Hospital level evaluation of the effect of a national quality improvement programme: Time-series analysis of registry data

T.J. Stephens, C.J. Peden, R.W. Haines, M.P.W. Grocott, D. Murray, D. Cromwell, C. Johnston, S.P. Hare, J. Lourtie, S. Drake, G. Martin and R.M. Pearse on behalf of the Enhanced Perioperative Care for High-risk patients (EPOCH) trial group.*

***see supplementary file**

T. J. Stephens, William Harvey Research Institute, Queen Mary University of London, UK
t.t.stephens@qmul.ac.uk
C.J. Peden, Keck School of Medicine, University of Southern California, USA. Carol.Peden@med.usc.edu
R.W. Haines, William Harvey Research Institute, Queen Mary University of London, UK. r.haines@qmul.ac.uk
M.P.W. Grocott, Southampton NIHR Biomedical Research Centre, University Hospital Southampton / University of Southampton, UK. mike.grocott@soton.ac.uk
D. Murray, James Cook University Hospital, Middlesbrough, UK. dave.murray1@nhs.net
D. Cromwell, London School of Hygiene & Tropical Medicine, London, UK. david.cromwell@lshtm.ac.uk
C. Johnston, St Georges University Hospital NHS Foundation Trust, London, UK. Carolyn.johnston1@nhs.net
S.P Hare, Medway Maritime Hospital, Kent, UK. sarah.hare1@nhs.net
J. Lourtie, National Emergency Laparotomy Audit, Royal College of Anaesthetists, London, UK.
JLourtie@rcoa.ac.uk
S. Drake, Royal College of Anaesthetists, London, UK. SDrake@rcoa.ac.uk
G.P.Martin, THIS Institute (The Healthcare Improvement Studies Institute), University of Cambridge, UK.
graham.martin@thisinstitute.cam.ac.uk
R.M. Pearse, William Harvey Research Institute, Queen Mary University of London, UK. r.pearse@qmul.ac.uk

Correspondence to:
T.J. Stephens
Critical Care and Perioperative Medicine Research Group,
Adult Critical Care Unit, Royal London Hospital, London, E1 1BB
United Kingdom

Keywords: Emergency surgery; Quality Improvement; Implementation; Evaluation;

Main text: 4542 words **Abstract:** 299 words

ABSTRACT

Background and objectives

A clinical trial in 93 NHS hospitals evaluated a quality improvement programme for emergency abdominal surgery, designed to improve mortality by improving the patient care pathway. Large variation was observed in implementation approaches and the main trial result showed no mortality reduction. Our objective therefore was to evaluate whether trial participation led to care-pathway implementation and to study the relationship between care-pathway implementation and use of six recommended implementation strategies.

Methods

We performed a hospital-level time-series analysis using data from the Enhanced Peri-Operative Care for High-risk patients (EPOCH) trial. Care-pathway implementation was defined as achievement of >80% median reliability in ten measured care-processes. Mean monthly process performance was plotted on run-charts. Process improvement was defined as an observed run-chart signal, using probability-based 'shift' and 'runs' rules. A new median performance level was calculated after an observed signal.

Results

Of 93 participating hospitals, 80 provided sufficient data for analysis, generating 800 process measure charts from 20,305 patient admissions over 27 months. No hospital reliably implemented all ten processes. Overall, only 279 of the 800 processes were improved (3 [2-5] per hospital) and 14/80 hospitals improved more than six processes. Mortality-risk documented (57/80 [71%]), lactate measurement (42/80 [53%]) and cardiac-output guided fluid therapy (32/80 [40%]) were most frequently improved. Consultant-led decision making (14/80 [18%]), consultant review before surgery (17/80 [21%]) and time to surgery (14/80 [18%]) were least frequently improved. In hospitals using ≥5 implementation strategies, 9/30 (30%) hospitals improved ≥6 care processes compared with 0/11 hospitals using ≤2 implementation strategies.

Conclusion

Only a small number of hospitals improved more than half of the measured care-processes, more often when at least 5 of 6 implementation strategies were used. In a longer-term project this understanding may have allowed us to adapt the intervention to be effective in more hospitals.

BACKGROUND

As the volume of surgical procedures performed worldwide continues to increase [1,2] the need for improvement in the quality and safety of surgical care has become a global healthcare priority [3–5]. This is of particular importance considering both the increasing age and complexity of the surgical population and the global mortality burden associated with surgery [6,7]. Emergency abdominal surgery is a commonly performed procedure worldwide, with high mortality rates, and wide variations in the standards of care [8–11]. The Enhanced Peri-Operative Care for High-risk patients (EPOCH) trial was performed to test whether a national quality improvement (QI) programme to implement a care-pathway could reduce 90-day mortality following emergency abdominal surgery [12].

The EPOCH trial intervention consisted of an evidence based care-pathway designed to improve patient outcomes and a QI programme promoting a set of implementation strategies designed to enable care-pathway implementation [13]. The main analyses were designed to evaluate the impact of the QI programme across a large cohort of 93 NHS hospitals, leveraging the large sample size to adequately power the trial. A different perspective is to view the EPOCH trial QI programme as an enabling factor in 93 separate hospital-level QI projects. The impact of local context on the effectiveness of QI efforts is increasingly understood, especially in relation to complex intervention delivery [14–16]. We observed wide variation in the approaches taken to implement the care-pathway, including differing ways of engaging colleagues and decisions regarding which parts of the pathway to implement first, as well variations in the challenges faced. More details are provided below and in our concurrent trial process and ethnographic evaluation papers [13,17]. Given the level of heterogeneity across participating hospitals, an analysis designed to understand changes in care processes at the individual hospital level is also needed.

In this paper, we explore how a form of simple time-series chart (the run-chart) might enable detailed hospital level analysis of process change over time when system improvements are attempted at a national level [18]. The primary objectives of this study were to: 1) evaluate, at the individual hospital level, whether participation in the EPOCH trial QI programme led to implementation of the EPOCH care-pathway and 2) to assess the relationship between care-pathway implementation and use of the implementation strategies. Our secondary objectives were: 1) to describe the number of improvements in care processes overall and 2) to describe which care processes were most commonly improved (or potentially degraded).

METHODS

This was a prospectively designed time series analysis of registry data provided by hospitals participating in the EPOCH trial, a stepped-wedge cluster randomised trial across 93 UK National Health Service (NHS) hospitals. The registry was the National Emergency Laparotomy Audit (NELA), funded separately by the UK Healthcare Quality Improvement Programme, which started collecting individual patient data on 1st December 2013.

Patients and hospitals

Patients were recruited from March 2014 to October 2015. Recruited hospitals were grouped into 15 clusters of six to eight geographically co-located hospitals; clusters were randomised to start the intervention at five-week intervals. QI leads from each stakeholder discipline (surgery, anaesthesia, and critical care) were tasked with leading hospital wide improvement to implement the care pathway with the support and guidance of the national EPOCH QI team. QI leads were informed of their hospitals start date 12 weeks in advance. The main EPOCH trial analysis found no effect on the interventions upon any of the trial outcomes measures; 90-day risk adjusted mortality, length of hospital stay or hospital readmission [12]. Analysis of trial process measures (see below) suggested little improvement had occurred as a result of the intervention across the entire cohort. These results did not differ significantly between hospitals activated earlier in the stepped-wedge design compared with those activated later. The EPOCH trial was approved by the Research Ethics Committee of the NHS (REC reference 13/EM/0415).

