BJA

British Journal of Anaesthesia, xxx (xxx): xxx (xxxx)

doi: 10.1016/j.bja.2019.06.003
Advance Access Publication Date: xxx
Editorial

EDITORIAL

The Enhanced Peri-Operative Care for High-risk patients trial: an independent discussion and commentary

Victoria Cornelius[1] and Carolyn L. Johnston[2,*]

[1]Imperial Clinical Trials Unit, School of Public Health, Imperial College London, UK and [2]Department of Anaesthesia, St Georges University Hospitals NHS Foundation Trust, London, UK

*Corresponding authors. E-mails: v.cornelius@imperial.ac.uk, Carolyn.johnston1@nhs.net

Emergency laparotomy has a high mortality rate and there is significant variation in care processes and outcomes around England and Wales.[1] The Enhanced Peri-Operative Care for High-risk patients (EPOCH) trial studied the effectiveness of a national quality improvement (QI) programme to implement a care pathway to reduce mortality in patients undergoing emergency laparotomy.[2] There have been several smaller studies looking at components of the emergency surgery care pathway with encouraging results. Subsequent to the publication of the EPOCH study, the emergency laparotomy collaborative (ELC) group have reported an uncontrolled before and after study of a QI collaborative approach in 28 English hospitals.[3] This confirmed their previous study in four hospitals that a collaborative approach to implement a simple care bundle led to a reduction in mortality, noting that many of the mortality benefits were not apparent until the second year of the programme. Other studies have also shown a reduction in mortality. In Denmark, a prospective study of staff education and a care protocol showed a reduction in mortality compared with historical controls in high-risk abdominal surgery and peptic ulcer disease.[4,5]

There is mixed evidence in peer reviewed publications on the effectiveness of QI interventions in general.[6] This may be attributable to poor fidelity of QI methods such as plan–do–study–act cycles or a mistaken emphasis on specific interventions over cultural change.[7,8] The EPOCH study sought to evaluate the effectiveness of a QI programme on the implementation of a care pathway spanning the whole perioperative pathway from diagnosis to discharge, across all of England and Wales. The study commenced in March 2014, shortly after the National Emergency Laparotomy Audit (NELA) commenced data collection in December 2013. The recommended care standards measured by NELA are based on consensus statements and publications from the Royal College of Surgeons of England, Association of Surgeons of Great Britain and Ireland, reports from the National Confidential Enquiry into Patient Outcome and Death and the National Institute for Health and Care Excellence. None of these standards of care are based on the findings of RCTs.[9]

Here, we consider the findings of the EPOCH study using the format of an independent discussion.[10] An independent discussion is one written by an author with expertise in research methodology, independent of the original research team. The author has access to the introduction, methods, results, and raw data, but not the discussion section. Highlighting the similarities and differences between the original discussion and the second, independent discussion may help readers contextualise the findings of a study and its limitations. This is intended to increase the inferential reproducibility of scientific research by mitigating some of the potential limitations of existing discussion sections.

Main findings of the trial

Original discussion

'The principal finding of this trial was that there was no survival benefit associated with a national quality improvement programme to implement an evidence-based care pathway for patients undergoing emergency abdominal surgery. Furthermore, there was no beneficial effect on 180 day mortality, hospital stay or hospital readmission. At a national level, there were only modest improvements amongst the ten measures selected to reflect key processes of care within the pathway. Experience from individual hospitals suggested wide variations in which of the 37 pathway elements local QI teams chose to tackle, the rate of change they achieved, and their eventual success.'

'Our findings reveal that implementation of such an extensive care pathway was a more complex challenge

than expected by our clinical community. It is important to interpret the results of this trial alongside those of the ethnographic study and process evaluation, which together suggest that quality improvement programmes designed to implement complex care pathways require more resources, with dedicated time for clinical teams to focus on making change happen.'

Independent discussion

The EPOCH QI programme was developed using a well-established consensus-based approach based on existing guidelines and published literature. The authors then undertook a large national trial to evaluate the policy of implementing the EPOCH programme. They carried out a randomised stepped-wedge cluster trial which is a recognised design to increase the efficiency of the trial with respect to required sample size.

