OPEN ACCESS

Retrospective evaluation of an intervention based on training sessions to increase the use of control charts in hospitals

Laura Kudrna ,[1] Paul Bird,[2,3] Karla Hemming,[1] Laura Quinn,[1] Kelly Schmidtke,[4] Richard Lilford [1]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjqs-2021-013514).

[1]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[2]University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
[3]West Midlands Academic Health Science Network, Birmingham, UK
[4]Warwick Medical School, University of Warwick, Coventry, UK

Correspondence to
Dr Laura Kudrna, University of Birmingham, Birmingham, UK; l.kudrna@bham.ac.uk

Received 29 April 2021
Accepted 29 March 2022

To cite: Kudrna L, Bird P, Hemming K, *et al*. *BMJ Qual Saf* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjqs-2021-013514

ABSTRACT

Background Statistical process control charts (SPCs) distinguish signal from noise in quality and safety metrics and thus enable resources to be targeted towards the most suitable actions for improving processes and outcomes. Nevertheless, according to a recent study, SPCs are not widely used by hospital boards in England. To address this, an educational training initiative with training sessions lasting less than one and a half days was established to increase uptake of SPCs in board papers. This research evaluated the impact of the training sessions on the inclusion of SPCs in hospital board papers in England.

Methods We used a non-randomised controlled before and after design. Use of SPCs was examined in 40 publicly available board papers across 20 hospitals; 10 intervention hospitals and 10 control hospitals matched using hospital characteristics and time-period. Zero-inflated negative binomial regression models and t-tests compared changes in usage by means of a difference in difference approach.

Results Across the 40 board papers in our sample, we found 6287 charts. Control hospitals had 9/1585 (0.6%) SPCs before the intervention period and 23/1900 (1.2%) after the intervention period, whereas intervention hospitals increased from 89/1235 (7%) before to 328/1567 (21%) after the intervention period; a relative risk ratio of 9 (95% CI 3 to 32). The absolute difference in use of SPCs was 17% (95% CI 6% to 27%) in favour of the intervention group.

Conclusions The results suggest that a scalable educational training initiative to improve use of SPCs within organisations can be effective. Future research could aim to overcome the limitations of observational research with an experimental design or seek to better understand mechanisms, decision-making and patient outcomes.

Key messages

What is already known on this topic

⇒ Statistical process control charts (SPCs) provide a basis for quality management and enable resources to be targeted effectively. Earlier research suggests that many hospital governing bodies, known as hospital boards in England, do not use SPCs.

What this study adds

⇒ An educational initiative with training sessions is ongoing to stimulate the demand for and supply of SPCs. This study reports positive findings of a controlled before and after study on the effectiveness of the intervention using naturally occurring observational data from board meeting papers.

How this study might affect research, practice and/or policy

⇒ Our results were not likely due to a 'rising tide' of greater use of SPCs, which suggests that focused interventions supporting uptake may still be required. Future research should consider mechanisms and use an experimental design.

INTRODUCTION

Rationale for the use of statistical process control charts (SPCs)

The principles underlying statistical process control charts (SPCs) have been fundamental tenets of safety science since they were promoted by Deming and Shewhart in the 1930s.[1 2] Originally developed to drive quality improvement in manufacturing, SPCs are now widely recommended for use in healthcare.[3] A key feature of SPCs is 'process' or 'control' limits (henceforth used interchangeably) that visualise statistical variation from a mean. SPCs thus distinguish signal from noise or, in Deming and Shewhart's original terminology, special cause from common cause variation. As a result, attention can be focused where

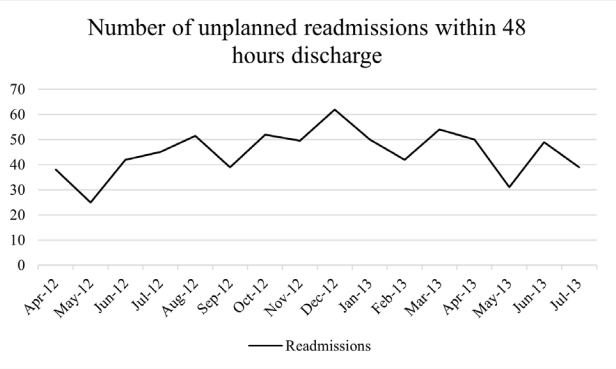


Figure 1 Drawn from real hospital data presented in Schmidtke *et al*.[17] Time series chart showing the number of unplanned readmissions within 48 hours of discharge from April 2012 until July 2013 at a single hospital.

it is needed. Presenting data in SPCs improve the ability of public advisors and hospital decision-makers to make good decisions given variation in the data, for instance, by not over-reacting to variation that is typical for a particular process of care.[4] Examples of charts without and with process limits are shown in figures 1 and 2, respectively. Including process limits can limit the influence of cognitive biases that may otherwise guide decision-making. For example, in 'anchoring bias', human attention anchors on the most extreme and recent data points in a time-series chart, regardless of whether these data lie within common cause variation.[4 5] A recent randomised trial showed that the use of SPCs was associated with fewer adverse surgical outcomes.[6] Thus, omitting information about statistical variation could compromise decision-making about process variation, instigate unnecessary intervention, and, consequently, lead to the inefficient allocation of resources.

Lack of adoption of control chart methods

Hospital boards in the English National Health Service (NHS) are made up of executive and non-executive members who have a duty to assure the quality and safety of services. Board papers, therefore, include charts displaying quality and safety metrics. A previous study investigated the prevalence of SPCs in the documents used by hospital board members (board papers) in England, UK. The findings showed that SPCs are not widely included in hospital board papers in England: in 30 randomly selected English acute care hospitals' quality and safety board papers, nearly half (14/30, 47%) of board papers did not contain any SPCs and only 12% (72 of 589) of the charts across papers were SPCs.[7] Although the inclusion of SPCs in board papers does not necessarily indicate that these charts are being used effectively, it does suggest engagement with aspects of the approach.

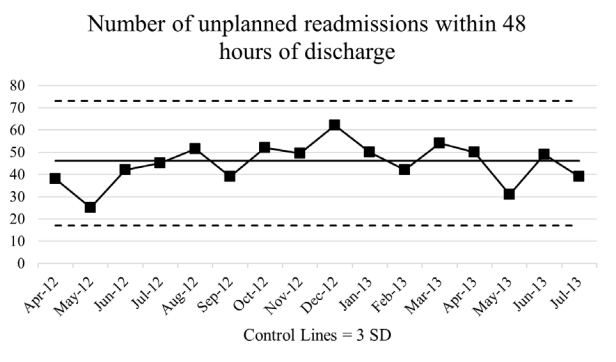


Figure 2 Drawn from real hospital data presented in Schmidtke *et al*.[17] SPC showing the number of unplanned readmissions within 48 hours of discharge from April 2012 until July 2013 at a single hospital. SPC, statistical process control chart.

