT						
	Total flap loss	Partial flap loss*	Fat necrosis*	Wound dehiscence and delayed wound healing*	Clavien-Dindo complication grade†		
					II	IIIa	IIIb
Chatterjee et al.[59]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Cooke et al.[60]	0 versus 0	9 versus 6	2 versus 1	3 versus 5	2 versus 4	n.a.	6 versus 1
O'Connell et al.[58]	0 versus 0	0 versus 0	1 versus 2	4 versus 9	3 versus 3	3 versus 3	1 versus 5
Peeters et al.[66]	n.a.	n.a.	6 versus 36	n.a.	n.a.	n.a.	n.a.
Rogers and Allen[61]	n.a.	n.a.	7 versus 0‡	11 versus 8	5 versus 7	7 versus 0	25 versus 26
Billig et al.[62]	0 versus n.a.	n.a.	4 versus n.a.	17 versus n.a.	8 versus n.a.	n.a.	5 versus n.a.
Huang et al.[63]	0 versus n.a.	n.a.	7 versus n.a.	n.a.	82 versus n.a.	5 versus n.a.	n.a.

*Complication grades were not always defined or classified. †Grade II, complications requiring pharmacological treatment with drugs other than those allowed for grade I complications (drugs other than antiemetics, antipyretics, analgesics, diuretics and electrolytes); grade IIIa, complications requiring surgical intervention not under general anaesthesia; grade IIIb, complications requiring surgical intervention under general anaesthesia. RT, radiotherapy; n.a. not applicable/available. ‡$P < 0.050$.

Results

A total of 697 studies were identified. Of these, 12 studies[58−69] (including 1756 patients) evaluated adjuvant radiotherapy (350 patients), neoadjuvant radiotherapy (723) and no radiotherapy (683) (*Fig. 1*). There were three prospective study designs[59,60,62] and nine that were retrospective[58,61,63−69], but no RCTs. There were two multicentre (1 prospective[62] and 1 retrospective[66]) and ten single-centre studies (2 prospective[59,60] and 8 retrospective[58,61,63−65,67−69]) (*Table 1*). Study quality (GRADE) was low in eight studies[58,59,61,63−66,68] and moderate in the other four[60,62,67,69], with an overall high risk of bias. A summary of baseline characteristics, including numbers of centres, country of origin, dates, patient numbers, breast cancer pathology and adjuvant medical treatments in comparative adjuvant and neoadjuvant radiotherapy groups, including non-comparative studies, is provided in *Table S1* (supporting information).

Clinical outcomes (*Tables 2–5*)

No study prospectively graded surgical complications according to an accepted classification such as CDC (fat necrosis, partial or total flap loss, infection and wound complications). One study[64] graded partial flap loss using a novel flap necrosis classification system, adapted from Kwok et al.[70]. Only 30 per cent of all surgical complications (30 of 99) reported across the 12 included studies were defined a priori.

Table 4 Surgical complications: delayed autologous breast reconstruction and neoadjuvant radiotherapy including non-comparative studies (neoadjuvant radiotherapy only)

Reference	GRADE	ROBINS-I	No. of patients		Follow-up (months)		Total no. of complications		No. of reoperations for complications	
			Neoadjuvant RT	No neoadjuvant RT	Neoadjuvant RT	No neoadjuvant RT	Neoadjuvant RT	No neoadjuvant RT	Neoadjuvant RT	No neoadjuvant RT
Modarressi *et al.*[64]	Low	Serious	60	45	1	1	20	9	n.a.	n.a.
Mull *et al.*[65]	Low	Serious	142	312	1	1	26	45	26	45
O'Connell *et al.*[58]	Low	Serious	38	80	50·3*	48·7*	12	20	3	8
Peeters *et al.*[66]	Low	Serious	77	109	≥ 12	≥ 12	n.a.	n.a.	n.a.	n.a.
Baumann *et al.*[69]	Moderate	Moderate	189	n.a.	11†	n.a.	88	n.a.	69	n.a.
Billig *et al.*[62]	Moderate	Moderate	67	n.a.	24	n.a.	37	n.a.	1	n.a.
Levine *et al.*[67]	Moderate	Moderate	50	n.a.	22·7†	n.a.	n.a.	n.a.	3	n.a.
Temple *et al.*[68]	Low	Serious	100	n.a.	≥ 12	n.a.	41	n.a.	18	n.a.

Values are *median and †mean. GRADE, Grading of Recommendation, Assessment, Development, and Evaluation (tool for grading the quality of evidence); ROBINS-I, Risk Of Bias In Non-randomised Studies – of Interventions (tool for assessing risk of bias); RT, radiotherapy; n.a., not applicable/available.

Table 5 Clavien–Dindo classification of surgical complications: delayed autologous breast reconstruction and neoadjuvant radiotherapy including non-comparative studies (neoadjuvant radiotherapy only)

	Neoadjuvant RT *versus* no neoadjuvant RT						
					Clavien-Dindo complication grade†		
Reference	Total flap loss	Partial flap loss*	Fat necrosis*	Wound dehiscence and delayed wound healing*	II	IIIa	IIIb
Modarressi *et al.*[64]	2 *versus* 1	12 *versus* 2	n.a.	n.a.	n.a.	n.a.	n.a.
Mull *et al.*[65]	5 *versus* 15	7 *versus* 5‡	n.a.	n.a.	n.a.	n.a.	26 *versus* 45
O'Connell *et al.*[58]	0 *versus* 0	0 *versus* 0	2 *versus* 2	7 *versus* 9	2 *versus* 3	0 *versus* 3	3 *versus* 5
Peeters *et al.*[66]	n.a.	n.a.	29 *versus* 36	n.a.	n.a.	n.a.	n.a.
Baumann *et al.*[69]	5 *versus* n.a.	14 *versus* n.a.	15 *versus* n.a.	22 *versus* n.a.	4 *versus* n.a.	n.a.	69 *versus* n.a.
Billig *et al.*[62]	0 *versus* n.a.	n.a.	7 *versus* n.a.	11 *versus* n.a.	4 *versus* n.a.	n.a.	1 *versus* n.a.
Levine *et al.*[67]	n.a.	1 *versus* n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Temple *et al.*[68]	2 *versus* n.a.	7 *versus* n.a.	16 *versus* n.a.	n.a.	n.a.	n.a.	18 *versus* n.a.

*Complication grades were not always defined or classified. †Grade II, complications requiring pharmacological treatment with drugs other than those allowed for grade I complications (drugs other than antiemetics, antipyretics, analgesics, diuretics and electrolytes); grade IIIa, complications requiring surgical intervention not under general anaesthesia; grade IIIb, complications requiring surgical intervention under general anaesthesia. RT, radiotherapy; n.a. not applicable/available. ‡$P < 0.050$.

Adjuvant post-mastectomy radiotherapy

Meta-analyses comparing PMRT (350 patients; mean follow-up 27·1 (range 12·0–54·0) months) and no radiotherapy (326 patients; mean follow-up 25·2 (12·0–48·7) months) showed no interstudy differences in rates of: overall complications (233 patients; odds ratio (OR) 1·52 (95 per cent c.i. 0·84 to 2·75), $Z = 1·40$, $P = 0·160$) (*Fig. 2a*); CDC grade III surgical complications (233 patients; OR 2·35 (0·63 to 8·81), $Z = 1·27$, $P = 0·200$)

(*Fig. 2b*); CDC grade II (293 patients; OR 0·94 (0·32 to 2·76), $Z = 0·11$, $P = 0·910$) (*Fig. 2c*); or fat necrosis (418 patients; OR 1·83 (0·67 to 5·00), $Z = 1·18$, $P = 0·240$) (*Fig. 2d*). There were no differences in rates of infection (293 patients; OR 0·94 (0·32 to 2·76), $Z = 0·11$, $P = 0·910$) (*Fig. S1a*, supporting information) or wound complications (293 patients; OR 1·16 (0·56 to 2·39), $Z = 0·40$, $P = 0·690$) (*Fig. S1b*, supporting information). There were no total flap losses.
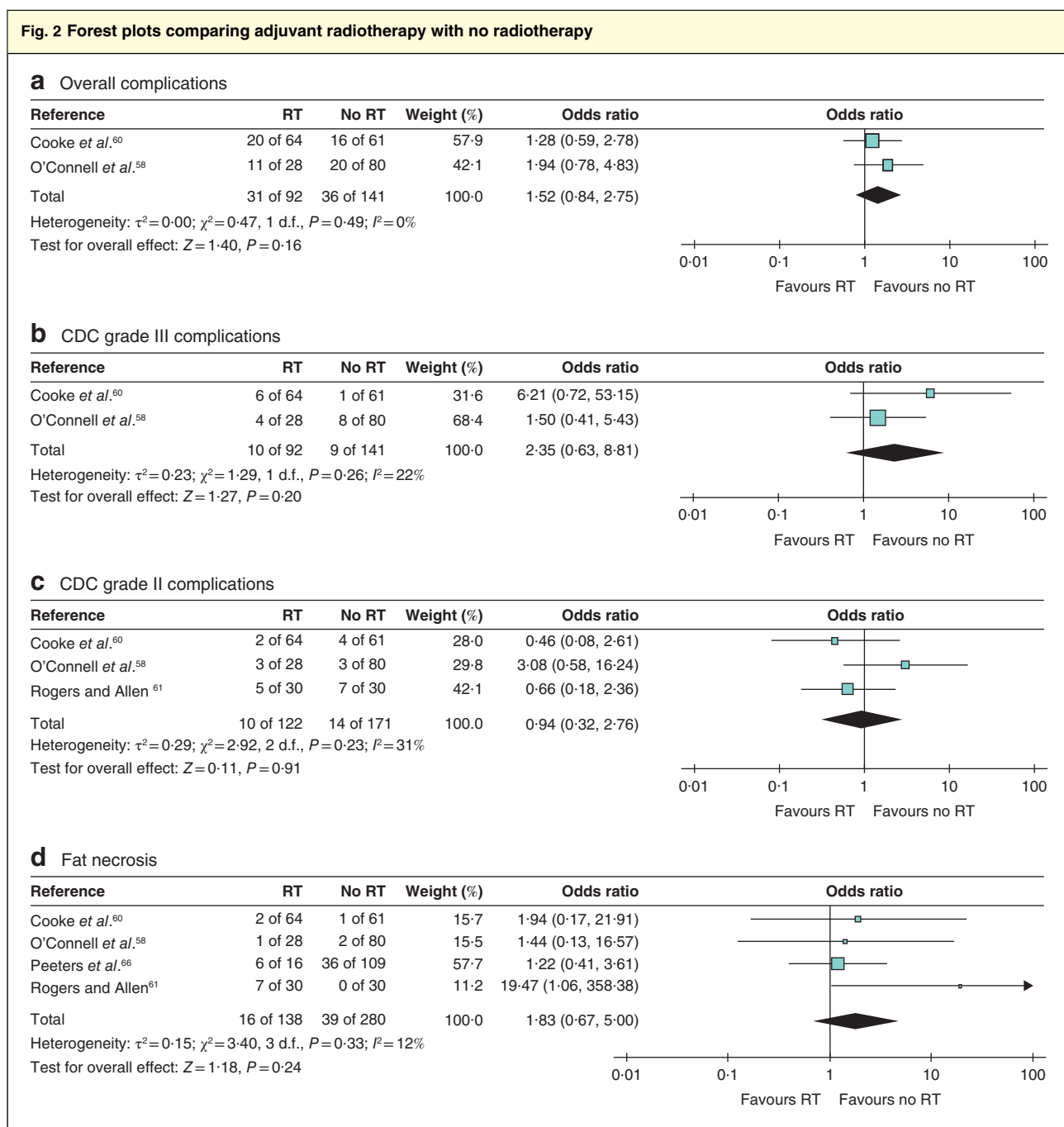
Fig. 2 Forest plots comparing adjuvant radiotherapy with no radiotherapy

a Overall complications, b Clavien–Dindo classification (CDC) grade III complications, c CDC grade II complications, d fat necrosis. A Mantel–Haenszel random-effects model was used for meta-analysis. Odds ratios are shown with 95 per cent confidence intervals. RT, radiotherapy.

Neoadjuvant radiotherapy

Comparisons between neoadjuvant radiotherapy (723 patients; mean follow-up 16·8 (range 1·0–50·3) months) and no radiotherapy (546 patients; mean follow-up 15·7 (1·0–48·7) months) showed no differences in overall complications (677 patients; OR 1·45 (95 per cent c.i. 0·97 to 2·18), $Z = 1·82$, $P = 0·070$) (Fig. 3a) and CDC grade III surgical complications (572 patients; OR 1·24 (0·76 to 2·04), $Z = 0·85$, $P = 0·390$) (Fig. 3b). One comparative study[58] reported similar CDC grade II

Fig. 3 Forest plot comparing neoadjuvant radiotherapy with no radiotherapy

a Overall complications

Reference	Neoadjuvant RT	No RT	Weight (%)	Odds ratio
Modarressi *et al.*[64]	20 of 60	9 of 45	19·8	2·00 (0·81, 4·95)
Mull *et al.*[65]	26 of 142	45 of 312	57·8	1·33 (0·78, 2·26)
O'Connell *et al.*[58]	12 of 38	20 of 80	22·4	1·38 (0·59, 3·24)
Total	58 of 240	74 of 437	100·0	1·45 (0·97, 2·18)

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.60$, 2 d.f., $P = 0.74$; $I^2 = 0\%$
Test for overall effect: $Z = 1.82$, $P = 0.07$

b CDC grade III complications

Reference	Neoadjuvant RT	No RT	Weight (%)	Odds ratio
Mull *et al.*[65]	26 of 142	45 of 312	87·3	1·33 (0·78, 2·26)
O'Connell *et al.*[58]	3 of 38	8 of 80	12·7	0·77 (0·19, 3·09)
Total	29 of 180	53 of 392	100·0	1·24 (0·76, 2·04)

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.52$, 1 d.f., $P = 0.47$; $I^2 = 0\%$
Test for overall effect: $Z = 0.85$, $P = 0.39$

c Fat necrosis

Reference	Neoadjuvant RT	No RT	Weight (%)	Odds ratio
O'Connell *et al.*[58]	2 of 38	2 of 80	8·5	2·17 (0·29, 16·00)
Peeters *et al.*[66]	29 of 77	36 of 109	91·5	1·23 (0·67, 2·25)
Total	31 of 115	38 of 189	100·0	1·29 (0·72, 2·30)

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.29$, 1 d.f., $P = 0.59$; $I^2 = 0\%$
Test for overall effect: $Z = 0.85$, $P = 0.40$

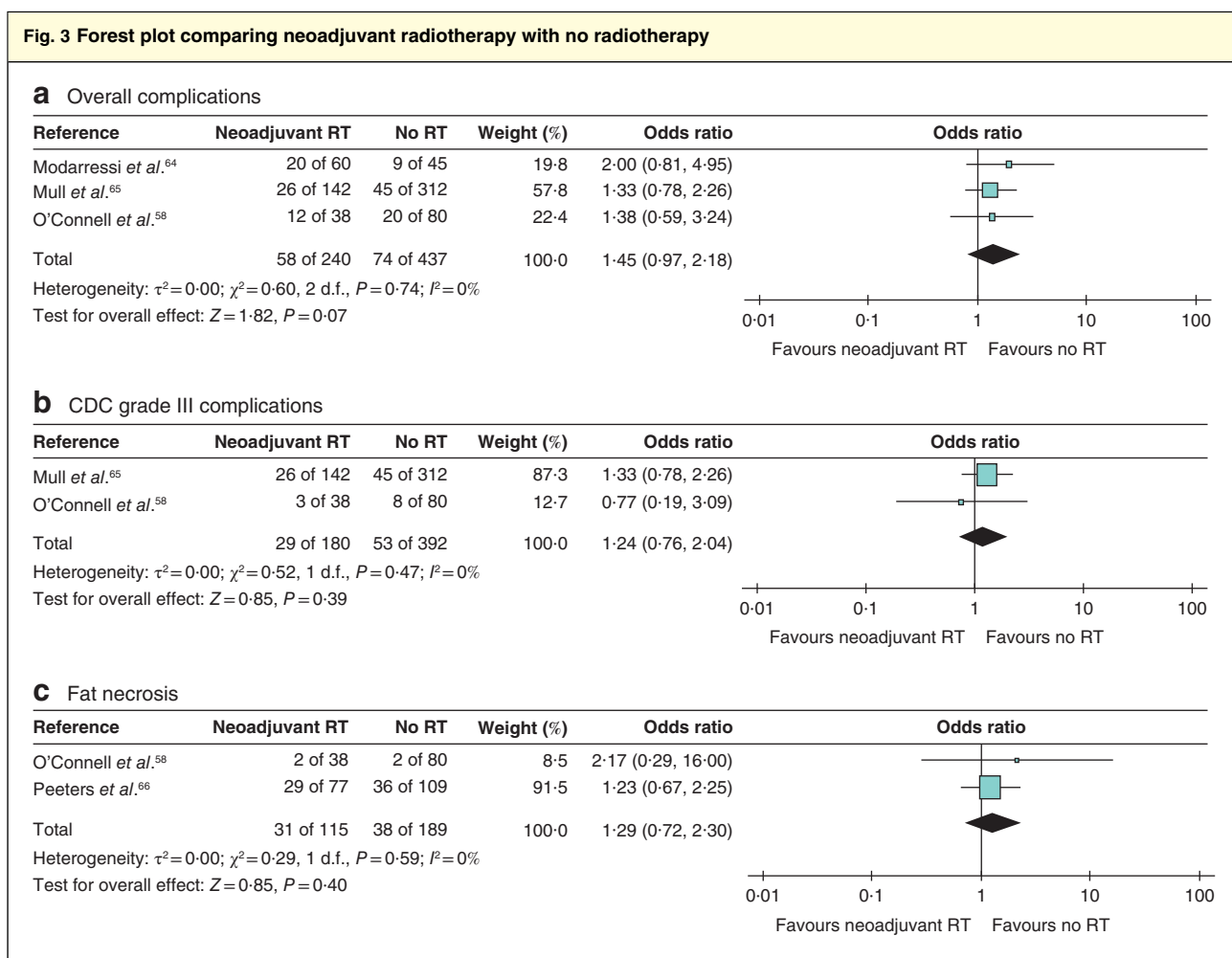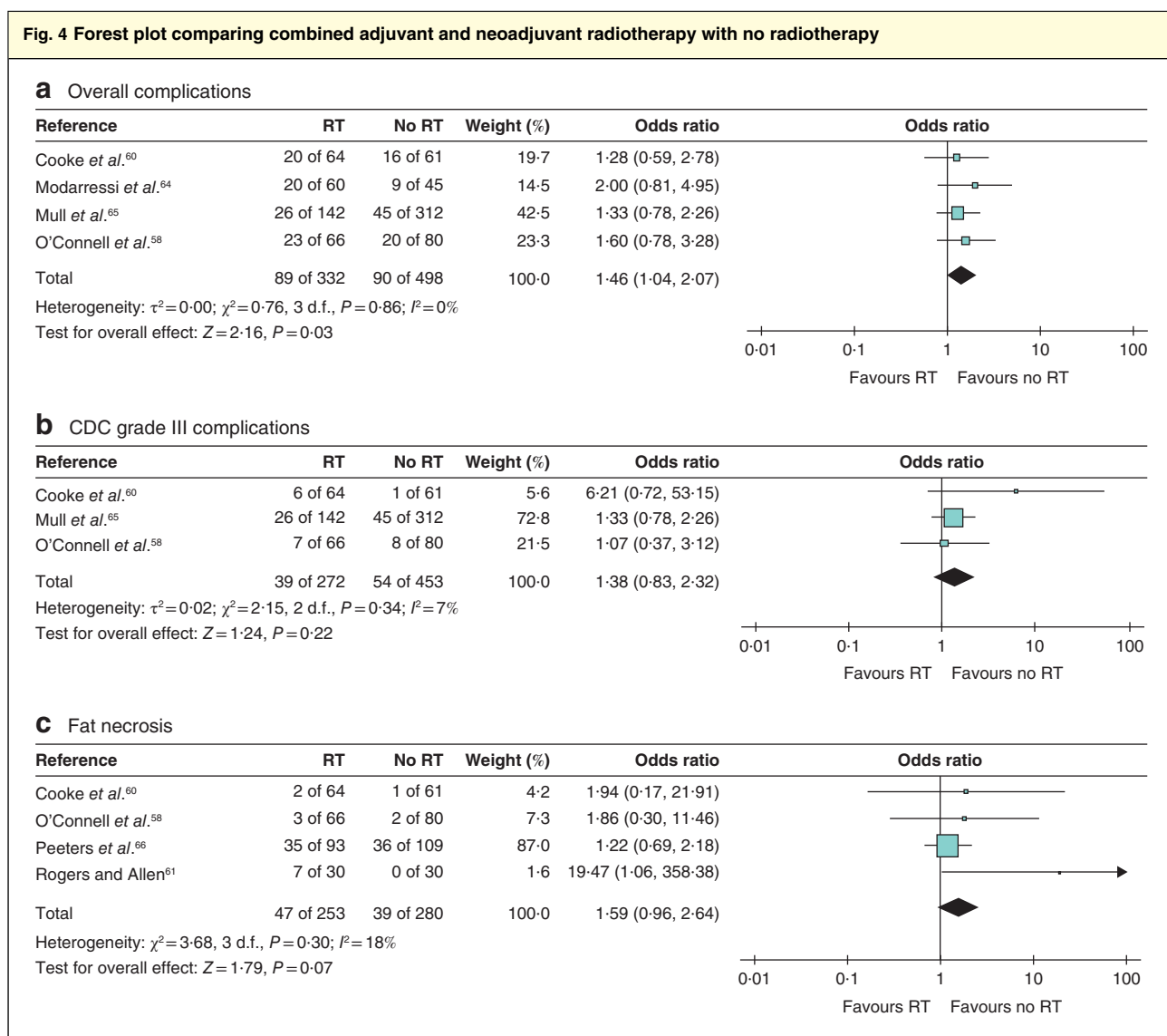**a** Overall complications, **b** Clavien–Dindo classification (CDC) grade III complications, **c** fat necrosis. A Mantel–Haenszel random-effects model was used for meta-analysis. Odds ratios are shown with 95 per cent confidence intervals. RT, radiotherapy.

complications between neoadjuvant and no radiotherapy (118 patients; OR 1·43 (0·23 to 8·91), $Z = 0.38$, $P = 0.700$). There were no differences in rates of fat necrosis (304 patients; OR 1·29 (0·72 to 2·30), $Z = 0.85$, $P = 0.400$) (*Fig. 3c*). Rates of partial flap loss were higher for neoadjuvant radiotherapy than for no radiotherapy (559 patients; OR 3·85 (1·51 to 9·76), $Z = 2.83$, $P = 0.005$) (*Fig. S2a*, supporting information), with no differences in rates of total flap loss (559 patients; OR 0·81 (0·31 to 2·09), $Z = 0.44$, $P = 0.660$) (*Fig. S2b*, supporting information).

Combined adjuvant and neoadjuvant radiotherapy

Meta-analyses of pooled PMRT and neoadjuvant radiotherapy compared with pooled no radiotherapy groups (mean follow-up 18·3 (range 1·0–48·7) months) were performed as a potential hypothesis-generating exercise. This showed significantly higher overall complications in the combined radiotherapy groups compared with no radiotherapy (830 patients; OR 1·46 (95 per cent c.i. 1·04 to 2·07), $Z = 2.16$, $P = 0.030$) (*Fig. 4a*). There were no interstudy differences in: CDC grade III complications (725 patients; OR 1·38 (0·83 to 2·32), $Z = 1.24$, $P = 0.220$) (*Fig. 4b*); CDC grade II complications (331 patients; OR 0·89 (0·37 to 2·10), $Z = 0.28$, $P = 0.780$) (*Fig. S3a*, supporting information); rates of fat necrosis (533 patients; OR 1·59 (0·96 to 2·64), $Z = 1.79$, $P = 0.070$) (*Fig. 4c*); or emergency reoperations for complications (725 patients; OR 1·38 (0·83 to 2·32), $Z = 1.24$, $P = 0.220$) (*Fig. S3b*, supporting information). Rates of partial flap loss were also higher in the combined *versus* no radiotherapy groups (684 patients; OR 2·59 (1·27 to 5·28), $Z = 2.63$, $P = 0.009$) (*Fig. S3c*, supporting

Fig. 4 Forest plot comparing combined adjuvant and neoadjuvant radiotherapy with no radiotherapy

a Overall complications

Reference	RT	No RT	Weight (%)	Odds ratio
Cooke *et al.*[60]	20 of 64	16 of 61	19·7	1·28 (0·59, 2·78)
Modarressi *et al.*[64]	20 of 60	9 of 45	14·5	2·00 (0·81, 4·95)
Mull *et al.*[65]	26 of 142	45 of 312	42·5	1·33 (0·78, 2·26)
O'Connell *et al.*[58]	23 of 66	20 of 80	23·3	1·60 (0·78, 3·28)
Total	89 of 332	90 of 498	100·0	1·46 (1·04, 2·07)

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.76$, 3 d.f., $P = 0.86$; $I^2 = 0\%$
Test for overall effect: $Z = 2.16$, $P = 0.03$

b CDC grade III complications

Reference	RT	No RT	Weight (%)	Odds ratio
Cooke *et al.*[60]	6 of 64	1 of 61	5·6	6·21 (0·72, 53·15)
Mull *et al.*[65]	26 of 142	45 of 312	72·8	1·33 (0·78, 2·26)
O'Connell *et al.*[58]	7 of 66	8 of 80	21·5	1·07 (0·37, 3·12)
Total	39 of 272	54 of 453	100·0	1·38 (0·83, 2·32)

Heterogeneity: $\tau^2 = 0.02$; $\chi^2 = 2.15$, 2 d.f., $P = 0.34$; $I^2 = 7\%$
Test for overall effect: $Z = 1.24$, $P = 0.22$

c Fat necrosis

Reference	RT	No RT	Weight (%)	Odds ratio
Cooke *et al.*[60]	2 of 64	1 of 61	4·2	1·94 (0·17, 21·91)
O'Connell *et al.*[58]	3 of 66	2 of 80	7·3	1·86 (0·30, 11·46)
Peeters *et al.*[66]	35 of 93	36 of 109	87·0	1·22 (0·69, 2·18)
Rogers and Allen[61]	7 of 30	0 of 30	1·6	19·47 (1·06, 358·38)
Total	47 of 253	39 of 280	100·0	1·59 (0·96, 2·64)

Heterogeneity: $\chi^2 = 3.68$, 3 d.f., $P = 0.30$; $I^2 = 18\%$
Test for overall effect: $Z = 1.79$, $P = 0.07$

a Overall complications, **b** Clavien–Dindo classification (CDC) grade III complications, **c** fat necrosis. A Mantel–Haenszel random-effects model was used for meta-analysis. Odds ratios are shown with 95 per cent confidence intervals. RT, radiotherapy.

information), with no differences in rates of total flap loss (559 patients; OR 0·81 (0·31 to 2·09), $Z = 0.44$, $P = 0.660$) (*Fig. S3d*, supporting information), infection (331 patients; OR 0·89 (0·37 to 2·10), $Z = 0.28$, $P = 0.780$) (*Fig. S3e*, supporting information) or wound complications (dehiscence/delayed wound healing) (331 patients; OR 1·29 (0·68 to 2·47), $Z = 0.78$, $P = 0.430$) (*Fig. S3f*, supporting 1information).

Assessment of heterogeneity and meta-analyses

Clinical outcomes within studies of PMRT *versus* no radiotherapy were homogeneous (I^2 values below 50 per cent). All remaining meta-analyses of outcomes were similar (neoadjuvant radiotherapy *versus* no radiotherapy, pooled PMRT and neoadjuvant radiotherapy *versus* no radiotherapy).

Quality of life

There was limited reporting of patient-reported QOL; outcomes were detailed in only two prospective studies[60,62] and one retrospective study[58], with small patient numbers and short follow-ups for the PMRT groups[58,60,62]. *A priori* hypothesis-driven selection of QOL domains was absent

from methods[58,60,62], with no reporting of missing data or how this problem was tackled[34].

Three studies[58,60,62] used the BREAST-Q and one[60] used the breast cancer-specific questionnaire (EORTC QLQ-BR23)[42]. One small study[58] reported significantly better 'satisfaction with breast' ($P = 0.008$) after a median follow-up of 27·5 months for PMRT compared with 48·7 months for no radiotherapy (*Table S2*, supporting information). The moderate-quality comparative prospective study[60] found a significant adverse impact of PMRT on breast symptoms at 1 year ($P < 0.001$) compared with no radiotherapy (*Table S2*, supporting information).

The third study[62] evaluated serial QOL outcomes, concluding a significant impact of PMRT on QOL domains (BREAST-Q) at 1 and 2 years, despite the absence of a control group (no radiotherapy). Moreover, clinical significance was defined as $P = 0.05$, which may not account for multiple variables (*Table S2*, supporting information)[43,62]. Highly significant abdominal adverse effects in a small patient group (108 patients) may be unrelated to PMRT, but rather an indication of donor site morbidity. Interestingly, when evaluating the impact of neoadjuvant radiotherapy in a small non-comparative study[62], significant time-related improvements in most QOL domains were observed, except lower physical well-being relating to the abdomen at 1 year (*Table S3*, supporting information).

Cosmetic outcomes

Three studies[58,61,63] evaluated PMRT and the effects on aesthetic outcomes (187 patients). There was no standardized evaluation of cosmetic outcomes, precluding meta-analyses. Studies lacked robust methodology.

Discussion

The mixture of underpowered observational studies included in this review were, in large part, lacking contemporaneous data to reflect current practice. Most were retrospective single-centre cohorts, demonstrating poor levels of clinical evidence (levels 3 and 4) with insufficient follow-up[11].