The trial did not find evidence that adoption of the EPOCH programme for emergency open major abdominal surgery in the NHS improved 90 day mortality. This result was not a consequence of the study not meeting its anticipated sample size (15 873 rather than 27 540) as, despite a lack of precision in the primary intervention effect estimate, there was no convincing evidence for survival benefit. The 95% confidence interval (CI) for the hazard ratio (HR) covered 0.96 to 1.28, where an HR <1 favours the EPOCH programme. Owing to the lack of precision in the estimate, the result is consistent with the conclusion that there is no difference between control and EPOCH, and also that there is an important increase in mortality in the control.

Commentary

Both the original investigators and the independent review provide the same summary of the quantitative findings—that there was no evidence that the intervention reduced 90 day mortality.

Relationship of main finding to previous studies independent discussion

Original discussion

'A review undertaken by the authors found previous studies evaluating either quality improvement components or packages did find some benefit on survival, but these were assessed as being at high risk of bias. The results of this large scale randomised trial differ as it suggests no benefit. The seemingly contrasting results with previous research may be due to this trial answering a more pragmatic question regarding adoption of the EPOCH programme, rather than implementation of quality improvement components, or it may be due to higher risk of bias in previous studies.'

Independent discussion

The EPOCH pathway was developed using a Delphi process to update the existing Royal College of Surgeons guidelines. The evidence underlying most of these recommendations is of lower grades. Of the 27 recommendations, 15 are graded as grade I, that is 'Scientific evidence is lacking, of poor quality, or conflicting, such that the risk *vs* benefit balance cannot be assessed. Clinicians should help patients understand the uncertainty surrounding the clinical service'. A further four recommendations are level C ('At least fair scientific evidence suggests that there are benefits provided by the clinical service, but the balance between benefits and risks are too close for making general recommendations. Clinicians need not offer it unless there are individual considerations'), seven are level B ('At least fair scientific evidence suggests that the benefits of the clinical service outweighs the potential risks'), and only one, on venous thromboprophylaxis, is level A ('Good scientific evidence suggests that the benefits of the clinical service substantially outweigh the potential risks').

Commentary

Both discussions indicate a high risk of bias in study design of the previous studies, and that before and after studies are potentially confounded by decreasing mortality in the patient population, which the stepped wedge cluster design is intended to mitigate. The original discussion highlights that the simpler interventions in previous studies were more readily achieved than the complex EPOCH intervention. The original discussion also lists studies of theoretical implementation models, commenting that none of these models emphasise institutional support, which the EPOCH ethnographic study found to be a key determinant of successful implementation. The original discussion does not cover the dearth of grade A evidence underlying the most important components of the care pathway, despite the efficacy of these interventions being crucial to the outcomes measured of the study. This is a key factor to consider when interpreting the study findings—did mortality remain unchanged because the QI intervention did not result in pathway adoption, or because the pathway was adopted, but this did not reduce mortality.

Additional (secondary) findings

Original discussion

'The primary outcome of 90 day mortality occurred in 1393 patients in the usual care group (16%) compared with 1210 patients in the QI group (16%) (Hazard ratio, QI vs usual care: 1.11 [0.96 to 1.28]). Results were similar for mortality within 180 days (HR 1.12 [0.98 to 1.28]). Patients in the QI group had a lower probability of hospital discharge (Hazard ratio for hospital discharge 0.90 [0.83 to 0.97]), leading to a marginally longer hospital stay (days in hospital, usual care: 8 [13 to 23] days *vs* QI: 8 [13 to 24] days), although this difference was not clinically meaningful. There was no difference between groups in hospital re-admission within 180 days (usual care 1618 (20%) vs. QI 1242 (18%); Hazard ratio for re-admission 0.87 [0.73 to 1.04]). In a secondary analysis, we found no evidence that the QI strategy became more effective the longer it had been adopted.'

'The extent to which the QI programme was delivered as intended, as well as enablers and barriers to change, are described in full in the report of the EPOCH trial process evaluation.'