An intervention to improve use of SPCs in board papers

The above findings underpinned the NHS Improvement/England (NHS I/E) (2019) initiative called 'Making Data Count' that encourages NHS institutions to adopt SPCs.[8] NHS I/E is the organisation responsible for driving up the standard of care in the NHS. The initiative is comprised of educational resources and training sessions which take less than one and a half days to deliver, as described below in the "intervention" section.

Study aims

The research aimed to assess the effect of the Making Data Count training sessions on the appearance of SPCs in publicly available board papers from NHS hospitals and to assess perceptions of the sessions among attendees. We conducted a systematic search for initiatives that aimed to improve use of SPCs for routine surveillance in healthcare. Our search strategy is laid out in figure 3 and discussed in the study protocol (online supplemental file 1). We looked for studies using SPCs in routine surveillance (rather than within an intervention to improve a given process)[9] and found no papers replicating our approach.

METHODS

A study protocol detailing the methods was published on the Open Science Foundation[10] (online supplemental file 1). The SQUIRE reporting guideline checklist[11] was completed (online supplemental file 2).

Context

NHS Improvement delivered Making Data Count training sessions to NHS hospital board members and hospital analysts from November 2017. Hospital recruitment was performed by snowball sampling, where information on the training sessions was disseminated using social media, email and word of mouth.

Intervention

The TIDier checklist[12] was completed (online supplemental file 3). The Making Data Count training

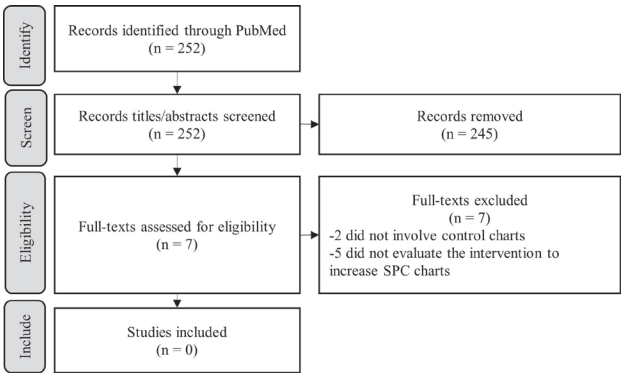


Figure 3 Results of systematic review seeking studies on training interventions to increase the use of SPCs for routine monitoring within institutions. SPC, statistical process control chart.

sessions were delivered to two groups of hospital staff. One group was board members who received sessions lasting around 90 min. The second group were quality improvement staff, including analysts, clinicians and operational staff, and their training took place over one working day. The training sessions aimed to improve knowledge about SPCs and increase their uptake (see online supplemental files 4 and 5 for training PowerPoints). Content included background on SPCs, when and how to use them, how they can be generated and how they can inform decision-making about process variation. Topics included identifying trends, special versus common cause variation and using icons to summarise trends. The limitations of other charts were discussed, and, importantly, each training session was personalised using hospitals' own data. No specific software platform was recommended for creating SPCs, but the training team provided tools in Excel and SQL software that could be adapted by the trainees. If trainees requested further tools, the training team provided details about other organisations that could provide information on other software tools such as Business Objects, Tableau and Qlik.

Study of the intervention

Sample size

Our sample size was based on detecting a 30 percentage-point improvement in the proportion of SPCs from 10% preintervention to 40% postintervention. Given that the effectiveness of the training intervention on patient safety is contingent on changes in the uptake of SPCs in board papers, we believed that at least a 'moderate' effect size[13] would be necessary to stimulate widespread adoption. Assuming 5% significance and 80% power, and assuming a correlation between preintervention and postintervention measures of 0.90 based on a t-test,[14] a minimum of 16 hospitals in total with preintervention and postintervention measures was required (eight in each arm). We included 20 hospitals to err on the side of caution.

Hospital selection

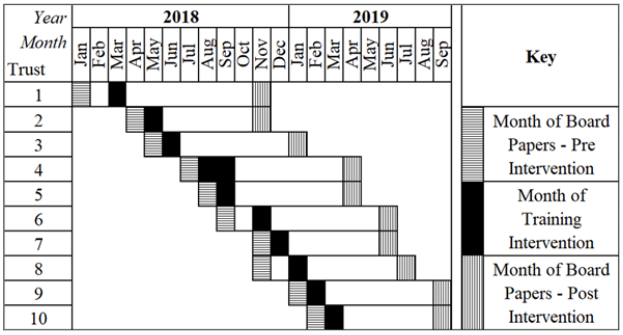
We selected 10 acute care hospitals that received the training after February 2018. To achieve temporal heterogeneity, we sampled one training intervention hospital per month. If more than one hospital received the training intervention in each month, we randomly selected one of the hospitals. We then selected matched control hospitals that had not received the training using the NHS Digital Peer Finder tool.[15] Hospitals were matched on the number of patient attendances, degree of specialisation and deprivation level. Degree of specialisation was defined as the divergence of individual trust Healthcare Resource Group activity profile from the national profile.[15] Deprivation level was obtained from the average 2010 Index of Multiple Deprivation score in Lower Super Output Areas (containing about 1500 people) where the hospitals' patients live.[16] Tiebreaker characteristics were number of full-time equivalent staff, urban location and whether the hospital had been classified as a 'foundation hospital' by NHS authorities.

Board paper selection

For the intervention hospitals, the preintervention board paper was the first paper published at least 1 month before the training intervention. The postintervention board paper was the first board paper published at least 6 months after the intervention. The papers from the control hospitals were selected at the closest month to their matched intervention hospitals. Figure 4 shows the study design with 20 observations for the intervention hospitals (10 preintervention and 10 postintervention) and 20 observations for the matched control hospitals (again 10 preintervention and 10 postintervention), giving a total sample of 40 board papers across 20 hospitals.

Quantitative measures: intervention versus control hospitals

In line with previous research on use of SPCs in board papers,[17] our main outcome measure was the



Notes: The black boxes show the month of the training intervention. The ten cells with horizontal lines are those months where we sample the preintervention board papers from before the training. The ten cells with vertical lines are those months where we sample the postintervention board papers from after the training, which occur at least six months after the training. Trust 4 received trainings in August and September.

Figure 4 Selected board papers for preintervention and postintervention periods, and month of training intervention, for 10 acute hospitals that received 'Making Data Count' training sessions.

proportion of all charts in the board papers made up of SPCs. There were three other outcomes: first, the proportion of quality and safety charts made up of SPCs; second, the proportion of time series charts made up of SPCs and third, the proportion of time series *and* between groups charts made up of SPCs (between group charts include funnel charts that show data between hospitals).