The EPOCH care pathway and implementation strategies

Details of the 37-component evidence based care pathway are provided in Figure 1, and a full summary of evidence is available on the trial website (www.epochtrial.org). The EPOCH programme theory was based on current evidence and learning from a range of other QI programmes [19–22]. Six specific implementation strategies were developed to facilitate care-pathway implementation (see Table 1). The EPOCH QI programme was designed to support local clinicians. The programme comprised: a one day face-to-face educational meeting; a half-day follow up; a virtual learning environment (VLE) and telephone / email support from the core EPOCH team. The QI intervention was designed to be “light touch”, recognising the limited resources of the study, of clinician time within the NHS and the fact that data collection through NELA was already taking place. Full details of the EPOCH QI programme are reported elsewhere [13].

Data collection

Data were collected through the NELA database (www.nela.org.uk). Inclusion and exclusion criteria for these analyses were the same as for the main trial [12]. Patients were eligible for inclusion in the data analysis if they were 40 years or older, and undergoing emergency open abdominal surgery in a participating hospital. Patients were excluded from the analysis if they were undergoing a simple appendicectomy, 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. We pre-defined a longer data collection period than the main trial, so that data from the 1st January 2014 to 31st March 2016 (six months following the end of the EPOCH trial) were analysed. The rationale for this is that the shift rule requires at least six data points (i.e. six months of data, see below) for change to be demonstrated. There is also evidence that the effects of QI may take longer than expected to show [23,24]. Therefore, we included this six-month wash-out period to provide clusters activated later in the trial adequate opportunity to demonstrate improvement. We used data from our process evaluation questionnaire to quantify recommended implementation strategy use in each hospital (Table 1). Full results of the process evaluation are reported elsewhere [13]. 77/93 (83%) of QI leads completed the exit questionnaire. For this study, we used binary responses related to implementation strategy usage e.g. we did / did not form a QI team.

Process measures

Process measures in this study are the same as those in the main trial, but now analysed at the individual hospital level rather than in a pooled analysis. The 10 key care processes of the EPOCH care-pathway for which process measure data was available via the NELA dataset were: 1) Consultant-led decision making; 2) Consultant review of patient before surgery; 3) Pre-op mortality risk documented; 4) Time from decision to operate to entrance to the operating theatre; 5) Entry to operating theatre within NCEPOD target timeframe; 6) Consultant delivered surgery; 7) Consultant delivered anaesthesia; 8) Cardiac output monitoring to guide fluid therapy; 9) Measurement of serum lactate intra-operatively; and 10) Admission to critical care post-operatively.

Data analysis

Process measure data were analysed for each hospital. Data for each calendar month were pooled and the mean calculated and plotted onto run charts, using a pre-programmed Excel Workbook designed specifically for the EPOCH trial (see Figure 2 for a worked example). A baseline median was constructed with the first ten data points (January 2014–October 2014) or from January 2014 up to and including the month of trial cluster activation, whichever provided the longer baseline period. To

increase the likelihood that any signals identified in the run charts were associated with the EPOCH trial, and not pre-existing improvement efforts (such as involvement in NELA), each hospital's baseline median was assessed for signals using the run chart rules. In particular the 'runs rule' was used to identify potential improvements in patient care processes before the improvement intervention started (see [18] for the reference chart for this). In line with recommended practice, if no signals were seen, the baseline median was fixed and extended forward creating the centre-line for all data points on the chart, to facilitate analysis of signals over time [18,25]. Where too few runs were seen, the median was not fixed and extended but instead continued with all data points in the chart contributing to this. The patterns of data points on the charts were visually inspected for signals compatible with accepted run-chart rules which are probability-based, predefined data patterns with a probability of <5% of occurring by chance [18]. The two run-chart rules used in this analysis are: 1) a shift, identified as a signal with 6 or more data points on one side of the median line and 2) too few runs, identified by counting the number of runs (groups of data points falling either above or below the median line), and then referring to the published guidance for the upper and lower limits [18]. The trend rule was not included due to evidence of lack of utility [25]. When a signal was identified in a care-process, a new median monthly delivery rate was calculated based upon the data contributing to the signal.

To answer objective 1, we considered the care-pathway to be implemented if the 10 measured processes improved to the extent that all had a median monthly delivery rate of >80% following activation to the intervention (or a sustained median of <6 hours to surgery for process measure four, as above). Eighty percent was chosen as it is considered a minimum level of process reliability and is used by NELA to define an acceptable standard of care [8,26]. Care processes already reliably delivered to >80% of patients were also included. To answer objective 2, we defined care-process improvement as any signal toward improvement identified on a hospital's run-chart, regardless of the magnitude of the improvement (unless followed by a subsequent signal toward process degradation). For each hospital we also calculated the proportion of patients before and after activation to the intervention who received each of the target care-processes and the median time from decision to operate to entry into the operating theatre (see care-process 4 above) pre and post activation. These were then aggregated for all trial hospitals included in the run chart analysis to show the overall effect size of process changes.

We report the relationship between care-pathway implementation and implementation strategy fidelity using descriptive statistics and analysed the relationship using a scatterplot and R^2 calculation. To explore this relationship further, we undertook post-hoc analysis comparing: fidelity to

implementation strategy usage (using 5 or 6 strategies vs. using <5 strategies), individual implementation strategy usage, NELA data collection method and care-process improvement between the least improved (≤2 care-processes improved; n=28) and the most improved (≥6 care-processes improved n=14) hospitals. We used Fishers Exact Test, with 2x2 contingency tables to compare groupings and a one-sided p-value, with significance set at $p<0.05$.

We undertook a validation exercise, with an independent reviewer (RH) analysing a random selection of 200 of the total 800 charts (25%). The reviewer repeated the analysis of each chart. Results for the 200 charts were compared with the original analyses and any inconsistencies of analysis discussed, and the final result agreed upon. An error rate of ≥5% was decided as the threshold for whether a further validation exercise would be necessary. We also undertook two post-hoc sensitivity analyses on the charts from the hospitals that improved more than half the process measures (14/80 hospitals improved ≥6 care processes) to test the different results obtained by using stricter analysis rules. These rules use thresholds for identifying signals (runs and shifts) based upon the total number of data points on the chart, rather than a fixed rule, which may provide more accurate findings [27]. In this group of most improved hospitals we also undertook analysis using a run chart centre line (median) based on all chart data, rather than the fix and extend method.

