Cost-effectiveness of a national quality improvement intervention to improve survival after emergency abdominal surgery

Study design: Health Economic Evaluation

Running title: cost-effectiveness of quality improvement for emergency surgery

Fan Yang[1], Simon Walker[1], Gerry Richardson[1], Tim Stephens[2], Mandeep Phull[3], Ann Thompson[4], and Rupert M. Pearse[2], on behalf of the Enhanced Peri-Operative Care for High-risk patients (EPOCH) trial group*.

[1]Centre for Health Economics, University of York, UK
[2]Barts & The London School of Medicine and Dentistry, Queen Mary University of London, UK
[3]Barking, Havering and Redbridge University Hospitals, NHS Trust, UK
[4]Pragmatic Clinical Trials Unit, Queen Mary University of London, UK

*The full list of EPOCH trial members is available in the Appendix.

Correspondence to:
Dr Fan Yang
Centre for Health Economics, University of York
Alcuin A Block, Heslington, York, UK
YO10 5DD
Telephone: +44 (0)1904 321491
ORCID: 0000-0003-4689-265X
Email: fan.bella.yang@york.ac.uk

Conflict of Interest: RP holds research grants and has given lectures and/or performed consultancy work for BBraun, Intersurgical, Glaxo Smithkline and Edwards Lifesciences, and is a member of the editorial board of the British Journal of Surgery. All other authors declare no conflict of interest.

Cost-effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery

Abstract

Background: Patients undergoing emergency abdominal surgery are exposed to high risk of death. A quality improvement (QI) programme to improve the survival for these patients was evaluated in the Enhanced Peri-Operative Care for High-risk patients (EPOCH) trial. This study aims to assess its cost-effectiveness versus usual care from a UK health service perspective.

Methods: Data collected in a subsample of trial participants were employed to estimate costs and quality-adjusted life years (QALYs) for the QI programme and usual care within the 180-day trial period, with results also extrapolated to estimate lifetime costs and QALYs. Cost-effectiveness was estimated using incremental cost-effectiveness ratios (ICERs). The probability of being cost-effective was determined for different cost-effectiveness thresholds (£13,000 to £30,000 per QALY). Analyses were performed for lower-risk and higher-risk subgroups based on the number of surgical indications (single vs multiple).

Results: Within the trial period, QI was more costly (£467) but less effective (-0.002 QALYs). Over a lifetime, it was more costly (£1395) and more effective (0.018 QALYs), but did not appear to be cost-effective (ICER: £77,792 per QALY, higher than all cost-effectiveness thresholds; probability of being cost-effective: 28.7% to 43.8% across the thresholds). For lower-risk patients, QI was more costly and less effective both within trial period and over a lifetime and it did not appear to be cost-effective. For higher-risk patients, it was more costly and more effective, and did not appear cost-effective within the trial period (ICER: £158,253 per QALY) but may be cost-effective over a lifetime (ICER: £14,293 per QALY).

Conclusion: The QI programme does not appear cost-effective at standard cost-effectiveness thresholds. For patients with multiple surgical indications, this programme is potentially cost-effective over a lifetime, but this is highly uncertain.

Key words: quality improvement, emergency abdominal surgery, cost effectiveness

1. Introduction

Patients undergoing emergency abdominal surgery are exposed to a much greater risk of death than the general surgical patient population (1, 2). In the National Health Service (NHS), there are around 30,000 patients undergoing this surgery each year, with one in ten dying within 30 days of having the operation, rising to 25% at 90 days (3). Therefore, there is an urgent need to consider interventions that might improve survival for these patients.

An evidence-based care pathway was proposed by the Royal College of Surgeons of England to improve the quality of care for these patients, representing a best practice standard of peri-operative care deliverable in all NHS hospitals (4). To implement this care pathway, the evidence-based quality improvement (QI) approach was used to change the current practice and culture of care for the patient group. QI initiatives have been found to be associated with improved survival in surgical patients (5-8). The QI programme included quality improvement training and support for the local QI leads, nominated by each participating hospital, to develop and implement action plans tailored for each hospital's needs to make the required improvements in patient care (9). The effectiveness of the QI programme was evaluated in the Enhanced Peri-Operative Care for High-risk patients (EPOCH) trial (9), but no survival benefits was observed from this programme at either 90 or 180 days after surgery.

It was expected that patients' health-related quality of life would be improved with the implementation of this QI programme and that it may reduce healthcare resource use in the long term, compared to usual care. Therefore, this study aimed to assess the cost-effectiveness of the QI programme compared to usual care for patients undergoing emergency laparotomy to provide evidence about whether this programme should be implemented widely in the NHS.

2. Methods

In this study, costs were estimated from the perspective of the NHS and health outcomes measured in quality adjusted life years (QALYs). The analysis was based on data from the EPOCH trial. A convenient subsample of trial participants were interviewed to collect the full range of data for the economic analysis and the cost-effectiveness was measured both within the 180-day trial period and extrapolated over the patients' lifetime. We also explored the cost-effectiveness in clinically relevant patient subgroups based on their risk of mortality, defined by the number of surgical indications (single vs multiple). A further analysis

predicted the within trial and lifetime results in all trial participants. Cost-effectiveness was estimated using incremental cost-effectiveness ratios (representing the incremental costs per additional QALY of one strategy compared to the other), where appropriate, and incremental net health benefit (the difference between the health generated with the intervention relative to its comparator and the health which would be generated elsewhere in the health care system if the required resources were used for other purposes) based on cost-effectiveness thresholds of £13,000 (10), £20,000 and £30,000 per QALY (11). Costs and QALYs beyond one year were discounted at 3.5% per annum in line with UK recommendations (11). The work has been reported in line with the CHEERS criteria (12).