A previous study[24] of over 40 000 women undergoing BRR in 134 studies found that only 20 per cent reported *a priori* surgical complications, as well as inconsistent interstudy definitions[24]. The present review found similar interstudy discrepancies, without uniform adoption of the CDC[26]. The present authors graded all reported surgical complications using the CDC. All surgical interventions were graded as CDC IIIa or IIIb, and surgical reoperations were differentiated according to whether

they were for complications or cosmetic revisions. Some complications were not amenable to retrospective grading in three studies[64,66,67]. In one[66], it was not possible to determine whether fat necrosis required surgical revision for each radiotherapy group (adjuvant or neoadjuvant), compared with no radiotherapy. A second[64] omitted individual abdominal complications relative to timings of radiotherapy, and the third[67] omitted overall numbers of complications. Reviewed studies also failed to define postoperative wound infections according to Centers for Disease Control and Prevention criteria[71].

The IDEAL (Idea, Development, Exploration, Assessment, Long-term study) Collaboration describes key methodological criteria for robust prospective cohort studies[72]: studies should be powered on the effect size of primary outcomes evaluating interventions of interest. The Mastectomy and Breast Reconstruction Outcomes Collaborative (MROC) is a multicentre prospective cohort study that provides IDEAL level 2b evidence for clinical safety and satisfactory QOL outcomes in the evaluation of surgical complications in immediate autologous reconstructions with PMRT *versus* no radiotherapy (delayed BRR) in 11 US centres[17,60]. The MROC cohort data were excluded from this systematic review based on its reporting of group-related summative data for all types of autologous reconstruction, as opposed to individual abdominal donor sites.

The MROC has reported all surgical complications at 2 years and demonstrated that PMRT (*versus* no radiotherapy) was significantly associated with a greater risk of developing any complication (OR 1·50 (95 per cent c.i. 1·20 to 1·86); $P < 0.001$), reoperative complications (OR 1·52 (1·17 to 1·97); $P < 0.002$) and wound infection (OR 2·77 (1·78 to 4·31); $P < 0.001$)[16]. Autologous BRR was done more commonly in irradiated than non-irradiated patients (38 *versus* 25 per cent respectively; $P < 0.001$), with similarly low rates (1–2·4 per cent) of reconstruction failure at 2 years[17].

Eligible studies in the present systematic review were significantly underpowered in comparison with the MROC study, which evaluated irradiated autologous BRR at 1 year (236 patients) and 2 years (199), and non-irradiated procedures at 1 year (1625) and 2 years (332). The MROC data showed no differences between radiotherapy and no radiotherapy groups in the rates of total complications (25·6 *versus* 28·3 per cent respectively), major complications (17·6 *versus* 22·9 per cent) or flap failure (1·0 *versus* 2·4 per cent) at 2 years after immediate autologous reconstruction[17]. Studies in the present review showed significantly lower rates of major complications after radiotherapy compared with the MROC results, suggesting

suboptimal overall reporting of surgical complications in the reviewed studies[24].

The retrospective grading of surgical complications in the two moderate-quality studies reported showed a rate of major complications (CDC grade IIIb) of 9 per cent (6 of 64) at 1 year, and 4·6 per cent (5 of 108) at 2 years[60,62]. These rates are also likely to reflect under-reporting compared with the MROC rates of 14·8 per cent (35 of 236) at 1 year and 17·6 per cent (35 of 199) at 2 years[17]. Despite its strengths, the MROC cohort is based on the review of complications from electronic patient records, potentially also underestimating true complication rates[17].

One way to measure what matters to patients is to use patient-reported outcome measures (PROMs) to assess the effects of disease or treatment on symptoms, functioning and health-related QOL[34]. In this systematic review, PROMs were poorly reported and underpowered for overall small effect sizes of individual QOL domains[43]. Preliminary conclusions regarding statistical significance were not substantiated by adequate patient numbers, lack of a comparator group or prospectively defined time points for questionnaire collection[58]. Standardized and objective evaluations of cosmetic outcome have also remained elusive with emerging adoption of newer technologies such as the Vectra® XT[58]. Robust study designs evaluating these innovations should be accompanied by surgery- and disease-specific questionnaires[34].

Clear recommendations for the optimal timing of radiotherapy in relation to autologous BRR will remain elusive until information from high-quality systematic reviews forms part of shared preoperative decision-making[73].

Adequately powered prospective studies and ongoing audits, to allow comparisons of postoperative radiotherapy with neoadjuvant radiotherapy, are warranted. Current evidence for irradiating autologous abdominal flaps remains weak, involving only two moderate-quality studies of the 12 included in this report. Future cohort studies should be designed and powered to take advantage of newly evolving study designs, such as multiple-cohort RCTs or trials within cohorts[74]. These designs permit collection of big data within registry or cohort platforms, and allow multiple synchronous randomized trials to be conducted in a cost-effective manner[74].

Acknowledgements

References

1 Ginsburg O, Bray F, Coleman MP, Vanderpuye V, Eniu A, Kotha SR *et al.* The global burden of women's cancers: a grand challenge in global health. *Lancet* 2017; **389**: 847–860.

2 Winters S, Martin C, Murphy D, Shokar NK. Breast cancer epidemiology, prevention, and screening. *Prog Mol Biol Transl Sci* 2017; **151**: 1–32.

3 Veronesi U, Cascinelli N, Mariani L, Greco M, Saccozzi R, Luini A *et al.* Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *N Engl J Med* 2002; **347**: 1227–1232.

4 van Maaren MC, le Cessie S, Strobbe LJA, Groothuis-Oudshoorn CGM, Poortmans PMP, Siesling S. Different statistical techniques dealing with confounding in observational research: measuring the effect of breast-conserving therapy and mastectomy on survival. *J Cancer Res Clin Oncol* 2019; **145**: 1485–1493.

5 Ho AY, Hu ZI, Mehrara BJ, Wilkins EG. Radiotherapy in the setting of breast reconstruction: types, techniques, and timing. *Lancet Oncol* 2017; **18**: e742–e753.

6 O'Halloran N, Potter S, Kerin M, Lowery A. Recent advances and future directions in postmastectomy breast reconstruction. *Clin Breast Cancer* 2018; **18**: e571–e585.

7 Yang TJ, Ho AY. Radiation therapy in the management of breast cancer. *Surg Clin North Am* 2013; **93**: 455–471.

8 Macdonald SM, Harris EE, Arthur DW, Bailey L, Bellon JR, Carey L *et al.* ACR Appropriateness Criteria® locally advanced breast cancer. *Breast J* 2011; **17**: 579–585.

9 EBCTCG (Early Breast Cancer Trialists' Collaborative Group), McGale P, Taylor C, Correa C, Cutter D, Duane F, Ewertz M *et al.* Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. *Lancet* 2014; **383**: 2127–2135.

10 Marks LB, Kaidar-Person O, Poortmans P. Regarding current recommendations for postmastectomy radiation therapy in patients with one to three positive axillary lymph nodes. *J Clin Oncol* 2017; **35**: 1256–1258.

11 Recht A, Comen EA, Fine RE, Fleming GF, Hardenbergh PH, Ho AY *et al.* Postmastectomy radiotherapy: an American Society of Clinical Oncology, American Society for Radiation Oncology, and Society of Surgical Oncology focused guideline update. *Ann Surg Oncol* 2017; **24**: 38–51.

12 Kunkler IH, Dixon JM, Maclennan M, Russell NS. European interpretation of North American post mastectomy radiotherapy guideline update. *Eur J Surg Oncol* 2017; **43**: 1805–1807.

13 Russell NS, Kunkler IH, van Tienhoven G. Determining the indications for post mastectomy radiotherapy: moving from 20th century clinical staging to 21st century biological criteria. *Ann Oncol* 2015; **26**: 1043–1044.

14 Donker M, van Tienhoven G, Straver ME, Meijnen P, van de Velde CJ, Mansel RE *et al.* Radiotherapy or surgery of the axilla after a positive sentinel node in breast cancer (EORTC 10981-22023 AMAROS): a randomised, multicentre, open-label, phase 3 non-inferiority trial. *Lancet Oncol* 2014; **15**: 1303–1310.

15 O'Halloran N, Lowery A, Kalinina O, Sweeney K, Malone C, McLoughlin R *et al.* Trends in breast reconstruction practices in a specialized breast tertiary referral centre. *BJS Open* 2017; **1**: 148–157.

16 Bennett KG, Qi J, Kim HM, Hamill JB, Pusic AL, Wilkins EG. Comparison of 2-year complication rates among common techniques for postmastectomy breast reconstruction. *JAMA Surg* 2018; **153**: 901–908.

17 Jagsi R, Momoh AO, Qi J, Hamill JB, Billig J, Kim HM *et al.* Impact of radiotherapy on complications and patient-reported outcomes after breast reconstruction. *J Natl Cancer Inst* 2018; **110**: 157–165.

18 Barry M, Kell MR. Radiotherapy and breast reconstruction: a meta-analysis. *Breast Cancer Res Treat* 2011; **127**: 15–22.

19 Potter S, Conroy EJ, Cutress RI, Williamson PR, Whisker L, Thrush S *et al.*; iBRA Steering Group; Breast Reconstruction Research Collaborative. Short-term safety outcomes of mastectomy and immediate implant-based breast reconstruction with and without mesh (iBRA): a multicentre, prospective cohort study. *Lancet Oncol* 2019; **20**: 254–266.

20 Santosa KB, Qi J, Kim HM, Hamill JB, Wilkins EG, Pusic AL. Long-term patient-reported outcomes in postmastectomy breast reconstruction. *JAMA Surg* 2018; **153**: 891–899.

21 Velikova G, Williams LJ, Willis S, Dixon JM, Loncaster J, Hatton M *et al.*; MRC SUPREMO trial UK investigators. Quality of life after postmastectomy radiotherapy in patients with intermediate-risk breast cancer (SUPREMO): 2-year follow-up results of a randomised controlled trial. *Lancet Oncol* 2018; **19**: 1516–1529.

22 Momoh AO, Colakoglu S, de Blacam C, Gautam S, Tobias AM, Lee BT. Delayed autologous breast reconstruction after postmastectomy radiation therapy: is there an optimal time? *Ann Plast Surg* 2012; **69**: 14–18.

23 Kelley BP, Ahmed R, Kidwell KM, Kozlow JH, Chung KC, Momoh AO. A systematic review of morbidity associated with autologous breast reconstruction before and after exposure to radiotherapy: are current practices ideal? *Ann Surg Oncol* 2014; **21**: 1732–1738.

24 Potter S, Brigic A, Whiting PF, Cawthorn SJ, Avery KN, Donovan JL *et al.* Reporting clinical outcomes of breast

reconstruction: a systematic review. *J Natl Cancer Inst* 2011; **103**: 31–46.

25 Potter S, Holcombe C, Ward JA, Blazeby JM; BRAVO Steering Group. Development of a core outcome set for research and audit studies in reconstructive breast surgery. *Br J Surg* 2015; **102**: 1360–1371.

26 Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004; **240**: 205–213.

27 Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg* 2009; **124**: 345–353.

28 Winters ZE, Benson JR, Pusic AL. A systematic review of the clinical evidence to guide treatment recommendations in breast reconstruction based on patient-reported outcome measures and health-related quality of life. *Ann Surg* 2010; **252**: 929–942.

29 Winters ZE, Thomson HJ. Assessing the clinical effectiveness of breast reconstruction through patient-reported outcome measures. *Br J Surg* 2011; **98**: 323–325.

30 Cano SJ, Klassen AF, Scott AM, Cordeiro PG, Pusic AL. The BREAST-Q: further validation in independent clinical samples. *Plast Reconstr Surg* 2012; **129**: 293–302.

31 Klassen AF, Pusic AL, Scott A, Klok J, Cano SJ. Satisfaction and quality of life in women who undergo breast surgery: a qualitative study. *BMC Womens Health* 2009; **9**: 11.

32 Tevis SE, James TA, Kuerer HM, Pusic AL, Yao KA, Merlino J *et al.* Patient-reported outcomes for breast cancer. *Ann Surg Oncol* 2018; **25**: 2839–2845.

33 Zinzindohoué C, Bertrand P, Michel A, Monrigal E, Miramand B, Sterckers N *et al.* A prospective study on skin-sparing mastectomy for immediate breast reconstruction with latissimus dorsi flap after neoadjuvant chemotherapy and radiotherapy in invasive breast carcinoma. *Ann Surg Oncol* 2016; **23**: 2350–2356.

34 Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. *BMJ* 2019; **364**: k5267.

35 Khajuria A, Winters Z, Mosahebi A. *A Systematic Review and Meta-Analysis of Clinical and Patient-Reported Outcomes (PROs) of Immediate* versus *Delayed Autologous Abdominal-Based Flap Breast Reconstruction in the Context of Post-Mastectomy Radiotherapy [PROSPERO 2017 CRD42017077945]*. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42017077945 [accessed 26 October 2019].

36 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; **339**: b2700.

37 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.

38 Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S *et al.*; GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; **328**: 1490.

39 Wagner IJ, Tong WM, Halvorson EG. A classification system for fat necrosis in autologous breast reconstruction. *Ann Plast Surg* 2013; **70**: 553–556.

40 Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J *et al.* COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018; **27**: 1159–1170.

41 Mokkink LB, Prinsen CA, Bouter LM, Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther* 2016; **20**: 105–113.

42 Winters ZE, Afzal M, Rutherford C, Holzner B, Rumpold G, da Costa Vieira RA *et al.*; European Organisation for Research and Treatment of Cancer Quality of Life Group. International validation of the European Organisation for Research and Treatment of Cancer QLQ-BRECON23 quality-of-life questionnaire for women undergoing breast reconstruction. *Br J Surg* 2018; **105**: 209–222.

43 Winters ZE, Afzal M, Balta V, Freeman J, Llewellyn-Bennett R, Rayter Z *et al.*; Prospective Trial Management Group. Patient-reported outcomes and their predictors at 2- and 3-year follow-up after immediate latissimus dorsi breast reconstruction and adjuvant treatment. *Br J Surg* 2016; **103**: 524–536.

44 Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. *J Clin Nurs* 2005; **14**: 798–804.

45 Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. *Pain* 1999; **83**: 157–162.

46 Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B *et al.*; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care* 2007; **45**: S3–S11.

47 Melzack R. The McGill pain questionnaire: major properties and scoring methods. *Pain* 1975; **1**: 277–299.

48 Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006; **166**: 1092–1097.

49 Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**: 606–613.

50 O'Connell RL, Khabra K, Bamber JC, deSouza N, Meybodi F, Barry PA *et al.* Validation of the Vectra XT three-dimensional imaging system for measuring breast volume and symmetry following oncological reconstruction. *Breast Cancer Res Treat* 2018; **171**: 391–398.

51 Liu Z, Yao Z, Li C, Liu X, Chen H, Gao C. A step-by-step guide to the systematic review and meta-analysis of diagnostic and prognostic test accuracy evaluations. *Br J Cancer* 2013; **108**: 2299–2303.

52 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**: 1539–1558.

53 DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. *Contemp Clin Trials* 2015; **45**: 139–145.

54 Ousmen A, Conroy T, Guillemin F, Velten M, Jolly D, Mercier M *et al.* Impact of the occurrence of a response shift on the determination of the minimal important difference in a health-related quality of life score over time. *Health Qual Life Outcomes* 2016; **14**: 167.

55 Musoro ZJ, Hamel JF, Ediebah DE, Cocks K, King MT, Groenvold M *et al.*; EORTC Quality of Life Group. Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol. *BMJ Open* 2018; **8**: e019117.

56 Cano SJ, Klassen AF, Scott A, Alderman A, Pusic AL. Interpreting clinical differences in BREAST-Q scores: minimal important difference. *Plast Reconstr Surg* 2014; **134**: 173e–175e.

57 Mundy LR, Homa K, Klassen AF, Pusic AL, Kerrigan CL. Normative data for interpreting the BREAST-Q: augmentation. *Plast Reconstr Surg* 2017; **139**: 846–853.

58 O'Connell RL, Di Micco R, Khabra K, Kirby AM, Harris PA, James SE *et al.* Comparison of immediate *versus* delayed DIEP flap reconstruction in women who require postmastectomy radiotherapy. *Plast Reconstr Surg* 2018; **142**: 594–605.

59 Chatterjee JS, Lee A, Anderson W, Baker L, Stevenson JH, Dewar JA *et al.* Effect of postoperative radiotherapy on autologous deep inferior epigastric perforator flap volume after immediate breast reconstruction. *Br J Surg* 2009; **96**: 1135–1140.

60 Cooke AL, Diaz-Abele J, Hayakawa T, Buchel E, Dalke K, Lambert P. Radiation therapy *versus* no radiation therapy to the neo-breast following skin-sparing mastectomy and immediate autologous free flap reconstruction for breast cancer: patient-reported and surgical outcomes at 1 year-a mastectomy reconstruction outcomes consortium (MROC) substudy. *Int J Radiat Oncol Biol Phys* 2017; **99**: 165–172.

61 Rogers NE, Allen RJ. Radiation effects on breast reconstruction with the deep inferior epigastric perforator flap. *Plast Reconstr Surg* 2002; **109**: 1919–1924.

62 Billig J, Jagsi R, Qi J, Hamill JB, Kim HM, Pusic AL *et al.* Should immediate autologous breast reconstruction be considered in women who require postmastectomy radiation therapy? A prospective analysis of outcomes. *Plast Reconstr Surg* 2017; **139**: 1279–1288.

63 Huang CJ, Hou MF, Lin SD, Chuang HY, Huang MY, Fu OY *et al.* Comparison of local recurrence and distant metastases between breast cancer patients after postmastectomy radiotherapy with and without immediate TRAM flap reconstruction. *Plast Reconstr Surg* 2006; **118**: 1079–1086.

64 Modarressi A, Müller CT, Montet X, Rüegg EM, Pittet-Cuénod B. DIEP flap for breast reconstruction: is abdominal fat thickness associated with post-operative complications? *J Plast Reconstr Aesthet Surg* 2017; **70**: 1068–1075.

65 Mull AB, Qureshi AA, Zubovic E, Rao YJ, Zoberi I, Sharma K *et al*. Impact of time interval between radiation and free autologous breast reconstruction. *J Reconstr Microsurg* 2017; **33**: 130–136.

66 Peeters WJ, Nanhekhan L, Van Ongeval C, Fabré G, Vandevoort M. Fat necrosis in deep inferior epigastric perforator flaps: an ultrasound-based review of 202 cases. *Plast Reconstr Surg* 2009; **124**: 1754–1758.

67 Levine SM, Patel N, Disa JJ. Outcomes of delayed abdominal-based autologous reconstruction *versus* latissimus dorsi flap plus implant reconstruction in previously irradiated patients. *Ann Plast Surg* 2012; **69**: 380–382.

68 Temple CL, Strom EA, Youssef A, Langstein HN. Choice of recipient vessels in delayed TRAM flap breast reconstruction after radiotherapy. *Plast Reconstr Surg* 2005; **115**: 105–113.

69 Baumann DP, Crosby MA, Selber JC, Garvey PB, Sacks JM, Adelman DM *et al*. Optimal timing of delayed free lower abdominal flap breast reconstruction after postmastectomy radiation therapy. *Plast Reconstr Surg* 2011; **127**: 1100–1106.

70 Lie KH, Barker AS, Ashton MW. A classification system for partial and complete DIEP flap necrosis based on a review of 17 096 DIEP flaps in 693 articles including analysis of 152 total flap failures. *Plast Reconstr Surg* 2013; **132**: 1401–1408.

71 Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR. Guideline for prevention of surgical site infection, 1999. Centers for Disease Control and Prevention (CDC) Hospital Infection Control Practices Advisory Committee. *Am J Infect Control* 1999; **27**: 97–134.

72 Hirst A, Philippou Y, Blazeby J, Campbell B, Campbell M, Feinberg J *et al*. No surgical innovation without evaluation: evolution and further development of the IDEAL framework and recommendations. *Ann Surg* 2019; **269**: 211–220.

73 Elwyn G, Durand MA, Song J, Aarts J, Barr PJ, Berger Z *et al*. A three-talk model for shared decision making: multistage consultation process. *BMJ* 2017; **359**: j4891.

74 Young-Afat DA, van Gils CH, van den Bongard H, Verkooijen HM; UMBRELLA Study Group. The Utrecht cohort for Multiple BREast cancer intervention studies and Long-term evaLuAtion (UMBRELLA): objectives, design, and baseline results. *Breast Cancer Res Treat* 2017; **164**: 445–450.

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the article.

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Review

The impact of mobile technology on teamwork and communication in hospitals: a systematic review

Guy Martin,[1] **Ankur Khajuria,**[1] **Sonal Arora,**[1] **Dominic King,**[1,2] **Hutan Ashrafian,**[1] **and Ara Darzi**[1]

[1]Institute of Global Health Innovation, Imperial College London, London, United Kingdom, and [2]DeepMind, London, United Kingdom

Corresponding Author: Guy Martin, Institute of Global Health Innovation, Imperial College London, 10th Floor, QEQM Building, St Mary's Hospital, Praed Street, London W2 1NY, UK (guy.martin@imperial.ac.uk)

ABSTRACT

Objectives: Effective communication is critical to the safe delivery of care but is characterized by outdated technologies. Mobile technology has the potential to transform communication and teamwork but the evidence is currently uncertain. The objective of this systematic review was to summarize the quality and breadth of evidence for the impact of mobile technologies on communication and teamwork in hospitals.

Materials and Methods: Electronic databases (MEDLINE, PsycINFO, EMBASE, CINAHL Plus, HMIC, Cochrane Library, and National Institute of Health Research Health Technology Assessment) were searched for English language publications reporting communication- or teamwork-related outcomes from mobile technologies in the hospital setting between 2007 and 2017.

Results: We identified 38 publications originating from 30 studies. Only 11% were of high quality and none met best practice guidelines for mobile-technology-based trials. The studies reported a heterogenous range of quantitative, qualitative, and mixed-methods outcomes. There is a lack of high-quality evidence, but nonetheless mobile technology can lead to improvements in workflow, strengthen the quality and efficiency of communication, and enhance accessibility and interteam relationships.

Discussion: This review describes the potential benefits that mobile technology can deliver and that mobile technology is ubiquitous among healthcare professionals. Crucially, it highlights the paucity of high-quality evidence for its effectiveness and identifies common barriers to widespread uptake. Limitations include the limited number of participants and a wide variability in methods and reported outcomes.

Conclusion: Evidence suggests that mobile technology has the potential to significantly improve communication and teamwork in hospital provided key organizational, technological, and security challenges are tackled and better evidence delivered.

Key words: medical informatics, communication, hospitals, smartphone

INTRODUCTION

Effective communication between healthcare professionals within hospitals is critical to the safe delivery of care but is frequently characterized by a reliance on outdated technologies. The delivery of high-quality care inherently relies on effective communication and the inaccurate, incomplete, or delayed transfer of information can result in avoidable errors and patient harm.[1–4] Failures in communication occur twice as often as those due to inadequate skill or knowledge[5] and contribute to more than half of all patient safety events.[3,6]

Interprofessional teamwork within hospitals is complex and around the world typically relies on a mix of technologies and

approaches including 1-way pagers, fixed telephones, face-to-face conversations, and newer technologies such as e-mail and smartphone messaging. Numerous problems have been highlighted with traditional pagers such as the fragmentation and burden of communication,[7,8] interruptive communication behaviors,[9–11] and limitations with 1-way data transfer and the supply of supporting contextual information,[12,13] all of which may contribute to harmful failures of care for patients.[14–16] These failings not only harm patients, but also lead to significant financial costs for healthcare providers.[17]

Outside of healthcare, there has been a technological revolution in handheld communication devices spawning new ways to effectively and reliably communicate, collaborate, and share information. The requirements for immediacy and accuracy of communication within healthcare, together with the potentially harmful consequences of communication failure, mean that emergent communication technologies must be studied robustly. Any change to clinical practice as a result of the deployment of new technology must be based on evidence and not on transient technology trends or individual preference. Despite this, hospital communication systems receive much less attention than other areas of healthcare innovation, and there is little robust empirical evidence on which to assess the relative advantages and disadvantages of new technologies.[18] There is a careful trade-off to be made between new technologies that lead to increased complexity and cognitive overload and those that deliver meaningful improvements in communication, teamwork, and patent safety.[19] The aim of this review was therefore to evaluate the current quality and breadth of evidence for the impact of mobile technologies on communication and teamwork within hospitals.

MATERIALS AND METHODS

This review was conducted in accordance with best practice principles as outlined in the PRISMA Statement.[20] The review protocol was prospectively registered with the PROSPERO Database as per best practice guidelines (CRD42017064128).[21]

Search strategy and study selection

In consultation with expert medical librarians at Imperial College London, MEDLINE, PsycINFO, EMBASE, CINAHL Plus, HMIC, the Cochrane Library, and National Institute of Health Research Health Technology Assessment Database were searched for relevant literature published in English online or in print between January 1, 2007, and January 1, 2017. The search strategy encompassed 3 broad categories: mobile technology teamwork and communication, and the hospital setting. The search terms and strategy employed for each respective database are summarized in Supplementary Appendix Table 1 and prespecified inclusion and exclusion criteria in Supplementary Appendix Table 2. This review focuses on the impact of mobile technology on communication and teamwork within real-life hospital settings. For the purposes of this review, mobile technology was defined as hand-held devices (mobiles, smartphones, tablets, or bespoke mobile devices) that facilitate 2-way communication or data transfer and which directly impact patient care. All studies evaluating the impact of mobile technologies were included, even if the intervention studied did not form part of the study protocol (eg, questionnaire studies reporting the impact of mobile technology at work in general). There were otherwise no restrictions on study de-

sign, intervention, or sample size, and both qualitative and quantitative studies were included.

Two reviewers (GM, AK) independently reviewed all titles and abstracts for eligibility against the specified inclusion and exclusion criteria with only those papers considered relevant advanced to full text review. Cohen's kappa agreement was calculated for each stage of screening and review with disagreements resolved through consensus. The PRISMA Diagram for study inclusion is outlined in Figure 1.

Data extraction and quality assessment

For each study, relevant data on study design, population, intervention, comparators, outcomes, and setting were extracted. A second independent investigator reviewed this data for quality and accuracy before analysis. A quality and risk-of-bias assessment was performed for all studies according to the appropriate National Institutes of Health Quality Assessment Tool[22] with findings confirmed by consensus. A further quality assessment of each interventional study was performed by assessing compliance to the mobile health (mHealth) evidence reporting and assessment (mERA) checklist.[23] The mERA checklist was compiled by the World Health Organization mHealth Technical Evidence Review Group and identifies a minimum set of information that is needed to define the content, context, and technical features of an mHealth intervention and standardize the quality of evidence reporting, essentially a CONSORT[24] or PRISMA[20] statement for mobile technology–based interventions.

Data synthesis and analysis

The data for each study were summarized and are presented in Table 1 together with the quality assessment outcome. Studies deemed to be of poor quality are typically excluded for the purposes of analysis; however, as they formed a large number of the identified studies in this instance, they were retained. For the purposes of the analysis, studies were grouped into 6 categories: quantitative interventional studies, qualitative interventional studies, mixed-methods interventional studies, quantitative noninterventional studies, qualitative noninterventional studies, and mixed-methods noninterventional studies.

RESULTS

A total of 8 072 studies were initially identified, and following removal of duplicates a total of 5 683 eligible papers remained for screening and review. From this, we identified 38 publications from 30 unique studies as outlined in Figure 1. Included studies originated from a broad range of countries: 15 from Canada; 4 each from the United States and United Kingdom; 2 each from Singapore, Saudi Arabia and New Zealand; and a single paper arising from each of Germany, Turkey, India, Australia, Israel, Malaysia, Taiwan, Sweden, and South Korea. Inter-rater agreement for inclusion and exclusion of papers was "very good" throughout, with a Cohen's kappa of 0.842-0.980 reported at each stage. Of note, 9 publications reported data related to the same study investigating the introduction of smartphones and web-based messaging across a small number teams within a single institution.[35,36,38,43,45,46,58,61,62] Table 1 summarizes the recorded data for each study. Quality assessments for all studies are summarized together with mERA Checklist compliance for the 22 interventional studies in Table 2.