RESULTS

Of the 93 hospitals enrolled in the EPOCH study, 86 hospitals had data available for analysis. However, six hospitals had data capture of insufficient quality (<50% case-ascertainment reported via NELA for either both Years 1 and 2 of the audit or for the year in which the hospital was activated to the QI intervention) to enable month-by-month analysis using run charts. Therefore, 80 hospitals were included in analyses resulting in the generation of 800 run charts for the 10 measures of interest, based upon analysis of data from 20,305 patient admissions (Figure 3). Table 2 displays key hospital characteristics of interest. In the validation exercise, six errors were identified, giving an inter-observer agreement of >95%. Of the six errors, three were Type 1 errors, where charts were marked as having signals toward improvement that were not there, and two were Type 2 errors, where signals toward improvement were missed. One was an error where a degraded care-process was missed (Type 2 error). In all cases, signals were marginal and overall, these errors did not substantially change our main findings or conclusions.

No hospital achieved implementation of the care-pathway according to our definition (all ten measured care processes improved to a ≥80% median delivery rate). Regarding objective 2 (describing all improvement, not just achievement of >80% reliability), 21/80 hospitals improved ≥ 5 of the 10 measured processes and 14/80 improved ≥ 6. Figure 4 displays the overall number of improved care processes per hospital. Pre-operative risk assessment (57/80 [71%]), intra-operative lactate measurement (42/80 [53%]) and cardiac-output guided fluid therapy (32/80 [40%]) were the most frequently improved care processes (Table 3). Consultant-led decision-making (14/80 [18%]), consultant review before surgery (17/80 [21%]) and time from decision to operate to surgery (14/80 [18%]) were the least likely care processes to improve (Table 3). Questionnaire data describing implementation strategy use showed that 10/77 (13%) of QI leads responding said that all six strategies had been used, 23/77 (30%) indicated five had been used, 21/77 (27%) indicated four had been used, 8/77 (10%) used three strategies, 10/77 (13%) used two and 5/77 (6%) just one. No QI lead reported zero implementation strategy usage. Table 1 shows the reported usage of each QI strategy. As no hospital achieved care-pathway implementation, we undertook analysis of the relationship between implementation strategy usage and the number of care-processes improved. We divided the cohort into tertiles of implementation strategy usage (1-2 strategies, 3-4 strategies and 5-6 strategies) and defined successful hospitals as those with six or more improved care processes (i.e. more than half of care processes improved). In hospitals that used 1-2 strategies, we found that no hospitals (0/11) improved six or more care processes, whilst among those that used 3-4 strategies 4/25 (16%) hospitals improved six or more care processes, and in those that used 5-6 strategies 9/30 (30%) of

hospitals improved ≥6 care processes. However, using a linear analysis model across the whole trial cohort, we found no correlation between implementation strategies used and the number of care-processes improved at individual hospitals (R^2 = 0.084, Supplementary File). Figure 5a-c, presents the post-hoc analysis findings, comparing least and most improved hospitals by implementation strategy usage and NELA data collection method. Prospective NELA data collection, by all members of the care team i.e. presenting a lower time-burden for QI leads, was positively associated with greater care process improvement (p=0.039). Details of further evaluation of the relationship between care-process improvement and implementation strategy usage are reported in the supplementary file.

During the analyses we identified the care processes in each hospital that were already reliably delivered, as defined by a baseline median of ≥80% delivery of a process measure. Consultant led decision making was the care-process most reliably delivered pre-EPOCH, with 71/80 hospitals achieving a median of ≥80% for this measure. Of these hospitals 11 (15%) further improved upon this performance during the EPOCH intervention period. Consultant delivered surgery was often already reliably delivered, with 70/80 hospitals already achieving a median of ≥80% for this measure. Nevertheless, 19 of these hospitals (27%) managed to further improve this care-process. Consultant delivered anaesthesia was the next most reliably delivered care-process at baseline (57/80 hospitals) and 16 of these hospitals (28%) demonstrated further improvement in consultant delivered anaesthesia. Conversely, only 2/80 (2.5%) hospitals had a median time from decision to operate to surgery of <6hrs before the EPOCH trial started; this was also the most challenging care-process to improve, although 17.5% (14/80) of hospitals did demonstrate an improvement on the run chart analysis. Process degradation was also observed during run-chart analysis. We found 43/800 (5.4%) care processes across 28 hospitals to be degraded after participation in the EPOCH QI programme i.e. a signal toward worse performance associated with activation to the EPOCH intervention. Despite being the 3rd most frequently improved care-process, use of cardiac output monitoring to guide fluid therapy was the most commonly degraded process (10/80 hospitals).

In the sensitivity analyses, using stricter run chart rules would have identified 78/140 care processes as improved in this group, rather than 90/140 using the standard rules, resulting in a group of 10 hospitals, rather than 14 that improved ≥6 care-processes. Regarding different approaches to the chart median, 6 / 140 (4%) of charts in the sensitivity analysis used a median based upon all data points in the original analysis (due to signals in the baseline period). Across the group of most improved hospitals, using a chart centre line based on all data points would have identified 57/140

(41%) care processes as improved, resulting in a group of 3 hospitals, rather than 14, that improved ≥6 care-processes (see supplementary Table 1).

DISCUSSION

The main finding of this analysis was that no hospital in the EPOCH trial reliably implemented the care-pathway within six months of the end of the intervention period. However, we did identify areas of improvement. In total, 279 (of a possible 800) care processes were improved by hospitals through participation in the EPOCH trial and a small group of hospitals (17.5%, 14/80) were successful in improving ≥6 care processes. Effect sizes overall were marginal, but with substantial variance for each process across trial hospitals. We specifically did not seek to evaluate changes in patient outcomes associated with the trial intervention, but it seems logical that if only a small proportion used all the recommended implementation strategies and only a sub-set of these hospitals were able to improve more than half the target care-pathway processes, then the causal mechanism we hypothesised would lead to outcome improvement was largely absent in the EPOCH cohort. This confirms the findings of the main, patient level, trial analysis. This supports the use of individual hospital level time-series analysis, both during a programme to monitor progress and support hospitals facing challenges, and as part of the evaluation strategy to provide granular understanding of cohort-level analyses. We used a prospectively defined run chart methodology, but in a sensitivity analysis we found our findings were sensitive to the use of alternative methods of run chart construction.

This study contributes to the growing literature on methods to better understand improvement and implementation research results in the face of complexity [24,28–30]. In particular, hospitals participating in multi-site cohorts may well achieve differing results; understanding this local level granularity enables a clearer understanding of what happened during a large-scale intervention and what led to, or hindered, overall success [31,32]. In line with evidence that a multi-faceted approach to change is more effective [33], the hospital teams in our study that achieved greater care-process improvement also reported using more of the implementation strategies recommended by the QI programme. Whilst the relationship was absent in the linear model, this approach may be poorly suited to the complexity of this issue, especially across a large and heterogenous cohort. Analysis by groupings, and in particular when comparing the least and most improved hospitals (Figures 5a-c and supplementary file), suggests that greater improvement was possible (but not guaranteed) with use of more of the recommended implementation strategies. This supports the hypothesis that the QI intervention could be effective, but only if used in full and deployed within a supportive context. Our concurrent process evaluation paper describes in detail the contextual factors, both enablers and barriers, faced by hospitals as they attempted improvement [13]. Major barriers included limited time and scarce resources to support clinicians leading improvement and, relatedly, an onerous burden of data collection which limited capacity to subsequently use these data for improvement. Related to

this specific factor, our post-hoc analysis also indicated that in hospitals where systems to collect data prospectively existed, minimising the data burden on NELA and EPOCH QI leads, the number of improved care processes increased. Lack of interest amongst colleagues and seniors was also reported as a problem in many hospitals. Almost universally, contextual enablers were the opposite of these and future improvement programmes will need to fully address these factors to be successful, including allocating job-planned time for frontline improvement leaders and additional funding for support functions such as data collection and analysis.