2.1. Study design and participants

The EPOCH was a multi-centre, stepped wedge randomised cluster trial conducted in 93 NHS hospitals across the UK over an 85-week period (13). Eligible hospitals were grouped into 15 geographical clusters which commenced the QI programme in random order over the 85-week trial period from 3rd March 2014 to 19th October 2015. Eligible patients were those aged 40 years and over undergoing non-elective open abdominal surgery. Patients were excluded if they were undergoing a simple appendectomy, surgery related to organ transplant, gynaecological surgery, laparotomy for traumatic injury, treatment of complications of recent elective surgery or if they had previously been included in the EPOCH trial. A total of 15,856 patients were included, referred to as "EPOCH population", with outcome data on mortality within 90 and 180 days following surgery, duration of inpatient stay after surgery and whether re-admitted to hospital (Yes/No) within 180 days of surgery extracted from the Office for National Statistics (ONS), Hospital Episodes Statistics (HES) and the Information Services Division of NHS Scotland databases. Of the 15,856 patients, a total of 680 patients were included in a subsample, referred to as "EPOCH subsample", with additional data for the economics analysis including patients' health-related quality of life (HRQoL), and healthcare resource use not captured by HES, e.g., outpatient visits, primary care consultations. The EPOCH subsample were selected based on a convenience sampling approach and the patients were from eight hospitals, which were amongst those which commenced the QI programme midway through the trial period. Before data analysis, the characteristics of patients in the EPOCH subsample were compared to those in the EPOCH population to check whether they were representative of the EPOCH population. The trial was approved by the East Midlands (Nottingham 1) Research Ethics Committee Research Ethics Committee (13/EM/0415).

2.2. Intervention

The QI programme has been described in detail elsewhere (9). Briefly, it was developed through an evidence-based Delphi consensus process to update existing guidelines published by the Royal College of Surgeons of England with the aim of changing the practice and culture of care for patients undergoing emergency abdominal surgery. The QI programme required an extensive care pathway with 37 components to be implemented (9) (a full summary of evidence grading is available on the trial website www.epochtrial.org). Nominated QI leads in each participating hospital were tasked with leading a hospital wide quality improvement programme to implement the care pathway with the support and guidance of the national EPOCH QI team. Before or during the first week of intervention activation, the QI team provided training for each geographical cluster to develop the knowledge, skills and attitudes that the QI leads required to achieve change. During the trial period, the EPOCH QI team provided advice and support by phone and email and there were two national meetings to facilitate shared learning. As a results of the stepped-wedge trial design, the duration of the QI programme period varied between clusters from 5 to 85 weeks.

2.3. Usual care

Usual care was defined as the patients undergoing emergency abdominal surgery in a participating hospital before the implementation of the QI programme.

2.4. Resource use and costs

Inpatient hospital length of stay after initial surgery was available for all patients in the EPOCH population from HES. Whether patients were readmitted to hospital within 180 days after surgery was extracted from HES for all trial patients, but there was no information for the entire EPOCH population on number of hospital readmissions or length of stay of readmissions. Therefore, resource use for readmissions was estimated by the data collected from the EPOCH subsample. Other resource use including outpatient appointments, accident and emergency (A&E) attendances, and primary care consultations was also collected at 180 days after surgery, from the EPOCH subsample. Staff resources associated with the QI programme through the trial period were recorded by the EPOCH QI team. Costs of the QI programme were estimated including salary costs for the QI leads and QI coordinator and non-salary costs of the QI programme (e.g., travel, meetings and training, etc.) and then an average cost per patient in the QI programme group was estimated. Total healthcare costs over the trial period were estimated by multiplying the amount of each resource used by

appropriate unit costs (14, 15). Costs are expressed in UK Sterling (GBP) at 2016/2017 prices.

2.5. Outcomes

Health outcomes were measured in QALYs, a generic measure of health which captures quality adjusted survival. QALYs were estimated based on patients' responses to the 3-level EQ-5D instrument (EQ-5D-3L) collected before surgery and at 90 and 180 days after surgery for the EPOCH subsample. The EQ-5D-3L asks patients to rate the severity of their problems (no problems, moderate problems or severe problems) in 5 domains (mobility, self-care, usual activity, pain/discomfort and anxiety/depression). Responses to EQ-5D-3L were transformed into health related quality of life (HRQoL) scores using the preference based UK tariff (16), on a scale where 0 represents death and 1 represents full health. QALYs were estimated using the HRQoL scores and survival based on the area under the curve method and linear interpolation between time points (17).

2.6. Analysis

2.6.1. Within-trial

Costs and QALYs over the trial period were calculated for patients in the EPOCH subsample. Where costs and EQ-5D-3L data were missing, multiple imputation using chained equations and predictive mean matching was used (18). The imputed data were then used to calculate per patient costs and QALYs for patients. Mean difference in healthcare costs incurred and QALYs accrued between trial arms were estimated using seemingly unrelated regression model (19), adjusted for baseline EQ-5D-3L score (20) (in QALYs regression), age and sex.

2.6.2. Lifetime extrapolation

Lifetime cost-effectiveness was estimated by extrapolating QALYs and costs over the remaining lifetime of patients in the EPOCH subsample based on HRQoL at the end of the trial period, general population survival data and general population health care costs by age and sex. Patients alive at the end of the trial period were assumed to return from their HRQoL score at day 180 to the population average for someone of their age and sex over the next 3 years based on linear interpolation and then have the population age- and sex-adjusted EQ-5D-3L scores until death (21). Patients whose life expectancy was less than 3 years were assumed to have the HRQoL at day 180 until death. Life expectancy was estimated to match the population average based on their age and sex, reported in the ONS National Life Tables

of UK 2012-2014 (22). We assumed that patients alive were attributed average healthcare costs by age and sex over the rest of their modelled lives (23). Lifetime costs and QALYs were calculated for all patients in the EPOCH subsample. Similar to the within-trial analysis , the mean difference in healthcare costs and QALYs between the QI programme and usual care was estimated using seemingly unrelated regression model (19), controlling for baseline EQ-5D-3L score (in QALYs regression), age and sex.