Table 1. Included studies with data for each by study design, comparator group, setting, intervention, findings, compliance with the mERA checklist, and quality assessment

Study	Study Design	Setting and Intervention	Key Findings (communication/teamwork)	Quality Assessment[22]	mERA Assessment[23]
Interventional Studies—Quantitative Outcomes					
Daruwalla et al 2014[25]	Prospective observational cohort study	Orthopaedic surgical team (25 participants)—Singapore MyDoc—HIPAA-compliant mobile application with messaging, case discussion, patient details and photo sharing functionality	- 23 of 25 (92%) agreed it should replace current communication methods - 23 of 25 (92%) agreed they could communicate easily using the application - 22 of 25 (88%) agreed that the potential for telerounding via the application may have advantages (eg, out-of-hours)	Poor	6/16
Duhm et al 2016[26]	Controlled prospective crossover study	University Hospital (14 participants)—Germany iPad with mobile eHR	- Application led to improvements in discussing clinical evidence with colleagues and streamlined clinical workflows	Fair	6/16
Gulacti et al 2016[27]	Retrospective observational cohort study	Tertiary hospital emergency department (628 consultations)—Turkey WhatsApp Messenger	- Message content: 510 images, 517 text messages, 59 videos, 10 voice messages across 519 patients - Median arrival time 3.94 min and response time 2.83 min - As a result of messaging 59.9% led to discharge of patient without a face-to-face specialty consultation and 71.6% out-of-hours consultations	Fair	5/16
Khanna et al 2015[28]	Pre/postobservational cohort study	Tertiary orthopaedic department (8 junior doctors, 25 consecutive patients pre/post intervention)—India Issued smartphone with WhatsApp messenger	- 100% felt WhatsApp improved the efficiency of handover and patient care - Use of WhatsApp led to significant improvement in quality of information transfer and recall	Poor	7/16
Lane et al 2012[29]	Pre/postobservational cohort study	University hospital (40 participants)—United States VigiVU—integrated mobile situational awareness application with monitoring, text and voice communication and access to eHR functionality	- Use of the application increased speed of communication compared with pagers (latency 18 s vs 22 s)	Poor	11/16
Motulsky et al 2017[30]	Prospective cross-sectional mixed-methods study	University hospital (124 participants)—Canada FLOW—in-house mobile application allowing free-text communication of 200 characters within eHR accessed through personal smartphones	- Number of "flows" created mean 26 per day, 8 per patient per day - Majority prefer to access information and communicate through a smartphone - Majority think application improves handover and patient care	Fair	6/16
Ng et al 2007[31]	Prospective observational cohort study	Neurosurgical team in University hospital (12 participants)—Singapore Issued smartphone with multimedia messaging and picture capability	- Senior doctor perspectives: frequently used, improved confidence and decision making, improved interteam communication, and reduced need for call-back - Junior doctor perspective: frequently used, facilitated increased involvement of senior decision making from home	Poor	3/16

(continued)

Table 1. continued

Study	Study Design	Setting and Intervention	Key Findings (communication/teamwork)	Quality Assessment[22]	mERA Assessment[23]
Patel et al 2016[32]	Pre/postobservational cohort study	4x clinical teams in large University hospital (229 multi-professional participant's preintervention, 210 participants' postintervention)—United States Cureatr—HIPAA compliant smartphone application with encrypted messaging and other applications accessed through personal and issued devices	- 708 456 messages across 130 073 patient threads - Junior doctors and nurses the largest senders: 5 (range, 2-12) and 6 (range, 2-13) per day - Messages sent by doctors shorter (28 vs 41; $P < .001$) - >50% of messages sent read in <1 min - All staff found the application to cause significantly less disruption to workflows than pagers, with more responsive physicians and better transfer of information	Fair	10/16
Power et al 2014[33]	Prospective observational cohort study	Pharmacy team in hospital setting (90 participants)—Canada Issued iPhone with multiple generic functionalities	- Principle use as a communication device - 98% found it useful, 87% improved performance, 68% improved efficiency, - Positive impact: accessibility, rapid communication, easier management of email and calendar - Negative comments: small screen size, connectivity	Fair	5/16
Przybylo et al 2014[34]	Controlled prospective cluster-randomized study	5 general medicine teams at a University hospital (26 control and 49 intervention participants)—United States Medigram—HIPAA compliant group messaging application accessed through institutional or personal smartphones	- Ineffective aspects of pagers: time wasted for responses, 1-way nature of communication, needing to find a computer/phone - Effective features of pagers: reliability, ease of use, responsiveness, brevity - At baseline majority (90.5%) already use text messaging - Compared with paging smartphones significantly more effective, allow clearer more efficient communication, and integrate better into workflow - Satisfaction with smartphone higher. 85% would recommend its use	Good	10/16
Smith et al 2012[35]	Prospective observational cohort study	4 medical teams in 2 large hospitals (34 participants—analysis of 13 717 e-mails)—Canada Issued team and individual Blackberry smartphones with messaging/email functionality	- 7 784 structured and 5 933 unstructured messages - Median response time 2.3 min, 50% did not get a response - 28.1% of emails requested an inappropriate response given content	Poor	3/16
Vaisman and Wu 2017[36]	Retrospective observational cohort study	8 clinical teams across 2 large academic hospitals (21 doctor participants over 18 months)—Canada Institutional smartphones with secure voice calls, messaging and e-mail functionality	- 187 049 interruptions identified - Peak of interruptions at 11 am to 12 pm and 2-3 pm - Average daily interruptions 42.3-51.4 per day per team - Crisis mode experienced 2.3 per day per team with a mean duration of 35.1 min	Fair	4/16

(continued)

Table 1. continued

Study	Study Design	Setting and Intervention	Key Findings (communication/teamwork)	Quality Assessment[22]	mERA Assessment[23]
Wani et al 2013[37]	Prospective observational cohort study	Plastic surgery department in academic hospital (116 communication events)—Saudi Arabia Institutional smartphone with WhatsApp	- Overall positive response to the efficacy of using WhatsApp as a means of communication - Led to elimination of redundant steps in vertical reporting within teams	Poor	6/16
Wu et al 2015[38]	Prospective observational cohort study	5 general medicine teams in 2 large academic hospitals (60 969 messages, 165 multiprofessional participants)—Canada Clinical Message—bespoke application with secure messaging and handover tools accessed through institutional smartphone	- On average, 14.8 messages per day per team with median response time 2.3 min - 76.5% requested a text reply, 7.7% a call back, and 15.7% no response - Majority of staff felt system improved care and speed of work, accountability, timeliness of communication, and interprofessional relationships - Not seen as effective for communicating complex issues - Doctors felt frequently interrupted with low-value information, nurses conversely perceived a lack of desired response	Fair	6/16

Interventional Studies—Qualitative Outcomes

Study	Study Design	Setting and Intervention	Key Findings (communication/teamwork)	Quality Assessment[22]	mERA Assessment[23]
Farrell 2016[39]	Retrospective cross-sectional interview study	Gynaecology ward (20 participants)—Australia iPhone with relevant generic medical applications (eg, MIMS drug information, MedCalc, Medscape)	- Overall positive impact on interprofessional interactions and communication - Primary use for interprofessional communication - Negative aspects: screen size, battery life, connectivity unprofessional to use at bedside	Poor	6/16
Lo et al 2012[40]	Retrospective cross-sectional questionnaire study	General internal medicine teams (31 participants) in teaching hospital—Canada Individual and team Blackberry smartphones with web-paging/email functionality	- Positive impact of smartphones: value in delivery of nonurgent information, aid in triage and prioritization, improvement in efficiency of communication and access to clinical staff, improved timeliness of replies compared with pagers - Negative impact of smartphones: conflict between nurses and doctors about correct communication method and subjective decision on urgency/priority, accessibility leads to increase in unnecessary communication, residents find increased calls disruptive	Fair	4/16

Interventional Studies—Mixed Methods Outcomes

Study	Study Design	Setting and Intervention	Key Findings (communication/teamwork)	Quality Assessment[22]	mERA Assessment[23]
Johnston et al 2015[41]	Prospective mixed-methods cohort study	Acute general surgery team in a teaching hospital (40 participants, 1140 hours of clinical communication with 1495 communication events)—United Kingdom WhatsApp messenger	- Median number of communication events within team 65.5 per week. - Message content: 39.3% communication events, 35.6% information giving, 60.5% administration - Juniors like the ability to send messages rather than voice calls, seniors like additional supervision; universal agreement that it led to the removal of communication barriers	Fair	8/16

(continued)

Table 1. continued

Study	Study Design	Setting and Intervention	Key Findings (communication/teamwork)	Quality Assessment[22]	mERA Assessment[23]
O'Connor et al 2009[42]	Prospective mixed-methods cohort study	Intensive Care Unit in community hospital (106 multi-professional participants)—Canada. Institutional Blackberry with messaging/e-mail functionality	- Staff sent a mean 5.2 messages and received 8.9 per day - Positive perceptions—usability, impact on communication, team relationships and patient care, fast and reliable, improved doctor response times, improved coordination and job satisfaction - Negative experiences reported: impact on quality of communication, reduced face-to-face communication, and inappropriate use of devices for personal reasons - 87% wanted to continue using the devices	Good	8/16
Quan et al 2013[43]	Pre-/post observational cohort study	Four general internal medicine teams in academic hospital (17 multiprofessional participants—5 doctors, 8 nurses, 2 pharmacists, 2 social workers)—Canada. Institutional Blackberry with email/messaging functionality	- Increase is number of messages 710 vs 2 196 - 233% increase in interruptions to clinical tasks - Increased interruptions due to elimination of traditional barriers (eg, waiting for phone), ease of access and impersonal nature of communication - Increased messaging from nurses due to push for accountability and reassurance, doctors saw this as nurses absolving themselves of responsibility - Nurses found to often exaggerate severity or urgency of issues to illicit a response, particularly at the end of a shift	Poor	5/16
Webb et al 2016[44]	Pre-/Post observational cohort study	2 academic hospitals and a satellite community hospital (104 multiprofessional iPhone users with 49 web console users)—Canada. Vocera Collaboration Suite—smartphone enabled application with call alerting, chat, voice calls	- Significant reduction in response times (5.5 min vs 3 min; P = .027) - 85% of staff used mobile for day-to-day communication - 35% of staff used mobile for communication with patients - 81% of doctors positive about system - Positive aspects of system: reduction in interruptions, ability to answer in own time, ability to send additional information, receipt confirmation, convenience - Negative aspects of system: battery life, having to enter password every time, balance between interruptions and missing messages when on do not disturb	Fair	6/16

(continued)

Table 1. continued

Study	Study Design	Setting and Intervention	Key Findings (communication/teamwork)	Quality Assessment[22]	mERA Assessment[23]
Wu et al 2011[45]	Prospective observational cohort study	General medicine teams in multiple academic hospitals (16 months data collection)—Canada Institutional Blackberry with email/messaging functionality	- Analysis of 13 717 calls and 12 936 emails - Efficiency: smartphones lead to faster response times and increased accessibility, and increase multidisciplinary communication - Interruptions: smartphones lead to increase in interruptions through overall increase in number of calls/messages - Interprofessional relationships: nurses think smartphones reduce face-to-face interactions which are valued; conversely, doctors felt there were no negative implications for team working - Professionalism: using phones during clinical activities seen to be unprofessional with negative perceptions from patients	Fair	5/16
Wu et al 2013[46]	Prospective observational cohort study	General medicine teams in multiple academic hospitals (16 months data collection)—Canada Institutional Blackberry with email/messaging functionality	- Impact on senders: frustrations with pagers (lack of response, wait for call back, no ability to identify caller, often need to re-page, lack of acknowledgement of receipt); benefits of smartphones (quicker resolution, no need to wait by phone, can page and continue to work, acknowledgement of receipt and ability to convey urgency) - Impact on receiver: ability to defer, smartphones facilitate triage and prioritization and make it easier to reply; pagers hugely disruptive due to need to find phone, smartphones disruptive due to increased message/call load; direct voice calls very disruptive	Fair	5/16
Noninterventional Studies—Quantitative Outcomes					
Avidan et al 2017[47]	Cross-sectional observational study	Operating theaters (7 207 min of observation across 52 surgical procedures)—Israel No intervention—impact of mobile phones on interruptions	- 100% of procedures interrupted by phone calls - Median 3 calls/procedure (interquartile range, 2-5 calls) - 0% of incoming calls related to patient undergoing the procedure - 14.7% of calls led to a stoppage of care (mean duration 43.6 s)	Fair	
Ganasegeran et al 2017[48]	Cross-sectional questionnaire study	General/Emergency Medicine (307 multiprofessional participants)—Malaysia No specific intervention—benefits of WhatsApp	- 68.4% perceived WhatsApp to be useful adjunct to clinical practice - 5.6 hours/day on WhatsApp during clinical practice - Common reasons for use: clinical questions, information transfer, instruction giving, patient administration	Fair	

Table 1. continued

Study	Study Design	Setting and Intervention	Key Findings (communication/teamwork)	Quality Assessment[22]	mERA Assessment[23]
Jamal et al 2016[49]	Cross-sectional questionnaire study	17 specialties across 2 large academic teaching hospitals (101 doctor participants)—Saudi Arabia No specific intervention—prevalence and perceptions of mobile phone use	- Those clinicians who have been using WhatsApp for longer and more frequently report greater perceived benefit from its use - 99% of staff mobile phone users - Work-related use: 65.3% text applications and 64.4% voice calls - 98% agree integrating smartphones with hospital systems is a good idea, and 89% say mobiles useful for staff communication - 79% support replacing existing pagers with hospital-provided mobiles - Key issues highlighted: short battery life, distractions caused by mobiles, confidentially and security	Fair	
Martin et al 2016[50]	Cross-sectional questionnaire study	Hospital doctors (206 doctor participants)—United Kingdom No specific intervention—prevalence and perceptions of mobile phone use	- 92% use their personal mobile for work and switchboard holds personal numbers for 64% - 77% discuss patient matters and 12% have sent a photo with PID - 32% contacted on a weekly basis, 21% on a daily basis when not at work - 73% feel pagers should be replaced with mobiles	Poor	
Menzies a*t al* 2012[51]	Cross-sectional questionnaire study	Hospital doctors (850 doctor participants)—New Zealand No specific intervention—prevalence and perceptions of mobile phone use	- 51% of participants use smartphones for work - 26% stored patient data, of which 31% were not password protected - Principal uses: emails/communication, informatics, sharing images - Issues with mobiles: cost, lack of institutional integration, battery life, screen size, user interface, dependency, lack of support, security concerns	Poor	
Mobasheri et al 2015[52]	Cross-sectional questionnaire study	Large academic hospital (718 participants—249 doctors and 469 nurses)—United Kingdom No specific intervention—prevalence and perceptions of mobile phone use	- 98.9% of doctors and 95.1% of nurses own a smartphone - 92.6% of doctors and 53.2% of nurses use a mobile for daily clinical practice - 93.8% of doctors and 28.5% of nurses communicate at work with smartphones, and 50.2% use a smartphone in place of issue pager - 27.5% of doctors and 3.6% of nurses have PID on their phones - 71.6% want a secure messaging platform for identifiable data	Fair	

(continued)

Table 1. continued

Study	Study Design	Setting and Intervention	Key Findings (communication/teamwork)	Quality Assessment[22]	mERA Assessment[23]
O'Connor et al 2014[53]	Cross-sectional questionnaire study	Junior doctors in national training network (108 participants)—Canada No specific intervention—prevalence and perceptions of smartphone use	- 94.4% own a smartphone (67% iPhone, 27% Android) - 83.3% use their smartphone for work-related calls, 87.2% for text messaging, 41.2% for emails, and 52.9% for pictures	Fair	
Prochaska et al 2015[54]	Cross-sectional questionnaire study	Two academic hospitals (132 doctor participants)—United States No specific intervention—prevalence and perceptions of mobile phone use	- 71.7% prefer text messaging to pagers/landlines, with 79.8% using it as their preferred method of communication - 82.5% though existing pagers better for security, but despite this 70.9% have received identifiable data on their mobile	Poor	
Wyber et al 2013[55]	Cross-sectional questionnaire study	Large academic hospital (208 doctors)—New Zealand No specific intervention—prevalence and perceptions of mobile phone use	- 95.7% carried mobile phones at work - Content of messages: clinical management (61%), logistics (55%), social arrangements (42%), results (34%) - Rationale for using mobiles at work: more convenient, less intrusive, less reliable, more efficient, less intimidating - Barriers: cost, ambiguity of communication, reliability, patient confidentiality, impolite/unsocial, slowness, unsure of others use	Fair	

Noninterventional Studies—Qualitative Outcomes

Study	Study Design	Setting and Intervention	Key Findings (communication/teamwork)	Quality Assessment[22]	mERA Assessment[23]
Hsiao and Chen 2012[56]	Cross-sectional questionnaire study	Hospital-based nursing staff (219 participants)—Taiwan No specific intervention—benefits of mNIS	- mNIS systems promote information identification, integration and interpretation - mNIS has a significant positive impact on message exchanges between healthcare professionals, facilitates communication with patients and improves overall performance and quality	Good	
Scholl and Groth 2012[57]	Cross-sectional ethnographic study	Department of surgery in academic hospital (25 participants, 360 h of data collection)—Sweden No specific intervention—ethnographic study of mobile phone use	- Advantages of mobiles over pagers: ease of contact, displays who is calling, no need to find phone for call back, reduced delays in answering - Disadvantages of mobiles: problematic contexts (busy environments, large number of devices, lack of usage policy), nonprofessional image in using in front of patients, interruption of work/life balance with interruptions and ease of contact when not at work - Design for ripple effect: improve awareness that mobiles may impact those not directly involved in the communication (eg, nurses in operating theater)	Good	

(continued)

Table 1. continued

Study	Study Design	Setting and Intervention	Key Findings (communication/ teamwork)	Quality Assessment[22]	mERA Assessment[23]
Wu et al 2014[58]	Cross-sectional ethnographic study	General medicine wards in 5 hospitals with text-based mobile messaging systems (108 interviews, 260 h of observation)—Canada No specific intervention— ethnographic study of text-based mobile messaging systems	- Decontextualization and depersonalization of communication highlighted - Mobile-based systems lead to increasing communication workload and asynchronous communication - Depersonalization of communication is a barrier to effective interprofessional teamwork through reduction in nonverbal	Fair	
Noninterventional Studies—Mixed Methods Outcomes					
Moon and Chang 2014[59]	Cross-sectional questionnaire study	Academic hospital (122 multiprofessional participants)—South Korea No specific intervention— prevalence and perceptions of mobile phone use	- 56.5% use hospital-issued smartphones - 51.4% receive regular work-related calls, 37.5% messages - Attitude toward smartphones influenced by cost, quality, ease of use, support, and security	Fair	
Moore and Jayewardene 2014[60]	Cross-sectional questionnaire study	161 hospital organizations (416 participants—82 nurses, 334 doctors)— United Kingdom No specific intervention— prevalence and perceptions of mobile phone use	- 81% of doctors and 58% of nurses use their smartphones for work - Perceptions of smartphones: easy to use, improve safety, useful, save time - Smartphones improve communication, access to information, efficiency, and decision making - Minority perform a risk assessment before using a phone (eg, for storing using identifiable data)	Poor	
Tran et al 2014[61]	Cross-sectional mixed-methods study	General medicine teams in 4 academic hospitals— Canada No specific intervention— mixed-methods study of mobile phone use	- 59% of respondents carry personal smartphones and use them as their primary method of communication - Acknowledgment of risk to security and confidentiality of information, but respondents favor efficiency and mobility over security - Minority of users observed using personal smartphones at work	Poor	
Wu et al 2013[62]	Cross-sectional ethnographic study	General medicine teams in 5 academic hospitals— Canada No specific intervention— mixed-methods study of mobile phone use	- Pagers are frustrating, slower and deliver less context to the message than smartphones; lack of response to pagers the major frustration - Smartphones make it easier to receive and respond to calls, and coordinate teams, but still highly disruptive; direct calls to phones are very disruptive Impact on privacy and security acknowledged – The use of hospital issued smartphones influences the adoption of informal communication (eg, adding 911 to bleeps). Informal communication methods can cause confusion	Fair	

eHR: electronic health record; MIMS: Monthly Index of Medical Specialties; mNIS: mobile nursing information system.

Table 2. Summary of quality assessment for each study included as per National Institutes of Health Quality Assessment Tools[22] and mERA Checklist[23] compliance for each interventional study

Study	Overall Quality Rating	1— Infrastructure	2— Technology platform	3— Interoperability	4— Intervention delivery	5— Intervention content	6— Usability	7— User feedback	8- Access of participants	9—Cost assessment	10— Adoption inputs	11— Delivery limitations	12— Adaptability	13— Replicability	14— Data security	15— Regulatory compliance	16— Fidelity	TOTAL
Avidan et al 2017[47]	Fair																	
Daruwalla et al 2014[25]	Poor	X	X	–	–	X	–	X	–	–	–	–	–	–	X	X	–	6/16
Duhm et al 2016[26]	Fair	–	X	–	X	–	–	X	X	–	–	–	–	–	X	–	X	6/16
Farrell 2016[39]	Poor	–	X	–	–	X	–	X	X	X	X	–	–	–	–	–	–	6/16
Ganasegeran et al 2017[48]	Fair																	
Gulacti et al 2016[27]	Fair	–	X	–	X	X	–	–	–	X	–	–	–	–	–	–	X	5/16
Hsiao and Chen 2012[56]	Good																	
Jamal et al 2016[49]	Fair																	
Johnston et al 2015[41]	Fair	X	X	–	X	X	–	X	–	X	–	–	–	–	X	–	X	8/16
Khanna et al 2015[28]	Poor	–	X	–	X	X	–	–	–	X	X	–	–	–	X	X	–	7/16
Lane et al 2012[29]	Poor	X	X	X	X	X	X	X	–	–	X	–	X	–	X	–	X	11/16
Lo et al 2012[40]	Fair	–	X	–	–	X	–	X	–	–	X	–	–	–	–	–	–	4/16
Martin et al 2016[50]	Poor																	
Menzies at al 2012[51]	Poor																	
Mobasheri al 2015[52]	Fair																	
Moon and Chang 2014[59]	Fair																	
Moore and Jayewardene 2014[60]	Poor																	
Motulsky et al 2017[30]	Fair	–	X	X	X	X	–	X	–	–	–	–	–	–	–	–	X	6/16
Ng et al 2007[31]	Poor	–	X	–	X	X	–	–	–	–	–	–	–	–	–	–	–	3/16
O'Connor et al 2009[42]	Good	–	X	–	X	X	–	X	–	X	–	–	–	–	X	X	X	8/16
O'Connor et al 2014[53]	Fair																	
Patel et al 2016[32]	Fair	–	X	–	X	X	–	X	X	X	X	–	–	–	X	X	X	10/16
Power et al 2014[33]	Fair	–	X	–	X	X	–	X	–	–	X	–	–	–	–	–	–	5/16
Prochaska et al 2015[54]	Poor																	
Przybylo et al 2014[34]	Good	X	X	–	X	X	–	X	X	–	X	–	–	–	X	X	X	10/16
Quan et al 2013[43]	Poor	–	X	–	X	–	–	X	–	–	X	–	–	–	–	–	X	5/16
Scholl and Groth 2012[57]	Good																	
Smith et al 2012[35]	Poor	–	X	–	X	–	–	–	–	–	–	–	–	–	–	–	X	3/16
Tran et al 2014[61]	Poor																	
Vaisman and Wu 2017[36]	Fair	–	X	–	X	X	–	–	–	–	–	–	–	–	–	–	X	4/16
Wani et al 2013[37]	Poor	–	X	–	X	X	–	X	–	X	X	–	–	–	–	–	–	6/16
Webb at al 2016[44]	Fair	X	X	–	X	X	–	X	–	–	–	–	–	–	–	X	–	6/16
Wu et al 2011[45]	Fair	–	X	–	X	X	–	X	–	–	–	–	–	–	–	–	X	5/16
Wu et al 2013[46]	Fair	–	X	–	–	–	–	X	X	–	–	–	–	–	X	–	X	5/16
Wu et al 2013[62]	Fair																	
Wu et al 2014[58]	Fair																	
Wu et al 2015[38]	Fair	–	X	X	X	X	–	X	–	–	–	–	–	–	–	–	X	6/16
Wyber et al 2013[55]	Fair																	

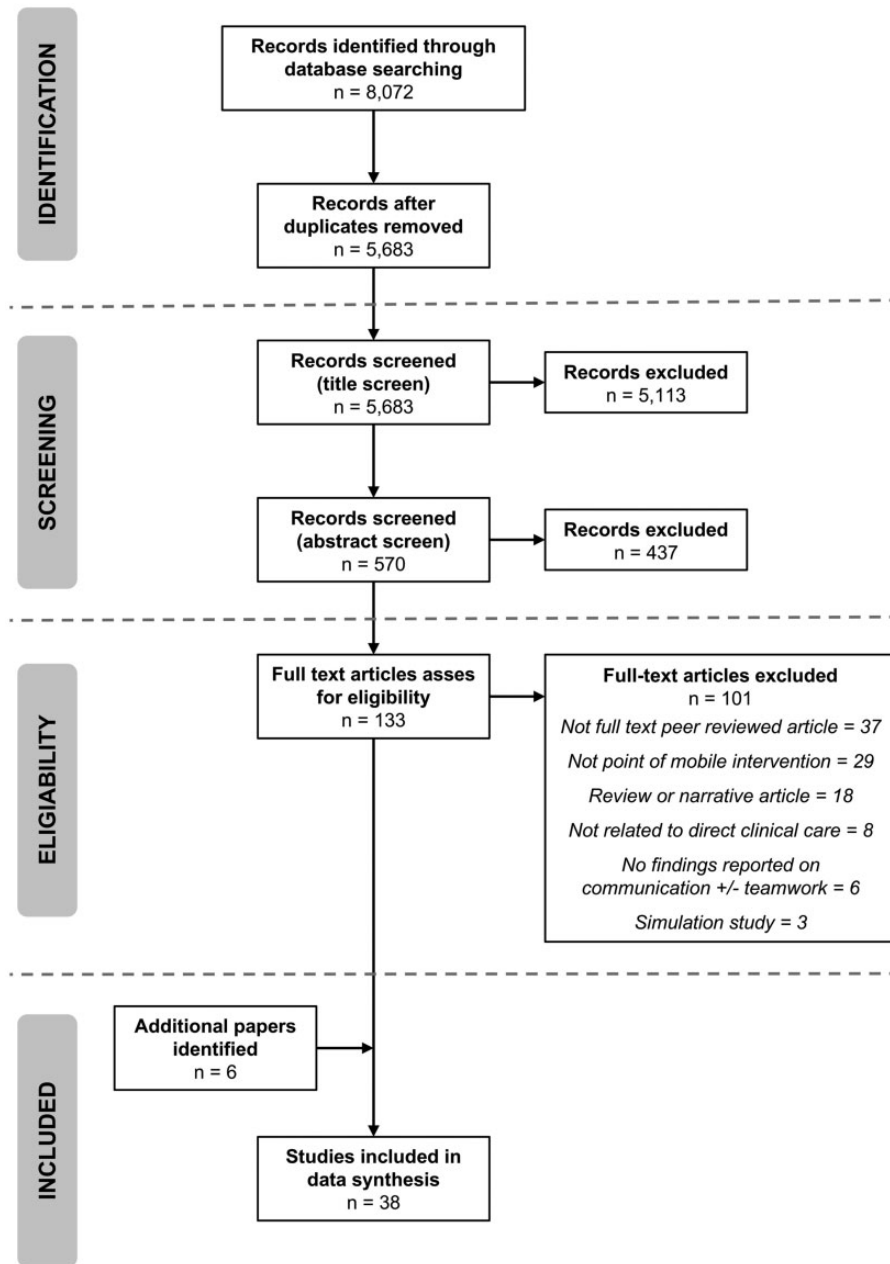**Figure 1.** PRISMA Diagram of study identification, screening, and inclusion.

Interventional studies

Description of studies

Twenty-two interventional studies—those with a specific technology deployed for the purpose of evaluation—were identified. Of these 14 reported quantitative outcomes,[25–38] 2 were qualitative outcomes[39,40] and a further 6 were mixed-methods outcomes.[41,43–46] Overall, the interventional study designs adopted were heterogeneous, with only 1 study involving any form of randomization,[34] and a further single study employing a crossover study design[26]; all other studies otherwise took the form of uncontrolled cohort studies. The populations studied were varied, but importantly were of limited size, with the mean number of participants being only 63 (range, 8-210). Seventeen discrete interventions were available for comparison; 8 bespoke mobile applications,[25,26,29,30,32,34,38,44]

4 WhatsApp messenger (WhatsApp Inc, Menlo Park, CA) services,[27,28,37,41] 3 generic smartphones,[31,33,39] and the remaining 2 interventions involved smartphones with a specific messaging or email communication functionality that were reported across multiple studies.[35,36,40,43,45,46] The 14 studies reporting quantitative results utilized a range of methodologies with all but 2[29,36] using questionnaires, and 7 using content analysis of mobile phone data.[27,30,32,35–38] Two studies used direct observational data, with one assessing the time taken to complete handover[28] and the other the speed and latency of communication.[29] Two further studies reported qualitative outcomes, with one using semistructured interviews and focus groups[39] and the other using an exploratory case study approach.[40] Six studies adopted a mixed-methods approach, with all including content analysis of messages sent or received

during the study period, 4 including additional structured interviews,[41,43,45,46] 2 including questionnaires,[42,44] and 2 including more direct observation.[45,46] Overall, the quality of studies as judged through compliance with the mERA Checklist[23] was poor, with a mean score of 6.1 of 16 (range, 3-11), and no study was fully compliant. The studies assessed a variety of mobile interventions with a range of cross-cutting themes being evident: improvements in workflow, efficiency, and the quality of communication; improvements in accessibility and interteam relationships; and the near universal acceptance that mobile devices should replace current methods of communication despite some key limitations being identified.

Workflow, efficiency and quality of communication

Broadly speaking, the introduction of mobile devices led to improvements in workflow, efficiency, and the quality of communication. A number of papers reported significant streamlining of clinical workflows and improvements in the quality of clinical discussion,[26,34] improvements in handover and patient care,[30] faster response times,[33,45] and the elimination of redundant steps in vertical communication within teams.[37] Significant improvements in the effectiveness of communication with greater efficiency and integration into existing workflows[34] and improvements in the quality of information transfer and recall[28] were also demonstrated. A further study reported that smartphones created additional value by facilitating the easy delivery of nonurgent information while also supporting the triage, prioritization, and timeliness of communication.[40] Some studies looked to quantify these improvements in efficiency and timeliness with a mean response time of 2-3 minutes with mobile devices,[27,38,41] and 1 study reported that >50% of email messages sent by smartphone were read in <1 minute.[32] Meanwhile, the use of mobile applications led to significantly less disruption to clinical workflows,[32] improvements in the speed of communication,[29] and significant reductions in response times, from 5.5 to 3 minutes.[44]

Accessibility and interteam relationships

In addition to improved efficiency and quality of communication the use of mobile devices also had a positive impact on accessibility, interprofessional interactions, and the involvement of senior decision makers in clinical care.[31,32,39] Many of the positive impacts of better communication on team relationships were highlighted in the previous section; however, improved accessibility and ease of communication can also be highly interruptive. One study identified an average of 42-51 interruptions per day and 35 minutes a day where the level of interruptions reached a potentially dangerous level.[36] Other studies identified that doctors frequently felt that they were regularly interrupted with low-value and unnecessary information[38] and that they were often overwhelmed by the volume of interruptions caused by their mobile device.[45] A further study identified that the introduction of mobile devices led to a large increase in the number of messages sent and a subsequent 233% increase in interruptions.[43] This increased communication burden may account for the 50% of messages that do not get a response.[35] Increases in the communication burden may also lead to the depersonalization of the clinical team. Nurses reported feeling that mobiles negatively impact interprofessional relationships via a reduction in the face-to-face interactions that they value in helping to build relationships; conversely, doctors felt this was a positive change.[45] One study reported that doctors felt the frequent interruptions they received were often inappropriate given the content and context,[38] and another found

interprofessional conflict due to the different subjective assessment of the urgency and priority of messages.[40] A further study reported that increased messaging by nurses to seek accountability and reassurance was perceived as an attempt to absolve themselves of responsibility by doctors who felt that nurses often exaggerated the severity or urgency of a issues to illicit a response.[43]

Limitations and professionals' views of mobile technology

In addition to the many positive influences reported, there were also some negative consequences of mobile devices identified. The physical limitations of mobile devices was commonly highlighted as a weakness, with small screen size, poor battery life, the requirement to enter a password on a regular basis, and unreliable connectivity all identified as limiting their effectiveness.[33,39,44] In addition to their practical limitations, mobile devices were also reported to be regarded as less effective than face-to-face or direct communication for complex patient issues,[38] potentially giving an unprofessional appearance if used at the bedside[39] and often used inappropriately for personal non–work-related reasons.[42] Despite these negative reports, there was universal agreement that the use of mobile devices acted to remove barriers to effective communication. In one study, 87% of participants wanted to continue using their devices at the end of the study period,[42] while in another the majority of users stated that they would prefer to access information and communicate through a smartphone.[30] A total of 85% of participants recommended the widespread use of mobiles,[34] and 92% agreed that mobile applications should replace traditional pagers and there is significant potential for the greater integration of mobiles in the hospital setting.[25]

Noninterventional studies

Description of studies

Sixteen noninterventional studies were identified. Of these 9 reported quantitative outcomes,[47–55] 3 reported qualitative outcomes,[56–58] and a further 4 reported mixed-methods outcomes.[59–62] All 16 studies adopted a cross-sectional study design, with 11 questionnaire-based studies,[48–56,59,60] 3 ethnographic study designs,[57,58,62] and 1 purely observational study.[47] The final study used a mixed-methods approach combining direct observation, interviews, and questionnaire data.[61] This group of noninterventional studies sampled a larger population with a mean number of participants of 220 (range, 25-718).