Independent discussion

There were three secondary clinical outcomes. The 180 day survival rate result was consistent with the 90 day result. Both results did not demonstrate benefit and the range of values in the 95% CI predominantly covered values that were in favour of control. The result was not statistically significant at the 5% level, but the likelihood of no difference or increased mortality in the control group were more probable than benefit for EPOCH. The 'hazard' of discharge from hospital was again in favour of control but this time the finding was statistically significant, although the absolute difference in time is unlikely to be meaningful to patients as both groups had a median in-hospital stay of 8 days and their interquartile ranges were similar. There was inconclusive evidence regarding hospital re-admission rates. The point estimate indicated a lower hazard of being readmitted with the EPOCH programme but the estimate was imprecise, and the 95% CI is consistent with both the conclusion that there is no difference, and that there is a meaningfully important reduction in readmission for EPOCH as the HR 95% CI ranged from 0.73 to 1.04.

The process measures have only been presented by authors as raw counts and percentages. There is a temptation to compare these absolute numbers across, but they are not adjusted for time or clustering effects, so this approach would not be valid. Interpreting the raw data at face value in Table 2[2] does suggests that the difference may be fairly small and not enough behaviour change to improve clinical outcomes for patients. The ethnographic sub-study identified multiple challenges in implementation of the EPOCH programme. If these challenges were experienced across all hospitals this would offer an explanation for not observing any improvement in mortality.

Commentary

Both the original and independent discussions agree on the interpretation of the secondary outcome measures of 180 day survival and length of stay. They differ on the interpretation of hospital readmission; the original discussion states the hazard of hospital readmission did not change, whereas the independent discussion states that this finding was inconclusive. This was attributable to the smaller-than-expected sample size resulting in imprecise estimates, that is both conclusions that there is no difference and a meaningful reduction in readmissions are possible in the results published.

The original paper also lists the representation of hospitals at initial and follow-up meetings as a process measure, showing a large decrease in attendance between the initial and follow-up meetings. Both discussions mention patient level process measures but do not test these for statistical significance as per the original EPOCH analysis plan, and the risk of bias in this type of before and after measures.

Limitations

Original discussion

'Despite the large sample, fewer patients than expected underwent emergency abdominal surgery, and the 90 day mortality rate was lower than anticipated. The sample size calculation was based on Hospital Episodes Statistics data which do not provide a specific diagnostic code for emergency abdominal surgery. Instead we identified a series of codes for relevant procedures. We chose to power the trial to detect a very modest treatment effect partly to accommodate the possibility that these data were poorly representative of the EPOCH trial population. However, the 95% confidence interval for our primary effect estimate was narrow, with a lower limit which indicates a maximum potential mortality reduction of 4%. Our findings are unlikely to change with a larger sample size.'

Independent discussion

Large pragmatic trials are primarily interested in determining the 'treatment policy strategy'.[11] In this trial the intervention effect estimates pertain to adoption of the EPOCH programme into NHS hospitals. With no intervention effect observed, the natural question turns to whether this is attributable to the EPOCH programme not being implemented and managing to achieve the required behaviour change as intended. The ethnographic sub-study and process evaluation identified that adherence and fidelity may have been lower than expected. Therefore this offers one reason as to why no benefit may have been seen. It was not the aim of the study to estimate the effect of the EPOCH programme if it was implemented as intended, but the authors did explore whether the length of time the programme was in place made a difference. The results did not reveal any improvements over time, and it remains unclear whether the lack of observed effect is attributable to the lack of implementation or the programme itself not making a difference.

This particular trial, where the results are more in favour of control rather than just uncertainty in either direction, provides an example that raises an interesting question around the absence of evidence within the null hypothesis testing framework, and when it can be regarded as evidence of absence. Controversies in the interpretation of null hypothesis significance testing have been put forward by Tijmstra.[12] In this discussion, we have placed our focus on the interpretation of the plausible range of values provided by the 95% CI of the intervention effect estimate. To resolve this uncertainty future work would need to ensure the implementation of QI components, rather than just the adoption of a policy to attempt implementation.

Stepped-wedge trials are more complex to operationalise than cluster RCTs, and can be vulnerable to bias caused by external factors that cause changes over time. Although these effects can technically be disentangled though the analysis, if the design is not implemented as planned and the analysis assumptions are not met, the result may be bias. The design was ambitious in the number of time periods and steps used, requiring 90 hospitals over an 85 week period, in 17 time periods of 5 weeks, where hospitals were grouped into 15 clusters based on location to reduce contamination bias. Whether these time steps were achieved with a suitable accuracy required for the design is not clear. The parametric survival analysis model appropriately adjusted for secular trend and varying secular trends across hospitals that would allow for changes in care delivery not caused by the intervention to vary over time and by hospital, but the time period was limited to a definition of before and after treatment rather than more periodic time points.[13] The model included a single term for intervention allowing only a constant shift in any intervention

effect which carries the assumption that the intervention effect would remain constant over the varied 5–80 week intervention period. This assumption was examined to a limited extend in secondary analysis providing some justification for this decision.