There are other possible reasons why we did not find greater care pathway implementation or care process change. Firstly, our definition of reliable pathway implementation may have been too stringent. The standards set by NELA only require consultant presence in the operating room and admission to critical care for patients with a ≥5% risk of mortality whereas the aspirational improvement goal of the EPOCH trial was for all patients requiring emergency abdominal surgery to be put onto the recommended care-pathway. Whilst 80% is an accepted threshold for defining minimum reliability [26], it is possible that hospitals were guided by the more pragmatic standards as set by NELA, thus reducing the chances of pathway implementation as defined in this paper. Second, three key care processes (consultant-led decision making, consultant-delivered surgery and consultant-delivered anaesthesia) were already being reliably delivered (to >80% patients) in most hospitals at the start of the intervention period, which may have limited the head-room for further improvement of these particular care-processes in some hospitals. Also, the value of one key process, cardiac output monitoring, was under debate in the UK during the time of the study, [34] and this may have meant some teams chose not to focus on it, or as our data shows, to move away from delivering this process completely. Third, system level care processes, such as reducing the time to get patients into the operating room, were harder to improve than processes that individual clinicians were able to improve by themselves, such as assessing mortality risk. Nevertheless, we did see that nearly 30% of hospitals improved their performance on getting patients to the operating room in a time-frame appropriate for their operative urgency. Considering the complexity of this system-level process, contingent on the actions of multiple stakeholders and on the other pressures faced by operating room suites in the UK, we feel this is a substantial achievement. This mirrors findings from previous QI work regarding the degree of difficulty in attempting to improve systems-level processes compared to more discrete, individual professional or small-team delivered processes [35], and supports the need to consider different, and potentially more intensive, strategies to improve system-level care processes. This may be of particular importance for this patient group given recent evidence

demonstrating the positive impact upon mortality of system level changes such as single pathways of care in emergency general surgery and dedicated emergency surgery units [36].

At the trial level, without these further analyses, the degree to which each hospital had implemented the care-pathway as intended or improved would have remained unclear, as each hospital's signal was obscured within the results of a large and heterogeneous cohort. The use of run-charts to evaluate QI programmes at scale remains rare, with some notable exceptions [37], yet they are ideally suited to this level of granular data analysis. The main strength of this analysis is that we have tested this approach experientially, alongside our main trial analyses, using the same dataset, and found it was largely congruent with, but added value to, our previous understanding of what happened during the EPOCH trial. We mitigated against human error, inherent in the visual inspection of run-charts (present even when using automated data analysis programmes), by undertaking a validation exercise to provide assurance of reliability, which we consider a strength of this work. We also tested different approaches to constructing and analysing run charts in sensitivity analyses and found that the approach used may have a substantial impact on findings. This analysis also had some limitations. Firstly, performance of hospitals in 6 of the 15 trial clusters was analysed using run-charts that had a baseline median constructed of 10 data points, which is the minimum acceptable number to use the probability based run-chart rules [18]. This was due to the trial and the data collection process, via NELA, starting nearly contemporaneously thus limiting baseline data in early clusters. Secondly, analyses requires decisions about the desired trade-off between sensitivity and the risk of false-positive signals being identified [18,25]. Both our sensitivity analyses, using stricter analysis rules and comparing different methods for creating the chart median, reduced the number of process improvements observed. In the latter analysis, this reduction was substantial. If stricter run chart rules, or a chart median based upon all data points, had been applied across all analyses, the level of care-process improvement identified would have been smaller than we found than in our pre-planned analysis. Third, variations in the denominator for the monthly plotted percentages sometimes interfered with signals in the data e.g. in a month with a small denominator, a few process failures may create a data point that breaks a signal that would otherwise indicate a move toward improvement. This, combined with the time-bound nature of the analyses, may have led to some real-world improvements not being identified using the run-charts (i.e. Type 2 errors). Our analysis may therefore ultimately have provided an overly conservative estimate of the volume of improvement associated with the EPOCH intervention. This problem may have been mitigated by using both run charts and Statistical Process Control (SPC) charts in a head to head comparison. Although it would have produced further valuable reflections on various types of time-series chart for evaluation, it was

beyond the scope of these pre-planned analyses to do this. Fourth, some of the analysis of the relationship between care process improvement and fidelity to implementation strategies was undertaken post hoc, as the lack of care-pathway implementation meant we could not complete our pre-planned primary objective. Finally, whilst our process and ethnographic evaluation identified several potential enabling strategies and influences, we were not able to quantify these to explore their relationship directly with process improvement in these analyses. There may therefore be some important missing strategies that we did not include in the original programme theory and were also not evaluated in this study.

CONCLUSION

The EPOCH QI intervention did not achieve reliable care-pathway implementation in any trial hospital but was associated with individual improvement of care processes across the cohort and substantial improvement in a minority of hospitals. Individual hospital performance analysis using time series charts can help granular analysis of data from large, heterogeneous cohorts. This approach allowed us to fully understand changes in the delivery of patient care in response to the EPOCH trial intervention, but findings may be sensitive to the chosen run-chart design. In a longer-term project this understanding may have allowed us to adapt the intervention to be more successful.

Acknowledgements

We would like to thank Dr Alex Fowler for his help with the additional data analysis during the revision of this manuscript.

Funding

The EPOCH trial was funded by the National Institute of Health Research (NIHR) of the United Kingdom (HS&DR - 12/5005/10). The funder had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; or in the preparation, review or approval of the manuscript. TS received a scholarship from the Florence Nightingale Foundation during the data analysis and writing of this manuscript.

Department of Health Disclaimer

The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HS&DR, NIHR, NHS or the Department of Health and Social Care.

Competing interests

TS,CP, RP and GM received grant funding to design, deliver and evaluate the EPOCH trial. RP holds research grants, and has given lectures and/or performed consultancy work for Nestle Health Sciences, BBraun, Medtronic, and Edwards Lifesciences, and is a member of the Associate editorial board of the British Journal of Anaesthesia. D.M, C.J and S.H. received programmed activities for the roles in the NELA Project Team. M.G. received programmed activities for their role in the NELA Project Team, is a medical adviser for Sphere Medical Ltd, and Director of Oxygen Control Systems Ltd, and received an honorarium and travel expenses from Edwards Lifesciences in 2016. TS received a scholarship from the Florence Nightingale Foundation during the data analysis and writing of this manuscript.