2.6.3. Cost-effectiveness analysis

The additional costs per QALY gained of QI programme compared to usual care, i.e., incremental cost-effectiveness ratio (ICER), was calculated where appropriate (where one intervention was more effective and more costly than another). The incremental net health benefit (NHB) measured in QALYs at cost-effectiveness thresholds of £13,000, £20,000 and £30,000 per QALY was also calculated. The NHB is the gain in health to the patient less the health lost as a result of any additional costs of the intervention which result in resources not being available for other patients' care (the health opportunity costs) (24). When the incremental NHB is positive, the intervention is cost-effective. Probabilistic sensitivity analysis was used to estimate decision uncertainty; that is, the probability that the joint uncertainty in costs and QALYs, estimated based on the seemingly unrelated regression, would result in QI not being cost-effective at a given cost-effectiveness threshold. These probabilities were presented visually on a cost-effectiveness acceptability curve (CEAC) (25, 26).

Two scenario analyses were performed for the EPOCH subsample to assess the robustness of the findings to alternative assumptions regarding HRQoL beyond the trial period. Patients alive at the end of trial were assumed to follow the group mean HRQoL until death (in scenario one) or retain their individual EQ-5D-3L scores at day 180 until death (in scenario two).

Subgroup analysis was established to explore whether cost-effectiveness varies between lower-risk versus higher-risk patients defined by the number of surgical indications. Patients with single indication for surgery were considered as lower-risk and those with multiple indications as higher-risk.

2.6.4. EPOCH population extrapolation

In the EPOCH trial, mortality and inpatient stay data were available for all the trial participants. We used these data to explore the cost-effectiveness of the QI programme in the

whole EPOCH population, combining these data with HRQoL, costs of readmissions, and costs of outpatient, A&E and primary care predicted based on the EPOCH subsample. Generalised linear models were used to predict costs (including age, sex and cluster as covariates) and baseline HRQoL (including age, sex and indication for surgery as covariates) and a generalised least square random effects model was used to predict HRQoL gain after surgery, controlling for baseline EQ-5D-3L score, age and sex. We then used the predicted HRQoL scores and the mortality data to calculate QALYs over the trial period using area under curve and linear interpolation method (17). The same approach employed to extrapolate lifetime costs and QALYs in the EPOCH subsample was also used in the EPOCH population. These regression models predicted the conditional means in costs and HRQoL scores for those not in the EPOCH subsample, and the potential variability in patient estimates was not captured and therefore only the mean values of difference in costs and QALYs are reported and ICER and NHB were calculated to assess the cost-effectiveness. All statistical analyses were performed using Stata 14 (StataCorp, College Station, Texas, USA).

3. Results

3.1. Patient characteristics

Baseline characteristics of patients in the EPOCH population are summarised in **Table S1**. A total of 15,856 eligible patients were included (8,482 in the usual care group and 7,374 in the QI group) and there were no marked differences in baseline characteristics between patients in the QI group and those in the usual care group. Of the 680 eligible patients in the EPOCH subsample, baseline characteristics did not markedly differ between the usual care group (n=415) and the QI group (n=265), with the exception of higher baseline HRQoL scores (0.262 vs 0.168) and slightly higher proportion of multiple surgical indications (higher-risk) patients (25.7% vs 21.5%) in the QI group (**Table S1**). The characteristics of patients in the EPOCH subsample and those in the EPOCH population were broadly similar, although patients in the subsample were slightly younger on average and a lower proportion were females. The 180-day all-cause mortality and the characteristics of patients alive at the end of the trial period are presented in **Table S2**. A higher proportion of patients died in the usual care group compared to QI group in both EPOCH subsample and EPOCH population. The mortality was lower in the EPOCH subsample compared to that in the EPOCH population.

3.2. Resource use and costs

The healthcare resources used over the 180-day follow-up period for the EPOCH subsample and the unit costs associated with each type of resource use are shown in **Table 1**. Resource use was highly variable between individuals, as evidenced by the large standard deviations, but was similar on average between patients in the QI programme and usual care. The total costs of the QI programme across the 93 trial hospitals with 7,383 patients in the QI group (9 patients excluded from the analysis as outcome data could not be calculated) were estimated to be £231,400, resulting in a cost per patient of £32 of the programme.

The mean costs of resource use in the EPOCH subsample based on the imputed data are shown in **Table 2**. Costs of inpatient stay after surgery is the principal costs driver, accounting for around 80% of the mean total costs. Overall, mean 180-day costs for a patient in the usual care group was £8216 while the costs for a patient in the QI group was higher, at £8675 (mean difference: 467, 95% CI: -800 to 1735).

3.3. Outcomes

The HRQoL scores at each time point and 180-day QALYs for the EPOCH subsample are summarised in **Table 3** (based on the imputed data). At baseline, the QI patients had higher HRQoL scores than those in usual care, but the follow-up HRQoL scores were similar between two groups. The mean QALYs over the 180 days after surgery were 0.274 (95% CI 0.260 to 0.288) for the usual care group and 0.287 (95% CI 0.269 to 0.306) for the QI programme group; patients in the intervention group had slightly higher QALYs than those in the usual care group (mean difference: 0.013, 95% CI: -0.009 to 0.036), however, after adjusting for baseline HRQoL, age and sex, the QI programme generated fewer QALYs compared to usual care, although the differences were very small (mean difference: -0.002, 95% CI: -0.022 to 0.017).