The primary endpoint had to be modified after the complication that 90 day mortality data could not be obtained for patients in Wales. Instead, information up to point of discharge was used for these patients after which their data were censored, meaning they had shorter follow-up time. This approach makes an assumption of uninformative censoring for these patients and affected 730 (4.8%) patients. This compromise is unlikely to have made any meaningful impact on study results or altered study conclusions.

The study was underpowered to detect the 3% difference in mortality that the investigators had set out to achieve, but a larger sample size would not have increased the precision such that survival benefit for EPOCH would have been found. The unadjusted data such as those in Table 2 and Figures 2 and 3[2] should not be over-interpreted, and focus should remain on the appropriately adjusted estimates that take account of time and clustering effects.

Commentary

The independent discussion highlights the potential limitations on the analysis if the ambitious time steps in implementation are not conducted as required by the study design; this is not discussed in the original paper. The independent discussion also describes that a limitation of the analysis model, whilst adjusting for secular trend in time and across hospitals, uses a model treating time as a single term rather than more periodic time points. This was examined in the EPOCH paper's secondary analysis, but not discussed in detail as a potential limitation.

Both discussions agree that the study was underpowered to detect a 3% reduction in mortality, as the number of patients having surgery was lower than expected, and the mortality rate was lower than expected. Both discussions state this was not a cause of the null finding. Both also include the limitation created by the lack of post-discharge survival data from Wales, but that this compromise is unlikely to have materially impacted on the results obtained.

The independent discussion also highlights that the data as presented do not fully answer the question as to whether the EPOCH programme did not reduce mortality caused by a lack of implementation or a lack of efficacy in the recommended pathway components. The original discussion comments on the lack of evidence behind some of the pathway recommendations, and although they were determined by a robust Delphi consensus process, the underlying evidence for adoption of some of the recommendations is weak.

Strengths

Main discussion

'The strengths of this trial include wide generalizability (large number of consecutive patients enrolled by many hospitals), robust trial design and the devolved leadership to local clinical QI teams. The EPOCH care pathway was developed through a Delphi consensus process to update national professional guidelines.[7] As with many evidence-based treatment guidelines, some recommendations were graded as strong although the available evidence was weak. The choice of component interventions such as intensive care admission and consultant-led care was primarily based on expert opinion; it is unclear how this evidence base could be improved. Partnership with the National Emergency Laparotomy Audit allowed an efficient trial design with no additional data collection for participating staff.'

Independent discussion

The EPOCH programme was developed by experts through a structured process based on existing well-established guidelines. It was a large scale randomised trial designed with attention to minimise contamination bias for workforce movement within local geographic regions. It includes a representative range of NHS hospitals and achieved an extremely low missing data rate with minimal hospital withdrawal, and all 93 hospitals that started the trial contributed to the final primary analysis. Its primary outcome of mortality is objective which is crucial in unblinded trials. The study protocol and detailed statistical analysis plan have been publicly available from early on in the conduct of the study via the study website. These documents demonstrate the consistency and intent to report a specified primary outcome and three clinical secondary outcomes, and 10 predefined process measures, all of which are reported in the publication.

Commentary

Both the original and independent discussions include similar strengths to the study design and implementation. The original discussion also lists the partnership with the National Emergency Laparotomy Audit as a strength, allowing for a large trial to be undertaken efficiently, requiring no additional data collection for participating staff.

Future directions

Original discussion

'These findings suggest future quality improvement programmes should implement fewer, more discrete changes and ensure leadership teams have adequate time to achieve sustained improvements in patient care. Undue emphasis on success stories from small early studies may lead us to under-estimate the requirements for successful quality improvement interventions.'