Fifteen of the studies looked to evaluate the prevalence, perception, or use of mobile technology on communication in hospitals, with a further study specifically characterizing the impact of mobile phones on interruptions in the operating theater.[47] Key findings from these studies were consistent; namely, the ubiquitous use of mobile technology by healthcare professionals, the predominance of personal devices being used for work-related activity, the clear benefits that mobile-based technologies bring despite well-articulated negative consequences, the potential risks to patient confidentiality and security, and the broad support for the formal adoption of mobile technologies by healthcare institutions.

Prevalence of mobile technology in hospitals

Mobile technologies are used on a daily basis by the vast majority of healthcare professionals. Doctors use their personal devices at work more frequently than other healthcare professionals do, with up to 95% reporting regular daily use and sharing of their personal number with other members of staff[50,52,55] compared with only around

50% of nurses.[52,60] The messaging and email functionality of mobile devices was consistently highlighted as the principal reason for their use. One study reported that around 65% of staff use text applications,[49] and another found that up to 88% use messaging or e-mails,[50] and a further study found that 87% of staff use text messaging and a further 41% emails[53] while at work. Indeed, 72% of staff prefer text messaging to traditional pagers, 80% cite it as their preferred method of communication,[54] and 68% believe that WhatsApp is a useful adjunct to clinical practice.[48] There were a number of advantages to be gained with the use of mobile communication devices, such as the ease of contact, ability to see who is calling, and reduced delays in answering.[57] Another study highlighted the promotion of better information identification, integration, and interpretation and the positive impact of this on overall performance and quality.[56] Further studies found mobile devices to be more convenient, less intrusive, more efficient, and less intimidating than traditional methods of communication,[55] while also helping deliver better context to messages and facilitating the easy coordination of teams[62] and enhancing access to information and improving decision-making.[60]

Negative impact of mobile technology

In addition to the benefits that mobile devices may bring, there were also a number of negative consequences identified. Studies described issues with mobile devices including the cost, lack of institutional integration and support, poor battery life, reliability, and small screen size.[49,51,55] The use of mobile phones was also seen as promoting a nonprofessional image and appearing rude or impersonal when used in front of patients.[55,57] One study described how the use of mobile devices depersonalizes and decontextualizes communication and introduces informal work-arounds compared with direct methods of communication such as face-to-face interactions or voice calls.[58] It was also observed that mobile devices can lead to unwanted ripple effects such as disturbing nurses in the operating theater, or by increasing the unwanted contact of doctors when not at work.[57] Indeed, one-third of doctors are contacted on a weekly basis, and over 1 in 5 on a daily basis when not at work.[50]

Patient confidentiality and data security

Importantly, a number of studies identified the potential risk to security and confidentiality of patient information with the use of personal devices.[49,51,54,55,59,61] Despite these security concerns, staff favor efficiency and mobility over security,[61] with only a minority performing any form of security risk assessment[60] and one-third of devices not password protected.[51] Crucially, 71% of staff have received[54] and a further 28% regularly store confidential patient information on their personal device.[52] Despite the potential negative consequences of mobile devices, the vast majority of studies found that clinical staff advocate their use and strongly support their wider deployment. One study reported that the overwhelming majority of clinical staff think mobile devices and secure messaging platforms should be integrated with current hospital systems and that existing pagers should be replaced with hospital-issued mobile phones.[49,50,52]

DISCUSSION

Delivering high-quality, safe healthcare is a complex endeavor requiring the careful and precise coordination of numerous professionals in the care of a single patient. This review has found that overall there is a lack of high-quality evidence evaluating the impact of mobile technologies on communication and teamwork in hospital settings. Only 11% of studies were deemed to be of high quality, no study complied with best practice guidelines for the conduct and reporting of trials involving mobile technology, and all examined small populations in restricted environments that do not truly represent complex real-world settings. Importantly, no studies sought to examine the impact on meaningful patient outcomes. Despite the relative lack of evidence, this review supports the assertion that mobile technology has the potential to significantly improve communication and teamwork within hospitals provided that concerns over the evolution of negative communication behaviors, technological flaws, and security and privacy concerns are adequately addressed and that greater evidence for safety and efficacy is delivered.

Mobile technology is ubiquitous across the world. This review has shown that these technologies are valued by healthcare staff for being more convenient and are preferred to existing modes of communication such as traditional pagers. They may act to improve and streamline clinical workflows and boost the efficiency and quality of communication. Mobile technology may also act to increase the accessibility and responsiveness of staff, improve interprofessional teamwork and relationships, and enhance access to information and better decision making. The review has also highlighted that that the negative aspects of mobile technology must be carefully considered. Clear physical and technological limitations have been identified including poor battery life, small screen size, unreliable connectivity, and the lack of consistent integration with other hospital systems. Making communication easier may result in a large increase in the communication burden that could stem from the elimination of traditional communication barriers such as the need to wait for a phone, the impersonal nature of message-based communication, and flattening of hierarchal team structures. This increased communication burden can lead to potentially harmful disruptions to care, cognitive overload, and conflict. It is crucial to align the content and purpose of a message against the process and mode of communication to mitigate against these risks.

One barrier to the adoption of mobile technology is the lack of high-quality evidence that supports the new investment hospitals need to make. It is difficult to draw clear conclusions due to methodological inadequacies including the lack of prospective randomization or assessment of matched comparator groups, the limited number of participants and truncated study lengths, and an inability to effectively pool results from multiple studies due to the substantial variability in methodologies and outcomes used. The majority of studies were based in single centers and the populations evaluated were small. Twenty-six of the studies included some form of questionnaire-based data collection yet only 6[30,42,49,52,56,59] discussed validity testing of the questionnaires used. While some of these methodological flaws may be put down to the inherent difficulty of assessing such interventions in complex hospital settings, few studies clearly set out to try and overcome these challenges in a meaningful way. Of the 22 interventional studies reviewed, only 2[26,34] had any form or randomization or prospective assessment of matched comparator groups, and in the remainder only 5[28,29,32,43,44] made reference to preintervention baseline data against which the mobile intervention was compared.

Despite the pervasive use of mobile technology outside of work, there are a number of diverse organizational, individual, and technological factors that are likely to impact the adoption of new communication technologies. The failure to adopt new technologies may be caused by a failure to plan for the complexity and cost, not gaining

buy-in and engagement from end users or failing to appreciate that new technology changes the work, the nature of work, and who does that work.[63] Additional technical, financial, legal, social, and ethical factors have also been identified that prevent the widespread uptake of new technologies.[64,65] In addition to these structural factors, it has also been suggested that the extent to which mobile devices deliver value is unclear and there is a need address explicit questions about how mobile technology will deliver real benefit.[66] However, it has been estimated that the use of mobile technology in healthcare has the potential to significantly improve productivity and reduce costs.[67] There is a need to promote the positives of a "mobile-first" culture within healthcare organizations and provide the required leadership and resource to deliver it while being cognizant of the potential risks. This focus must come hand-in-hand with a need to target future research on understanding the broader sociotechnical aspects of new mobile technology, and how it complements and enables new pathways and processes of care to improve outcomes for patients and the working life of staff.

Concerns of privacy and security were highlighted in this review, particularly when personal mobile devices are used for the transmission of patient identifiable data. In both the United States[68] and Europe[69] the need to comply with stringent legislation has undoubtedly limited the deployment of smartphone-based messaging, and the use of SMS messaging for in-hospital communication has been discouraged by the Joint Commission due to security concerns.[70] Improving the awareness and training of staff with regard security and privacy hand-in-hand with developing security compliant technology has the potential to greatly accelerate the uptake of new mobile technologies. Many of the negative aspects of mobile devices relate to the technology itself including poor battery life, small screen size, and lack of connectivity. To address these concerns, there is a need to design and develop technology specifically for the healthcare context and adapt work practices to alleviate some of these technological limitations. As devices become increasingly complex and data heavy, the importance of the underlying supporting infrastructure that is needed to securely and reliability store, process, and transmit huge volumes of clinical and communication data becomes increasingly important.[71]

CONCLUSION

Healthcare professionals use innovative mobile technology on a daily basis outside of work, yet have to cope with outdated and inadequate technology to coordinate and deliver care at work. Mobile technology can deliver very real benefits, but there is a lack of high-quality evidence, and the poor experience of institutional technology results in the development of a potentially harmful patchwork of informal workarounds and ad hoc technology adoptions. An evidence-based approach to the development, deployment and evaluation of new mobile communication devices is therefore required. To secure the "right" technology it is important to recognize and understand both the advantages and disadvantages of any particular technology and how it is used in real-world settings. Mobile technology has the potential to transform communication and teamwork in hospitals and deliver very real benefits provided a pragmatic and evidence-based approach is taken to its design, deployment and evaluation.

REGISTRATION

The review protocol was prospectively registered with the PROSPERO Database (CRD42017064128).

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

Conflict of interest statement. None declared.

REFERENCES

1. Ash J, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc* 2003; 11 (2): 104–12.
2. Coiera E. When conversation is better than computation. *J Am Med Inform Assoc* 2000; 7 (3): 277–86.
3. Sutcliffe K, Lewton E, Rosenthal M. Communication failures: an insidious contributor to medical mishaps. *Acad Med* 2004; 79 (2): 186–94.
4. Alvarez G, Coiera E. Interdisciplinary communication: an uncharted source of medical error? *J Crit Care* 2006; 21 (3): 236–42.
5. Zinn C. 14,000 preventable deaths in Australian hospitals. *Br Med J* 1995; 310 (6993): 1487–8.
6. The Joint Commission's sentinel event policy: ten years of improving the quality and safety of health care. *Jt Comm Perspect* 2005; 25 (1): 3–5.
7. Tipping MD, Forth VE, O'Leary KJ, et al. Where did the day go?—a time-motion study of hospitalists. *J Hosp Med* 2010; 5 (6): 323–8.
8. Patel S, Lee J, Ranney D, et al. Resident workload, pager communications and quality of care. *World J Surg* 2010; 34 (11): 2524–9.
9. Parker J, Coiera E. Improving clinical communication: a view from psychology. *J Am Med Inform Assoc* 2000; 7 (5): 453–61.
10. Bailey B, Konstan J. On the need for attention-aware systems: measuring effects of interruption on task performance, error rate and affective state. *Comput Hum Behav* 2006; 22 (4): 685–708.
11. Westbrook J, Woods A, Rob M, et al. Association of interruptions with increased risk and severity of medication administration errors. *Arch Intern Med* 2010; 170 (8): 683–90.
12. Nguyen T, Battat A, Longhurst C, et al. Alphanumeric paging in an academic hospital setting. *Am J Surg* 2006; 191 (4): 561–5.
13. Espino A, Cox D, Kaplan B. Alphanumeric paging: a potential source of problems in patient care and communication. *J Surg Educ* 2011; 68 (6): 447–51.
14. Coiera E, Tombs V. Communication behaviours in a hospital setting: an observational study. *J Br Med* 1998; 316: 673–6.
15. Westbrook JI, Coiera E, Dunsmuir WTM, et al. The impact of interruptions on clinical task completion. *Qual Saf Health Care* 2010; 19 (4): 284–9.
16. Weigl M, Muller A, Vincent C, et al. The association of workflow interruptions and hospital doctors' workload: a prospective observational study. *BMJ Qual Saf* 2012; 21 (5): 399–407.
17. Agarwal R, Sands D, Schneider J. Quantifying the economic impact of communication inefficiencies in U.S. hospitals. *J Healthc Manag* 2010; 55 (4): 265–81.
18. Wu R, Tran K, Lo V, et al. Effects of clinical communication interventions in hospitals: a systematic review of information and communication technology adoptions for improved communication between clinicians. *Int J Med Inform* 2012; 81 (11): 723–32.

19. McElroy L, Ladner D, Holl J. The role of technology in clinician-to-clinician communication. *BMJ Qual Saf* 2013; 22 (12): 981–3.

20. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-anayses: the PRISMA statement. *Ann Intern Med* 2009; 151 (4): 264–70.

21. Centre for Reviews and Dissemination. *PROSPERO—International Prospective Register of Systematic Reviews*. 2017. Available at: https://www.crd.york.ac.uk/prospero/. Accessed Jan 29, 2018.

22. National Institutes of Health. *National Institutes of Health Study Quality Assessment Tools*. 2017. Available at: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools. Accessed August 23, 2018.

23. Agarwal S, LeFevre A, Lee J, et al.; WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *Br Med J* 2016; 352: i1174.

24. Schulz K, Altman D, Moher D; CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med* 2010; 24(8):18.

25. Daruwalla Z, Wong K, Thambiah J. The application of telemedicine in orthopedic surgery in singapore: a pilot study on a secure, mobile telehealth application and messaging platform. *J Med Internet Res* 2014; 2 (2): e28.

26. Duhm J, Fleischmann R, Schmidt S, et al. Mobile electronic medical records promote workflow: physicians' perspective from a survey. *JMIR Mhealth Uhealth* 2016; 4 (2): e70.

27. Gulacti U, Lok U, Hatipoglu S, et al. An analysis of WhatsApp usage for communication between consulting and emergency physicians. *J Med Syst* 2016; 40: 130. doi:10.1007/s10916-016-0483-8.

28. Khanna V, Sambandam S, Gul A, et al. WhatsApp"ening in orthopedic care: a concise report from a 300-bedded tertiary care teaching center. *Eur J Orthop Surg Traumatol* 2015; 25 (5): 821–6.

29. Lane J, Sandberg W, Rothman B. Development and implementation of an integrated mobile situational awareness iPhone application VigiVU at an academic medical center. *Int J CARS* 2012; 7 (5): 721–35.

30. Motulsky A, Wong J, Cordeau J-P, et al. Using mobile devices for inpatient rounding and handoffs: an innovative application developed and rapidly adopted by clinicians in a pediatric hospital. *J Am Med Inform Assoc* 2017; 24: e69–78.

31. Ng WH, Wang E, Ng I. Multimedia messaging service teleradiology in the provision of emergency neurosurgery services. *Surg Neurol* 2007; 67 (4): 338–41.

32. Patel N, Siegler J, Stromberg N, et al. Perfect Storm of inpatient communication needs and an innovative solution utilizing smartphones and secured messaging. *Appl Clin Inform* 2016; 7: 777–89.

33. Power J, Spina S, Forbes D, et al. Integration of smartphones into clinical pharmacy practice: an evaluation of the impact on pharmacists' efficiency. *Health Policy Technol* 2014; 3 (4): 296–305.

34. Przybylo J, Wang A, Loftus P, et al. Smarter hospital communication: secure smartphone text messaging improves provider satisfaction and perception of efficacy, workflow. *J Hosp Med* 2014; 9 (9): 573–8.

35. Smith C, Quan S, Morra D, et al. Understanding interprofessional communication: a content analysis of email communications between doctors and nurses. *Appl Clin Inform* 2012; 3: 38–51.

36. Vaisman A, Wu R. Analysis of smartphone interruptions on academic general internal medicine wards. *Appl Clin Inform* 2017; 8: 1–11.

37. Wani S, Rabah S, AlFadil S, et al. Efficacy of communication amongst staff members at plastic and reconstructive surgery section using smartphone and mobile WhatsApp. *Indian J Plast Surg* 2013; 46 (3): 502–5.

38. Wu R, Lo V, Morra D, et al. A smartphone-enabled communication system to improve hospital communication: usage and perceptions of medical trainees and nurses on general internal medicine wards. *J Hosp Med* 2015; 10 (2): 83–9.

39. Farrell M. Use of iPhones by Nurses in an acute care setting to improve communication and decision-making processes: qualitative analysis of nurses' perspectives on iPhone yse. *JMIR Mhealth Uhealth* 2016; 4 (2): e43.

40. Lo V, Wu R, Morra D, et al. The use of smartphones in general and internal medicine units: a boon or a bane to the promotion of interprofessional collaboration? *J Interprof Care* 2012; 26 (4): 276–82.

41. Johnston MJ, King D, Arora S, et al. Smartphones let surgeons know WhatsApp: an analysis of communication in emergency surgical teams. *Am J Surg* 2015; 209 (1): 45–51.

42. O'Connor C, Friedrich J, Scales D, et al. The use of wireless e-mail to improve healthcare team communication. *J Am Med Informatics Assoc* 2009; 16: 705–13.

43. Quan S, Wu R, Rossos P, et al. It's not about pager replacement: an in-depth look at the interprofessional nature of communication in healthcare. *J Hosp Med* 2013; 8 (3): 137–43.

44. Webb C, Spina S, Young S. Integrating smartphone communication strategy and technology into clinical practice: a mixed methods research study. *Health Policy Technol* 2016; 5 (4): 370–5.

45. Wu R, Rossos P, Quan S, et al. An evaluation of the use of smartphones to communicate between clinicians: a mixed-methods study. *J Med Internet Res* 2011; 13 (3): e59.

46. Wu R, Tzanetos K, Morra D, et al. Educational impact of using smartphones for clinical communication on general medicine: more global, less local. *J Hosp Med* 2013; 8 (7): 365–72.

47. Avidan A, Yacobi G, Weissman C, et al. Cell phone calls in the operating theater and staff distractions: an observational study. *J Patient Saf* 2017 Jan 9. doi:10.1097/PTS.0000000000000351.

48. Ganasegeran K, Renganathan P, Rashid A, et al. The m-Health revolution: exploring perceived benefits of WhatsApp use in clinical practice. *Int J Med Inform* 2017; 97: 145–51.

49. Jamal A, Temsah M-H, Khan SA, et al. Mobile phone use among medical residents: a cross-sectional multicenter survey in Saudi Arabia. *J Internet Res*. *JMIR Mhealth Uhealth* 2016; 4 (2): e61.

50. Martin G, Janardhanan P, Withers T, et al. Mobile revolution: a requiem for bleeps? *Postgrad Med J* 2016; 92 (1091): 493–6.

51. Menzies O, Thwaites J. A survey of personal digital assistant use in a sample of New Zealand doctors. *N Z Med J* 2012; 125 (1352): 48–59.

52. Mobasheri M, King D, Johnston M, et al. The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study. *BMJ Innov* 2015; 1 (4): 174. doi:10.1136/bmjinnov-2015-000062.

53. O'Connor P, Byrne D, Butt M, et al. Interns and their smartphones: use for clinical practice. *Postgrad Med J* 2014; 90: 75–9.

54. Prochaska M, Bird A-N, Chadaga A, et al. Resident use of text messaging for patient care: ease of use or breach of privacy? *JMIR Med Inform* 2015; 3: e37.

55. Wyber R, Khashram M, Donnell A, et al. The Gr8est good: use of text messages between doctors in a tertiary hospital. *J Commun Healthc* 2013; 6 (1): 29–34.

56. Hsiao J, Chen R. An investigation on task-technology fit of mobile nursing information systems for nursing performance. *Comput Inform Nurs* 2012; 30 (5): 265–73.

57. Scholl J, Groth K. Of organization, device and context: interruptions from mobile communication in highly specialized care. *Interact Comput* 2012; 24 (5): 358–73.

58. Wu R, Appel L, Morra D, et al. Short message service or disService: Issues with text messaging in a complex medical environment. *Int J Med Inform* 2014; 83 (4): 278–84.

59. Moon B, Chang H. Technology acceptance and adoption of innovative smartphone uses among hospital employees. *Healthc Inform Res* 2014; 20 (4): 304–12.

60. Moore S, Jayewardene D. The use of smartphones in clinical practice. *Nurs Manage* 2014; 21 (4): 18–22.

61. Tran K, Morra D, Lo V, et al. The use of smartphones on general internal medicine wards: a mixed methods study. *Appl Clin Inform* 2014; 5 (3): 814–23.

62. Wu R, Lo V, Morra D, et al. The intended and unintended consequences of communication systems on general internal medicine inpatient care delivery: a prospective observational case study of five teaching hospitals. *J Am Med Inform Assoc* 2013; 20 (4): 766–77.

63. Wachter R. *Making IT Work: Harnessing the Power of Health IT to Improve Care in England*. 2016. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf. Accessed August 23, 2018.

64. Stroetmann K, Artmann J, Dumortier J, *et al*. United in diversity: legal challenges on the road towards interoperable eHealth eolutions in Europe. *Eur J Biomed Inform* 2012; 8: 3–10.

65. Currie W, Seddon J. A cross-national analysis of eHealth in the European Union: Some policy and research directions. *Inf Manag* 2014; 51 (6): 783–97.

66. Prgomet M, Georgiou A, Westbrook J. The impact of mobile handheld technology on hospital physicians' work practices and patient care: a systematic review. *J Am Med Inform Assoc* 2009; 16 (6): 792–801.

67. Thomairy N, Mummaneni M, Alsalamah S, *et al*. Use of smartphones in hospitals. *Health Care Manag (Frederick)* 2015; 34 (4): 297–307.

68. US Department of Health and Human Services. *Summary of the HIPAA Privacy Rule*. 1996. Available at: http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/. Accessed Jan 31, 2018.

69. Information Commissioner's Office. *Overview of the General Data Protection Regulation (GDPR)*. 2017. Available at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/. Accessed May 22, 2017.

70. The Joint Commission. *To Text or Not to Text?* 2015. Available at: https://www.jointcommission.org/clarification_use_of_secure_text_messaging/. Accessed Jan 31, 2018.

71. Baig M, GholamHosseini H, Moqeem A, *et al*. Clinical decision support systems in hospital care using ubiquitous devices: current issues and challenges. *Health Informatics J* 2017 Nov 1. doi:10.1177/1460458217740722.

Systematic Reviews

PROTOCOL   **Open Access**

Protocol for a systematic review and meta-analysis on the clinical outcomes and cost of deep inferior epigastric perforator (DIEP) flap versus implants for breast reconstruction

Ankur Khajuria[1,2]*, Oliver J. Smith[2], Maxim Prokopenko[2], Maximillian Greenfield[2] and Afshin Mosahebi[2]

Abstract

Background: Mastectomy in the context of breast malignancy can have a profoundly negative impact on a woman's self-image, impairing personal, sexual and social relationships. The deep inferior epigastric perforator (DIEP) flap and implants are the two commonest reconstructive modalities that can potentially overcome this psychological trauma. The comparative data on clinical outcomes and costs of the two modalities is limited. We aim to synthesise the current evidence on DIEP versus implants to establish which is the superior technique for breast reconstruction, in terms of clinical outcomes and cost-effectiveness.

Methods: A comprehensive search will be undertaken of EMBASE, MEDLINE, Google Scholar, CENTRAL and Science citation index databases (1994 up to August 2017) to identify studies relevant for the review. Primary human studies evaluating clinical outcomes and cost of DIEP and implant-based reconstruction in context of breast malignancy will be included. Primary outcomes will be patient satisfaction and cosmetic outcome from patient-reported outcome measures (scores from validated tools, e.g. BREAST-Q tool), complications and cost-analysis. The secondary outcomes will be duration of surgery, number of surgical revisions, length of stay, availability of procedures and total number of clinic visits.

Discussion: This will be the first systematic review and meta-analysis in available literature comparing the clinical outcomes and cost-effectiveness of DIEP and implants for breast reconstruction. This review is expected to guide worldwide clinical practice for breast reconstruction.

Systematic review registration: PROSPERO CRD42017072557.

Keywords: Breast implant, DIEP, Cost-effectiveness, Autologous flap reconstruction, Deep inferior epigastric artery perforator flap

* Correspondence: ankur.khajuria09@imperial.ac.uk
[1]Department of Surgery and Cancer, Imperial College London, London, UK
[2]Department of Plastic & Reconstructive Surgery, Royal Free Hospital NHS Trust, London, UK

Khajuria *et al. Systematic Reviews* (2017) 6:232

Page 2 of 5

Background

Breast cancer is the commonest malignancy in women and a major cause of cancer-related mortality [1]. While mastectomy is the primary treatment modality for these patients, it can have a profoundly negative impact on their lives, impairing personal, sexual and social relationships. Fifty percent of women post-mastectomy suffer from negative self-image with negative changes in their sexuality [2, 3]. The demand for reconstructive procedures has risen, not only as a consequence of advancing cancer treatment but also because of the demonstrated functional, psychological and social benefits for patients, overcoming the psychological trauma associated with mastectomy [4–9]. The rates of post-mastectomy breast reconstruction doubled from 13 to 26% between 1998 and 2007 [10].

A number of reconstructive techniques exist for breast reconstruction. The two most frequently employed techniques include the autologous deep inferior epigastric perforator (DIEP) flap and implant-based reconstruction [11]. The choice of treatment is determined by combination of patient factors (individual preference, age, body image) and surgeon factors (team experience, availability of resources) [8, 9]. Despite this, many plastic surgery units worldwide regard autologous flap reconstruction as the superior technique as it follows the paradigm of replacing 'like with like' [10]. Indeed, there is growing evidence to support increased aesthetic patient satisfaction with autologous flap reconstruction compared to implants, as well as increased suppleness and resiliency, especially in irradiated recipient beds [11–19].

On first glance, implant-based reconstruction is a simpler technique compared to free flap reconstruction, requires less training and time and can be performed by many more surgeons [15]. However, implant-based reconstruction has complications. These include migration, implant rupture, infection, exposure/extrusion and patient dissatisfaction with edge visibility and implant animation [20]. Capsular contracture can result in pain, asymmetry, increased palpability and requirement of implant removal [21]. The placement of the implant itself can lead to reduced or absent sensation at the nipple in 1 in 7 women [20]. Allergan's 10-year cumulative risk study found that 24.6% of patients who underwent implant-based reconstruction developed capsular contracture necessitating implant removal and/or replacement [22].

Conversely, DIEP flap is often now considered the gold-standard autologous flap reconstructive technique. This is because it results in less abdominal donor site morbidity compared to the traditional transverse rectus abdominus myocutaneous (TRAM) flap, by preserving the continuity of the rectum abdominis muscles [23]. Compared to implant-based reconstruction, some authors have argued that DIEP flap reconstruction is more cost-effective and results in fewer complications [11, 24]. Modern healthcare aims to provide cost-effective treatment, and thus, discussion on reconstructive modalities warrants scrutiny on cost associated with autologous and implant-based reconstruction. While some North American and European centres have published cost-effectiveness analysis on DIEP versus implants, the data is sparse and there is a relative scarcity of inclusion of data from public and free universal healthcare system settings [11].

Thus, an extensive search will be undertaken in the MEDLINE (Ovid SP), EMBASE (OvidSP), Google Scholar, Cochrane Central Register of Controlled Trials and Scientific Citation Index databases to identify primary studies on DIEP (intervention) and implant-based reconstruction (comparator) in context of patients with breast malignancy. Data extracted will be used to evaluate which technique is superior in terms of clinical outcomes and cost and thus inform worldwide clinical practice.

Methods

Objective

This systematic review is intended to evaluate the current evidence on the clinical outcomes and cost of deep inferior epigastric perforator (DIEP) flap versus implants for breast reconstruction post cancer-related mastectomy, to determine which technique is more cost-effective and clinically superior.

General methods

This protocol has been registered with the National Institute of Health Research (NIHR) Prospective Register of Systematic Reviews PROSPERO CRD42017072557. We have adhered to and completed the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement [25] (please see Additional file 1, PRISMA-P checklist). If no randomised controlled trials (RCT) are available, the review will be reported according to the Meta-Analysis of Observational Studies in Epidemiology (MOOSE) guidelines [26].

Search strategies

We will conduct a comprehensive search of the MEDLINE (OVID SP), EMBASE (OVID SP), Google Scholar, CENTRAL and Science citation index databases from 1994 up to August 2017 to identify studies relevant to the review. A combination of Medical Subject Headings (MeSH) terms and free text will be used, combined with Boolean logical operators to construct the search strategy. Explode function will be used to capture narrower terms. No language or study design restrictions will be applied. The reference list of all included articles will also be screened for relevance. A

Khajuria *et al. Systematic Reviews* (2017) 6:232

Page 3 of 5

sample search strategy, for EMBASE (OVID SP), is shown below and a similar strategy will be adapted for the other databases:

- (1) exp Breast Neoplasms/ OR ((breast adj6 cancer*) or (breast adj6 neoplasm*) or (breast adj6 carcinoma*) or (breast adj6 tumour*) or (breast adj6 tumor*) or (breast* adj4 reconstruct*))
- (2) exp deep inferior epigastric perforator flap/ OR DIEP flap* OR DIEAP flap* OR ((Deep and inferior and epigastric and perforator) adj2 flap*) OR Deep and inferior and epigastric and perforator and flap*)
- (3) exp breast implant/ OR breast adj3 implant* OR exp silicone prosthesis/
- [(1) AND (2)] OR [(1) AND (3)]; publication date: January 1994–August 2017

Identification and selection of studies

Studies extracted following database searching will be populated into an Endnote X7 library (Clarivate Analytics, USA). The screening will be carried out in two stages using pre-specified screening criteria by two independent reviewers. Inter-rater reliability will be calculated using Cohen's kappa score.