Quantitative measures: examination of SPCs in intervention and control hospitals

We examined SPCs included in board papers of the intervention and control hospitals for inclusion of certain specific factors included in the training for intervention hospitals (see PowerPoint slides in online supplemental file 4). One factor was icons (slide 47) that summarise statistical variation visually using colours and letters that indicate special or common cause variation or indicate performance relative to a target. Another was if the control limits were labelled (slides 32, 34). See online supplemental file 6 for the coding frame.

Quantitative coding

Four independent reviewers (R1, R2, R3, R4) conducted the quantitative coding. In step one, R1 and R2 independently identified charts and classified them according to whether they were quality and safety charts. In step two, R2 removed information regarding the hospital and the board meeting date. In step three, R3 and R4 identified the types of charts and specific elements of SPCs if identified. Any deblinding was reported.

Qualitative measures

The qualitative measures were four questions asked after the training sessions in feedback forms: 'What went well today?', 'What could have been done differently?', 'What are your key takeaways?' and 'Any other comments about today?'. These forms were designed and administered by NHS-I/E and made available to the research team.

Analysis

Hospital characteristics were summarised using means and SD. Inter-rater reliability was calculated using kappa statistics. Information regarding the type of charts and features of SPCs (online supplemental file 6) was summarised using counts and proportions. Next, we examined the effect of the training intervention on the main outcomes. For all hospitals, we first summarised the number of SPCs (outcome), the total number of charts and the proportion of SPCs out of all charts. The difference in the proportion of SPCs between preintervention and postintervention was computed for each hospital. This information was stratified by intervention and control hospitals, compared using a t-test and represented as a difference in difference with 95% CI.

To determine the relative effect (risk ratio) of the intervention, we fit a cluster-level analysis using zero-inflated negative Binomial regression model (as outcome data contain a high number of zero counts and there was overdispersion), with the outcome the number of SPCs in the postintervention period, fixed categorical effects for the intervention, the proportion of SPCs in preintervention period and an exposure of all charts in the postintervention period.

In sensitivity analyses (see online supplemental file 7), we explored other models. The analyses presented as our primary analysis (zero-inflated negative Binomial) differed to that planned (Poisson) due to many hospitals having no SPCs (high number of zero counts).

For the qualitative responses, a thematic analysis was conducted to identify barriers to and facilitators of using SPCs.[18] We used an inductive, semantic and (critical) realist approach. One researcher coded each response into the main theme present in the data. These were reviewed by a second researcher who discussed the codes with the first researcher.

RESULTS

Hospital characteristics

Information about the 20 hospitals from the NHS Digital Peer Finder Tool[15] at baseline is summarised in table 1. On average, there were slightly more patient attendances per year in the intervention hospitals (1.7 mil, SD=0.5 mil) than in the matched control hospitals (1.3 mil, SD=0.75 mil). The degree of specialisation score was lower on average in the intervention group (83 739, SD=80 639) than in the matched control group (138 747, SD=135 068). The average 2010 Index of Multiple Deprivation was similar, at 24 (SD=7) in the intervention and 23 (SD=5) in the matched control sample.

Table 1 Hospital characteristics. means with SD in parentheses

	Intervention	Matched control	Overall
	N=10	N=10	N=20
Attendances	1 167 058 (506 825)	1 341 442 (750 439)	1 254 250 (646 233)
Degree of specialisation	83 739 (80 639)	138 747 (135 068)	105 623 (113 366)
Deprivation	24 (7)	23 (5)	23 (6)

Further details available from NHS Digital Per Finder Tool.[15]

Table 2 Chart characteristics (all charts)

Type of chart	All charts (n=6287) n (%)
Quality and safety chart	3003 (47.7)
Time series, between group or both	6287 (100)
Time series only	4741 (75.4)
Between group only	640 (10.2)
Time series and between group	906 (14.4)

Further details available in online supplemental file 6.

Inter-rater reliability and blinding

Percentage agreement was 99.6% (Cohen's k=0.97) for SPCs, 98.5% (Cohen's k=0.94) for time series charts, 89.0% (Cohen's k=0.61) for time series and between group charts, and 89.9% (Cohen's k=0.80) for quality and safety charts. In no cases was a rater 'de-blinded' such that they could discern whether a board paper arose before or after the salient intervention period. There were 12 images referred to the chief project investor because it was unclear whether they were charts (eg, the resolution may have been too poor to tell) and agreement on the appropriate decision was reached in all cases.

Chart characteristics for all charts in intervention and control hospitals

There were 6318 charts identified. However, 31 were either educational SPCs with example data, illustrative data not about the hospital, or they were icons without any data. These charts were removed from the analyses. After excluding these charts, 6287 charts were retained for analyses (see table 2). Nearly one-half of charts (3003/6287, 48%) were quality and safety charts. Time series charts were more common (4741/6287, 75%) than between group charts (640/6287, 10%) and 906/6287 (14%) charts were comprised of both time series and between group presentations (combined). Of all 6287 charts, 449 (7%) were SPCs. Of the 449 SPCs, 63/449 (14%) had a summary icon displayed on them, and the control limits were labelled for 342/449 (76%) of the SPCs. For most charts with labelled limits (191/342, 56%), the label was UCL ('upper confidence limit') or LCL ('lower confidence limit') rather than specifying where the limit was set (see online supplemental file 6 for further description of the SPCs).

Effects of training intervention (intervention versus control hospitals)

All charts

The raw numbers and proportions of SPCs used by group (control and intervention), hospital and time-period (preintervention and postintervention) for all charts are shown in table 3 and figure 5. On average in the control group, there was very little change in use of SPCs from before (9/1585, 0.6%) to after (23/1900, 1.2%) the intervention period (average difference 0%, 95% CI −2% to 2%). In the training intervention

Table 3 SPC usage by group, hospital and period (all charts)

Control group				Intervention group			
	Preintervention	Postintervention	Post-Pre		Preintervention	Postintervention	Post-Pre
Hospital	SPC/chart (%)	SPC/chart (%)	% difference	Hospital	SPC/chart (%)	SPC/chart (%)	% difference
1	2/62 (3)	0/81 (0)	−3	11	1/206 (0)	9/225 (4)	4
2	0/87 (0)	0/127 (0)	0	12	0/149 (0)	0/131 (0)	0
3	0/13 (0)	2/119 (2)	2	13	0/123 (0)	0/84 (0)	0
4	0/643 (0)	0/687 (0)	0	14	3/140 (2)	91/256 (36)	34
5	0/158 (0)	0/170 (0)	0	15	52/116 (45)	47/67 (70)	25
6	0/101 (0)	15/179 (8)	8	16	0/70 (0)	58/81 (72)	72
7	0/157 (0)	0/151 (0)	0	17	0/18 (0)	27/67 (40)	40
8	0/104 (0)	0/101 (0)	0	18	18/176 (10)	42/457 (9)	−1
9	2/153 (1)	6/200 (3)	2	19	0/89 (0)	27/86 (31)	31
10	5/107 (5)	0/85 (0)	−5	20	15/148 (10)	27/113 (24)	14
Total	9/1585 (0.6)	23/1900 (1.2)	0.6	Total	89/1235 (7)	328/1567 (21)	14
Average difference in control group (95% CI)			0 (−2 to 2)	Average difference in intervention group (95% CI)			22 (2 to 42)
				Average difference between intervention and control group* (95% CI)			17 (6 to 27)
				Average relative change between intervention and control group† (95% CI)			9 (3 to 32)

For each hospital in preintervention and postintervention periods, the number of SPCs, the number of all charts and percentage of SPCs out of all charts are reported.
*T-test comparing average difference in proportions between intervention and control group. Percentage difference and 95% CI are reported.
†Zero-inflated negative Binomial regression models. Outcome is number of SPCs in postintervention period, adjusting for preintervention proportion of SPCs . Exposure is all charts. Risk ratios and 95% CI are reported.
SPC, statistical process control chart.