Independent discussion

The process evaluation identified difficulties in engaging with the EPOCH programme which is a complex intervention of many QI components. Future work to understand whether implementing QI components rather than just the adoption of the policy to implement a programme would provide valuable insight as to whether these offer potential interventions to help reduce the much-needed reduction in mortality in emergency abdominal surgery patients.

Commentary

Both discussions agree that future directions should focus on fewer individual pathway components rather than a complex pathway. As the aim of the study was to test the QI intervention, future work with similar aims should focus on pathway components with stronger evidence of patient benefit, so any lack of impact on patient outcomes could be more clearly attributed on lack of efficacy of the QI intervention. The original discussion also indicates that further focus is needed on leadership time and resources in addition to pathway components.

Conclusion

Original discussion

'In this stepped-wedge cluster randomised trial, we did not identify any survival benefit from a national quality improvement programme to implement an enhanced pathway of care for patients undergoing emergency abdominal surgery. This is likely attributable to variation between hospitals in fidelity of implementation, prioritisation of pathway components, and the time required to achieve effective change. These findings suggest future quality improvement programmes should implement fewer, more discrete changes and ensure leadership teams have adequate time to achieve sustained improvements in patient care. Undue emphasis on success stories from small early studies may lead us to underestimate the requirements for successful quality improvement interventions.'

Independent discussion

There is no evidence that adopting the EPOCH programme for patients undergoing emergency abdominal surgery will provide survival benefit in an NHS hospital setting.

Commentary

The conclusion from both commentaries is the same. The original conclusion expands on the reasons behind the findings with the ethnographic data reported in papers currently under review: fidelity of implementation, prioritisation of pathway components, and the time required to achieve effective change.

Inferential reproducibility

There is excellent concordance between the two discussions, with good inferential reproducibility. The independent discussion lists more limitations of the complex trial design than the original discussion and commentary on the underlying evidence base for the pathway components, but both agree on the key findings, their interpretation, and future directions of study.

Authors' contributions

Writing of independent discussion: VC
Writing of commentary: CJ
Selection of excerpts from the original study: CJ
Additions to independent discussion: CJ
Both authors reviewed and commented on the final draft.

Declaration of interest

The authors declare that they have no conflicts of interest.

References

1. NELA Project Team. *Fourth patient report of the national emergency laparotomy Audit*. RCoA London; 2018
2. Peden CJ, Stephens T, Martin G, 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
3. Aggarwal G, Peden CJ, Mohammed MA, et al. Evaluation of the collaborative use of an evidence-based care bundle in emergency laparotomy. *JAMA Surg* 2019, e190145
4. Tengberg LT, Bay-Nielsen M, Bisgaard T, Cihoric M, Lauritsen ML, Foss NB, AHA study group. Multidisciplinary perioperative protocol in patients undergoing acute high-risk abdominal surgery. *Br J Surg* 2017; **104**: 463–71
5. Møller MH, Adamsen S, Thomsen RW, Møller AM. Peptic Ulcer Perforation (PULP) trial group. Multicentre trial of a perioperative protocol to reduce mortality in patients with peptic ulcer perforation. *Br J Surg* 2011; **98**: 802–10
6. Dixon-Woods M, Martin GP. Does quality improvement improve quality? *Future Hosp J* 2016; **3**: 191–4
7. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan–do–study–act method to improve quality in healthcare. *BMJ Qual Saf* 2014; **23**: 290–8
8. Bosk CL, Dixon-Woods M, Goeschel CA, Pronovost PJ. The art of medicine. Reality check for checklists. *Lancet* 2009; **374**: 444–5
9. Odor PM, Grocott MP. From NELA to EPOCH and beyond: enhancing the evidence base for emergency laparotomy. *Periop Med* 2016; **5**: 23
10. Avidan MS, Ioannidis JP, Mashour GA. Independent discussion sections for improving inferential reproducibility in published research. *Br J Anaesth* 2019; **122**: 413–20
11. European Medicines Agency. *ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials* August 2017. [Accessed 16 April 2019]. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical_en.pdf
12. Tijmstra J. Why checking model assumptions using null hypothesis significance tests does not suffice: a plea for plausibility. *Psychon Bull Rev* 2018 Apr 1: 1–2
13. Hemming K, Taljaard M, Forbes A. Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. *Trials* 2017; **18**: 101