Stage 1: Title and abstract screening will be carried out by two researchers acting independently. Any discrepancies will be resolved by consensus. If any doubt remains, the article would usually proceed to full-text review.

Stage 2: The full-text versions of the studies included in Stage 1 will be downloaded and screened for eligibility by two researchers independently. Discrepancies will again be resolved by consensus. If this is not possible, a third author will be consulted.

Study design

All primary human studies evaluating clinical outcomes and cost of DIEP and implant-based reconstruction in context of breast malignancy will be included. The intervention is the DIEP flap and the comparator is implant-based reconstruction. The inclusion and exclusion criteria are highlighted below.

Inclusion criteria

a Studies involving adult patients between 18 and 90 years old.

b Unilateral DIEP flap or implant-based breast reconstruction due to breast cancer (bilateral reconstruction is usually after bilateral prophylactic mastectomy).

c Clinical studies (randomised controlled trials, prospective and retrospectives cohort studies, case series).

Exclusion criteria

a) Review articles, case reports, simulation studies, clinical studies in non-human subjects, patients with segmental or partial mastectomy, technical descriptions of operative repair with no outcome measures, breast reconstruction not related to cancer, other autologous flap techniques.

b) Duplicates will be excluded and studies will be screened for bias. The Cochrane's risk of bias tool will be used for randomised controlled trials [27]. Bias will be assessed and judged as being high, low or unclear for individual elements from five domains (selection, performance, attrition, reporting and other) [27]. For non-randomised comparative studies, ROBINS-I (Risk Of Bias In Non-randomised Studies—of Interventions) by Cochrane will be used [28]. ROBINS-I covers seven distinct domains from which bias may be introduced, with 'signalling questions' that facilitate judgements about the risk of bias. The judgements within each domain will be carried forward for an overall risk of bias judgement across bias domains [28]. Studies affected by bias will be excluded.

Outcomes

The primary outcomes will be:

1. Patient satisfaction and cosmetic outcome from patient-reported outcome measures (PROMs, scores from validated tools, e.g. BREAST-Q tool)
2. Complications (arterial thrombosis, fat necrosis, venous congestion, infection, partial/full flap loss, donor site complications, haematoma/seroma, return to theatre, capsular contracture, scarring, implant deflation/rupture/displacement)
3. Cost-analysis

The secondary outcomes will be:

1. Duration of surgery
2. Number of surgical revisions
3. Length of stay
4. Availability of procedures
5. Total number of clinic visits

If the data is appropriate for quantitative synthesis, then risk ratio with 95% confidence interval (CI) will be used to determine dichotomous outcomes (complications). Continuous outcomes (cost, PROMs [BREAST-Q],

Khajuria *et al. Systematic Reviews* (2017) 6:232

Page 4 of 5

secondary outcomes excluding availability of procedures) will be determined by weighted or standardised mean differences with 95% CI. Subgroup analysis may be performed for patients with different breast cancer types and for breast implant materials, dependent on sufficient data sets. Where possible, we will utilise results from an intention to treat analysis.

Data extraction, collection and management

Data, from the full-text articles, will be extracted by two independent authors using a standardised extraction form. Any discrepancy will be resolved by consensus or with referral to a third author. If any data is missing or further information is required, the primary authors of the manuscript will be contacted directly. The following data will be extracted:

- first author
- year of publication
- study design
- study setting
- study population
- participant demographics (sex, mean age, BMI, comorbidity)
- complications (arterial thrombosis, venous congestion, infection, fat necrosis, partial/full flap loss, haematoma/seroma, donor site complications, return to theatre, capsular contracture, scarring, implant, deflation/rupture/displacement)
- measures of patient satisfaction (PROMs e.g. BREAST-Q)
- economic data

An assessment of heterogeneity will be performed using Review Manager 5.3 provided by The Cochrane Collaboration [29]; if the studies are relatively homogenous in terms of methodology and outcomes, meta-analyses of the data will be performed. If there is high heterogeneity, a narrative synthesis will be performed instead, without meta-analysis.

Statistical heterogeneity will be quantified by the I^2 statistic [30]. If the I^2 statistic is high, indicating high heterogeneity, a random effects model will be employed. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [31] will be utilised to assess the methodological quality of the studies. Cochrane has produced GRADE tables that identify the basis for judgements about evidence quality. An overall GRADE score (from 4 to 0) is calculated based on quality of overall evidence. The tables specify why points may be added or deducted to obtain the final score [31]. Sensitivity analysis maybe performed based on the quality of the studies, with analyses repeated after removal of poor quality studies to evaluate any change in the overall effect estimate.

Discussion

The aim of this review is to evaluate the clinical outcomes and cost of DIEP flap versus implants for breast reconstruction in context of breast malignancy. Despite many centres ascribing DIEP flap as the gold-standard reconstructive modality, data on clinical outcomes and cost-effectiveness is limited. Therefore, it is important to determine which of the two techniques is clinically superior and more cost-effective as this will guide clinical management. To our knowledge, this is the first systematic review to compare the clinical outcomes and cost of DIEP versus implants.

Dissemination

Based on the results of this systematic review, independent recommendations will be made to plastic surgeons, researchers, policy makers and plastic surgery societies. The results will be disseminated at international meetings in the fields of Plastic, Reconstructive and Aesthetic Surgery and published in a high-impact peer-reviewed journal.

Additional file

Additional file 1: PRISMA-P checklist. (DOCX 509 kb)

Khajuria *et al. Systematic Reviews* (2017) 6:232

Page 5 of 5

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Ganz PA. Psychological and social aspects of breast cancer. Oncology (Williston Park). 2008;22:642–6. 650. discussion 650, 653
2. Helms RL, O'Hea EL, Corso M. Body image issues in women with breast cancer. Psychol Health Med. 2008;13:313–25.
3. Fobair P, Stewart SL, Chang S, et al. Body image and sexual problems in young women with breast cancer. Psychooncology. 2006;15:579–94.
4. Eltahir Y, Werners LL, Dreise MM, et al. Quality-of-life outcomes between mastectomy alone and breast reconstruction: comparison of patient-reported BREAST-Q and other health-related quality-of-life measures. Plast Reconstr Surg. 2013;132:201e–9e.
5. Guyomard V, Leinster S, Wilkinson M. Systematic review of studies of patients' satisfaction with breast reconstruction after mastectomy. Breast. 2007;16:547–67.
6. Yueh JH, Slavin SA, Adesiyun T, Nyame TT, Gautam S, Morris DJ, et al. Patient satisfaction in postmastectomy breast reconstruction: a comparative evaluation of DIEP, TRAM, latissimus flap, and implant techniques. Plast Reconstr Surg. 2010;125:1585–95.
7. Dean C, Chetty U, Forrest AP. Effects of immediate breast reconstruction on psychosocial morbidity after mastectomy. Lancet. 1983;1:459–62.
8. Filiberti A, Tamburini M, Murru L, et al. Psychologic effects and aesthetic results of breast reconstruction after mastectomy. Tumori. 1986;72:585–8.
9. Rowland JH, Holland JC, Chaglassian T, Kinne D. Psychological response to breast reconstruction: expectations for and impact on postmastectomy functioning. Psychosomatics. 1993;34:241–50.
10. Sisco M, Du H, Warner JP, Howard MA, Winchester DP, Yao K. Have we expanded the equitable delivery of postmastectomy breast reconstruction in the new millennium? Evidence from the National Cancer Data Base. J Am Coll Surg. 2012;215:658–66. discussion 666
11. Lagares-Borrego A, Gacto-Sanchez P, Infante-Cossio P, Barrera-Pulido F, Sicilia-Castro D, Gomez-Cia TA. Comparison of long-term cost and clinical outcomes between the two-stage sequence expander/prosthesis and autologous deep inferior epigastric flap methods for breast reconstruction in a public hospital. J Plast Reconstr Aesthet Surg. 2016;69:196–205.
12. Fischer J, Nelson J, Cleveland E. Breast reconstruction modality outcome study: a comparison of expander/implants and free flaps in select patients. Plast Reconstr Surg. 2013;131:928–34.
13. Liu C, Momeni A, Zhuang Y, et al. Outcome analysis of expander/implant versus microsurgical abdominal flap breast reconstruction: a critical study of 254 cases. Ann Surg Oncol. 2014;21:2074–82.
14. Pirro O, Mestak O, Vindigni V, Sukop A, Hromadkova V, Nguyenova A, et al. Comparison of patient-reported outcomes after implant versus autologous tissue breast reconstruction using the BREAST-Q. Plast Reconstr Surg Glob Open. 2017;5:e1217.
15. Spear S, Newman M, Bedford S, et al. A retrospective analysis of outcomes using three common methods for immediate breast reconstruction. Plast Reconstr Surg. 2008;122:340–7.
16. Alderman A, Wilkins E, Lowery J, et al. Determinants of patient satisfaction in postmastectomy breast reconstruction. Plast Reconstr Surg. 2000;106:769–76.
17. Tønseth K, Hokland B, Tindholdt T, et al. Quality of life, patient satisfaction and cosmetic outcome after breast reconstruction using DIEP flap or expandable breast implant. J Plast Reconstr Aesthet Surg. 2008;61:1188–94. Epub 2007 Jul 2
18. Grover R, Padula WV, Van Vliet M, Ridgway EB. Comparing five alternative methods of breast reconstruction surgery: a cost-effectiveness analysis. Plast Reconstr Surg. 2013;132:709e–23e.
19. Atherton DD, Hills AJ, Moradi P, Muirhead N, Wood SH. The economic viability of breast reconstruction in the UK: comparison of a single surgeon's experience of implant; LD; TRAM and DIEP based reconstructions in 274 patients. J Plast Reconstr Aesthet Surg. 2011;64:710–5.
20. Agha RA, Fowler AJ, Herlin C, Goodacre TE, Orgill DP. Use of autologous fat grafting for breast reconstruction: a systematic review with meta-analysis of oncological outcomes. J Plast Reconstr Aesthet Surg. 2015;68:143–61.
21. Steiert AE, Boyce M, Sorg H. Capsular contracture by silicone breast implants: possible causes, biocompatibility, and prophylactic strategies. Med Devices (Auckl). 2013;6:211–8.
22. Spear SL, Murphy DK. Allergan silicone breast implant U.S. Core clinical study group. Natrelle round silicone breast implants: Core study results at 10 years. Plast Reconstr Surg. 2014;133:1354–61.
23. Bajaj AK, Chevray PM, Chang DW. Comparison of donor-site complications and functional outcomes in free muscle-sparing TRAM flap and free DIEP flap breast reconstruction. Plast Reconstr Surg. 2006; 117:737–46. discussion 747
24. Matros E, Albornoz CR, Razdan SN, Mehrara BJ, Macadam SA, Ro T, et al. Cost-effectiveness analysis of implants versus autologous perforator flaps using the BREAST-Q. Plast Reconstr Surg. 2015;135:937–46.
25. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred reporting items for systematic review and metaanalysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4:1–9.
26. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283:2008–12.
27. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
28. Sterne JA, Hernan MA, Reeves BC, Savovic 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.
29. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan) version 5.2; 2012
30. Higgins JP, Thompson S, Deeks J, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557–60.
31. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ 2004;328: 1490

HHS Public Access

A Randomized Controlled Trial to Assess the Effects of Competition on the Development of Laparoscopic Surgical Skills

Daniel A. Hashimoto, M.D., M.S.[1,3], **Ernest D. Gomez, M.D., M.S.**[1], **Laura Beyer-Berjot, M.D.**[2], **Ankur Khajuria, B.Sc.**[3], **Noel N. Williams, MBBCh, FRCSI**[1], **Ara Darzi, KBE, PC, M.D., FMedSci, FRCS, FACS**[3], and **Rajesh Aggarwal, MBBS, Ph.D., M.A., FRCS**[4,5]

[1]Perelman School of Medicine, University of Pennsylvania

[2]Center for Surgical Teaching and Research (CERC), Université de la Méditerranée, Marseille, France

[3]Department of Biosurgery and Surgical Technology, St. Mary's Hospital, Imperial College London

[4]Department of Surgery, Faculty of Medicine, McGill University

[5]Arnold & Blema Steinberg Medical Simulation Centre, Faculty of Medicine, McGill University

Abstract

Background—Serious games have demonstrated efficacy in improving participation in surgical training activities, but studies have not yet demonstrated the effect of serious gaming on performance. This study investigated whether competitive training affects laparoscopic surgical performance.

Methods—Twenty novices were recruited, and 18 (2 drop-outs) were randomized into control or competitive (CT) groups to perform 10 virtual reality (VR) laparoscopic cholecystectomies (LC). Competitiveness of each participant was assessed. The CT group was informed they were competing to outperform one another for a prize; performance ranking was shown prior to each session. The control group did not compete. Performance was assessed on time, movements, and instrument path length. Quality of performance was assessed with a global rating score (GRS).

Results—There were no significant intergroup differences in baseline skill or measured competitiveness. Time and GRS, at final LC, were not significantly different between groups; however, the CT group was significantly more dexterous than control and had significantly lower variance in number of movements and instrument path length at the final LC (p=0.019). Contentiousness was inversely related to time in the CT group.

Corresponding Author Daniel Hashimoto, MD MS, Department of Surgery, Massachusetts General Hospital, 55 Fruit Street, GRB 425, Boston, MA 02114, Telephone: 617-726-2800, Facsimile: 617-724-3499, dahashimoto@partners.org.

Conclusion—This was the first randomized controlled trial to investigate if competitive training can enhance performance in laparoscopic surgery. Competitive training may lead to improved dexterity in laparoscopic surgery but yields otherwise similar performance to standard training in novices. Competition may have different effects on novices versus experienced surgeons, and subsequent research should investigate competitive training in experienced surgeons as well.

Keywords

Surgical education; Virtual reality simulation; Student education; Competitive training; Minimally invasive

Introduction

Surgical training programs are working to adapt their curricula to improve the efficiency of surgical education by augmenting didactic training and intraoperative education with simulation to remain in compliance with Accreditation Council for Graduate Medical Education (ACGME) requirements.[1] Many questions remain, however, about how best to implement simulation into curricula to maximize the efficiency of training.

While prior comparisons to aviation have yielded a fair amount of knowledge regarding the utility of simulation,[2] other high performance industries, such as sports, may also provide valuable insight into potential training strategies to elicit superior performance. Some pedagogical techniques identified in sports have already been investigated in surgery, including warm-up,[3] mental practice,[4] and deliberate practice.[5,6] Competition has been found to lead to improved performance in sports including golf, weight lifting, and basketball.[7-9] Gamification, the use of game mechanics such as competition, has been successfully utilized to improve motivation to participate in surgical simulation training and to teach and assess clinical decision making[10,11]; however, no studies have investigated the effects of competition on technical skills performance in a randomized, controlled manner.

We hypothesized that competition would lead to improved performance in trainees. This study investigated the effects of competition on performance during successive virtual reality (VR) laparoscopic cholecystectomy (LC) cases.

Methods

Participant Selection

Due to the educational nature of the study, this protocol was exempted from further ethics review. Informed consent was obtained from all participants, and participants were informed that their participation, or lack thereof, would in no way impact their medical training or medical care they might receive. Medical students from London hospitals with an interest in surgery were invited to participate in the study. Based on power analysis and cost constraints, twenty (n=20) medical students were recruited. All trainees had limited surgical experience (performed 0 but observed > 1 LCs in the operating room). All participants were offered a certificate of completion in a basic laparoscopic skills course if they completed all sessions of the study. At recruitment, participants were randomized into one of two equal

groups – Competitive Training (CT) group or Control group – using a random number generator (STATA, College Station, TX) (Figure 1).

Baseline Assessment

Each participant underwent a validated baseline skills assessment on the LapMentor VR (Simbionix; Cleveland, OH) laparoscopic simulator on Basic Skills tasks 5 and 6. For Basic Skills task 5, time to completion was assessed. For Basic Skills task 6, time to completion and number of movements was recorded as these metrics have been shown to be construct valid.[12]

Participants were also asked to complete the Revised Competitiveness Index, a psychometric questionnaire designed to assess individuals' trait of competitiveness along two domains – enjoyment of competition and contentiousness (desire to outperform others).[13] Each domain is tested on its own subscale within the Revised Competitiveness Index and can be considered as an individual factor that makes up a person's trait of competitiveness.

Didactic and Proficiency Training

Participants underwent a modified laparoscopic skills training program based on a previously validated curriculum.[12] Participants were trained to proficiency in basic skills and were given video instruction on performing a full procedure LC on the simulator.

Competitive Training Group Sessions

Participants in the CT group underwent 10 training sessions comprising a total of 10 VR LCs. Participants in the CT group were told to perform each procedure as safely and efficiently as possible but were also informed that the top performer after 10 sessions would be awarded a gift card for a flight simulator experience (valued at approximately $150). Each session, participants completed a VR LC on the LapMentor simulator, and their performance was assessed in real time by a trained observer using a previously validated rating scale of surgical technical skill [Objective Structured Assessment of Technical Skills Global Rating Scale (OSATS GRS)].[14,15] They were then given immediate post-procedure feedback on their performance by being shown time to complete the VR LC, number of movements, total path length (cm) of instrument tips, and OSATS GRS. At the conclusion of each session, participants were shown a leader board demonstrating their performance and rank compared to others in the CT group (Figure 2).

Ranking was based on a formula (Formula 1) that weighted quality of performance (OSATS GRS) greater than time or dexterity (as measured by number of movements and path length) based on the recommendations of surgical educators at Imperial College London. Similar to golf, a lower score was considered to have a higher rank.

$$[\text{Time}\,(sec) + \text{Movements} + \text{Path Length}\,(cm)]\,/\,OSATSGRS \quad \text{Formula 1}$$

Before and after each VR LC, participants were asked to complete a short form State Trait Anxiety Inventory (STAI), a validated tool to assess state anxiety.[16] After each VR LC,

participants were also asked to complete the NASA Task Load Index (TLX), a validated, subjective multidimensional assessment tool that allows participants to rate perceived workload.[17] TLX was utilized to assess for any increased workload that participants may experience from competition. Participants in the CT group were asked to not disclose their status in the study to prevent a potential effect on motivation in the control group.

Control Group Training Sessions

The control group similarly underwent 10 training sessions comprising a total of 10 VR LCs, but no mention of a prize was made. They were only instructed to perform each procedure as safely and efficiently as possible. Their performance was also assessed using the same metrics as the CT group by the same trained observer. The control group was given immediate post-procedure feedback on their performance by being shown time to complete the VR LC, number of movements, total path length (cm) of instrument tips, and OSATS GRS. Participants in the control group were not ranked against one another and were not shown the performance of other participants.

Control group participants were also asked to complete a short form State Trait Anxiety Inventory before and after each VR LC. After each VR LC, participants were asked to complete the NASA TLX.

Participants in both groups were not allowed to practice laparoscopic skills outside of the scheduled study sessions. Participants were allowed no more than two sessions per day with each trial separated by one hour to prevent fatigue. Scheduling of sessions was controlled to allow for accurate comparison of performance amongst the CT group based on session number.

Statistical Analysis

Statistical analysis was performed using STATA Intercooled 12 (College Station, TX). Shapiro-Wilk test showed the nature of the data to be nonparametric. Mann-Whitney U-test was employed to compare intergroup baseline laparoscopic performance and VR training session performance. Wilcoxon signed-rank test was utilized for intra-group comparison. Data are reported as median (interquartile range). Levene's test was utilized to compare the consistency in performance of the CT group versus the control group. Multivariate regression was used to assess the effect of competitiveness and contentiousness on surgical performance. Nonlinear regression was utilized to assess the learning curve of participants.[18]

In addition to live ratings, videos of VR LCs from both groups were assessed by an independent, blinded rater using the OSATS GRS. Intraclass correlation coefficient (ICC) was calculated to assess inter-rater reliability of the OSATS GRS. A $p <0.05$ was be considered statistically significant.

Sample size was based on detecting at least a 25% difference in time and dexterity with alpha of 0.05 and beta of 0.8 as based on preliminary data collected prior to the study.

Results

Subjects

Eighteen of the twenty recruited participants completed the study. Two participants dropped out during proficiency training and cited scheduling conflicts for their inability to complete the study. All participants were right handed. Two of the participants in the control group and three in the CT group were female.

Baseline Assessment of Laparoscopic Skill

There were no significant differences in baseline laparoscopic skill between control and the CT groups (Table 1).

Competitiveness Index

Analyzing the results of the Revised Competitiveness Index, there was no significant difference in enjoyment of competition between groups (Control: 27.44 ± 2.51, CT: 27.11 ± 3.78; p=0.3). However, the CT group (16.89 ± 5.39) was significantly more contentious than the Control group (13 ± 2.59; p<0.001).

Virtual Reality LC Performance

Both the control and CT groups improved over the course of 10 LCs in time, movements, path length, and OSATS GRS (Table 2). The intraclass correlation coefficient for OSATS GRS was ICC=0.858.

At the first LC, the CT group was significantly faster, made fewer movements, and had lower path length than the control group. There was no significant difference between groups in quality of surgical performance as assessed by the OSATS GRS at first LC (Table 2). By the tenth and final LC, there were no significant differences between groups in time to complete the procedure or OSATS GRS score. With regards to dexterity, the CT group made significantly fewer movements and had lower path length than the control group (Table 2).

The CT group had significantly lower variance in number of movements and instrument path length than the control group at the tenth and final LC (p=0.019).

Virtual Reality LC Learning Curves

After 5 cases, the control group plateaued at an average procedure completion time of 345.8 seconds (p<0.001), total number of movements of 308 (p<0.001), and path length of 482 cm. (p<0.001). The control group plateaued at an average OSATS GRS of 20 after 5 cases (p<0.001) (Figure 3A-D).

The CT group plateaued at 365.1 seconds (p<0.001) after 8 cases, total movements of 288 (p<0.001) after 9 cases, and path length of 427.4 cm (p<0.001). The CT group on average plateaued at an OSATS GRS of 21 after 5 cases (p<0.001) (Figure 3A-D). There were no significant differences in the plateau levels of the two groups for any of the metrics.

Virtual Reality LC Performance and Psychometrics

The CT group reported higher mean state anxiety after completing a VR LC compared to state anxiety just prior to performing a VR LC while the control group reported no difference in state anxiety either before or after VR LC performance (Table 2). For both groups, state anxiety as assessed after VR LC had a negative effect on quality of surgical performance (Table 3).

There was no significant difference in perceived workload between groups after each VR LC. There was no effect of perceived workload on either group for any of the performance metrics.

For both groups, there was no effect of competitiveness or contentiousness on quality of performance. For the CT group, contentiousness was inversely related to time to complete a VR LC ($\beta=-4.56 \pm 2.23$, $R^2=0.57$, $p=0.044$) and number of movements ($\beta=-5.6 \pm 2.41$, $R^2=0.65$, $p=0.023$). There was no relationship between contentiousness and time or movements for the control group.

Discussion

Trainees engaging in competitive training developed greater dexterity when performing VR LC. Although the time taken to complete the procedure and quality of surgical performance were similar between trainees in CT and those in standard training, the decreased movements and instrument path length suggest that the CT group was able to complete a VR LC with greater efficiency. Furthermore, the CT group was more consistent in movements made and instrument path length as suggested by the decreased variance of these metrics in the CT group at the final LC (Figure 3B-C).

Gamification in surgical education has predominantly been rooted in the utilization of serious games.[10] Serious games are interactive, scored computer games that are fun, engaging, yet challenging with the goal of improving skills or knowledge applicable to real world scenarios. While many games investigated in the literature have focused on teaching decision-making skills or cognitive knowledge, surgical skills competitions are often held at various society meetings and within institutions as a fun exercise in skills practice. To our knowledge, this was the first randomized controlled trial investigating the effect of competitive training on the acquisition of surgical skill.

As the assignment of participants to CT versus control training was random, we did not intend to have more contentious people in the CT group. However, the CT group in this study reported being more contentious than the control group, and a regression model of the performance data suggests that contentiousness in the CT group relates to faster and more dexterous performance. While we interpret these findings with caution, one potential explanation is that contentious participants, when placed in a competitive environment, have improved performance in dexterity that may have been driven by their desire to outperform others.

Competition is not the only tool that may exist to promote improved performance in novice surgeons. Delivery of surgical care requires a coordinated team effort involving surgeons, nurses, technicians and other ancillary staff; and cooperation in the team setting may be the more appropriate means of improving performance as suggested in both the surgical and sports science literature.[9,19] However, a study in sports science suggests that in situations where individuals are involved in a structured, fair competition and are able to gauge the progress of opponents, competition can lead to higher levels of individual motor performance.[20] Acquisition of skill in the simulation center tends to occur in the individual setting, and a structured competition may promote improved performance in domains such as dexterity as suggested in this study. These skills may then be translated to the team environment to provide the best care for patients. Being able to anonymously gauge the progress of other trainees in a program may have provided a target for which the participants in the CT group could strive to outperform, thus leading to improved dexterity versus the control group.

Analysis of the state anxiety in participants found that high state anxiety after completion of VR LC negatively affected quality of surgical performance as assessed by the OSATS GRS. Due to the design of the study, we are not able to conclude whether increased anxiety during the VR LC resulted in a decreased OSATS GRS score or if the increased anxiety was a result of technical errors that were reflected in a decreased OSATS GRS score. Previous research suggests either explanation may be plausible as the literature has reported technical errors to be a source of stress for surgeons but also that some individuals may be less capable of skilled surgical performance under pressure.[21,22] While there was no difference in state anxiety before or after a VR LC for the control group, the CT group reported significantly higher state anxiety after completing a VR LC. The participants were not surveyed on why they may have felt more anxious; however, competition may have heightened levels of anxiety as the CT group awaited the results of their ranking. While it is certainly not ideal to stress trainees to the point of negatively impacting performance, surgeon stress is present and measurable in an operation.[23] Competition in the simulation environment may provide a safe avenue through which to expose trainees to stress that may be present in a real operative setting.

The results of this study are not without limitations. Although participants received their OSATS GRS scores, they did not receive feedback on specific steps to improve quality of surgical performance. Previous work has demonstrated that specific feedback is necessary to improve performance quality; thus, quality of performance may have been limited in this study.[5] To control elements such as complexity of the case, the competition was limited to VR in this initial study. As the ultimate goal of simulation training is to improve performance in the operating room, individuals who participate in CT should undergo evaluation in a live operative case. Since this study was conducted with medical students, no attempt was made to assess the participants in a live operating room. However, future work with residents should assess the effect of CT in a real clinical setting. Steps should be taken to ensure that competition remains confined to the simulation environment, as patient care and safety should not be treated as a game.

Implementation of competitive training for programs with pre-existing simulation programs can be low cost. We utilized an in-house programmed webpage; however, a similar study could be conducted by manually placing values into a spreadsheet and sorting values to calculate rank. Future work should survey participants on motivation prior to each session to determine whether participants subjectively report being motivated by competition, and studies can investigate a tailored approach to learning that compares competitive and non-competitive training based on the motivational preferences of the trainees. Such work may help elucidate whether competition motivates trainees to varying levels depending on individual motives.

Conclusion

Surgical skills competitions have been held at surgical meetings and at institutional levels, but competition as a novel training strategy had not previously been investigated. The results of this study suggest that competition in surgical education for medical students may lead to improved dexterity in laparoscopic cholecystectomy but has otherwise equivalent effects as standard, repetition-based training on time and quality of performance. Additional research is needed to determine if similar effects are seen in residents who receive feedback on their performance.

Acknowledgements

Abbreviations

ACGME	Accreditation Council for Graduate Medical Education
VR	Virtual reality
LC	Laparoscopic cholecystectomy
CT	Competitive Training
OSATS GRS	Objective Structured Assessment of Technical Skills Global Rating Scale
STAI	State Trait Anxiety Inventory
TLX	Task Load Index

References

1. ACGME. ACGME Program Requirements of Graduate Medical Education in Surgery. Committee RR. , editor. Accreditation Council for Graduate Medical Education; Chicago, IL: 2008.

2. Satava RM. Historical review of surgical simulation--a personal perspective. World journal of surgery. 2008; 32(2):141–148. [PubMed: 18097716]

3. Calatayud D, Arora S, Aggarwal R, et al. Warm-up in a virtual reality environment improves performance in the operating room. Ann Surg. 2010; 251(6):1181–1185. [PubMed: 20485133]

4. Arora S, Aggarwal R, Sirimanna P, et al. Mental practice enhances surgical technical skills: a randomized controlled study. Ann Surg. 2011; 253(2):265–270. [PubMed: 21245669]

5. Crochet P, Aggarwal R, Dubb SS, et al. Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills. Ann Surg. 2011; 253(6):1216–1222. [PubMed: 21516035]

6. Hashimoto DA, Sirimanna P, Gomez ED, et al. Deliberate practice enhances quality of laparoscopic surgical performance in a randomized controlled trial: from arrested development to expert performance. Surg Endosc. 2014

7. Rhea MR, Landers DM, Alvar BA, Arent SM. The effects of competition and the presence of an audience on weight lifting performance. J Strength Cond Res. 2003; 17(2):303–306. [PubMed: 12741867]

8. Tanaka Y, Sekiya H. The influence of audience and monetary reward on the putting kinematics of expert and novice golfers. Res Q Exerc Sport. 2010; 81(4):416–424. [PubMed: 21268465]

9. Tauer JM, Harackiewicz JM. The effects of cooperation and competition on intrinsic motivation and performance. J Pers Soc Psychol. 2004; 86(6):849–861. [PubMed: 15149259]

10. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg. 2012; 99(10):1322–1330. [PubMed: 22961509]

11. Kerfoot BP, Kissane N. The use of gamification to boost residents' engagement in simulation training. JAMA surgery. 2014; 149(11):1208–1209. [PubMed: 25229631]

12. Aggarwal R, Crochet P, Dias A, Misra A, Ziprin P, Darzi A. Development of a virtual reality training curriculum for laparoscopic cholecystectomy. Br J Surg. 2009; 96(9):1086–1093. [PubMed: 19672934]

13. Houston J, Harris P, McIntire S, Francis D. Revising the Competitiveness Index Using Factor Analysis. Psychological Reports. 2002; 90:31–34. [PubMed: 11899003]

14. Martin J, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997; 84:273–278. [PubMed: 9052454]

15. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A. Toward feasibe, valid, and reliable video-based assessments of technical surgical skills in the operating room. Ann Surg. 2008; 247(2):372–379. [PubMed: 18216547]

16. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol. 1992; 31(Pt 3):301–306. [PubMed: 1393159]

17. Hart, S.; Staveland, L. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.; Meshkati, N., editors. Development of NASA-TLX (Task Load Index). Elsevier; Amsterdam: 1988.