3.4. Cost-effectiveness analysis

3.4.1. Within-trial analysis

Cost-effectiveness results are presented in **Table 4**. For the EPOCH subsample, within the trial period, the QI programme was associated with incremental costs of £467 but fewer QALYs by -0.002, i.e., the QI programme was less effective and more costly and so is dominated by usual care. At the cost-effectiveness thresholds of £13,000, £20,000 and £30,000 per QALY, the incremental NHB of the QI programme (and the probability of being cost-effective) was -0.038 QALYs (23.5%), -0.025 QALYs (23.7%) and -0.018 QALYs

(24.2%) respectively (**Table 4**). The probabilities of QI being cost-effective across the cost-effectiveness threshold range are shown in **Figure 1a**.

3.4.2. Lifetime extrapolation

For the EPOCH subsample, when the lifetime perspective was adopted, the QI programme was associated with incremental costs (mean difference: £1395, 95% CI: -1083 to 3873) and more QALYs (mean difference: 0.018, 95% CI: -0.396 to 0.432), yielding an ICER of £77,792 per QALY (**Table 4**), which is higher than the cost-effectiveness thresholds considered. At the cost-effectiveness thresholds of £13,000, £20,000 and £30,000 per QALY, the incremental NHB of the QI programme (and the probability of being cost-effective) was -0.089 QALYs (28.7%), -0.052 QALYs (38.2%) and -0.029 QALYs (43.8%) respectively (**Table 4**). The cost-effectiveness acceptability curve is shown in **Figure 1b**. The QI programme did not appear to be cost-effective for patients in the subsample over the longer term. The estimated lifetime costs and lifetime QALYs are summarised in **Table S3**.

3.4.3. Scenario analysis

Table 4 also presents the results of scenario analyses for the EPOCH subsample. Using different assumptions of HRQoL beyond trial period, the QI programme was associated with more costs but fewer QALYs than usual care and was therefore dominated by usual care.

3.4.4. Subgroup analysis

For lower-risk patients (single surgical indication), the QI programme was associated with increased costs and fewer QALYs both within the trial period and over the lifetime horizon and all incremental NHBs were negative (**Table 4**). For higher-risk patients (multiple surgical indications), within the trial period the QI programme was associated with incremental costs and more QALYs, resulting in ICERs of £158,253 per QALY (**Table 4**), higher than the cost-effectiveness thresholds considered. The incremental NHB was negative for all three thresholds considered. However, over the lifetime horizon, the QI programme was associated with incremental costs and more QALYs, yielding an ICER of £14,293 per QALY (**Table 4**). At the cost-effectiveness thresholds of £13,000, £20,000 and £30,000 per QALY, the incremental NHB (and the probability of being cost-effective) was -0.031 QALYs (45.6%), 0.089 QALYs (61.5%) and 0.163 QALYs (69.3%). Therefore, the QI programme may be cost-effective for patients with multiple indications for surgery over the longer term at the thresholds of £20,000 and £30,000 per QALY, but this is subject to substantial decision uncertainty.

3.4.5. EPOCH population analysis

The results of the analyses for the EPOCH population are also shown in **Table 4**. Both within the trial period and over the lifetime horizon, the QI programme was associated with incremental costs and fewer QALYs. The incremental NHB was negative for all three thresholds considered. The QI programme was therefore dominated by usual care in all trial patients. Details of the estimated costs and regression models are available in **Table S4-S7**.

4. Discussion

The analysis shows that, for patients undergoing emergency laparotomy, a QI programme aiming to improve the quality of care for them does not appear cost-effective within the 180 days after the surgery or over their remaining lifetime, when compared with usual care. For patients with multiple indications for surgery, it may be cost-effective over the lifetime, but this is highly uncertain.

Our results are consistent with the primary clinical analysis showing no survival benefit from the programme within the trial period (9). The lack of the effectiveness of the QI programme observed in the EPOCH trial may be partially explained by the stepped wedged cluster design of the study resulting in limited time for the QI efforts to affect change (between 5 and 85 weeks in total). More importantly, this may be as a result of the QI programme not being implemented as successfully as expected in the EPOCH trial. Recent studies of quality improvement programmes in other clinical areas (5-8, 27-31) suggest that more focused, discrete clinical interventions may be more successfully implemented than interventions that include larger numbers of care processes and clinician behaviours. The QI programme in the EPOCH trial required an extensive care pathway with 37 components to be implemented (9) and there were wide variations in these elements at individual hospitals and thus local adaptations in each hospital were needed. According to the ethnographic evaluation of this QI programme (32), hospital staff often had little or no additional time to improve patient care and consequently, the objective of the EPOCH QI programme was not readily achieved because the care pathway was not implemented as fully as intended (33). As such the lack of cost-effectiveness identified here may be due, to a large extent, on implementation failure resulting in a loss of intervention effectiveness. It remains unknown whether the EPOCH trial interventions could be effective, and therefore potentially cost-effective, if implemented more fully for example in circumstances where more time and support was allocated to implementation.

When a lifetime horizon was adopted, in the EPOCH subsample this programme was associated with slightly more QALYs, which may be explained by that more females patients survived in the QI group at the end of the trial period than those in usual care (QI: 49.6% and usual care: 48.7%, **Table S2**). But the resultant ICER was much higher than the cost-effectiveness thresholds considered and it was therefore dominated by usual care. Using other assumptions of HRQoL beyond the trial period, QI was associated with fewer QALYs, suggesting that the lifetime estimates were highly sensitive to the long-term assumptions. Nevertheless, the use of different assumptions beyond the trial period did not affect the conclusion that the QI programme was unlikely to be cost-effective in this subsample. In the EPOCH population, the QI programme was associated with fewer QALYs both within the trial period and over lifetime horizon and was dominated by usual care.