Ethics approval

The EPOCH trial was approved by the Research Ethics Committee of the National Health Service (REC reference 13/EM/0415).

Authors' contributions

RP, CP, GM, TS and the EPOCH Trial group all contributed to protocol development and design of the EPOCH trial. DM, MG, SH, DC, SD, CJ and JL supported the EPOCH trial (data acquisition) through the NELA database. TS conceived the idea for this paper, collated the data and led the analysis and

write-up of results. RH contributed to the data analysis. All authors critically revised the content of the manuscript. All authors read and approved the final manuscript.

Patient consent for publication

Not applicable

Availability of data and materials

The datasets analysed during the current study are available from NELA by application. The full data-analysis procedure for this study is available from the corresponding author.

Exclusive licence

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in BMJ Quality & Safety and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

REFERENCES

1. Abbott TEF, Fowler AJ, Dobbs TD, Harrison EM, Gillies MA, Pearse RM. Frequency of surgical treatment and related hospital procedures in the UK: a national ecological study using hospital episode statistics. BJA Br. J. Anaesth. [Internet]. The Author(s); 2017;119:249–57. Available from: http://academic.oup.com/bja/article/119/2/249/4049141/Frequency-of-surgical-treatment-and-related

2. Weiser TG, Haynes AB, Molina G, Lipsitz SR, Esquivel MM, Uribe-leitz T, et al. Surgical Services : Access and Coverage Estimate of the global volume of surgery in 2012 : an. 2012;94305.

3. Pearse RM, Moreno RP, Bauer P, Pelosi P, Metnitz P, Spies C, et al. Mortality after surgery in Europe: A 7 day cohort study. Lancet [Internet]. 2012;380:1059–65. Available from: http://dx.doi.org/10.1016/S0140-6736(12)61148-9

4. The International Surgical Outcomes Study group. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. Br. J. Anaesth. [Internet]. 2016;117:601–9. Available from: http://bja.oxfordjournals.org/lookup/doi/10.1093/bja/aew316

5. Meara JG, Leather AJM, Hagander L, Alkire BC, Alonso N, Ameh EA, et al. Global Surgery 2030: Evidence and solutions for achieving health, welfare, and economic development. Lancet. 2015;386:569–624.

6. Fowler AJ, Abbott T, Prowle J, Pearse RM. The age of patients undergoing surgical procedures in England: An ecological study using Hospital Episode Statistics. Br. J. Surg. 2019;

7. Nepogodiev D, Martin J, Biccard B, Makupe A, Bhangu A. Global burden of postoperative death. Lancet. 2019;393:401.

8. The National Emergency Laparotomy Project Team. The first patient report of the national emergency laparotomy audit [Internet]. R. Coll. Anaesth. 2015. Available from: http://www.nela.org.uk/All-Patient-Reports#pt

9. The National Emergency Laparotomy Project Team. Fourth Patient Report of the National Emergency Laparotomy Audit (NELA). 2018.

10. Vester-Andersen M, Lundstrom LH, Moller MH, Waldau T, Rosenberg J, Moller AM. Mortality and postoperative care pathways after emergency gastrointestinal surgery in 2904 patients: A population-based cohort study. Br. J. Anaesth. 2014;112:860–70.

11. GlobalSurg Collaborative. Mortality of emergency abdominal surgery in high-, middle- and low-income countries. Br. J. Surg. [Internet]. 2016;103:971–88. Available from: http://dx.doi.org/10.1002/bjs.10151
12. Peden CJ, Stephens TJ, Martin G, Kahan B, Thomson A, Rivett K, et al. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery: A stepped-wedge cluster randomised trial. Lancet. 2019;393.
13. 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. [Internet]. 2018;13:142. Available from: https://implementationscience.biomedcentral.com/articles/10.1186/s13012-018-0823-9
14. Stevens DP, Shojania KG. Tell me about the context, and more. BMJ Qual. Saf. 2011;20:557–9.
15. Taylor SL, Dy S, Foy R, Hempel S, McDonald KM, Ovretveit J, et al. What context features might be important determinants of the effectiveness of patient safety practice interventions? BMJ Qual. Saf. 2011;20:611–7.
16. Kaplan HC, Brady PW, Dritz MC, Hooper DK, Linam WM, Froehle CM, et al. The influence of context on quality improvement success in health care: a systematic review of the literature. Milbank Q. [Internet]. 2010;88:500–59. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3037175&tool=pmcentrez&rendertype=abstract
17. Martin GP, Kocman D, Stephens T, Peden CJ, Pearse RM. Pathways to professionalism? Quality improvement, care pathways, and the interplay of standardisation and clinical autonomy. Sociol. Heal. Illn. 2017;39:1314–29.
18. Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. BMJ Qual. Saf. 2011;20:46–51.
19. 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;102:57–66.
20. Dixon-Woods M, McNicol S, Martin G. Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature. BMJ Qual. Saf. [Internet]. 2012;21:876–84. Available from: http://qualitysafety.bmj.com/content/21/10/876
21. Dixon-Woods M, Bosk CL, Aveling EL, Goeschel C, Pronovost PJ. Explaining Michigan: Developing an Ex post theory of a quality improvement program. Milbank Q. 2011;89:167–205.
22. Dixon-Woods M, Leslie M, Tarrant C, Bion J. Explaining Matching Michigan: an ethnographic study of a patient safety program. Implement. Sci. [Internet]. Implementation Science; 2013;8:70.

Available from: http://www.implementationscience.com/content/8/1/70
23. Eveleigh MO, Howes TE, Peden CJ, Cook TM. Estimated costs before, during and after the introduction of the emergency laparotomy pathway quality improvement care (ELPQuIC) bundle. Anaesthesia. 2016;71:1291–5.
24. Braithwaite J, Churruca K, Long JC, Ellis LA, Herkes J. When complexity science meets implementation science : a theoretical and empirical analysis of systems change. BMC Medicine; 2018;1–14.
25. Anhoej J, Olesen AV. Run charts revisited: A simulation study of run chart rules for detection of non-random variation in health care processes. PLoS One. 2014;9:1–13.
26. Resar RK. Making noncatastrophic health care processes reliable: Learning to walk before running in creating high-reliability organizations. Health Serv. Res. 2006;41:1677–89.
27. Anhoej J. Diagnostic value of run chart analysis: Using likelihood ratios to compare run chart rules on simulated data series. PLoS One [Internet]. 2015;10:1–9. Available from: http://dx.doi.org/10.1371/journal.pone.0121349
28. Reed JE, Howe C, Doyle C, Bell D. Simple rules for evidence translation in complex systems: A qualitative study. Int. J. Qual. Heal. care. BMC Medicine; 2018;In Press:1–20.
29. Lawton R, Taylor N, Clay-Williams R, Braithwaite J. Positive deviance: a different approach to achieving patient safety. BMJ Qual. Saf. [Internet]. 2014;23:880–3. Available from: http://qualitysafety.bmj.com/content/early/2014/07/21/bmjqs-2014-003115.full
30. Gabbay RA. A Positive Deviance Approach to Under- standing Key Features to Improving Diabetes Care in the Medical Home. 2013;99–108.
31. Stephens TJ, Bamber JR, Beckingham IJ, Quiney NF, Duncan E, Martin G. Understanding the influences on successful quality improvement in emergency general surgery: learning from the RCS CholeQuIC project. Implement. Sci. 2019;i.
32. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, A-H.S. B. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. New Engl. J. Med. 2009;360:491–9.
33. Johnson MJ, May CR. Promoting professional behaviour change in healthcare: What interventions work, and why? A theory-led overview of systematic reviews. BMJ Open. 2015;5.
34. MacDonald N, Pearse RM, Gillies MA, Hinds C, Grocott MPW, Harrison DA, et al. Effect of a Perioperative, Cardiac Output–Guided Hemodynamic Therapy Algorithm on Outcomes Following Major Gastrointestinal Surgery. Jama. 2014;311:2181.
35. Power M, Tyrrell PJ, Rudd AG, Tully MP, Dalton D, Marshall M, et al. Did a quality improvement collaborative make stroke care better? A cluster randomized trial. Implement. Sci. [Internet]. 2014;9:40. Available from: http://www.implementationscience.com/content/9/1/40