18. Feldman LS, Cao J, Andalib A, Fraser S, Fried GM. A method to characterize the learning curve for performance of a fundamental laparoscopic simulator task: defining "learning plateau" and "learning rate". Surgery. 2009; 146(2):381–386. [PubMed: 19628099]

19. Hull L, Arora S, Aggarwal R, Darzi A, Vincent C, Sevdalis N. The impact of nontechnical skills on technical performance in surgery: a systematic review. J Am Coll Surg. 2012; 214(2):214–230. [PubMed: 22200377]

20. Stanne M, Johnson D, Johnson R. Does competition enhance or inhibit motor performance: A meta-analysis. Psychological Bulletin. 1999; 125:133–154. [PubMed: 9990847]

21. Sami A, Waseem H, Nourah A, et al. Real-time observations of stressful events in the operating room. Saudi J Anaesth. 2012; 6(2):136–139. [PubMed: 22754439]

22. Malhotra N, Poolton JM, Wilson MR, Ngo K, Masters RS. Conscious monitoring and control (reinvestment) in surgical performance under pressure. Surg Endosc. 2012; 26(9):2423–2429. [PubMed: 22350243]

23. Jones KI, Amawi F, Bhalla A, Peacock O, Williams JP, Lund JN. Assessing surgeon stress when operating using Heart Rate Variability and the State Trait Anxiety Inventory: will surgery be the death of us? Colorectal disease: the official journal of the Association of Coloproctology of Great Britain and Ireland. 2014
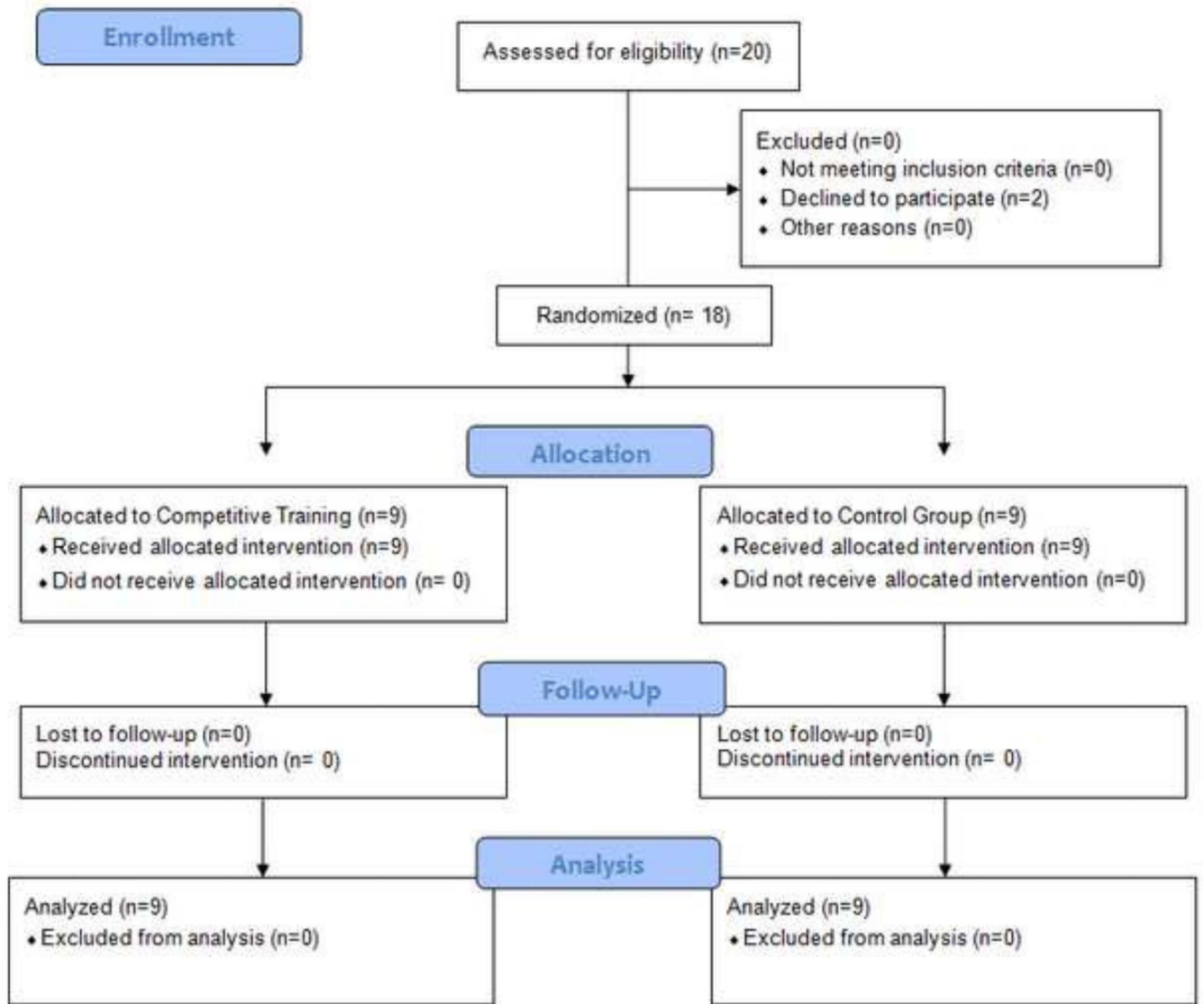
Figure 1.
Flow chart of study protocol with recruited subjects and drop-outs

These are the rankings by Total Weighted Score

User	LC Number	Time (seconds)	Total Movements	Total Path Length (cm)	Global OSATS Score (out of 35)	Total Weighted Score
	10	214	163.000	264.800	21	30.5619048
	10	318	199.000	300.400	20	40.8700000
	10	317	256.000	393.800	23	42.0347826
	10	308	239.000	343.000	21	42.3809524
1	10	304	258.000	444.700	23	43.7695652
	10	367	253.000	370.200	21	47.1523810
	10	357	230.000	376.700	20	48.1850000
	10	344	274.000	403.500	21	48.6428571
	10	404	295.000	420.700	22	50.8954545

Please close this tab when done.

Figure 2.
Leader board demonstrating participant performance and rank compared to others in the CT group.
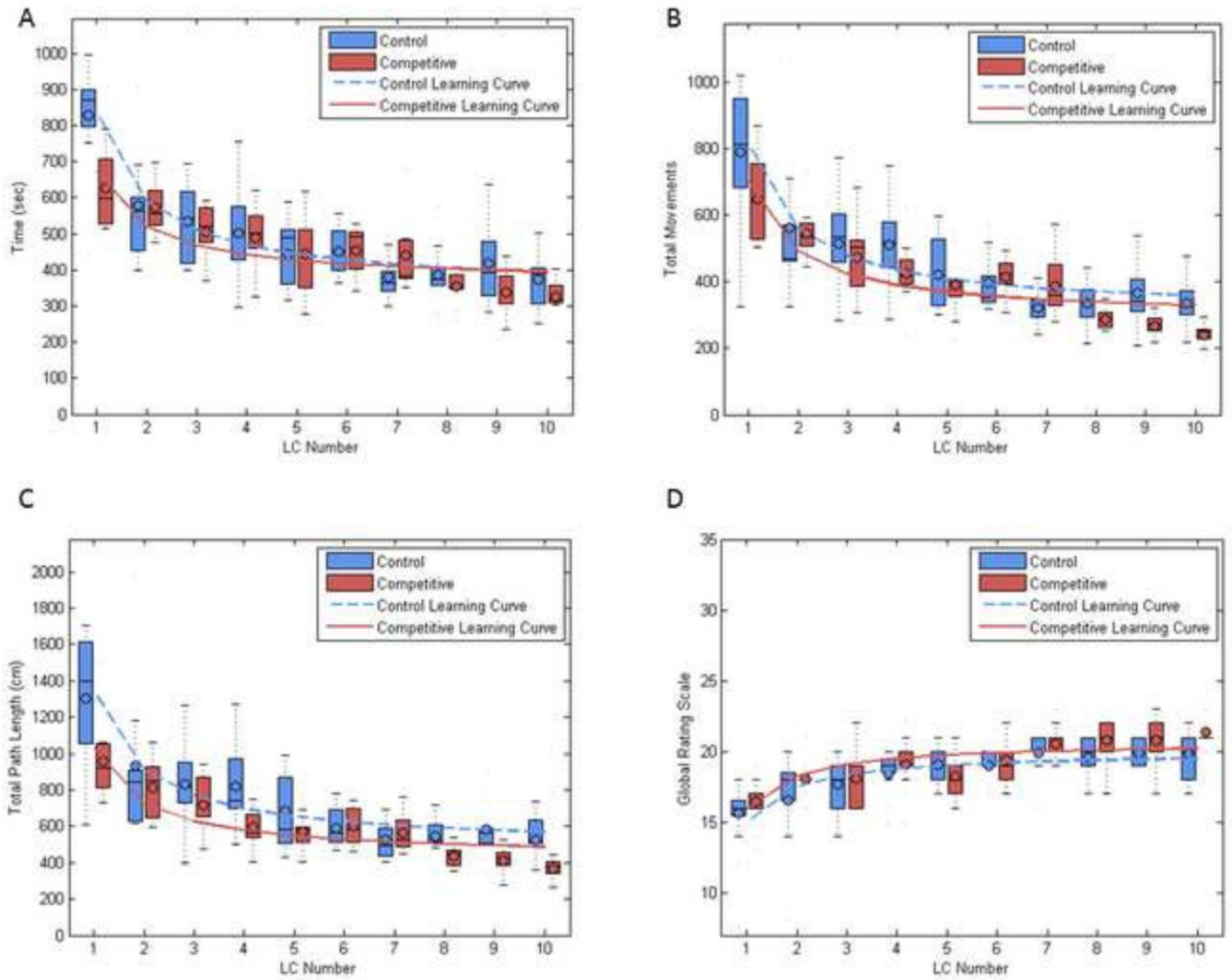
Figure 3.
(A-D) Virtual Reality learning curves and performance box-and-whisker plots of control and CT groups for time to complete procedure (A), total number of movements (B), total path length (cm) (C), and OSATS GRS (D).

Table 1

Pre-test baseline skills assessment of control and CT groups. Values as median (interquartile range).

	Control	**CT**	**p-value**
Task 5			
Time (sec)	132.7 (122-142)	134 (123-146)	0.70

	Control	**CT**	**p-value**
Task 6			
Time (sec)	164.3 (132-177)	170.5 (139-198)	0.31
Movements	208 (169-253)	264 (229-307)	0.10
Path Length (cm)	510 (405-631)	668.5 (525-726)	0.30

Table 2

Comparison of first and final VR LC time, movements, path length, and quality of surgical performance. Comparison of pre- and post-LC mean State Trait Anxiety Index between groups. Values as median (interquartile range).

		Control	CT	p-value
Time (sec)				
First LC		871 (797-898)	598 (529-708)	**0.01**
Final LC		390 (305-405)	319 (208-357)	0.40
	p-value	**0.008**	**0.008**	
Movements				
First LC		816 (684-952)	643 (529-756)	**0.04**
Final LC		327 (301-373)	253 (230-258)	**0.02**
	p-value	**0.008**	**0.008**	
Path Length (cm)				
First LC		1397 (1053-1617)	920.7 (810-1057)	**0.04**
Final LC		518 (509-633)	376.7 (343-404)	**0.02**
	p-value	**0.008**	**0.008**	
OSATS GRS				
First LC		16 (15-17)	16 (16-18)	0.47
Final LC		20 (18-21)	21 (21-22)	0.16
	p-value	**0.013**	**0.008**	
STAI				
Pre-LC		8 (7-11)	11 (10-12)	**<0.001**
Post-LC		9 (7-11)	12 (10-14)	**<0.001**
	p-value	0.118	**<0.001**	

Table 3

Multiple linear regression model for VR LC quality of performance and state anxiety. State anxiety after VR LC had a negative effect on quality of surgical performance.

Control Group State Anxiety					
Variable	**Coefficient**	**SE**	**95% CI**	**R²**	**p-value**
Intercept	20.66	0.92			
Pre-STAI	0.11	0.09	−0.08 to 0.29	0.13	0.253
Post-STAI	−0.33	0.09	−0.51 to −0.14	0.13	**0.001**

CT Group State Anxiety					
Variable	**Coefficient**	**SE**	**95% CI**	**R²**	**p-value**
Intercept	20.61	1.45			
Pre-STAI	0.16	0.12	−0.08 to 0.39	0.1	0.19
Post-STAI	−0.24	0.08	−0.4 to −0.09	0.1	**0.003**

RESEARCH ARTICLE

JOURNAL OF
MEDICAL VIROLOGY WILEY

The effect of smoking on COVID-19 severity: A systematic review and meta-analysis

Rohin K. Reddy[1] 🟢 | Walton N. Charles[1] | Alexandros Sklavounos[2] | Atul Dutt[1] |
Paul T. Seed[2] | Ankur Khajuria[1,3]

[1]Department of Surgery and Cancer, Imperial College London, London, UK

[2]Department of Women and Children's Health, King's College London, London, UK

[3]Kellogg College, University of Oxford, Oxford, UK

Correspondence
Ankur Khajuria, Kellogg College, University of Oxford, 60-62 Banbury Road, Oxford OX2 6PN, UK.
Email: ankur.khajuria09@imperial.ac.uk

Abstract

Various comorbidities represent risk factors for severe coronavirus disease 2019 (COVID-19). The impact of smoking on COVID-19 severity has been previously reported in several meta-analyses limited by small sample sizes and poor methodology. We aimed to rigorously and definitively quantify the effects of smoking on COVID-19 severity. MEDLINE, Embase, CENTRAL, and Web of Science were searched between 1 December 2019 and 2 June 2020. Studies reporting smoking status of hospitalized patients with different severities of disease and/or at least one clinical endpoint of interest (disease progression, intensive care unit admission, need for mechanical ventilation, and mortality) were included. Data were pooled using a random-effects model. This study was registered on PROSPERO: CRD42020180920. We analyzed 47 eligible studies reporting on 32 849 hospitalized COVID-19 patients, with 8417 (25.6%) reporting a smoking history, comprising 1501 current smokers, 5676 former smokers, and 1240 unspecified smokers. Current smokers had an increased risk of severe COVID-19 (risk ratios [RR]: 1.80; 95% confidence interval [CI]: 1.14-2.85; $P = .012$), and severe or critical COVID-19 (RR: 1.98; CI: 1.16-3.38; $P = .012$). Patients with a smoking history had a significantly increased risk of severe COVID-19 (RR: 1.31; CI: 1.12-1.54; $P = .001$), severe or critical COVID-19 (RR: 1.35; CI: 1.19-1.53; $P < .0001$), in-hospital mortality (RR: 1.26; CI: 1.20-1.32; $P < .0001$), disease progression (RR: 2.18; CI: 1.06-4.49; $P = .035$), and need for mechanical ventilation (RR: 1.20; CI: 1.01-1.42; $P = .043$). Patients with any smoking history are vulnerable to severe COVID-19 and worse in-hospital outcomes. In the absence of current targeted therapies, preventative, and supportive strategies to reduce morbidity and mortality in current and former smokers are crucial.

KEYWORDS

coronavirus, epidemiology, pandemics, pathogenesis, respiratory tract, virus classification, zoonoses

Rohin K. Reddy and Walton N. Charles are co-first authors and contributed equally to this work.

1 | INTRODUCTION

As of 28 July 2020, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has infected 16 341 920 patients, with 650 805 deaths across 188 countries.[1,2] Risk factors for poor outcome in patients with coronavirus disease 2019 (COVID-19) include older age, male sex, hypertension, diabetes, cardiovascular disease, and respiratory disease.[3-5] Remarkably, current peer-reviewed data surrounding the effect of smoking tobacco on the clinical severity of COVID-19 has thus far been controversial, and there is an urgent need for definitive answers.[6]

An early systematic review without meta-analysis concluded that smoking is most likely associated with negative progression and outcomes in COVID-19,[7] however, a preliminary meta-analysis showed that active smoking is not significantly associated with increased risk of severe disease.[8] Four subsequent meta-analyses have shown an increased risk of severe COVID-19 associated with smoking.[9-12] A summary of the six previously published systematic reviews[7-12] alongside assessment of their methodological quality using A Measurement Tool to Assess systematic Reviews 2[13] (AMSTAR 2) is provided in the Appendix (Appendix pp2-3). The articles ranged from critically poor to moderate quality, indicating that significant methodological flaws in critical domains exist with all six currently published reviews assessing the impact of smoking on COVID-19 severity. It is therefore likely that the true effect of smoking on COVID-19 severity reported in these analyses is clouded by considerable bias.

Furthermore, as a result of several nonpeer reviewed preprint articles falsely equating the prevalence of smoking in COVID-19 study populations with population estimates for smoking prevalence, there has been widespread attention paid to recent mass media reports that smoking may exert a protective effect against COVID-19 infection.[14] This led to the World Health Organization releasing a statement on 11 May urging caution with regards to these claims, and emphasizing the lack of evidence confirming a link between smoking or nicotine in the prevention or treatment of COVID-19.[15] Consequently, there remains a distinct lack of clarity and high-quality evidence regarding the relationship between smoking and the severity of COVID-19. Therefore, to address this important clinical question, this systematic review and meta-analysis aimed to evaluate the effect of smoking status, including current smoking and a history of smoking, on the clinical severity of COVID-19.

2 | METHODS

2.1 | Search strategy and selection criteria

This systematic review and meta-analysis adhered to PRISMA guidelines[16] and was AMSTAR 2 compliant (Appendix pp8-12).[13] Two authors independently searched MEDLINE, Embase, CENTRAL, and Web of Science for studies published between 1 December 2019 and 2 June 2020. The search strategy is provided in the Appendix (p13). No language restrictions were applied. COVID-19 resource centers of *The Lancet*, *The Lancet Respiratory Medicine*, *The New England Journal of Medicine*, and *The BMJ* were also hand searched up to 5 July 2020.

Reference lists of included studies and previous systematic reviews were additionally screened for their relevance.

To capture all available relevant evidence, randomized, and observational studies reporting the smoking status of hospitalized patients presenting with different severities of disease and/or at least one clinical endpoint of interest were deemed eligible for inclusion. Smoking history included current and former tobacco smokers or e-cigarette users. Disease severity, including severe or critical cases, was defined a priori and based on the COVID-19 diagnostic criteria issued by the Chinese National Health Commission (Appendix p13).[17] Other acceptable criteria included the Infectious Diseases Society of America/American Thoracic Society (IDSA/ATS) criteria for severe community-acquired pneumonia.[18] Clinical endpoints of disease progression, intensive care unit (ICU) admission, mechanical ventilation requirement, and/or mortality were used as surrogate markers for in-hospital severity. We excluded studies on other coronaviruses or if there was insufficient information to distinguish disease severity based on smoking status. Case series involving less than 20 patients, review articles, editorials, conference abstracts, and nonclinical studies were also excluded. Preprints were not assessed for eligibility due to their preliminary nature.

Two authors (WNC, AS) independently screened the titles and abstracts of retrieved studies, with full-texts of all potentially eligible papers subsequently assessed for inclusion. Any discrepancy was resolved by consensus discussion with the senior author (AK).

2.2 | Data analysis

Data from studies that fulfilled our inclusion criteria were extracted independently by three authors (WNC, AS, and RKR). Main datapoints included: study details (author, journal, date, country, study design, study period, and funding), total numbers of patients, and their clinical outcomes by smoking status.

Two authors (AS and AD) independently assessed the quality of included studies using the Newcastle-Ottawa Scale modified for case series, cohort studies, and cross-sectional studies.[19] Scores were then classified by the Agency for Healthcare Research and Quality standards as good, fair, or poor. Any discrepancies in quality assessment were resolved by a third author (WNC).

As per our prespecified analysis plan, random-effects meta-analyses of pooled raw data were employed using the DerSimonian and Laird method for each outcome with sufficient data to account for anticipated differences across countries and study design over time. Current smokers were compared to former and never-smokers, and patients with a smoking history were compared to never-smokers. Where available, adjusted effect estimates were combined and in the absence of adjustment for confounders, raw effect estimates were combined. The results are presented in forest plots as risk ratios (RR) and corresponding 95% confidence intervals (CI) for each outcome. I^2 estimates of heterogeneity, representing the variability across studies, are classified as low (<30%), moderate (30%-60%), or high (>60%). Sensitivity analyses included only good-quality studies and, for severity outcomes, studies using the

COVID-19-specific criteria for grading severity. Subgroup analyses were completed by country. Funnel plots were used to check for publication bias and tested for asymmetry using Harbord's test,[20] with studies with no events in either exposed or unexposed arms excluded from this analysis. P values <.05 were considered significant.

Data were analyzed using Stata (version 15). The study protocol was prospectively registered with PROSPERO, number CRD42020180920.[21]

2.3 | Role of the funding source

This study received no funding. All authors had full access to all of the data and took responsibility for the decision to submit for publication.

3 | RESULTS

The search identified 1038 papers, of which 339 were duplicates. After screening the titles and abstracts of the remaining 699 papers, 350 full-texts were reviewed. Overall, 35 studies met the inclusion criteria, with a further 12 identified from the references of included studies or by the reviewer team (Figure 1). The 47 included studies[4,22-67] represented a total of 32 849 hospitalized COVID-19 patients: 8417 (25.6%) with any reported smoking history, comprising 1501 current smokers, 5676 former smokers, and 1240 unspecified smokers; 22 420 (68.3%) never-smokers; and a further 2012 (6.1%) patients who did not currently smoke, though it was unclear whether they were former or never-smokers (Table 1).

There were 25 multicentre studies (three prospective[31,49,66] and 22 retrospective)[4,22,25-27,29,32-34,38,39,41-44,47,51,52,54,60,61,63] and 22 single-centre studies (two prospective,[36,55] two with prospective and retrospective components,[45,56] and 18 retrospective[23,24,28,30,35,37,40,46,48,50,53,57-59,62,64,65,67]). The majority of studies investigated a Chinese population (32/47, 68%), with the United States contributing 10 studies. Overall, study quality was good in 22 studies, fair in six and poor in 19 (Appendix p17). Of 38 studies disclosing funding status, 28 received funding.

Three studies[32,56,64] reported smoking index or pack-years by outcome of interest. Six studies[23,25,39,49,54,61] reported outcomes for tobacco smokers, including one[25] that had pooled outcomes with those of e-cigarette users. The remaining studies did not specify the substance of smoking.

Disease severity was graded according to the Chinese COVID-19-specific criteria in 14 studies,[24,28,32,35,38,46,48,53,55,57,63-66] the IDSA/ATS criteria in three studies[34,45,59] and a locally devised criteria in one study (Appendix p13).[54] Two studies[52,62] did not specify the criteria utilized.

Current smokers, whose outcomes were evaluated in 27 studies, had an overall prevalence of 6.2% (specifically, China: 8.7%, United States: 4.6%). They had a significantly increased risk of presenting with severe disease (RR: 1.80; 95% CI: 1.14-2.85; $P = .012$; $I^2 = 76\%$; Figure 2A), as well as severe or critical disease (RR: 1.98; 95% CI: 1.16-3.38; $P = .012$; $I^2 = 87\%$; Figure 2B), compared to former or never-smokers. Effects were


```
┌─────────────────────────────────────┐
│  1038 studies identified through    │
│  database search and abstracts      │
│  screened                           │
└─────────────────────────────────────┘
                  │
                  │      ┌──────────────────────────────────────┐
                  │      │ 688 studies excluded                 │
                  ├─────▶│   339 duplicates                     │
                  │      │   88 concerning other diseases       │
                  │      │   42 concerning other topics         │
                  │      │   167 reviews/editorials             │
                  │      │   24 conference abstracts/comments   │
                  │      │   28 non-clinical studies            │
                  │      └──────────────────────────────────────┘
                  │
┌─────────────────────────────────────┐
│  350 full-texts assessed for        │
│  eligibility                        │
└─────────────────────────────────────┘
                  │
                  │      ┌──────────────────────────────────────────┐
                  │      │ 315 studies excluded                     │
                  ├─────▶│   143 with no outcomes by smoking status │
                  │      │   30 with no outcomes of interest        │
                  │      │   1 with unclear outcomes for smoking    │
                  │      │   17 with less than 20 patients          │
                  │      │   112 reviews/editorials                 │
                  │      │   8 non-clinical studies                 │
                  │      │   2 full-texts not available in English  │
                  │      │   2 retracted since publication          │
                  │      └──────────────────────────────────────────┘
                  │
┌─────────────────────────────────────┐
│  35 studies included from database  │
│  search                             │
└─────────────────────────────────────┘
      ▲           │
┌─────────────────┐
│ 12 eligible studies identified      │
│ through handsearch by reviewers     │
└─────────────────┘
                  │
┌─────────────────────────────────────┐
│  47 studies included overall        │
└─────────────────────────────────────┘
```
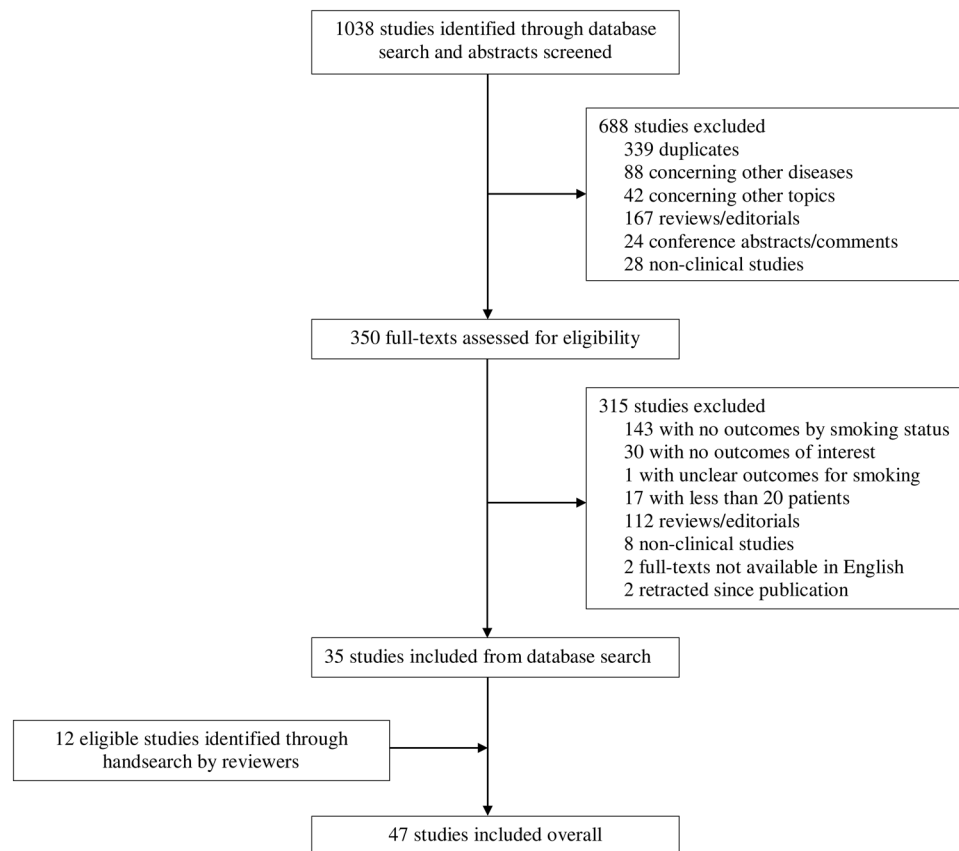
FIGURE 1 Flow diagram of selection of included studies