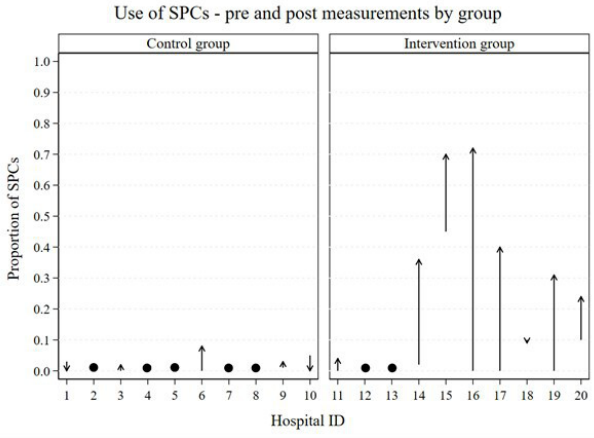


Figure 5 Use of SPCs—premeasurements and postmeasurements by group. SPC, statistical process control chart.

group, use of SPCs increased from 89/1235 (7%) to 328/1567 (21%), and the average difference was 22% (95% CI 2% to 42%). On average, the absolute difference in use of SPCs was 17% (95% CI 6% to 27%) higher in the intervention group compared with the control group. Use of SPCs in the postintervention period was nine times higher (95% CI 3 to 32) in the intervention group compared with the control group, adjusting for the preintervention (baseline) proportion of SPCs.

Subset of quality and safety charts only

As planned, we carried out an analysis restricted to quality and safety charts. The raw number and proportions of SPCs used by group (control, intervention), hospital, and time-period (preintervention versus postintervention) for quality and safety charts are shown in table 4. In the control group, there was very little change in use of SPCs before (7/657, 1%) to after (12/741, 2%) the training intervention period (average difference 0%, 95% CI −3% to 4%). In the training intervention group, use of SPCs was 71/684 (10%) before and 213/921 (23%) after the training, and the average difference was 21% (95% CI 0% to 42%). On average, the difference in use of SPCs was 18% (95% CI 7% to 29%) higher in the intervention group compared with the control group. In model-based analyses, use of SPCs in the postintervention period was nine times higher (95% CI 2 to 41) in the intervention group compared with the control group.

Subset of time series charts

Further analyses regarding changing the exposures to time series charts and between group charts are reported in online supplemental file 7, Tables S7-2 and S7-3. For the model with the time series chart exposure, the results were broadly similar to the main analysis.

Table 4 SPC usage by group, hospital and period (planned subgroup analysis—quality and safety charts only)

Control group				Intervention group			
	Preintervention	Postintervention	Post-Pre		Preintervention	Postintervention	Post-Pre
Hospital	SPC/chart (%)	SPC/chart (%)	% difference	Hospital	SPC/chart (%)	SPC/chart (%)	% difference
1	2/23 (9)	0/16 (0)	-9	11	1/130 (0)	3/125 (2)	2
2	0/56 (0)	0/95 (0)	0	12	0/87 (0)	0/71 (0)	0
3	0/13 (0)	2/26 (8)	8	13	0/38 (0)	0/29 (0)	0
4	0/189 (0)	0/198 (0)	0	14	3/95 (3)	49/152 (32)	29
5	0/80 (0)	0/86 (0)	0	15	37/70 (53)	33/47 (70)	17
6	0/50 (0)	9/98 (9)	9	16	0/47 (0)	26/41 (63)	63
7	0/86 (0)	0/86 (0)	0	17	0/11 (0)	25/48 (52)	52
8	0/60 (0)	0/52 (0)	0	18	16/74 (22)	35/285 (12)	−10
9	0/40 (0)	1/44 (2)	2	19	0/50 (0)	19/46 (41)	41
10	5/60 (8)	0/40 (0)	−8	20	14/82 (17)	23/77 (30)	13
Total	7/657 (1)	12/741 (2)	1	Total	71/684 (10)	213/921 (23)	13
Average difference in control group (95% CI)			0 (−3 to 4)	Average difference in intervention group (95% CI)			21 (0 to 42)
				Average difference between intervention and control group* (95% CI)			18 (7 to 29)
				Average relative change between intervention and control group† (95% CI)			9 (2 to 41)

For each hospital in preintervention and postintervention periods, the number of SPCs, the number of all charts and percentage of SPCs out of all charts are reported. Subgroup analysis safety and quality charts only.
*T-test comparing average difference in proportions between intervention and control group. Percentage difference and 95% CI are reported.
†Zero-inflated negative Binomial regression models. Outcome is number of SPCs in postintervention period, adjusting for preintervention proportion of SPCs . Exposure is all charts. Risk ratios and 95% CI are reported. Subgroup analysis safety and quality charts only.
SPC, statistical process control.

Subset of time series and between group charts

For the model with the times series and between group exposure, the average difference in use of SPCs was 10% (95% CI 0% to 20%) higher in the intervention group compared with the control group. The zero-inflated negative binomial model did not converge for these data, possibly due to the high number of zero cells in the outcome (37/40 observations).

Thematic analysis of qualitative data

Written responses from the feedback forms were available for 7 out of 10 hospitals in the training intervention sample, including two hospitals that increased the SPCs in board papers by less than 10%. Most comments consisted of a few words or one sentence. The main themes relating to responses to the question about what went well were the general format, content and delivery of the training (n=21/66), such as 'Topic relevant and timely'; practical and personal examples that use own hospitals' data (n=19/66), such as 'trust (hospital) data brought it alive'; conversation, discussion and interaction (n=10/66), such as 'interactive opportunity to discuss examples'; formatting, use and insights (n=10/66), such as 'good explanation of SPC rules' and other general comments (n=6/66).

The question about what could have been done differently during the training elicited fewer responses overall (n=32) than did the question about what went well (n=66); this was true across hospitals, including those that changed their use of SPCs both more and less than 10%. The main themes relating to what could have been done differently were the session format (n=15/32), such as 'more time for discussion' and 'break out into groups'; no suggestions for doing anything differently (5/32); the training content (4/32), such as having a 'technical supplement' and 'more on the calculation of control limits' and requests for more examples using own hospital data (3/32), providing handouts (3/32) and other (2/32).