36. Oliver CM, Bassett MG, Poulton TE, Anderson ID, Murray DM, Grocott MP, et al. Organisational factors and mortality after an emergency laparotomy: multilevel analysis of 39 903 National Emergency Laparotomy Audit patients. Br. J. Anaesth. [Internet]. Elsevier Ltd; 2018;121:1346–56. Available from: https://doi.org/10.1016/j.bja.2018.07.040
37. Clarke CM, Cheng T, Reims KG, Steinbock CM, Thumath M, Milligan RS, et al. Implementation of HIV treatment as prevention strategy in 17 Canadian sites : immediate and sustained outcomes from a 35-month Quality Improvement Collaborative. 2015;1–10.
38. NELA Project Team. The Second Patient Report of the National Emergency Laparotomy Audit (NELA). London; 2016.

TABLE 1: EPOCH implementation strategies; desired outcomes, resources and use in individual hospitals

(QuIP, Quality Improvement Programme; VLE, Virtual Learning Environment; NELA, National Emergency Laparotomy Audit)

Desired outcomes	**Proposed implementation strategies**	**QuIP activities and resources**	**Implementation strategy usage during the intervention period – questionnaire items**	**Implementation strategy usage during the intervention period – questionnaire responses**
Motivation for change created amongst stakeholders and improvement goals clearly understood	QI leads hold a stakeholder meeting after activation (**Strategy 1**)	1. Pre-activation checklist (providing guidance for planning of stakeholder meeting) 2. Evidence for QI and need for change provided 3. Presentation on achieving engagement	**Stakeholder meeting** Did you hold a stakeholder meeting as one of your QI activities? E.g. a meeting for all professionals involved in patient care	• 55% (41/75) : Yes • 45% (34/75) : No
Inter-professional collaboration (IPC) fostered	Each hospital to form an inter-professional improvement team (**Strategy 2**)	4. Team approach promoted 5. QI leads encouraged to invite colleague to EPOCH meetings 6. EPOCH VLE open to all local QI team members	**QI team formation** At your site, was a formal team created to work on QI activities related to EPOCH?	• 60% (46/77) : Yes • 27% (21/77) : No • 13% (10/77) : Other (comments indicated informal teams often existed)
Shared view of current performance created ('situational awareness')	QI leads analyse their own data (NELA data +/- case note reviews and local audit data) and feed this back to	7. Case-note review tool 8. Training on data for improvement 9. Training on how to access and analyse NELA data 10. Excel workbook programmed to create run charts from NELA data 11. Secure data sharing site created on VLE	**Data collection and analysis** After starting EPOCH did you or your colleagues download and analyse your local NELA data? If yes, how frequently did you do this?	• 79% (61/77) : Yes • 21% (16/77) : No • 43% (26/61) : Analysing data monthly or bi-monthly

Desired outcomes	Proposed implementation strategies	QuIP activities and resources	Implementation strategy usage during the intervention period – questionnaire items	Implementation strategy usage during the intervention period – questionnaire responses
	colleagues regularly (**Strategy 3**)			• 57% (35/61) : Analysing data less frequently
Frontline teams develop and use basic QI skills to effect change	QI leads and other team members: Use time-series charts ("run-charts") (**Strategy 4**) Use the Plan-Do-Study-Act (PDSA) cycles (**Strategy 5**) Segment the patient pathway **(Strategy 6)**	12. Introduction to QI skills training provided 13. Links to further reading and training resources for QI 14. Telephone and email support	**Run-charts** When analysing data, did you use run-charts? **PDSA approach** Did you or your colleagues use the "Plan Do Study Act" (PDSA) cycle approach during your QI activities? **Pathway segmentation** Please indicate statement most closely fits your hospitals improvement or implementation activity during EPOCH	• 92% (56/61) : Used run-charts to analyse data • 61% (45/74) : Yes, sometimes • 5% (4/74) : Yes, often • 34% (25/74) : No • 22% (17/77) : We introduced a single pathway of care (across Pre, Intra and Post-operative phases) • 32% (25/77) : We introduced separate pathways or care bundles for the peri-op phases • 40% (31/77) : We focused on introducing individual / separate interventions • 5% (4/77) : Other

Table 2: Key characteristics of data set.

Data presented median (IQR), mean (SD) or n (%).

Key Hospital Characteristics [38]	
No. of operating theatres / 100 hospital beds	2.5 (2.1 – 3.0)
No. of surgical critical care beds / 100 beds	2.7 (2.1 – 3.5)
Emergency laparotomy volume (Years 2014/15)	271 (204 – 371)
No. of secondary / tertiary referral hospitals	Secondary 58/80 Tertiary 22/80
Key Patient Characteristics	
Age	68 (13)
Sex – Female	11101 (53%)
P-Possum Score	7.6 (2.9 – 22.7)

Table 3: Process measure improvement per hospital temporally associated with participation in the EPOCH trial QI programme

	Data signals (shifts and runs) identified on run chart analysis			
Care Processes	**Number (%) of hospitals with care process improvement Observed n N=80**	**Number (%) of hospitals with median baseline care process delivery ≥80% n N=80**	**Number (%) of hospitals with degraded care-process after activation to EPOCH n N=80**	**% difference post intervention vs. pre intervention (median, IQR, range)**
1. Consultant led decision making	14 (17.5%)	71 (88.8%)	6 (7.5%)	0.44 (2.53 - 3.35, -16.19 - 20.33)
2. Consultant review of patient before surgery	17 (21.25%)	14 (17.5%)	4 (5%)	2.4109 (-3.67 - 6.37, -18.19 - 17.73)
3. Pre-op risk assessment documented	57 (71.25%)	6 (7.5%)	3 (3.8%)	13.66 (3.25 - 23.48, - 21.75 - 52.15)
4. Time from decision to operate (DTO) to entrance to the operating theatre	14 (17.5%)	2 (2.5%)	5 (6.3%)	*Time in hrs* -0.500 (-1.30 - 0.37, -8.25 - 3.4083)
5. Time to Theatre within NCEPOD timeframe	22 (27.5%)	32 (40%)	2 (2.5%)	8.391 (1.65 - 12.18, -7.81 - 25.65)
6. Consultant delivered surgery	24 (30.0%)	70 (87.5%)	2 (2.5%)	1.913 (-1.96 - 6.52, -13.86 - 18.66)