Subgroup analyses demonstrated QALYs benefits from the intervention for higher-risk patients both within trial period and over lifetime horizon and it was likely to be cost-effective over the long-term horizon at the cost-effectiveness threshold of £20,000 to £30,000 per QALY. These patients were with multiple indications for surgery and more likely to be sicker; there appeared to be benefits of the QI programme in a sicker population. However, the probability of being cost-effective was lower than 70% at the cost-effectiveness thresholds considered (**Table 4**) and the sample size was small (usual care: 89, QI: 68) (**Table S1**), and therefore these results should be seen as indicative.

The finding of this study potentially have important implications for policy makers. Although the national QI programme for patients undergoing emergency abdominal surgery did not show effectiveness or cost-effectiveness for all patients, in higher-risk patients, it was associated with health benefits at an acceptable cost and therefore the QI programme may have the potential for widespread implementation for sicker patients. Considering the high uncertainty with the results and the small sample size, generating further evidence assessing its cost-effectiveness may be worthwhile.

It should be noted that the costs of the QI programme were very low, at £32 per patient, and because of a lack of time to implement changes to improve care, this may not accurately reflect the costs if it was fully implemented. If the intervention was delivered as intended, the amount of staff time should have increased markedly and the intervention costs should be higher, which may lead to different cost-effectiveness results (impacting not only the costs,

but also potentially the effectiveness). Another important consideration is that the costs attributed to the intervention were only split across all patients in the intervention arm in this trial. If the QI programme could benefit new patients beyond the trial period, the average costs per intervention recipient would fall, which would also affect the cost-effectiveness.

There are several limitations in this study. First, there were large differences in baseline EQ-5D-3L scores between the two groups. This may be due to the unblinded cluster RCT design of the EPOCH trial with individual recruitment would be susceptible to selection bias (34). Although we have adjusted for it, this would still be a potential limitation. Second, we only captured the costs of the training, time spent implementing the QI programme, but not the potential costs associated with increased effort to comply with the programme, e.g., consultant time in the operating theatre. If the programme was implemented as successfully as expected, these costs would be increased and affect the cost-effectiveness of the programme. Last, the mortality in the EPOCH subsample was slightly lower than that in the EPOCH population in both arms (**Table S2**), which might indicate that the patient care (both usual care and the QI programme) was better in these eight hospitals than other hospitals, so the approach to extrapolating HRQoL and costs from the EPOCH subsample to the whole EPOCH population may still be an concern. In our analysis, data on survival and the main costs component, i.e., inpatient stay, were available for the EPOCH population and used to explore the cost-effectiveness to minimise the estimation bias. Results show that the QI programme was associated with fewer QALYs and the conclusion that it did not appear to be cost-effective was not affected.

5. Conclusion

Using data from a large number of patients enrolled by many hospitals with an efficient trial design, the national quality improvement programme to implement an enhanced pathway of care for patients undergoing emergency abdominal surgery is unlikely to be cost-effective. For a sicker patient subgroup with multiple indications for surgery, the intervention may be cost-effective over the lifetime, but this is highly uncertain.

Funding Statement: This was an investigator initiated study funded by the National Institute for Health Research (UK) Health Services & Delivery Research programme.

Data availability: Due to information governance restrictions imposed by organisations governing data access, we are unable to share the trial data unless applicants secure the relevant permissions. All trial materials are freely available on the trial website (www.epochtrial.org).

References

1. Findley GP, Outcome NCEiP, Death, Outcome NCEiP, Staff D. Knowing the Risk: A Review of the Peri-Operative Care of Surgical Patients: Summary: National Confidential Enquiry into Perioperative Deaths; 2011.
2. Pearse RM, Harrison DA, James P, Watson D, Hinds C, Rhodes A, et al. Identification and characterisation of the high-risk surgical population in the United Kingdom. Crit Care. 2006;10(3):R81.
3. Saunders DI, Murray D, Pichel AC, Varley S, Peden CJ, Network UKEL. Variations in mortality after emergency laparotomy: the first report of the UK Emergency Laparotomy Network. Br J Anaesth. 2012;109(3):368-75.
4. Anderson I, Eddlestone J, Lees N. The Higher Risk General Surgical Patient - Towards Improved Care for a Forgotten Group. London: Royal College of Surgeons of England; 2011.
5. Huddart S, Peden CJ, Swart M, McCormick B, Dickinson M, Mohammed MA, et al. Use of a pathway quality improvement care bundle to reduce mortality after emergency laparotomy. Br J Surg. 2015;102(1):57-66.
6. Vester-Andersen M, Lundstrom LH, Moller MH, Waldau T, Rosenberg J, Moller AM, et al. Mortality and postoperative care pathways after emergency gastrointestinal surgery in 2904 patients: a population-based cohort study. British journal of anaesthesia. 2014;112(5):860-70.
7. Moller MH, Adamsen S, Thomsen RW, Moller AM, Peptic Ulcer Perforation trial g. Multicentre trial of a perioperative protocol to reduce mortality in patients with peptic ulcer perforation. Br J Surg. 2011;98(6):802-10.
8. Tengberg LT, Bay-Nielsen M, Bisgaard T, Cihoric M, Lauritsen ML, Foss NB, et al. Multidisciplinary perioperative protocol in patients undergoing acute high-risk abdominal surgery. Br J Surg. 2017;104(4):463-71.
9. Peden CJ, Stephens T, Martin G, Kahan BC, Thomson A, Rivett K, et al. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery (EPOCH): a stepped-wedge cluster-randomised trial. Lancet. 2019;393(10187):2213-21.
10. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess. 2015;19(14):1-503, v-vi.
11. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. NICE; 2013.
12. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value Health. 2013;16(2):231-50.
13. Protocol 13PRT/7655:Enhanced PeriOperative Care for High-risk patients (EPOCH) Trial (ISRCTN80682973) [Internet]. The Lancet. 2013. Available from: https://www.thelancet.com/protocol-reviews/13PRT-7655.
14. National Health Service (NHS). Reference Costs 2015-16. 2016.
15. Personal Social Services Research Unit (PSSRU). Unit Costs of Health and Social Care 2016. 2016.
16. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095-108.
17. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. Bmj. 1990;300(6719):230-5.

18. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics. 2014;32(12):1157-70.
19. Gomes M, Ng ES, Grieve R, Nixon R, Carpenter J, Thompson SG. Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. Med Decis Making. 2012;32(2):350-61.
20. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ. 2005;14(5):487-96.
21. Janssen B, Szende A. Population Norms for the EQ-5D. In: Szende A, Janssen B, Cabases J, editors. Self-Reported Population Health: An International Perspective based on EQ-5D. Dordrecht: Springer Netherlands; 2014. p. 19-30.
22. ONS. National Life Tables, United Kingdom: 2012–2014 2015 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2015-09-23.
23. Asaria M. Health care costs in the English NHS: reference tables for average annual NHS spend by age, sex and deprivation group. In: Centre for Health Economics UoY, editor. CHE Research Paper2017.
24. Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation: OUP Oxford; 2006.
25. van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ. 1994;3(5):309-19.
26. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. Health Econ. 2004;13(5):405-15.
27. Bion J, Richardson A, Hibbert P, Beer J, Abrusci T, McCutcheon M, et al. 'Matching Michigan': a 2-year stepped interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. BMJ Qual Saf. 2013;22(2):110-23.
28. Dixon-Woods M, Leslie M, Tarrant C, Bion J. Explaining Matching Michigan: an ethnographic study of a patient safety program. Implement Sci. 2013;8:70.
29. Pannick S, Athanasiou T, Long SJ, Beveridge I, Sevdalis N. Translating staff experience into organisational improvement: the HEADS-UP stepped wedge, cluster controlled, non-randomised trial. BMJ Open. 2017;7(7):e014333.
30. Williams L, Daggett V, Slaven JE, Yu Z, Sager D, Myers J, et al. A cluster-randomised quality improvement study to improve two inpatient stroke quality indicators. BMJ Qual Saf. 2016;25(4):257-64.
31. Presseau J, Mackintosh J, Hawthorne G, Francis JJ, Johnston M, Grimshaw JM, et al. Cluster randomised controlled trial of a theory-based multiple behaviour change intervention aimed at healthcare professionals to improve their management of type 2 diabetes in primary care. Implement Sci. 2018;13(1):65.
32. Stephens TJ, Peden CJ, Pearse RM, Shaw SE, Abbott TEF, Jones EL, et al. Improving care at scale: process evaluation of a multi-component quality improvement intervention to reduce mortality after emergency abdominal surgery (EPOCH trial). Implement Sci. 2018;13(1):142.
33. Stephens T, Peden C, Haines R, Grocott M, Murray D, Johnston C, et al. Hospital level evaluation of the effect of a national quality improvement programme: Time-series analysis of registry data. BMJ Q&S. 2019.
34. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ. 2015;350:h391.

Table 1. Unit costs and resources used by the subsample over the trial period

Item	Unit	Unit cost (£)*	**Usual care (n=415)**	**QI (n=265)**
			Mean units (SD)	Mean units (SD)
Inpatient care				
Inpatient stay	Day	359	18.1 (20.5)	19.5 (19.5)
Hospital readmission	Event	3434	0.27 (1.11)	0.26 (0.79)
Outpatient and primary care				
Outpatient appointment	Visit	130	4.14 (6.19)	3.49 (4.78)
A&E visit	Visit	148	0.41 (1.54)	0.47 (1.78)
GP surgery visit (GP)	Visit	38	2.47 (4.01)	2.35 (3.27)
GP surgery visit (nurse)	Visit	21	1.90 (4.67)	1.41 (2.88)
GP home visit	Visit	120	0	0.01 (0.10)
Stoma nurse visit	Visit	55	0.33 (1.38)	0.45 (1.59)
Occupational therapy	Visit	45	0.19 (0.73)	0.37 (1.97)
Physiotherapy	Visit	49	0.38 (2.41)	1.10 (6.73)
Psychotherapy	Visit	194	0.14 (1.47)	0.12 (0.87)
Dietetics	Visit	71	0.04 (0.52)	0.12 (1.65)
QI programme	-	32	0	1

*Rounded up to nearest pound sterling.
SD: standard deviation

Table 2. Costs of resources used by EPOCH subsample over the trial period, costs of intervention and total costs*

Costs (£)	**Usual care (n=415)**				**QI (n=265)**			
	Missing	Mean (SE)	95% CI	% total costs	Missing	Mean (SE)	95% CI	% total costs
Inpatient stay	1 (<1%)	6486 (360)	5778-7194	78.9	0	7001 (431)	6152-7850	80.7
Readmissions	17 (4.10%)	934 (189)	561-1305	11.4	11 (4.15%)	881 (170)	547-1215	10.2
Outpatient, A&E and primary care	97 (23.4%)	797 (58)	682-912	9.7	70 (26.4%)	760 (60)	642-878	8.8
QI programme	-	0			-	32		0.37
Total		8216 (411)	7409-9024	100		8675 (497)	7696-9684	100
ΔCosts (£)[a]		-	-			**458**	**(-812 to 1728)**	
ΔCosts (£)[b]		-	-			**467**	**(-800 to 1735)**	