Most participants mentioned awareness of SPCs themselves as a key takeaway (n=29/70). Others commented on the general use of SPCs (n=23/70), such as trend lines, tools and templates, and understanding 'how poor presentation can lead to poor decisions'. Several participants commented that the training changed how they interpret data (n=6/70), intend to report data (6/70) or generally think about data and reporting (4/70). The other comments (n=2/70) were about encouraging others and timelines for implementation.

Finally, when asked for any other comments, most participants made generally positive comments on the training (25/26). Only one (1/26) participant suggested that 'next steps are important', which may reference the need to consider implementation steps in training.

DISCUSSION

Summary of main results

This study investigated whether an educational training intervention increased the use of SPCs in NHS hospitals. We studied the board papers of 10 hospitals that received the training before and after the intervention, along with those from 10 control hospitals that did not receive training over the same time-period. The results showed that most hospitals increased the proportion of SPCs in their board papers after the training intervention, while there was almost no change in the proportion of SPCs among the controls. In model-based analyses, trained hospitals increased their uptake nine-fold relative to controls. The intervention consisted of a day of training for quality improvement staff and 90 min for board members. As this is not a highly intensive intervention, it should be scalable across most contexts.

Interpretation of main results

Interpretation with reference to prior literature

These results are important for several reasons. First, many hospitals do not depict statistical variation in the documents used to inform decision-making about process variation.[7] Second, the use of SPCs enables management's recommendations to align with statistical findings.[4] A recent trial in France found that surgical departments using SPCs had better patient outcomes than controls. Notably, the French intervention appeared more intensive than the training intervention that we evaluated. It provided departments with SPCs from publicly available data, encouraged structured meetings and supplied logbooks for completion. These activities were all in addition to 3 days of training.[6] Our results suggest that a simpler approach can effect change in the prevalence of charts in board papers, although it is a matter of opinion as to whether the change in the hospitals that improved was sufficient to influence improvements in processes and outcomes. Evidence on generalisable mechanisms linking the appearance of charts to quality improvement would more fully inform such opinions, such as perceptions of decisions taken based on the charts and hospital culture.

Interpretation of heterogeneity of the results

Improvement was not uniform across intervention hospitals. The qualitative data do not explain why some hospitals improved but not others, as nearly all respondents reported positive perceptions of the training—including in hospitals that did not change their use of SPCs in board papers. However, these positive responses may have been shaped by social desirability bias.[19] Moreover, some respondents requested more information, including a technical supplement and more on calculating control limits, suggesting that not all training needs had been fulfilled and further sessions or re-engagement may be required.

Interpretation of proportional changes

There are several mechanisms by which the proportional changes in this study could be brought about. As intended, many charts that were previously not produced as SPCs could be transformed into SPCs. However, the total number of charts in the denominator could have decreased because of the intervention, thereby exaggerating improvement in the proportion of hospitals using SPCs (see detailed discussion in online supplemental file 8). Note that this mechanism is possible even in a randomised trial, as the intervention could have prompted changes in the number of charts presented to boards. On balance, we interpret our results as supporting the increased adoption of SPCs while acknowledging the alternative mechanisms. We also note that there is no agreed proportion of SPCs in board papers that would indicate sufficient usage after training, and the need for SPCs could vary by context as topics of concern may change over time.

Issues related to the presentation of SPCs in board papers

The presentation of SPCs could be further improved. Nearly half of SPCs did not state where the control limit had been set, either not mentioning the limit or simply recording 'UCL' and 'LCL' without specifying the limit (eg, three SD). Without labels on limits, the degree of uncertainty that they represent is unclear. We did not compare the labelling and limits of intervention and control hospitals due to the small number of identified SPCs.

Issues related to the implementation of SPCs in hospitals

The use of SPCs takes place within broader organisational contexts. It is possible that SPCs are not included in board papers but are used elsewhere—such as in quality and safety subcommittees. We believe this is unlikely given the explicit quality assurance function of hospital boards. Training alone may be insufficient to encourage adoption of SPCs if the organisational context is not supportive. Importantly, SPC usage is not a sufficient condition for improvement, just as checklists cannot, by themselves, effect safe practice.[20] There must be a supportive implementation context: a team of analysts to create the charts, board members who view and interpret charts, managers who discuss and act on the information presented in the chart and staff at the front line. SPCs are but one element in a chain of events influencing the safety and quality of patient care.

Limitations

Limitations of our study

Our research design does not fully permit a causal interpretation of the results. However, the use of contemporaneous controls showed that our results are not likely due to a 'rising tide' of greater use of SPCs among all NHS hospitals.[21] Although control hospitals were selected to be as similar as possible to intervention hospitals, clear differences were observed at baseline, including in use of SPCs (Hospitals 15, 18, 20). We adjusted for observed differences between hospitals and the before and after design allows us to adjust for differences in baseline rates of the outcome variable (use of SPCs). However, especially given baseline differences, we must suspect unobserved confounders; for example, the intervention hospitals might have been more motivated to change in response to the training.

Limitations of research in the area

Future research should consider an investigation that randomly assigns hospitals to training interventions to balance these factors between groups. Other investigations might also research effects for other forms of hospitals, such as mental health or community care hospitals, to explore generalisability. Studies could explore which aspects of the training are effective, such as the personalisation element, trainers themselves and trainees' understanding and confidence.[20] Importantly, the causal chain linking the prevalence of charts in board papers to patient outcomes should be evidenced, including by qualitatively understanding decision-making related to patient care.

Limitations of qualitative research

A limitation of our qualitative data is that it came from feedback solicited only shortly after the intervention, which restricts the investigation of mechanisms like confidence in the longer term.

CONCLUSION

Certainly, not all the charts within board papers could or should be SPCs. SPCs are not a panacea for understanding data related to all quality improvement issues. However, the high proportion charts with time series information in the board papers (90%), combined with lack of use of SPCs, suggests substantial scope to better visualise chance variation in the data presented to decision-makers. Our results suggest that educational training initiatives may bolster progress towards this aim.

Twitter Richard Lilford @rjlilford

Acknowledgements Thank you to Alan Girling for providing feedback on the overall research design, Bernard Crump for providing feedback on the coding frame for the charts in the board papers, Beth Ann Brackett for her input, the public contributors for their feedback and Sam Riley and Karen Hayllar from NHS-Improvement/England for collaborating and sharing their training resources.

Contributors LK wrote initial draft of protocol and manuscript, revised subsequent drafts of protocol and manuscript and attended NHS-I/E SPC training. PB attended NHS-I/E SPC training and commented on drafts of the protocol and manuscript. KS contributed to protocol development, conducted the literature review and informed and revised the content of the protocol and manuscript. LQ and KH

contributed to the design of the study, developed the statistical analysis plan, commented on drafts of the protocol, wrote the results section of the manuscript and revised manuscript drafts. RL initiated idea, commented on drafts of protocol and manuscript and provided conceptual guidance. All authors have read and approved the manuscript. LK acts as guarantor.