7. Consultant delivered anaesthesia	29 (36.25%)	57 (71.3%)	4 (5%)	3.8416 (-0.74 - 8.68, -22.948 - 30.30)
8. Cardiac output monitoring to guide fluid therapy	32 (40.0%)	3 (3.8%)	11 (13.8%)	4.766 (-1.10- 13.25, -29.21 - 50.72)
9. Measurement of serum lactate intra-operatively	42 (52.5%)	3 (3.8%)	3 (3.8%)	9.270 (2.14 - 17.52, -28.09 - 39.86)
10. Admission to critical care post-operatively	28 (35.0%)	15 (18.8%)	3 (3.8%)	2.222 (-3.62 - 7.33, -17.69 - 26.88)

Figure 1. The EPOCH trial recommended care-pathway

Legend: SIRS, Systemic Inflammatory Response Syndrome; Sepsis Six, a protocolised treatment for sepsis; CT, Computer-aided Tomography; WHO, World Health Organisation; ABG, Arterial Blood Gas; NMB, Neuro-muscular Blockade; CCOT, Critical Care Outreach Team; NEWS, National Early Warning Score; VTE, Venous Thrombo-embolism

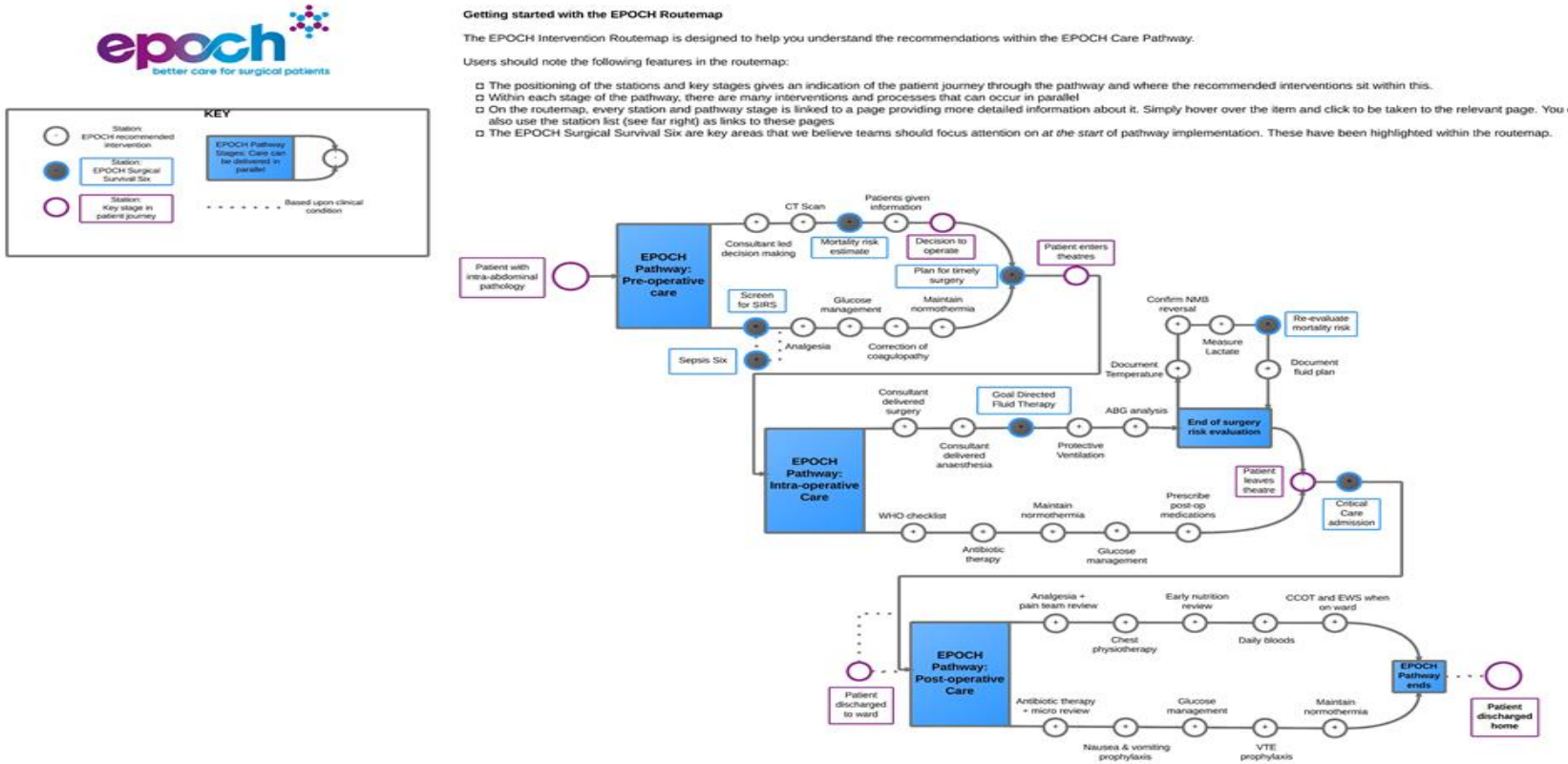

Figure 2. Run chart analysis process and worked example

Legend: NELA = National Emergency Laparotomy Audit / CO = Cardiac Output / QI = Quality Improvement

Explanation of steps in run-chart analysis: 1). Baseline median from month 1 - 11. There are 10 useful data point (excluding 1 point on the median) and 8 runs (groups of data points on either side of the median line). Referring to published guidance, this indicates the baseline performance is exhibiting normal (random) variation. 2). As the baseline period has no signals, baseline median is projected forward as centre-line for chart, against which to assess new data 3). Intervention period runs from month 12 - 27. Two signals of interest: i) number of runs, in this case 5, which when referring to published guidance would indicate a non-random signal and ii) shift, ≥6 data points on one side of the median line, which would also indicate a non-random signal. Therefore, a signal demonstrating an improved process is shown from month 19 onwards. 4) A new median performance is calculated from these data points (71% new median process delivery).