*Based on the imputed data and rounded up to nearest pound sterling.
[a] Unadjusted difference
[b] Adjusted for age and sex

Table 3. HRQoL scores of the subsample and QALYs over the trial period*

	Usual care (n=415)			**QI (n=265)**		
	Missing	Mean (SE)	95% CI	Missing	Mean (SE)	95% CI
Baseline HRQoL	6 (1.45%)	0.168 (0.022)	0.124-0.211	4 (1.51%)	0.262 (0.029)	0.205-0.319
90 days HRQoL	79 (19.0%)	0.649 (0.017)	0.615-0.684	51 (19.2%)	0.669 (0.023)	0.623-0.715
180 days HRQoL	101 (24.3%)	0.706 (0.018)	0.670-0.741	81 (30.6%)	0.677 (0.025)	0.628-0.725
QALYs	-	0.274 (0.007)	0.260-0.288		0.287 (0.009)	0.269-0.306
ΔQALYs[a]		-	-		**0.013 (0.012)**	**(-0.009 to 0.036)**
ΔQALYs[b]		-	-		**-0.002**	**(-0.022 to 0.017)**

*Based on the imputed data.
[a] Unadjusted difference
[b] Adjusted for baseline EQ-5D, age and sex

Table 4. Cost-effectiveness analysis results

	ΔCosts (£)[a], mean (95% CI)	**ΔQALYs (£)[b], mean (95% CI)**	**ICER (£/QALY)**	**Incremental NHB in QALYs at (£/QALY)**			**Probability of cost-effectiveness at (£/QALY)**		
				13,000	20,000	30,000	13,000	20,000	30,000
EPOCH subsample									
Within trial period	467 (-800 to 1735)	-0.002 (-0.022 to 0.017)	Less effective	-0.038	-0.025	-0.018	23.5%	23.7%	24.2%
Lifetime horizon	1395 (-1083 to 3873)	0.018 (-0.396 to 0.432)	77,792	-0.089	-0.052	-0.029	28.7%	38.2%	43.8%
Scenario 1* Lifetime horizon	1395 (-1083 to 3873)	-0.250 (-0.622 to 0.123)	Less effective	-0.357	-0.320	-0.297	0.4%	1.6%	3.3%
Scenario 2* Lifetime horizon	1395 (-1083 to 3873)	-0.804 (-1.596 to -0.013)	Less effective	-0.911	-0.874	-0.851	0.9%	1.3%	1.5%
Lower-risk patients (with single indication)									
Within trial period	257 (-1283 to 1797)	-0.007 (-0.029 to 0.016)	Less effective	-0.027	-0.020	-0.016	33.9%	32.6%	31.0%
Lifetime horizon	394 (-2520 to 3308)	-0.088 (-0.578 to 0.401)	Less effective	-0.118	-0.108	-0.101	26.6%	30.0%	32.1%
Higher-risk patients (with multiple indications)									
Within trial period	1101 (-896 to 3097)	0.007 (-0.029 to 0.043)	158,253	-0.078	-0.048	-0.030	18.5%	21.1%	25.0%
Lifetime horizon	4441 (-120 to 9002)	0.311 (-0.420 to 1.041)	14,293	-0.031	0.089	0.163	45.6%	61.5%	69.3%
EPOCH population									
Within trial period	302	-0.014	Less effective	-0.037	-0.029	-0.024	-	-	-
Lifetime horizon	454	-0.072	Less effective	-0.338	-0.245	-0.187	-	-	-

[a] Adjusted for age and sex

[b] Adjusted for baseline EQ-5D, age and sex

*****Scenario 1: Patients alive at the end of trial were assumed to follow the group mean HRQoL until death. Scenario 2: Patients alive at the end of trial were assumed to retain their individual EQ-5D-3L scores at day 180 until death.

Figure 1. Cost-effectiveness acceptability curve for the QI intervention in the EPOCH subsample