Funding The current project was supported by the National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) West Midlands, grant number NIHR200165.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This research was approved by the University of Warwick Biomedical and Scientific Research Ethics Committee (BSREC 116/18-19). Data come from publicly available hospital board meeting paper documents. No data can identify individual patients.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iDs
Laura Kudrna http://orcid.org/0000-0002-8163-7112
Richard Lilford http://orcid.org/0000-0002-0634-984X

REFERENCES

1 Dekker S. *Foundations of safety science : a century of understanding accidents and disasters*. Boca Raton, Florida: Taylor & Francis Group, CRC Press, 2019.
2 Shewhart WA. *Statistical methods from the viewpoint of quality control. Dover 1986 edition with Foreword by W. Edwards Deming, originally published 1939*. Mineola, N.Y.: Dover Publications, 1986.
3 Mohammed MA, Cheng KK, Rouse A, *et al*. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001;357:463–7.
4 Schmidtke KA, Watson DG, Vlaev I. The use of control charts by laypeople and hospital decision-makers for guiding decision making. *Q J Exp Psychol* 2017;70:1114–28.
5 Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science* 1974;185:1124–31.
6 Duclos A, Chollet F, Pascal L, *et al*. Effect of monitoring surgical outcomes using control charts to reduce major adverse events in patients: cluster randomised trial. *BMJ* 2020;371:m3840.
7 Bird P. Understanding the right and wrong time for intervention. *Health Serv J* 2017 https://www.hsj.co.uk/quality-and-performance/understanding-the-right-and-wrong-time-for-intervention/7015608.article
8 NHS England. Making data count. Available: https://www.england.nhs.uk/publication/making-data-count/ [Accessed 29 Jan 2021].
9 Juran JM. The quality trilogy. A universal approach management fot quality. *Qual Prog* 1986;19:19–24.
10 Kudrna L, Bird P, Hemming K, *et al*. Protocol for a retrospective evaluation of an intervention based on training sessions to increase the use of control charts in the NHS 2021.
11 Ogrinc G, Davies L, Goodman D, *et al*. Squire 2.0 (standards for quality improvement reporting excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf* 2016;25:986–92.
12 Hoffmann TC, Glasziou PP, Boutron I, *et al*. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687.
13 Lilford RJ, Chilton PJ, Hemming K, *et al*. Evaluating policy and service interventions: framework to guide selection and interpretation of study end points. *BMJ* 2010;341:c4413–20.
14 Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med* 1992;11:1685–704.
15 NHS Digital. Nhs trust peer finder tool. Available: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/innovative-uses-of-data/multi-dataset-analysis/nhs-trust-peer-finder-tool [Accessed 29 Jan 2021].
16 Department for Communities and Local Government. The English indices of deprivation, 2010. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/6320/1870718.pdf [Accessed 8 Nov 2021].
17 Schmidtke KA, Poots AJ, Carpio J, *et al*. Considering chance in quality and safety performance measures: an analysis of performance reports by boards in English NHS trusts. *BMJ Qual Saf* 2017;26:61–9.
18 Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? *Int J Qual Stud Health Well-being* 2014;9:26152.
19 Groves R JFF, Couper M, *et al*. *Survey methodology*. Hoboken, N.J.: John Wiley & Sons, 2009.
20 Dixon-Woods M, Bosk CL, Aveling EL, *et al*. Explaining Michigan: developing an ex post theory of a quality improvement program. *Milbank Q* 2011;89:167–205.
21 Chen Y-F, Hemming K, Stevens AJ, *et al*. Secular trends and evaluation of complex interventions: the rising tide phenomenon. *BMJ Qual Saf* 2016;25:303–10.

Supplementary 1 - Protocol

Protocol for a retrospective evaluation of an intervention based on training sessions to increase the use of control charts in the NHS

Supplementary 1 - Protocol

Abstract

Background. Hospital board members use data to inform their decision-making. The way that these data are presented can impact whether hospital board members choose to intervene. Existing research shows that the process limits on statistical process control (SPC) charts improve the ability of hospital decision-makers to align their investigative recommendations with statistical findings. Yet SPCs are not widely used within the UK National Health Services (NHS). An educational training initiative called 'Making Data Count' was established by NHS Improvement/England (NHS-I/E) to improve the uptake of SPC charts in NHS institutions. The present research will evaluate the impact of NHS-I/E training sessions on SPC chart usage.

Methods. A controlled before and after design. SPC chart usage will be examined in a sample of 40 board papers across 20 trusts. The sample will include 20 board papers across ten trusts that have completed the training intervention (ten pre and ten post intervention), and 20 board papers across ten different trusts that have not completed the training intervention that will be external controls. These external controls will be matched according to trust characteristics and time period. Poisson regression will be used to compare rates of control chart usage pre and post intervention, and between the intervention and control groups, using a difference in difference approach. Qualitative thematic analysis of feedback forms will be conducted.

Discussion. The present research will evaluate the impact of NHS-I/E training sessions on the use of SPC charts by examining whether SPC charts appear in NHS trust board papers before and after trainings. The results will contribute to our understanding of whether and why educational initiatives are effective in changing how data are used within healthcare settings.

Supplementary 1 - Protocol

Background

Problem description

Consider the following scenario: you are on the board of an NHS trust and have just received new data showing that average waiting times increased last week. Although you have not yet exceeded the national target for waiting times, you are inching ever-closer. You are uneasy. You do not want to be in breach of the target, but you are not sure that the increase from last week is meaningful enough to take any action. What steps would you take in order to decide whether the increase is meaningful?

While there are many aspects of this scenario that you could investigate, such as how far you are from the target and whether there have been any clinic cancellations recently, an important consideration is whether last week's increase is due to chance. In other words, is the variation within the bounds of what would be expected due to random fluctuations in the data that naturally occur over time? Despite the importance of this question, the data presented to boards do not always contain sufficient information for board members to consider how chance influences key indicators over time (Schmidtke et al., 2017). Omitting the role of chance could lead to sub-optimal decision-making and, consequently, inefficient allocation of resources. Adverse consequences might manifest through unnecessary intervention for a metric that has been incorrectly interpreted as deteriorating performance when it is in fact expected (or 'common-cause') variation.

Available knowledge

In the United Kingdom, the term 'trusts' refers to organisations within the National Health Service (NHS) that provide healthcare services. These trusts have boards that are comprised of executive and non-executive members who collaboratively review documents and make decisions about ongoing performance. The documents associated with these meetings are published as publicly available 'board papers', which contain text and charts. Some of the charts are statistical process control (SPC) charts, whereas others are not SPC charts.