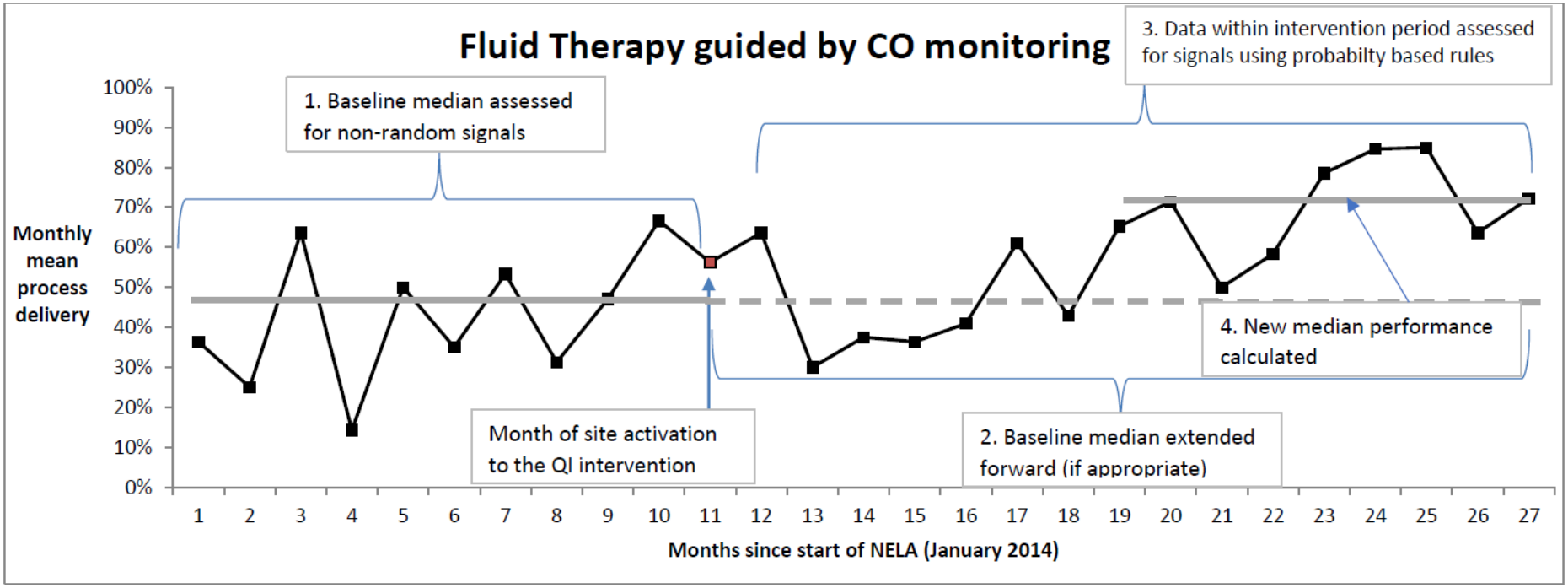

Figure 3. Inclusion of hospitals and patients in the run-chart analysis

Legend: NELA = National Emergency Laparotomy Audit

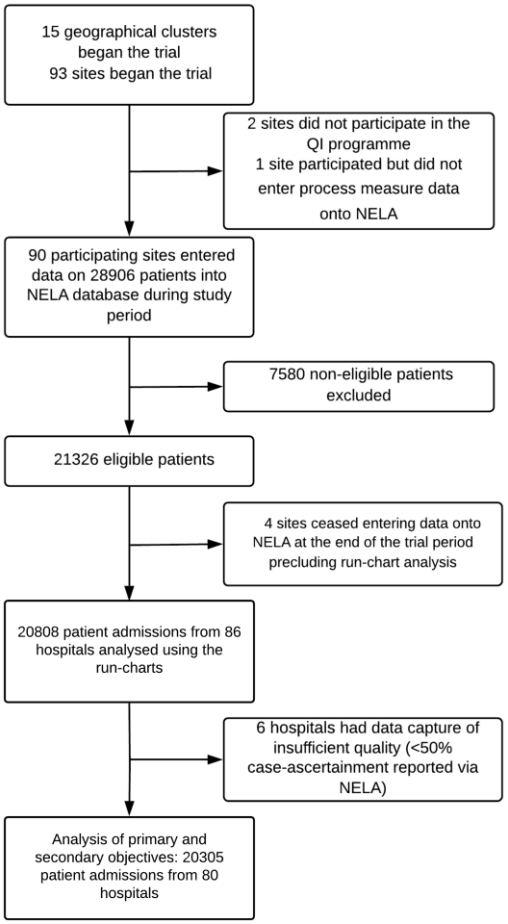

Figure 4. Number of care processes improved by each hospital during EPOCH trial (n= 80 hospitals)

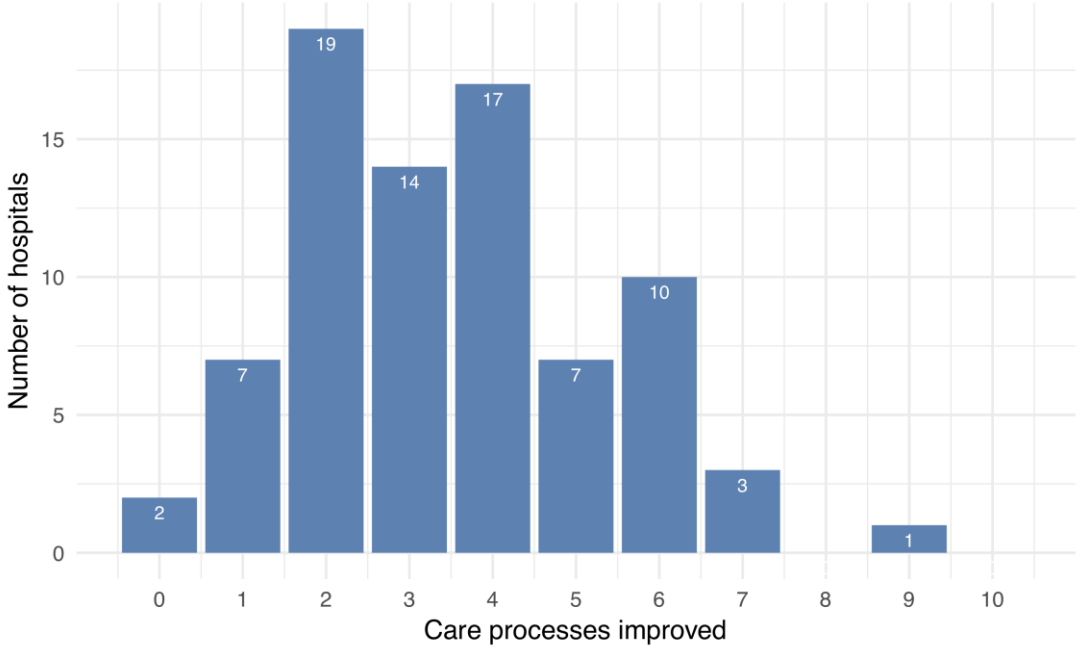

FIGURE 5a-c: FIGURE 5a : Difference in implementation strategy use between least and most improved hospitals, FIGURE 5b: Fidelity to implementation intervention, comparing least and most improved hospitals by strategy usage, FIGURE 5c: Comparison of least and most improved hospitals by NELA data collection process

Legend: NELA = National Emergency Laparotomy Audit / QI = quality improvement /PDSA = Plan do Study Act cycles