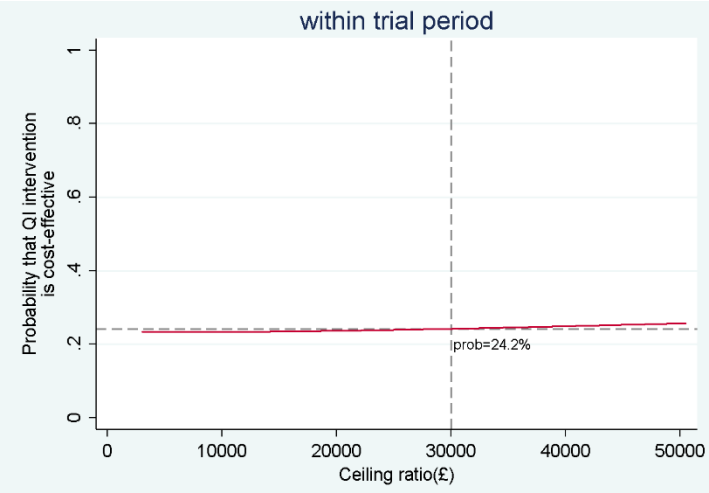

Figure 1a. within trial period

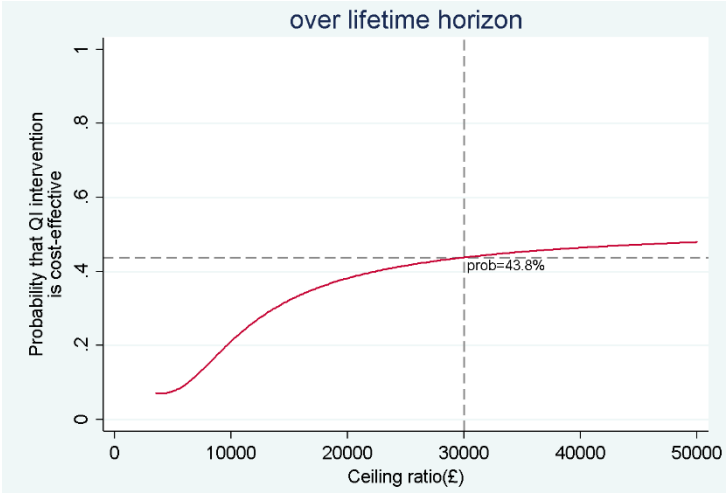

Figure 1b. over lifetime horizon

Table S1. Baseline characteristics of patients in EPOCH population and EPOCH subsample

	EPOCH subsample (n=680)		EPOCH population (n=15856)	
	Usual care (n=415)	QI (n=265)	Usual care (n=8482)	QI (n=7374)
Age – mean (SD)	66.7 (13.3)	66.4 (12.5)	68.5 (13.1)	68.2 (13.1)
Female	204 (49.2%)	131 (49.4%)	4547 (53.6%)	3935 (53.4%)
Indication for surgery				
Lower-risk (with single indication)	326 (78.5%)	197 (74.3%)	6360 (75.0%)	5512 (74.7%)
Higher-risk (with multiple indications)	89 (21.5%)	68 (25.7%)	2122 (25.0%)	1862 (25.3%)
Baseline EQ-5D-3L – mean (SD)	0.168 (0.448)	0.262 (0.466)	-	-

Table S2. Characteristics of patients alive at the end of trial period in EPOCH population and EPOCH subsample

	EPOCH subsample (n=614)		EPOCH population (n=12718)	
	Usual care (n=372)	QI (n=242)	Usual care (n=6784)	QI (n=5934)
All-cause mortality within 180 days	43 (10.4%)	23 (8.7%)	1108 (13.1%)	850 (12.5%)
Age – mean (SD)	66 (13)	66 (12)	67 (13)	67 (13)
Female	181 (48.7%)	120 (49.6%)	3664 (54.0%)	3168 (53.4%)
180 days EQ-5D-3L, mean (SE)	0.787 (0.015)	0.741 (0.023)	-	-

Table S3. Lifetime costs and QALYs of the subsample

	Mean (95% CI)	
	Usual care (n=415)	**QI (n=265)**
Over lifetime horizon		
Lifetime costs (£)	47749 (46063-49436)	49258 (47294-51221)
Lifetime QALYs	9.434 (8.938-9.929)	9.564 (8.986-10.143)
ΔCosts (£)[a]	-	1395 (-1083 to 3873)
ΔQALYs[b]	-	0.018 (-0.396 to 0.432)

[a] Adjusted for age and sex

[b] Adjusted for baseline EQ-5D, age and sex

Table S4. Estimated costs, EQ-5D-3L scores, and QALYs for EPOCH population over the trial period and over lifetime horizon

	Usual care (n=8482)		**QI (n=7374)**	
	Mean (SE)	95% CI	Mean (SE)	95% CI
Within trial period				
Costs of inpatient stay after surgery	6225 (75)	6078-6372	6347 (79)	6193-6501
Costs of readmissions	1116 (23)	1071-1160	1309 (28)	1254-1364
Costs of outpatient, A&E and primary care	789 (1)	787-791	738 (1)	736-740
QI intervention	0		32	
Total costs (£)	8129 (79)	7974-8284	8425 (85)	8258-8591
Baseline EQ-5D-3L	0.202 (0.001)	0.200-0.203	0.203 (0.001)	0.201-0.204
90 days EQ-5D-3L	0.593 (0.003)	0.587-0.598	0.588 (0.003)	0.582-0.594
180 days EQ-5D-3L	0.629 (0.003)	0.622-0.636	0.587 (0.003)	0.581-0.594
QALYs	0.238 (0.001)	0.235-0.240	0.225 (0.001)	0.222-0.228
Over lifetime horizon				
Lifetime costs (£)	42778 (222)	42344-43213	43379 (237)	42914-43844
Lifetime QALYs	8.115 (0.059)	7.998-8.231	8.129 (0.063)	8.005-8.253

Table S5. Marginal effect on costs of readmission and costs of outpatient and primary care

Using GLM with a Gamma family and log link	**Costs of readmissions**	**Costs of outpatient and primary care**
QI intervention	240	-4
Age	-119*	-7*
Female	2025	-23
Hospital cluster	55	-8

*P<0.05

Table S6. Marginal effect on HRQoL decrement at baseline

Using GLM with a Gamma family and log link	**HRQoL decrement**
Age	-0.001
Female	0.086*
Indication for surgery	
Peritonitis	ref
Perforation	-0.149
Intestinal obstruction	-0.238*
Haemorrhage	-0.212
Ischaemia	-0.102
Abdominal infection	-0.261*
Multiple indications	-0.148*
Other	-0.207*

*P<0.05

Table S7. Changes in HRQoL from randomisation based on the subsample

Covariate	**HRQoL change**
Randomised to usual care group (90 days follow-up)	0.746
Randomised to QI group (90 days follow-up)	0.739
Randomised to usual care group (180 days follow-up)	0.823
Randomised to QI group (180 days follow-up)	0.766
Baseline EQ-5D-3L	-0.857
Age	-0.0007
Female	-0.028