Historically, SPCs charts were first developed for the manufacturing industry and their use in the health sector is widely recommended (Mohammed, Cheng, Rouse & Marshall, 2001).

Supplementary 1 - Protocol

SPC charts can help decision-makers consider the role of chance by displaying 'process limits' that depict statistically informed thresholds, such as how far away a data point is from the mean. Examples of charts without and with process limits are shown in Figures 1 and 2, respectively. These are fictitious and stylised charts displaying 'diagnostic assessment compliance' rates for a disease from April 2016 to October 2017.

Figure 1: Run chart showing monthly changes in diagnostic assessment compliance – without process limits (stylised example)

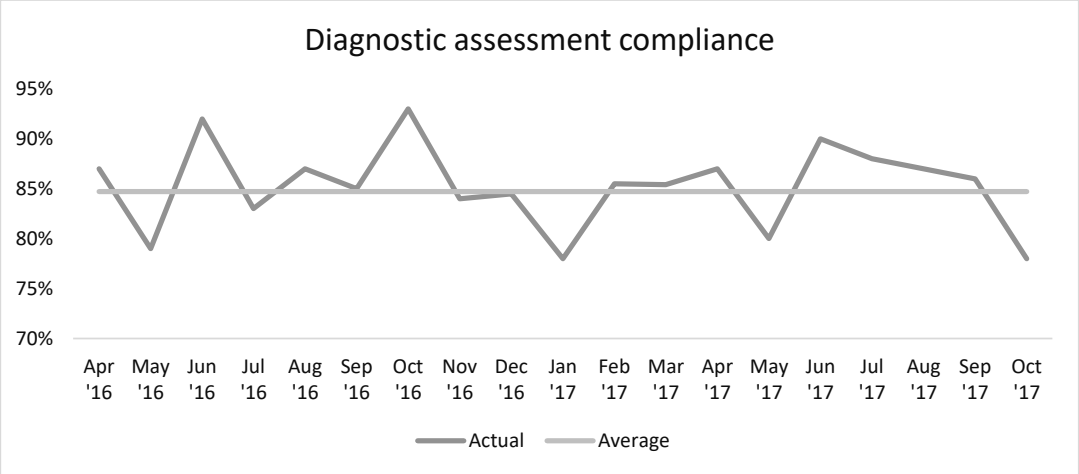
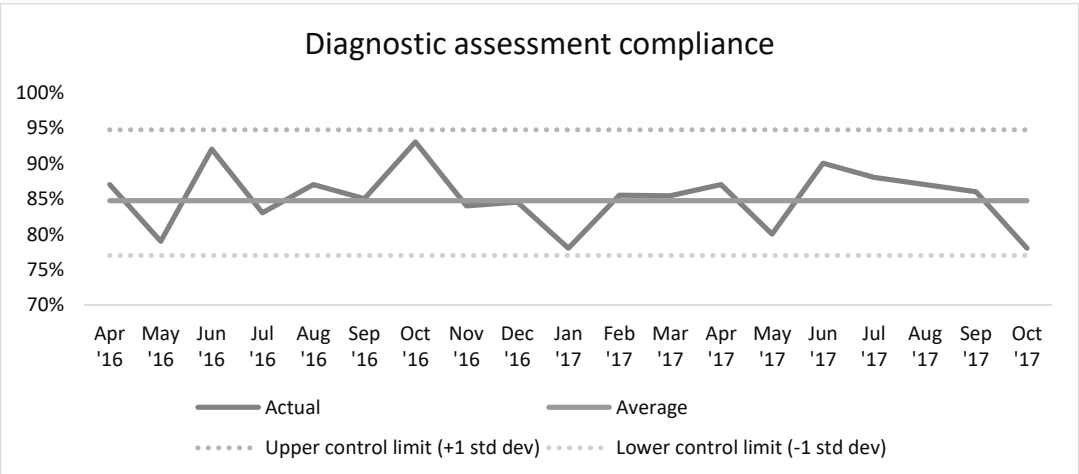


Figure 2: Statistical process control chart showing monthly changes in diagnostic assessment compliance – with process limits (stylised example)



In Figure 1, where the data do not have process limits, it is difficult to ascertain whether monthly compliance rates that are above and/or below the mean are departures from natural

variation over time. In Figure 2, with process limits, it is possible to see that the variations can be predicted by chance, at least within the specified process limits displayed as dashed lines. Further examples of SPC charts are contained in Appendices A and B (available here, https://bit.ly/3j0N4Iu), which are discussed in more detail in the Methods.

Despite recommendations to use SPC charts to monitor performance measures, SPC charts are still sparsely used in healthcare (Schmidtke et al., 2017). Other data presentation methods that do not include the role of chance are prevalent, such as R-A-G charts. R-A-G charts are typically tables of data colour coded to indicate whether data fail to meet a specific target (*r*ed), are in danger of not meeting that target (*a*mber), or are achieving and meeting that target (*g*reen). These targets are seldom informed[1] by the data, and, therefore, are not always well suited to guide quality improvement (Anhøj & Hellesøe, 2017). In contrast, the process limits in SPC charts are data-driven, such as two or three sigma or standard deviations above or below the mean (Wheeler, 2013).

SPC charts can improve people's abilities to identify outliers and align their investigative recommendations with statistical findings (Schmidtke, Watson & Vlaev, 2016). One of the reasons that incorporating process limits into run charts assists with interpreting the data is that they make sample size more salient, thus mitigating a cognitive bias called 'base-rate neglect' (Tversky & Kahneman, 1974; Schmidtke et al, 2017). However, whether SPC charts improve decision-making through automatic or meaningfully reflective cognitive processes may depend on various factors, including what other information is presented in the chart. One factor may be whether the chart includes a label describing where the process limits are set, such as the use of one standard deviation in Figure 2. Labelling enables decision-makers to more accurately understand what it means if data are outside the control limits. Without these limits, decision-makers choices may still align with statistical recommendations, but only in an automatic cognitive capacity brought about by what the chart dictates as a statistical aberration using the r-a-g method.[2]