TABLE 1 Characteristics of included studies

	Setting	Study design	Number of centers	Study period	Number of patients, current smokers vs former/never-smokers	Number of patients, any smoking history vs never-smokers	Study quality
Azar et al[22]	United States	Cohort	24	Jan-Apr	10 vs 216	73 vs 153	Fair
Bhargava et al[23]	United States	Cohort	1	Mar-Apr	11 vs 186	…	Good
Bi et al[24]	China	Cohort	1	Jan-Mar	8 vs 105	…	Good
Brenner et al[25]	International	Cohort	1+	-Apr	11 vs 150	…	Poor
Buckner et al[26]	United States	Case series	3	Mar-May	…	22 vs 64	Poor
CDC COVID-19 Response Team[27]	United States	Cohort	1+	Feb-Mar	27 vs 1467	105 vs 1389	Poor
Chen et al[28]	China	Case series	1	Jan-Mar	…	15 vs 130	Poor
Chen et al[29]	China	Cohort	575	-Jan	…	111 vs 1479	Good
Chen et al[30]	China	Case series	1	Jan-Feb	12 vs 262	…	Poor
Docherty et al[31]	UK	Cohort[‡]	208	Feb-May	852 vs 13 332	5216 vs 8968	Good
Feng et al[32]	China	Cohort	3	Jan-Mar	…	44 vs 410	Good
Goyal et al[33]	United States	Case series	2	Mar-Apr	20 vs 373	98 vs 295	Poor
Guan et al[34]	China	Cohort	552	Dec-Jan	137 vs 948	158 vs 927	Poor
Hu et al[35]	China	Case series	1	Jan-Mar	…	38 vs 285	Good
Huang et al[36]	China	Case series[‡]	1	Dec-Jan	3 vs 38	…	Poor
Huang et al[37]	China	Cohort	1	Jan-Mar	56 vs 288	…	Good
Huang et al[38]	China	Case series	8	Jan-Feb	…	16 vs 186	Good
Hur et al[39]	United States	Cohort	10	Mar-Apr	16 vs 470	163 vs 323	Good
Inciardi et al[40]	Italy	Cohort	1	Mar-Mar	…	17 vs 82	Poor
Ji et al[41]	China	Cohort	2	Jan-Mar	…	19 vs 189	Good
Kalligeros et al[42]	United States	Cohort	3	Feb-Apr	12 vs 91	48 vs 55	Good
Klang et al[43]	United States	Cohort	5	Mar-May	…	793 vs 2613	Good
Kuderer et al[44]	International[a]	Cohort	1+	Mar-May	25 vs 406	226 vs 205	Fair
Li et al[45]	China	Cohort[†]	1	Jan-Mar	41 vs 503	92 vs 452	Good
Li et al[46]	China	Case series	1	Jan-Feb	…	7 vs 18	Poor
Liu et al[47]	China	Cohort	3	Dec-Jan	…	5 vs 73	Good
Petrilli et al[49]	United States	Cohort[‡]	4	Mar-May	141 vs 2145	702 vs 1584	Good
Qin et al[48]	China	Cohort	1	Jan-Feb	…	7 vs 445	Poor
Rastrelli et al[50]	Italy	Case series	1	…	1 vs 30	12 vs 19	Poor
Shi et al[51]	China	Cohort	2	Jan-Mar	…	16 vs 290	Good
Shi et al[52]	China	Cohort	1+	-Feb	…	40 vs 434	Good
Sun et al[53]	China	Cohort	1	Feb-Mar	…	12 vs 45	Good
Toussie et al[54]	United States	Cohort	1+	Mar-Mar	…	29 vs 94	Fair
Wan et al[55]	China	Case series[‡]	1	Jan-Feb	9 vs 126	…	Poor
Wang et al[56]	China	Cohort[†]	1	…	41 vs 503	92 vs 452	Poor
Wang et al[57]	China	Cohort	1	Jan-Feb	16 vs 109	16 vs 109	Poor
Yang et al[58]	China	Cohort	1	Dec-Feb	…	2 vs 50	Poor
Yao et al[59]	China	Cohort	1	Jan-Mar	4 vs 104	…	Good

TABLE 1 (Continued)

	Setting	Study design	Number of centers	Study period	Number of patients, current smokers vs former/never-smokers	Number of patients, any smoking history vs never-smokers	Study quality
Yu et al[60]	China	Cohort	24	Jan-Mar	13 vs 408	…	Good
Yu et al[61]	China	Cross-sectional	2	Jan-Feb	…	5 vs 65	Good
Yu et al[62]	China	Cohort	1	Jan-Mar	…	16 vs 76	Poor
Yu et al[63]	China	Cohort	1+	Dec-Feb	…	26 vs 265	Fair
Zhang et al[64]	China	Case series	1	Jan-Feb	2 vs 138	9 vs 131	Poor
Zhang et al[65]	China	Cohort	1	Jan-Feb	6 vs 114	…	Fair
Zheng et al[66]	China	Cohort[‡]	3	Jan-Feb	8 vs 58	…	Fair
Zheng et al[67]	China	Case series	1	Jan-Feb	8 vs 65	8 vs 65	Poor
Zhou et al[4]	China	Cohort	2	Dec-Jan	11 vs 180	…	Good

Note: All studies are retrospective except: [†]ambispective (includes prospective and retrospective components) and [‡]prospective.
[a]Contains data from the United States, Canada, and Spain.

consistent when only analyzing studies using the COVID-19-specific criteria (Appendix p22). On sensitivity analysis, including only good-quality studies resulted in these effects becoming nonsignificant. There were no significant effects on in-hospital outcomes, including disease progression (RR: 1.54; 95% CI: 0.52-4.58; $P = .439$; $I^2 = 81\%$; Appendix p19), ICU admission (RR: 0.72; 95% CI: 0.42-1.24; $P = .237$; $I^2 = 40\%$; Appendix p20), mechanical ventilation requirement (RR: 1.13; 95% CI: 0.75-1.72; $P = .561$; $I^2 = 32\%$; Appendix p21) or mortality (RR: 1.46; 95% CI: 0.83-2.60; $P = .192$; $I^2 = 81\%$; Figure 2C). There were no differences in outcomes by country of origin (Appendix p23). A meta-analysis was not performed for critical disease alone as only one study reported this outcome.

The overall prevalence of a smoking history, including current, former, and/or unspecified smokers, was 26.9% (specifically, China: 10.3%, United States: 23.6%). Their outcomes were investigated in 35 studies. Compared to never-smokers, a history of smoking significantly increased the risk of presenting with severe disease (RR: 1.31; 95% CI: 1.12-1.54; $P = .001$; $I^2 = 12\%$; Figure 3A), as well as severe or critical disease (RR: 1.35; 95% CI: 1.19-1.53; $P < .0001$; $I^2 = 19\%$; Figure 3B). However, only the effect on severe or critical disease remained significant when limiting the analysis to only studies using the COVID-19-specific criteria for grading severity (Appendix p29). The effect on critical disease alone was not statistically significant (RR: 1.44; 95% CI: 0.95-2.17; $P = .085$; $I^2 = 0\%$; Appendix p25). However, a smoking history significantly increased mortality risk (RR: 1.26; 95% CI: 1.20-1.32; $P < .0001$; $I^2 = 0\%$; Figure 3C) in addition to other in-hospital outcomes, such as disease progression (RR: 2.18; 95% CI: 1.06-4.49; $P = .035$; $I^2 = 69\%$; Appendix p26) and mechanical ventilation requirement (RR: 1.20; 95% CI: 1.01-1.42; $P = .043$; $I^2 = 0\%$; Appendix p28). There was no statistically significant difference in ICU admission (RR: 1.12; 95% CI: 0.96-1.31; $P = .157$; $I^2 = 0\%$; Appendix p27). Sensitivity analyses excluding lower-quality studies supported the primary analyses for all outcomes of interest (Appendix p29). Only the mortality analysis facilitated comparison by country, in which significant detrimental effects were observed in publications from China,

United States, and the UK, but not Italy, which contributed one study only for this outcome (Appendix p29).

Overall, there was a moderate-to-high degree of heterogeneity between studies evaluating the effects of current smoking and a low degree of heterogeneity between studies investigating a history of smoking. The potential for publication bias was only detected in the comparison of disease progression in patients with a smoking history (Appendix p26), though heterogeneity was high for this outcome.

4 | DISCUSSION

To our knowledge, this is the largest meta-analysis amongst peer-reviewed literature assessing the effect of smoking tobacco on the severity of COVID-19. Principally, the present analysis found that current smokers have an increased risk of presenting to hospital with severe COVID-19 and are approximately twice as likely to experience severe or critical COVID-19 as former or never-smokers. While this risk became nonsignificant following sensitivity analysis of good-quality studies only, there were only two studies for each outcome and none graded disease severity by COVID-19-specific criteria, thus precluding meaningful interpretation. Overall, there was a high degree of heterogeneity amongst studies evaluating current smoking, even when analyzing good-quality studies only. For patients with a smoking history, there is an increased risk of presentation to hospital with severe, as well as severe or critical, COVID-19 and subsequent increased risk of in-hospital mortality. Additionally, these patients were more likely to experience disease progression and require mechanical ventilation. That all outcomes remained significant on inclusion of only good-quality studies suggests these analyses represent true effects. A high level of heterogeneity was only observed in assessing the effect of smoking history on disease progression, which is likely secondary to substantial inter-study variation in defining progression. This outcome also displayed potential for publication bias, however, none was found in other analyses, indicating the low impact of publication bias on our results.
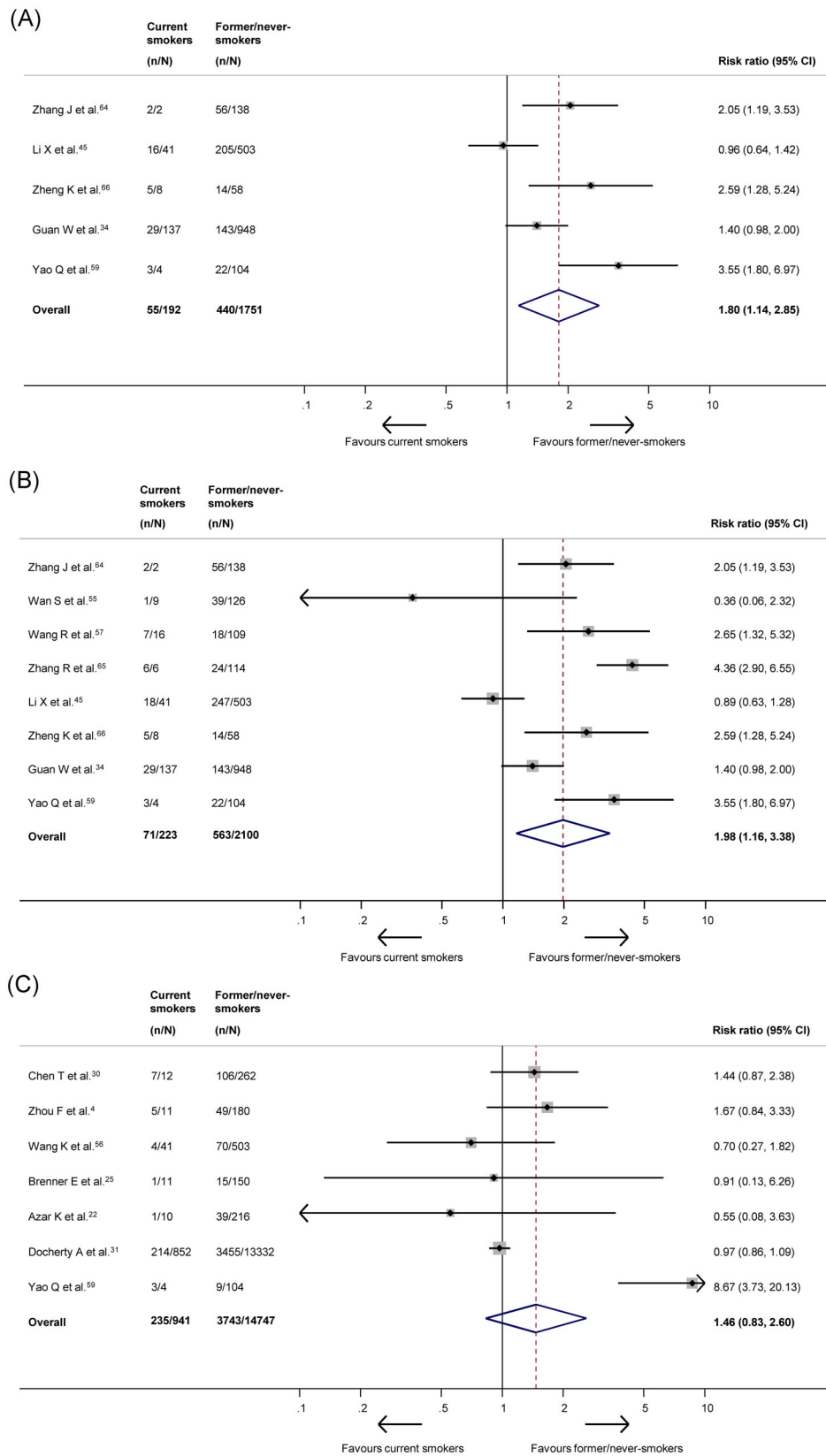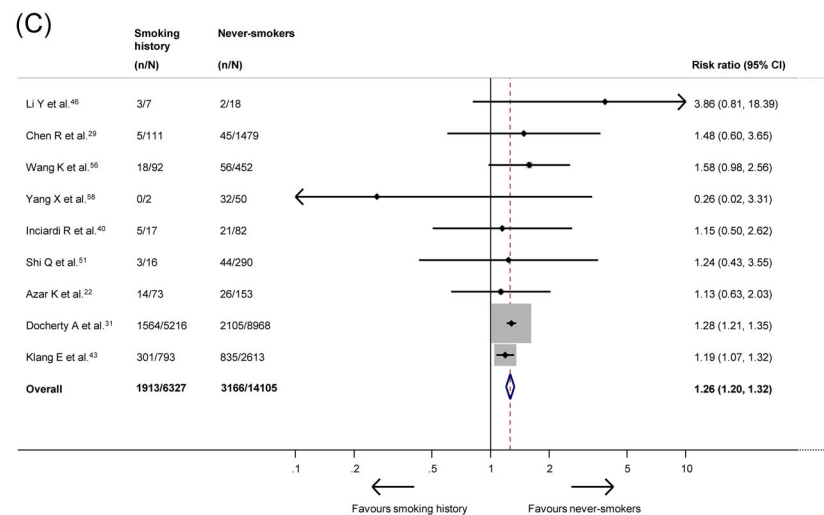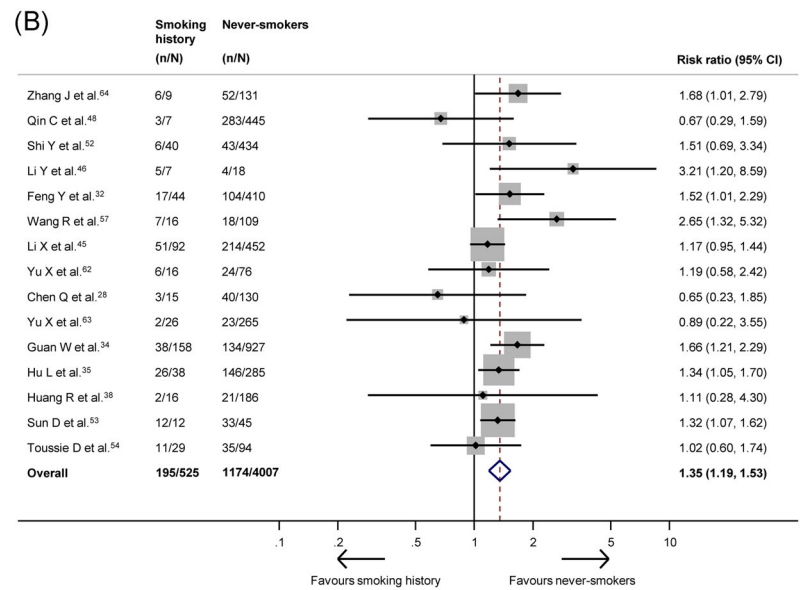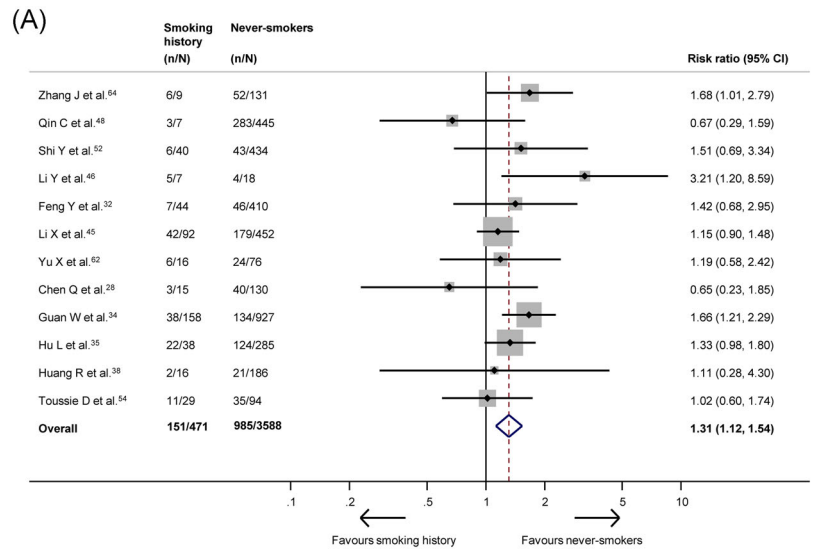
FIGURE 2 A, Forest plot showing the effect of current smoking on severe COVID-19. B, Forest plot showing the effect of current smoking on severe or critical COVID-19. C, Forest plot showing the effect of current smoking on mortality. COVID-19, coronavirus disease 2019

FIGURE 3 A, Forest plot showing the effect of a smoking history on severe COVID-19. B, Forest plot showing the effect of a smoking history on severe or critical COVID-19. C, Forest plot showing the effect of a smoking history on mortality. COVID-19, coronavirus disease 2019

(A)

	Smoking history (n/N)	Never-smokers (n/N)	Risk ratio (95% CI)
Zhang J et al.[64]	6/9	52/131	1.68 (1.01, 2.79)
Qin C et al.[48]	3/7	283/445	0.67 (0.29, 1.59)
Shi Y et al.[52]	6/40	43/434	1.51 (0.69, 3.34)
Li Y et al.[46]	5/7	4/18	3.21 (1.20, 8.59)
Feng Y et al.[32]	7/44	46/410	1.42 (0.68, 2.95)
Li X et al.[45]	42/92	179/452	1.15 (0.90, 1.48)
Yu X et al.[62]	6/16	24/76	1.19 (0.58, 2.42)
Chen Q et al.[28]	3/15	40/130	0.65 (0.23, 1.85)
Guan W et al.[34]	38/158	134/927	1.66 (1.21, 2.29)
Hu L et al.[35]	22/38	124/285	1.33 (0.98, 1.80)
Huang R et al.[38]	2/16	21/186	1.11 (0.28, 4.30)
Toussie D et al.[54]	11/29	35/94	1.02 (0.60, 1.74)
Overall	**151/471**	**985/3588**	**1.31 (1.12, 1.54)**

Favours smoking history · Favours never-smokers

(B)

	Smoking history (n/N)	Never-smokers (n/N)	Risk ratio (95% CI)
Zhang J et al.[64]	6/9	52/131	1.68 (1.01, 2.79)
Qin C et al.[48]	3/7	283/445	0.67 (0.29, 1.59)
Shi Y et al.[52]	6/40	43/434	1.51 (0.69, 3.34)
Li Y et al.[46]	5/7	4/18	3.21 (1.20, 8.59)
Feng Y et al.[32]	17/44	104/410	1.52 (1.01, 2.29)
Wang R et al.[57]	7/16	18/109	2.65 (1.32, 5.32)
Li X et al.[45]	51/92	214/452	1.17 (0.95, 1.44)
Yu X et al.[62]	6/16	24/76	1.19 (0.58, 2.42)
Chen Q et al.[28]	3/15	40/130	0.65 (0.23, 1.85)
Yu X et al.[63]	2/26	23/265	0.89 (0.22, 3.55)
Guan W et al.[34]	38/158	134/927	1.66 (1.21, 2.29)
Hu L et al.[35]	26/38	146/285	1.34 (1.05, 1.70)
Huang R et al.[38]	2/16	21/186	1.11 (0.28, 4.30)
Sun D et al.[53]	12/12	33/45	1.32 (1.07, 1.62)
Toussie D et al.[54]	11/29	35/94	1.02 (0.60, 1.74)
Overall	**195/525**	**1174/4007**	**1.35 (1.19, 1.53)**

Favours smoking history · Favours never-smokers

(C)

	Smoking history (n/N)	Never-smokers (n/N)	Risk ratio (95% CI)
Li Y et al.[46]	3/7	2/18	3.86 (0.81, 18.39)
Chen R et al.[29]	5/111	45/1479	1.48 (0.60, 3.65)
Wang K et al.[56]	18/92	56/452	1.58 (0.98, 2.56)
Yang X et al.[58]	0/2	32/50	0.26 (0.02, 3.31)
Inciardi R et al.[40]	5/17	21/82	1.15 (0.50, 2.62)
Shi Q et al.[51]	3/16	44/290	1.24 (0.43, 3.55)
Azar K et al.[22]	14/73	26/153	1.13 (0.63, 2.03)
Docherty A et al.[31]	1564/5216	2105/8968	1.28 (1.21, 1.35)
Klang E et al.[43]	301/793	835/2613	1.19 (1.07, 1.32)
Overall	**1913/6327**	**3166/14105**	**1.26 (1.20, 1.32)**

Favours smoking history · Favours never-smokers

Our finding that current smoking is associated with increased disease severity in COVID-19 patients validates previous findings from several smaller meta-analyses in a much larger patient population, achieved through a more rigorous, prospectively registered methodology.[9-12,21] The finding that patients with a smoking history are at increased risk of more severe disease, and increased mortality, also confirms previous findings of a smaller meta-analysis.[11] The association of both current smoking and smoking history with greater severity of COVID-19 is biologically plausible for a wealth of reasons. Smoking tobacco is the primary cause of preventable disease, disability, and death in the United States, and is responsible for over 8 million deaths worldwide per year.[68] Smoking is a major risk factor for adverse respiratory and cardiovascular outcomes, in addition to a wide range of malignant and nonmalignant disease.[68] In addition, smoking increases severity and mortality of both bacterial and viral infections through the induction of mechanical and structural changes in the respiratory tract and alteration of cell- and humoral-mediated immune responses.[69,70] In the context of respiratory viruses, smoking has been reported to cause increased hospital and ICU admissions with influenza infection, greater severity with respiratory syncytial virus bronchiolitis and increased mortality with viral pneumonia.[71-73]

With regard to coronaviruses, in particular, smoking is associated with increased susceptibility and mortality in MERS-CoV infection, potentially due to upregulation of dipeptidyl peptidase-IV, the host receptor for MERS-CoV, in smokers.[74,75] The angiotensin-converting enzyme-2 (ACE-2), previously shown to be the host receptor for SARS-CoV, has also been proven to be the host receptor for SARS-CoV-2, facilitating initial intracellular entry via interactions with viral spike glycoproteins.[76] Subsequent studies have confirmed that ACE-2 expression is upregulated in human lung tissue samples taken from both current and past smokers, likely mediated by the α-7 subtype of the nicotinic acetylcholine receptor.[77-81] In a series of elegant in vitro experiments, Smith et al[80] report a consistent correlation between smoking history and increased ACE-2 expression that was dose-dependent, detectable in both bulk and single-cell analyses, and remained significant after multivariate linear regression controlling for age, sex, race, and body mass index. It is, therefore, plausible that smokers are exposed to higher SARS-CoV-2 loads as a result of increased expression of ACE-2, which may provide a mechanistic explanation for the increased risk of severe disease and mortality associated with smoking in COVID-19 patients that we report. Moreover, the inhibition of SARS-CoV-2 progression in vitro by human recombinant soluble ACE-2, a neutralizing agent, holds therapeutic potential and is currently in phase II clinical trials (ClinicalTrials.gov Identifier: NCT04335136).[82] However, to complicate matters, previous studies also report that postentry viral-mediated downregulation of ACE-2 played a major role in the pathogenesis of SARS-CoV-associated acute lung injury.[83,84] Smoking itself has been postulated as having varying, organ-specific effects on ACE-2 levels, with specific cigarette components, such as nicotine, potentially exerting a different effect to whole smoke.[80] Therefore, further studies characterizing the complex relationship of smoking and ACE-2 in COVID-19 are warranted.

That smoking history is associated with a significantly increased risk of in-hospital mortality in COVID-19 patients, whilst current smoking is not, is a surprising finding. Reductions in morbidity and all-cause mortality following smoking cessation are well characterized and thus former smokers would be expected to have better baseline health status and improved outcomes.[68] A systematic review assessing prevalence of current smokers who were hospitalized for COVID-19 reported a pooled prevalence of 6.5% and propose that in view of the lower than expected prevalence of current smokers compared to population estimates, current smoking is not a predisposing factor for hospitalization and smoking and/or nicotine may exert a protective effect against severe COVID-19.[85] The idea that smoking and/or nicotine may be protective against COVID-19 is echoed by several preprint studies that gained widespread media attention.[14] Although these hypotheses may explain the nonsignificant association of current smoking and increased mortality that we report, since the majority of included studies did not statistically adjust the effect of smoking for baseline covariates, it is not appropriate to compare the prevalence of smoking in hospitalized COVID-19 patients with overall population estimates, as the two populations are inherently different with regards to demographic factors. We believe there are far more credible reasons for the nonsignificant association between current smoking and mortality that we report and the low prevalence of smoking among patients with COVID-19 in published studies.

Predominantly, in the context of a global pandemic, accurately recording smoking history is likely to be of low priority for frontline clinicians whose principal focus is stabilizing severely and critically ill patients. Therefore, patients may have been too acutely unwell to answer questions or clinicians may not enquired directly about smoking status, leading to misclassification of smokers as nonsmokers. Similarly, collateral history collected from family members or referring clinicians is likely to be less accurate than ascertainment of patient-reported smoking status. Additionally, in an example of reverse causality, hospitalized patients are more likely to have quit smoking on admission, resulting in additional potential misclassification of current smokers as former or nonsmokers. Given the well-known scarcity of ICU resources such as ventilators, it is also possible that social desirability bias may have contributed to patients not reporting current smoking for fear of being denied access to such interventions, further exacerbating misclassification bias.[14,86] Finally, given the association of smoking with lower socioeconomic status,[87] it is possible that current smokers are exhibiting worse health-seeking behaviors and either self-treating or deteriorating in the community. Thus, they would not be accounted for in the reported studies which assessed hospitalized patients, leading to survivorship bias and lower event rates for in-hospital mortality. Due to these factors, the summary estimate for in-hospital mortality we report has likely been biased towards a null result for current smokers. Similarly, the twofold increase in risk of severe or critical disease for current smokers is likely an underestimate of the effect size.

With no targeted therapies against COVID-19 currently available, as a scientific community, we must focus on prevention, particularly for those at risk of severe or critical disease. Frontline clinicians must conscientiously record accurate smoking histories in all confirmed COVID-19 patients, both for triage of vulnerable patients and to support future research efforts. During the current pandemic,

independent surveys have reported increased smoking frequency in current smokers and high rates of relapse in former smokers,[88-90] which is unsurprising given the stress, isolation, and other adverse psychosocial repercussions of life during a global pandemic.[91,92] Considering our finding that current smoking and smoking history are associated with increased COVID-19 severity, urgent public health measures emphasizing smoking cessation advice, support and pharmacotherapy must be provided to reduce overall disease burden, despite a currently altered social landscape. Good-quality studies have proven the benefits of mobile phone-based interventions,[93] highlighting that even during periods of social distancing and self-isolation, remote methods of smoking cessation may be feasible and efficacious. Furthermore, as countries begin easing lockdown restrictions, it is imperative that governments and policymakers protect vulnerable populations, such as current and former smokers, through adequate shielding measures and appropriate legislation.

The present analysis has several limitations, principally the use of aggregate data for our meta-analysis, which precludes adjustment for certain covariates reported to be predictive of disease severity, such as age, gender, and comorbidities,[3-5] and prevents examination of heterogeneity and subgroup analysis at the patient level. The use of individual patient data may have addressed this, however, considering the urgency of our research question and direct applicability to patient care, the considerable time burden associated with conducting an individual patient data meta-analysis was deemed inappropriate. Also, with most studies reporting on Chinese populations, we cannot exclude the possibility of ethnic differences in smoking and susceptibility to severe COVID-19 caused by smoking, which may have confounded our analysis. However, this reflects the current landscape of peer-reviewed literature, which at the present time consists mainly of data from China. We were also unable to assess the effect of e-cigarettes on COVID-19 as no studies collected separate data on their usage, which would have been informative considering rises in popularity of these products. Finally, as discussed, the high likelihood of misclassification bias concerning current smoking status across included studies suggests that our analysis potentially underestimates the impact of current smoking on both disease severity and mortality, creating an even more compelling argument for urgent public health measures to support smoking cessation during the present time.

In conclusion, in the largest meta-analysis available amongst peer-reviewed literature, we report that both current smoking and a smoking history significantly increased COVID-19 severity, whilst smoking history also significantly increased mortality risk. Due to problems with potential misclassification of current smokers among included studies, the analysis likely underestimates the likelihood of severity in this patient population. As the COVID-19 pandemic continues to burden societies worldwide, our analysis suggests that smoking represents one of the most immediately modifiable risk factors to reduce the substantial morbidity associated with the disease. In light of this finding, governments, policymakers, and other key stakeholders must ensure that appropriate measures are taken to support and maintain smoking cessation to protect vulnerable populations and reduce the strain placed on healthcare systems working at full capacity during this global crisis.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

AK conceptualized the work. RKR, WNC, AS, AD, PTS, and AK were responsible for acquisition, analysis, and interpretation of data. RKR, WNC, and AK drafted the manuscript. AS, AD, PTS, and AK provided critical revisions. All authors gave final approval of the version to be published and agree to be accountable for all aspects of the work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Rohin K. Reddy http://orcid.org/0000-0002-4726-157X

REFERENCES

1. WHO. Coronavirus disease (COVID-19) Situation Report - 190. July 28 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200728-covid-19-sitrep-190.pdf?sfvrsn=fec17314_2. Accessed July 29, 2020.
2. Johns Hopkins University and Medicine Coronavirus Resource Center. COVID-19 World Map. July 29 2020. https://coronavirus.jhu.edu/map.html. Accessed July 29, 2020.
3. Wu Z, McCoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention [published online ahead of print February 24, 2020]. *JAMA.* https://doi.org/10.1001/jama.2020.2648
4. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* 2020;395:1054-1062.
5. Alqahtani JS, Oyelade T, Aldhahir AM, et al. Prevalence, severity and mortality associated with COPD and smoking in patients with COVID-19: a rapid systematic review and meta-analysis. *PLoS One.* 2020;15: e0233147.
6. van Zyl-Smit RN, Richards G, Leone FT. Tobacco smoking and COVID-19 infection. *Lancet Respir Med.* 2020;8:664-665. https://doi.org/10.1016/S2213-2600(20)30239-3
7. Vardavas CI, Nikitara K. COVID-19 and smoking: a systematic review of the evidence. *Tob Induc Dis.* 2020;18:20.
8. Lippi G, Henry BM. Active smoking is not associated with severity of coronavirus disease 2019 (COVID-19). *Eur J Intern Med.* 2020;75:107-108.
9. Zhao Q, Meng M, Kumar R, et al. The impact of COPD and smoking history on the severity of COVID-19: a systemic review and meta-analysis [published online ahead of print April 15, 2020]. *J Med Virol.* https://doi.org/10.1002/jmv.25889
10. Zheng Z, Peng F, Xu B, et al. Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. *J Infect.* 2020;81:16. https://doi.org/10.1016/j.jinf.2020.04.021

11. Patanavanich R, Glantz S. Smoking is associated with COVID-19 progression: a meta-analysis [published online ahead of print May 13, 2020]. *Nicotine Tob Res.* https://doi.org/10.1093/ntr/ntaa082

12. Karanasos A, Aznaouridis K, Latsios G, et al. Impact of smoking status on disease severity and mortality of hospitalized patients with COVID-19 infection [published online ahead of print June 20, 2020]. *Nicotine Tob Res.* https://doi.org/10.1093/ntr/ntaa107

13. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017;358:j4008.

14. Wilson C. Smokers are actually at a higher risk of dying from covid-19. *New Sci.* 2020;246:8-9.

15. WHO. WHO statement: Tobacco use and COVID-19. May 11 2020. https://www.who.int/news-room/detail/11-05-2020-who-statement-tobacco-use-and-covid-19. Accessed July 18, 2020.

16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.* 2009;339:b2535.

17. National Health Commission & National Administration of Traditional Chinese Medicine. Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 7). *Chin Med J.* 2020;133:1087-1095.

18. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med.* 2020;200:e45-e67.

19. GA Wells BS, O'Connell D, Peterson, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomised studies in meta-analyses. 2019. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed July 18, 2020.

20. Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med.* 2006;25:3433-57.

21. Khajuria A, Charles W, Sklavounos A, Reddy R The effects of smoking on COVID-19 severity: a systematic review and meta-analysis. April 27 2020. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=180920. Accessed July 18, 2020.

22. Azar KMJ, Shen Z, Romanelli RJ, et al. Disparities in outcomes among COVID-19 patients in a large health care system In California. *Health Aff (Millwood).* 2020;39:1253-1262. https://doi.org/10.1377/hlthaff.2020.00598

23. Bhargava A, Fukushima EA, Levine M, et al. Predictors for severe COVID-19 infection [published online ahead of print May 30, 2020]. *Clin Infect Dis.* https://doi.org/10.1093/cid/ciaa674

24. Bi X, Su Z, Yan H, et al. Prediction of severe illness due to COVID-19 based on an analysis of initial Fibrinogen to Albumin Ratio and Platelet count. *Platelets.* 2020;31:674-679. https://doi.org/10.1080/09537104.2020.1760230

25. Brenner EJ, Ungaro RC, Gearry RB, et al. Corticosteroids, but not TNF antagonists, are associated with adverse COVID-19 outcomes in patients with inflammatory bowel diseases: results from an international registry. *Gastroenterology.* 2020;S0016-5085(20):30655-30657. https://doi.org/10.1053/j.gastro.2020.05.032

26. Buckner FS, McCulloch DJ, Atluri V, et al. Clinical features and outcomes of 105 hospitalized patients with COVID-19 in Seattle, Washington [published online ahead of print May 22, 2020]. *Clin Infect Dis.* https://doi.org/10.1093/cid/ciaa632

27. CDC COVID-19 Response Team. Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019—United States, February 12-March 28, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69:283-386.

28. Chen Q, Zheng Z, Zhang C, et al. Clinical characteristics of 145 patients with corona virus disease 2019 (COVID-19) in Taizhou, Zhejiang, China. *Infection.* 2020;48:543-551. https://doi.org/10.1007/s15010-020-01432-5

29. Chen R, Liang W, Jiang M, et al. Risk factors of fatal outcome in hospitalized subjects with coronavirus disease 2019 From a nation-wide analysis in China. *Chest.* 2020;158:97-105.

30. Chen T, Wu D, Chen H, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ.* 2020;368:m1091.

31. Docherty AB, Harrison EM, Green CA, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: prospective observational cohort study. *BMJ.* 2020;369:m1985.

32. Feng Y, Ling Y, Bai T, et al. COVID-19 with different severities: a multicenter study of clinical features. *Am J Respir Crit Care Med.* 2020;201:1380-1388.

33. Goyal P, Choi JJ, Pinheiro LC, et al. Clinical characteristics of Covid-19 in New York City. *N Engl J Med.* 2020;382:2372-2374.

34. Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* 2020;382:1708-1720.

35. Hu L, Chen S, Fu Y, et al. Risk factors associated with clinical outcomes in 323 COVID-19 hospitalized patients in Wuhan, China [published online ahead of print May 3, 2020]. *Clin Infect Dis.* https://doi.org/10.1093/cid/ciaa539

36. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395:497-506.

37. Huang J, Cheng A, Lin S, Zhu Y, Chen G. Individualized prediction nomograms for disease progression in mild COVID-19 [published online ahead of print May 5, 2020]. *J Med Virol.* https://doi.org/10.1002/jmv.25969

38. Huang R, Zhu L, Xue L, et al. Clinical findings of patients with coronavirus disease 2019 in Jiangsu Province, China: a retrospective, multi-center study. *PLoS Negl Trop Dis.* 2020;14:e0008280.

39. Hur K, Price CPE, Gray EL, et al. Factors associated with intubation and prolonged intubation in hospitalized patients with COVID-19. *Otolaryngol Head Neck Surg.* 2020;163:170-178.

40. Inciardi RM, Adamo M, Lupi L, et al. Characteristics and outcomes of patients hospitalized for COVID-19 and cardiac disease in Northern Italy. *Eur Heart J.* 2020;41:1821-1829.

41. Ji D, Zhang D, Xu J, et al. Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score [published online ahead of print April 9, 2020]. *Clin Infect Dis.* https://doi.org/10.1093/cid/ciaa414

42. Kalligeros M, Shehadeh F, Mylona EK, et al. Association of obesity with disease severity among patients with coronavirus disease 2019. *Obesity (Silver Spring).* 2020;28:1200-1204.

43. Klang E, Kassim G, Soffer S, Freeman R, Levin MA, Reich DL. Morbid obesity as an independent risk factor for COVID-19 mortality in hospitalized patients younger than 50 [published online ahead of print May 23, 2020]. *Obesity (Silver Spring).* https://doi.org/10.1002/oby.22913

44. Kuderer NM, Choueiri TK, Shah DP, et al. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. *Lancet.* 2020;395:1907-1918.

45. Li X, Xu S, Yu M, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. *J Allergy Clin Immunol.* 2020;146:110-118. https://doi.org/10.1016/j.jaci.2020.04.006

46. Li YK, Peng S, Li LQ, et al. Clinical and transmission characteristics of Covid-19—a retrospective study of 25 cases from a Single Thoracic Surgery Department. *Curr Med Sci.* 2020;40:295-300.

47. Liu W, Tao ZW, Wang L, et al. Analysis of factors associated with disease outcomes in hospitalized patients with 2019 novel coronavirus disease. *Chin Med J.* 2020;133:1032-1038.

48. Qin C, Zhou L, Hu Z, et al. Dysregulation of immune response in patients with COVID-19 in Wuhan, China. *Clin Infect Dis.* 2020;71:762-768. https://doi.org/10.1093/cid/ciaa248

49. Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ*. 2020;369:m1966.

50. Rastrelli G, Di Stasi V, Inglese F, et al. Low testosterone levels predict clinical adverse outcomes in SARS-CoV-2 pneumonia patients [published online ahead of print May 20, 2020]. *Andrology*. https://doi.org/10.1111/andr.12821

51. Shi Q, Zhang X, Jiang F, et al. Clinical characteristics and risk factors for mortality of COVID-19 patients with diabetes in Wuhan, China: a two-center, retrospective study. *Diabetes Care*. 2020;43:1382-1391.

52. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. *Crit Care*. 2020;24:108.

53. Sun DW, Zhang D, Tian RH, et al. The underlying changes and predicting role of peripheral blood inflammatory cells in severe COVID-19 patients: a sentinel? *Clin Chim Acta*. 2020;508:122-129. https://doi.org/10.1016/j.cca.2020.05.027

54. Toussie D, Voutsinas N, Finkelstein M, et al. Clinical and chest radiography features determine patient outcomes in young and middle age adults with COVID-19 [published online ahead of print May 14, 2020]. *Radiology*. https://doi.org/10.1148/radiol.2020201754

55. Wan S, Xiang Y, Fang W, et al. Clinical features and treatment of COVID-19 patients in northeast Chongqing. *J Med Virol*. 2020;92:797-806.

56. Wang K, Zhang Z, Yu M, Tao Y, Xie M. 15-day mortality and associated risk factors for hospitalized patients with COVID-19 in Wuhan, China: an ambispective observational cohort study. *Intensive Care Med*. 2020;46:1472-1474. https://doi.org/10.1007/s00134-020-06047-w

57. Wang R, Pan M, Zhang X, et al. Epidemiological and clinical features of 125 Hospitalized Patients with COVID-19 in Fuyang, Anhui, China. *Int J Infect Dis*. 2020;95:421-428.

58. Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med*. 2020;8:475-481.

59. Yao Q, Wang P, Wang X, et al. A retrospective study of risk factors for severe acute respiratory syndrome coronavirus 2 infections in hospitalized adult patients. *Pol Arch Intern Med*. 2020;130:390-399.

60. Yu Q, Wang Y, Huang S, et al. Multicenter cohort study demonstrates more consolidation in upper lungs on initial CT increases the risk of adverse clinical outcome in COVID-19 patients. *Theranostics*. 2020;10:5641-5648.

61. Yu T, Cai S, Zheng Z, et al. Association between clinical manifestations and prognosis in patients with COVID-19. *Clin Ther*. 2020;42:964-972.

62. Yu X, Sun S, Shi Y, Wang H, Zhao R, Sheng J. SARS-CoV-2 viral load in sputum correlates with risk of COVID-19 progression. *Crit Care*. 2020;24:170.

63. Yu X, Sun X, Cui P, et al. Epidemiological and clinical characteristics of 333 confirmed cases with coronavirus disease 2019 in Shanghai, China. *Transbound Emerg Dis*. 2020;67:1697-1707. https://doi.org/10.1111/tbed.13604

64. Zhang JJ, Dong X, Cao YY, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy*. 2020;75:1730-1741. https://doi.org/10.1111/all.14238

65. Zhang R, Ouyang H, Fu L, et al. CT features of SARS-CoV-2 pneumonia according to clinical presentation: a retrospective analysis of 120 consecutive patients from Wuhan city. *Eur Radiol*. 2020;30:4417-4426. https://doi.org/10.1007/s00330-020-06854-1

66. Zheng KI, Gao F, Wang XB, et al. Letter to the Editor: obesity as a risk factor for greater severity of COVID-19 in patients with metabolic associated fatty liver disease. *Metabolism*. 2020;108:154244.

67. Zheng Y, Xiong C, Liu Y, et al. Epidemiological and clinical characteristics analysis of COVID-19 in the surrounding areas of Wuhan, Hubei Province in 2020. *Pharmacol Res*. 2020;157:104821.

68. Substance Abuse and Mental Health Services Administration (US), Office of the Surgeon General (US). *Smoking Cessation: a Report of the Surgeon General*. Washington (DC): US Department of Health and Human Services; 2020.

69. Arcavi L, Benowitz NL. Cigarette smoking and infection. *Arch Intern Med*. 2004;164:2006-16.

70. Huttunen R, Heikkinen T, Syrjänen J. Smoking and the outcome of infection. *J Intern Med*. 2011;269:258-269.

71. Han L, Ran J, Mak YW, et al. Smoking and influenza-associated morbidity and mortality: a systematic review and meta-analysis. *Epidemiology*. 2019;30:405-17.

72. Bradley JP, Bacharier LB, Bonfiglio J, et al. Severity of respiratory syncytial virus bronchiolitis is affected by cigarette smoke exposure and atopy. *Pediatrics*. 2005;115:e7-e14.

73. Guo L, Wei D, Zhang X, et al. Clinical features predicting mortality risk in patients with viral pneumonia: the MuLBSTA Score. *Front Microbiol*. 2019;10:2752.

74. Park JE, Jung S, Kim A, Park JE. MERS transmission and risk factors: a systematic review. *BMC Public Health*. 2018;18:574.

75. Seys LJM, Widagdo W, Verhamme FM, et al. DPP4, the Middle East respiratory syndrome coronavirus receptor, is upregulated in lungs of smokers and chronic obstructive pulmonary disease patients. *Clin Infect Dis*. 2018;66:45-53.

76. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-273.

77. Brake SJ, Barnsley K, Lu W, McAlinden KD, Eapen MS, Sohal SS. Smoking upregulates angiotensin-converting enzyme-2 receptor: a potential adhesion site for novel coronavirus SARS-CoV-2 (Covid-19). *J Clin Med*. 2020;9:841.

78. Leung JM, Yang CX, Tam A, et al. ACE-2 expression in the small airway epithelia of smokers and COPD patients: implications for COVID-19. *Eur Respir J*. 2020;55:2000688.

79. Cai G, Bossé Y, Xiao F, Kheradmand F, Amos CI. Tobacco smoking increases the lung gene expression of ACE2, the receptor of SARS-CoV-2. *Am J Respir Crit Care Med*. 2020;201:1557-1559.

80. Smith JC, Sausville EL, Girish V, et al. Cigarette smoke exposure and inflammatory signaling increase the expression of the SARS-CoV-2 receptor ACE2 in the respiratory tract. *Dev Cell*. 2020;53:514-529. https://doi.org/10.1016/j.devcel.2020.05.012

81. Russo P, Bonassi S, Giacconi, Malavolta M, Tomino C, Maggi F. COVID-19 and Smoking: is nicotine the hidden link? *Eur Respir J*. 2020;55:2001116.

82. Monteil V, Kwon H, Prado P, et al. Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell*. 2020;181:905-913.

83. Imai Y, Kuba K, Rao S, et al. Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature*. 2005;436:112-116.

84. Kuba K, Imai Y, Rao S, et al. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med*. 2005;11:875-879.

85. Farsalinos K, Barbouni A, Niaura R. Systematic review of the prevalence of current smoking among hospitalized COVID-19 patients in China: could nicotine be a therapeutic option? [published online ahead of print May 9, 2020]. *Intern Emerg Med*. https://doi.org/10.1007/s11739-020-02355-7

86. Klein JD, Thomas RK, Sutter EJ. Self-reported smoking in online surveys: prevalence estimate validity and item format effects. *Med Care*. 2007;45:691-695.

87. Flint AJ, Novotny TE. Poverty status and cigarette smoking prevalence and cessation in the United States, 1983-1993: the independent risk of being poor. *Tob Control*. 1997;6:14-18.

88. Sun Y, Li Y, Bao Y, et al. Brief Report: increased addictive internet and substance use behavior during the COVID-19 pandemic in

China. *Am J Addict.* 2020;29:268-270. https://doi.org/10.1111/ajad.13066

89. Sidor A, Rzymski P. Dietary choices and habits during COVID-19 lockdown: experience from Poland. *Nutrients.* 2020;12:E1657.

90. Stanton R, To QG, Khalesi S, et al. Depression, anxiety and stress during COVID-19: associations with changes in physical activity, sleep, tobacco and alcohol use in Australian adults. *Int J Environ Res Public Health.* 2020; 17:4065.

91. Luo M, Guo L, Yu M, Jiang W, Wang H. The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public—a systematic review and meta-analysis. *Psych Res.* 2020; 291:113190.

92. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: a systematic review and meta-analysis. *Brain Behav Immun.* 2020;88:901-907. https://doi.org/10.1016/j.bbi.2020.05.026

93. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev.* 2020;4:CD006611.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

PROTOCOL　　　　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

The applications of machine learning in plastic and reconstructive surgery: protocol of a systematic review

Angelos Mantelakis[1,2*] and Ankur Khajuria[3,4†]

Abstract

Background: Machine learning, a subset of artificial intelligence, is a set of models and methods that can automatically detect patterns in vast amounts of data, extract information and use it to perform various kinds of decision-making under uncertain conditions. This can assist surgeons in clinical decision-making by identifying patient cohorts that will benefit from surgery prior to treatment. The aim of this review is to evaluate the applications of machine learning in plastic and reconstructive surgery.

Methods: A literature review will be undertaken of EMBASE, MEDLINE and CENTRAL (1990 up to September 2019) to identify studies relevant for the review. Studies in which machine learning has been employed in the clinical setting of plastic surgery will be included. Primary outcomes will be the evaluation of the accuracy of machine learning models in predicting a clinical diagnosis and post-surgical outcomes. Secondary outcomes will include a cost analysis of those models. This protocol has been prepared using the Preferred Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) guidelines.

Discussion: This will be the first systematic review in available literature that summarises the published work on the applications of machine learning in plastic surgery. Our findings will provide the basis of future research in developing artificial intelligence interventions in the specialty.

Systematic review registration: PROSPERO CRD42019140924

Keywords: Artificial intelligence, Machine learning, Deep learning, Plastic surgery, Big data

Background

In the era of big data, the plethora of efforts towards gathering and analysing patient data in large scale is rapidly increasing [1]. Amongst others, these efforts try to improve the diagnosis of diseases and the prediction of post-treatment outcomes using large amounts of data from past cases. The analysis of this vast amount of information, however, is beyond the capabilities of traditional statistical methods previously used in academic medicine [2].

Machine learning, a subset of artificial intelligence, is the set of models and methods that can automatically detect patterns in vast amounts of data, extract information and use it to perform various kinds of decision-making under uncertain conditions [3]. These models have the potential of two principally distinct functions: supervised and unsupervised learning (termed "deep learning"). Supervised learning involves the creation and optimisation of statistical models which aim to predict an outcome using information from past cases [2, 4]. In contrast, unsupervised learning aims to identify patterns

---

* Correspondence: aggelosmantelakis@yahoo.com
†Angelos Mantelakis and Ankur Khajuria contributed equally to this work.
1Lewisham and Greenwich NHS Trust, London, UK
2London, UK
Full list of author information is available at the end of the article

in previously seemingly random data and generate novel associations [2, 4, 5].

Healthcare professionals have been quick to adopt these emerging technologies to improve patient outcomes [5]. Examples include machine learning models created to identify clinical diagnoses, which perform to the level of expert clinicians in identifying acute cerebral ischaemia, malignant skin lesions and lung cancer subtypes [6–8]. In the field of surgery, this technology has demonstrated a unique potential in predictive post-operative success and complication rate in procedures such as traumatic brain injury, cervical spine fusion and glioma removal, amongst others [9–11].

This technology has the potential to provide clinically relevant information across many areas of plastic surgery. In burn surgery, machine learning has been used to predict whether complete wound healing will require more or less than 14 days with an accuracy rate of 86% [12]. In the field of microsurgery, authors have been able to predict surgical site infections following free flap reconstruction in head and neck cancer with a sensitivity of 81% and specificity of 61% through using artificial intelligence neural networks [13]. Further, machine learning has also been applied in aesthetic surgery research, where using supervised learning, the authors were able to extract potential beauty-determining facial features to guide pre-operative planning [14].

The aim of this review is to systematically analyse the available literature on the applications of deep learning in plastic surgery. Data collected will be used to provide an up-to-date overview of the potential utility of this technology in the specialty and suggest future directions of further research.

Methods
Aim
This systematic review is intended to evaluate the clinical applications of machine learning models in the field of plastic and reconstructive surgery and to determine areas of future research on this technology.

Protocol and registration
This protocol is registered in the Prospective Register of Ongoing Systematic Reviews (PROSPERO) CRD42019140924 and adheres to the Preferred Reporting Items for Systematic Review guidelines and Meta-Analysis Protocols (PRISMA-P 2015) [15] [Additional File].

Search strategies
All studies published between 1990 and the date of the search will be considered for review.

We will perform a comprehensive search of MEDLINE (OVID SP), EMBASE (OVID SP), Science Citation Index, ClinicalTrials.gov and CENTRAL. A combination

of free text and Medical Subject Headings (MeSH) terms will be used. An example search strategy in MEDLINE is the following:

1	("deep learning" OR "artificial intelligence" OR "machine learning" OR "decision trees" OR "random forests" OR SVM OR "support vector machine")
2	exp "NEURAL NETWORKS (COMPUTER)"/ OR exp "DEEP LEARNING"/
3	exp "ARTIFICIAL INTELLIGENCE"/
4	(1 OR 2 OR 3)
5	(microsurgery OR (surgery AND (plastic OR reconstructive OR esthetic OR aesthetic OR burns OR hand OR craniofacial OR "peripheral nerve")))
6	exp "SURGERY, PLASTIC"/ OR exp "RECONSTRUCTIVE SURGICAL PROCEDURES"/
7	(5 OR 6)
8	(4 AND 7)

Identification and selection of studies
Following database searching, studies will be populated into Endnote X7 library (Clarivate Analytics, USA). There will be two stages of screening, carried by two independent reviewers using pre-specified criteria. The search results, including abstracts, full-text articles and record of reviewer's decisions, including reasons for exclusion, will be recorded.

1. Stage 1: Title and abstract review. This will be carried out by the two independent researchers by adhering to the set eligibility criteria. Any discrepancies will be resolved through a consult by a third reviewer.
2. Stage 2: Studies included will undergo full-text review by the same independent reviewers. Any discrepancies will be resolved through a consult to a third reviewer.

Eligibility criteria
Types of studies
Any primary studies (including case reports), which assess the prediction rate of deep learning models in diagnosis of disease or post-operative outcomes in plastic surgery, either on its own or compared to other techniques, will be included. There will be no geographical restriction. Our exclusion criteria include studies utilising machine learning without clinical data, non-English language articles and review articles.

Types of study participants
We will include clinical data from adult participants (> 18 years old) with conditions requiring plastic or reconstructive surgery. Data from animal studies will be excluded.

Types of interventions

The studies considered will present artificial intelligence models utilising deep learning as an intervention with the aim to provide a diagnosis of a clinical presentation, or a clinical prognosis of a plastic surgery intervention. The intervention may be used by itself or in combination with other methods. Since this technology is new, there is no single best deep learning model, and because of the versatility of conditions treated in plastic surgery, it is expected that various different models will be identified in our review.

Outcomes

Primary outcomes

The primary outcomes will be the evaluation of deep learning models on two distinct functions. The first function is the accuracy of providing a clinical diagnosis. Studies must have a defined clinical condition for which the model is designed to identify. The accuracy of performing this task (either on its own or in assistance with a clinician) will be collected.

The second primary outcome will be the accuracy of prediction of post-operative outcomes and complications of plastic surgery interventions. In order to qualify, studies will need to have created a model to predict a particular clinical outcome (for example, probability of post-operative wound infection), with data collected prospectively or retrospectively.

In both settings, the model's accuracy will be assessed by the reported specificity, sensitivity, positive predictive value and negative predictive value of performing the named task.

Secondary outcomes

The secondary outcomes will include cost analysis of the deep learning models. Further, outcomes of studies that have utilised deep learning models as a treatment for a clinical condition (for example, neuroprosthesis) will also be collected.

Data extraction, collection and management

After the study selection is completed, the two reviewers will independently extract data using a standardised data extraction form. Any disagreements and differences will be resolved by discussion with a third reviewer.

The following data will be extracted:

1. Study characteristics (authors, year of publication, study design)
2. Patient demographics (number of participants, sex, mean age)
3. Indication of application of the software model (prediction of a diagnosis or treatment outcome)
4. Software characteristics

5. Outcomes (specificity, sensitivity, positive predictive value and negative predictive value of forming a diagnosis; predicting rates of overall survival, treatment success, post-operative function, aesthetic outcome, complications and recurrence)
6. Complications or adverse events reported

Risk of bias

The risk of bias in the selected randomised controlled trials will be evaluated by the two independent reviewers through utilising the Cochrane Collaboration Risk of Bias tool [16]. The methodological quality will be assessed based on appropriate participant selection and randomisation, blinding of participants and reviewers, attrition, selective reporting and others. An overall grading of low, medium or high risk of bias will then be allocated. For non-randomised trials, the ROBINS-I (Risk of Bias in Non-randomised Studies-of Interventions) will be utilised [17]. For quantitative studies in which the ROBINS-I is not applicable, risk of bias assessment will be undertaken using the Quality Assessment Tool for Quantitative studies [18]. Case reports will be included as part of screening for all available evidence; however, they are inherently at high risk of bias and this will be considered during the assessment of the quality of overall evidence.

The risk of bias in the performance of deep learning models will be evaluated using the QUADAS-2 (Quality Assessment for Diagnostic Accuracy Studies) tool [19]. This will examine the process of patient selection and the conduction and interpretation of the index test and reference standard. An overall risk of bias will be subsequently allocated (high, low, or unclear).

Data analysis

The two independent reviewers will explore the heterogeneity between the studies using the Review Manager 5.3 provided by the Cochrane Collaboration (1). Potential sources of heterogeneity include the deep learning software, its intention (diagnosis or treatment), the treatment indication and population. A narrative review will be carried out structured around the intervention and outcome of interest. A quantitative analysis (meta-analysis) will be performed if sufficient homogeneous studies in terms of design and outcomes measures are identified.

Statistical heterogeneity will be assessed using the I^2 statistic [20]. A random-effects model will be employed for heterogenous cohorts ($I^2 > 50\%$). The quality of overall evidence will be assessed using The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [21]. Sensitivity analysis will be attempted based on the study quality. This may be repeated after removal of poor-quality studies that may affect the overall effect estimate.

Discussion

Due to the incredible potential of machine learning to process vast amounts of patient information and provide clinically relevant predictions, it is important for plastic surgeons to be informed with the up-to-date applications of this technology in the specialty. The aim of our review is to systematically evaluate the current evidence of this technology in the clinical setting and to discuss the future prospects of machine learning in guiding patient management. To the authors' knowledge, this is the first systematic review to evaluate the applications of artificial intelligence in plastic surgery.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13643-020-01304-x.

> **Additional file 1.**

Abbreviations
CENTRAL: Cochrane Central Register of Controlled Trials; EMBASE: Excerpta Medica Database; GRADE: Grading of Recommendations Assessment, Development and Evaluation; PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols; QUADAS-2: Quality Assessment for Diagnostic Accuracy Studies; ROBINS-I: Risk of Bias in Non-randomised Studies-of Interventions

Authors' contributions
Both authors contributed equally to the conception of the protocol and study design, reviewed this report and approved the final manuscript.

Availability of data and materials
The datasets generated and/or analysed during the current study are available in the MEDLINE (OVID SP), EMBASE (OVID SP), Science Citation Index, ClinicalTrials.gov and CENTRAL repositories.
https://ovidsp.ovid.com/
http://mjl.clarivate.com/cgi-bin/jrnlst/jloptions.cgi?PC=K
https://clinicaltrials.gov/
https://www.cochranelibrary.com/central/about-central

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Lewisham and Greenwich NHS Trust, London, UK. [2]London, UK. [3]Kellogg College, University of Oxford, Oxford, UK. [4]Department of Surgery and Cancer, Imperial College London, London, UK.

References
1. Lee CH, Yoon HJ. Medical big data: promise and challenges. Kidney Res Clin Pract. 2017;36(1):3.
2. Kanevsky J, Corban J, Gaster R, Kanevsky A, Lin S, Gilardino M. Big data and machine learning in plastic surgery: a new frontier in surgical innovation. Plastic Reconstr Surg. 2016;137(5):890e–7e.
3. Murphy KP. Machine learning: a probabilistic perspective. Cambridge: MIT press; 2012.
4. Celtikci E. A systematic review on machine learning in neurosurgery: the future of decision-making in patient care. Turk Neurosurg. 2018 Jan 1;28(2):167–73.
5. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of healthcare. Am J Med. 2019;31.
6. Abedi V, Goyal N, Tsivgoulis G, et al. Novel screening tool for stroke using artificial neural network. J Stroke. 2017;48:1678–81.
7. Cruz-Roa AA, Arevalo Ovalle JE, Madabhushi A, González Osorio FA. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. Med Image Comput Comput Interv. 2013;16:403–10.
8. Lehman CD, Yala A, Schuster T, et al. Mammographic breast density assessment using deep learning: clinical implementation. Radiology. 2019;290:52–8.
9. Shi HY, Hwang SL, Lee KT, Lin CL. In-hospital mortality after traumatic brain injury surgery: a nationwide population-based comparison of mortality predictors used in artificial neural network and logistic regression models. J Neurosurg. 2013;118:746–52.
10. Arvind V, Kim JS, Oermann EK, Kaji D, Cho SK. Predicting surgical complications in adult patients undergoing anterior cervical discectomy and fusion using machine learning. Neurospine. 2018;15(4):329.
11. Macyszyn L, Akbari H, Pisapia JM, Da X, Attiah M, Pigrish V, Pal S, Davuluri RV, Roccograndi L, Dahmane N, Martinez-Lage M, Biros G, Wolf RL, Bilello M, O'Rourke DM, Davatzikos C. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. Neuro Oncol. 2016;18:417–25.
12. Yeong EK, Hsiao TC, Chiang HK, Lin CW. Prediction of burn healing time using artificial neural networks and reflectance spectrometer. Burns. 2005;31:415–20.
13. Kuo PJ, Wu SC, Chien PC, Chang SS, Rau CS, Tai HL, Peng SH, Lin YC, Chen YC, Hsieh HY, Hsieh CH. Artificial neural network approach to predict surgical site infection after free-flap reconstruction in patients receiving surgery for head and neck cancer. Oncotarget. 2018;9(17):13768.
14. Gunes H, Piccardi M. Assessing facial beauty through proportion analysis by image processing and supervised learning. Int J Human Comput Stud. 2006;64(12):1184–99.
15. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4:1.
16. Higgins Julian P T, Altman Douglas G, Gøtzsche Peter C, Jüni Peter, Moher David, Oxman Andrew D, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials BMJ. 2011;343:d5928.
17. Sterne JA, Hernan MA, Reeves BC, Savovic 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.
18. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract. 2012;18(1):12–8.
19. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3(1):25.
20. Higgins JP, Thompson S, Deeks J, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557–60.
21. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.