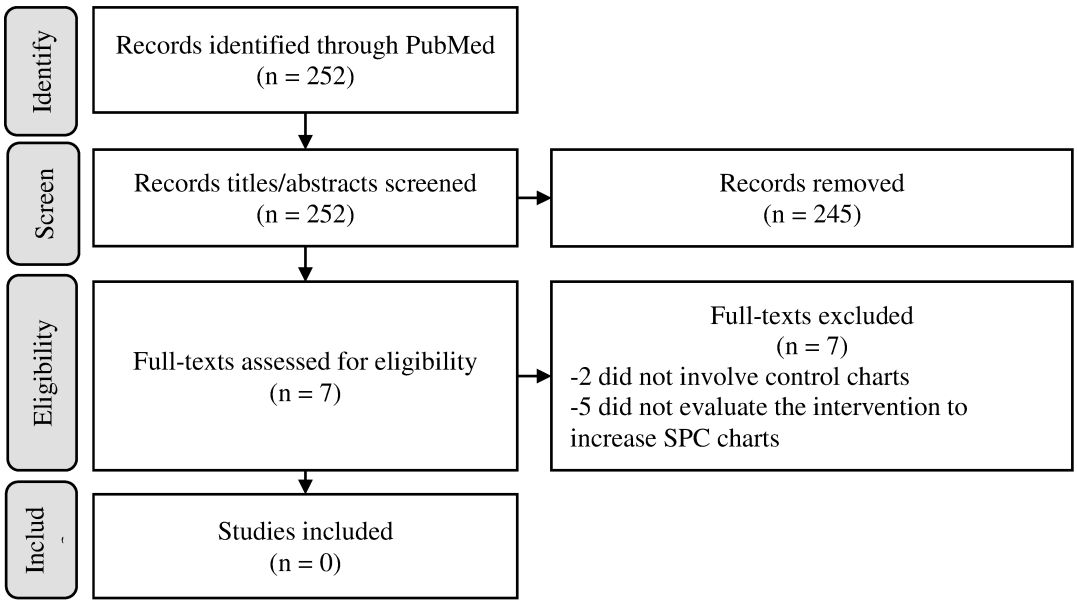
[1] The thresholds at which the RAG limits are set are sometimes user-defined. For example, if national target is to be above 90%, one Trust may define Amber as being performance below 94% - another may decide on 92%.
[2] These are not the only criteria that may influence whether decision-makers engage in reflective and/or automatic thinking. For example, decision-makers also need to have sufficient skills and knowledge to interpret the process data being modelled within the chart, in addition to the opportunity to so (Michie et al, 2011)

Of course, many other factors can influence whether board members can use control charts effectively. In order for any behaviour to occur, people must possess the relevant capabilities (psychological, physical), opportunities (social, physical) and motivations (reflective, automatic; Michie, et al. 2011). For example, board members need to know how to interpret the information presented in the control chart (a capability factor), have the motivation to engage with the data at a deeper level than they would with target-focused evaluations (a motivational factor), and have access to satisfactorily constructed SPC charts in their board papers (an opportunity factor). The present study focuses on the capability and opportunity factors: explaining the use of SPC charts to board members and increasing the number of control charts present in NHS board papers, respectively.

There are a large number of studies about specific quality improvement methodologies such as Lean, Six-sigma and Plan-Do-Action cycles that may use SPC methodology as part of the improvement process (Deblois & Lepanto, 2016). We are, however, not interested in the use of SPC methods as part and parcel of an intervention to improve a given process. Rather, we are interested in SPC methods being used in routine surveillance to identify processes to be improved. To understand if any similar studies had already been conducted, we therefore carried out a systematic literature search for methods to improve the use of SPC for routine surveillance. Our search strategy is laid out in Figure 3 and discussed in detail in Appendix C (available here: https://bit.ly/3j0N4Iu). We found no papers that replicated our study, and we can assert that this is the first study to examine the effectiveness of an intervention to increase the use of SPC charts across a range of routine monitoring programs at the institutional level.

Figure 3: Results of systematic review seeking studies on training interventions to increase the use of SPC charts for routine monitoring within institutions

Supplementary 1 - Protocol



Rationale

A previous review of 30 randomly selected English acute care trusts' board papers found that 13 (43%) lacked even a single chart depicting the role of chance (Schmidtke et al., 2017). This problem was summarised in a popular health management magazine (Bird, 2017), and underpinned the NHS Improvement/England (NHS I/E) (2019) initiative called 'Making Data Count' to increase the use of SPC charts in NHS institutions. This initiative has involved a set of educational resources along with a series of training sessions on the implementation and use of SPCs. The effectiveness of this training initiative has not yet been captured in a scientifically rigorous manner.

Specific aims

We set out to evaluate the effects of the Making Data Count training sessions on the use of SPC charts by NHS trusts, and to assess how the trainees perceived the training sessions.

Methods

Supplementary 1 - Protocol

Our objectives were to identify if the training resulted in an increase in the proportion of SPC charts in board papers, and to thematically analyse participants' reactions to the training sessions. This research protocol was prepared according to SQUIRE guidelines (Ogrinc, Davies, Goodman, Batalden, Davidoff & Stevens, 2016).

Context

The Making Data Count training sessions were delivered to NHS trust boards and to teams of hospital analysts by NHS Improvement from November 2017 onwards. NHS Improvement uses social media, email and word of mouth to invite trusts to participate. Thus, there is self-selection into the training sessions, and the approach to recruitment into the training sessions is effectively snowball sampling. All trusts that received a training session that we will investigate are based in England.

Intervention

This intervention is described here according to TIDieR guidelines (Hoffman et al, 2014). Our completed TIDier checklist is in Appendix D (https://bit.ly/3j0N4Iu). The brief name of the intervention is 'Making Data Count SPC training sessions'. The training sessions are conducted to improve knowledge about SPC charts and increase their uptake. Two examples of PowerPoints slides used in the training sessions are shown in Appendices A and B (https://bit.ly/3j0N4Iu). The two Making Data Count guidebooks that supplement the training are available online (NHS Improvement, 2019). The training sessions cover the strengths and weaknesses of presenting data in different ways, and include background on what SPC charts are, when and how to use them, why they should be used, and how they can improve decision-making. Topics include identifying trends (e.g. seven points in one direction), special versus common cause variation, and summarising data using icons (see Appendix B, Slide 47, https://bit.ly/3j0N4Iu). The limitations of *R-A-G* systems are discussed, and, importantly, each training is personalised: trusts' data from their board papers are presented using SPCs in order to demonstrate the value of using SPCs.

The Making Data Count training sessions are delivered at each trust to up to two groups of people separately, as mentioned above: board members and analysts. The training sessions

for board members are delivered over about 90 minutes, while training sessions for analysts are delivered over one working day. Training sessions are delivered by two experienced trainers from NHS Improvement with higher educational backgrounds in statistics and work experience in data analytics. One trainer visits each trust to deliver the training face-to-face to board members and, separately, to teams of analysts. Board and analyst trainings are not necessarily given on the same day and can be separated by around a month.

Study of the intervention (evaluation design)

Design

This study will conduct a quantitative and qualitative evaluation of the training sessions. The quantitative evaluation will be a controlled before and after design that uses data from ten acute care trusts that received the training, as well as ten different acute care trusts that will be external matched controls. Board papers from before and after the training dates will be selected. The qualitative evaluation will thematically analyse responses to feedback forms from some of the trusts. Overall, the study design is pragmatic and determined by resource capacity to find and extract data from the board meeting papers.

Selection of acute care trusts

Due to resource constraints, we will be unable to examine board papers in all trusts that received a training from the start of the intervention period in March 2018. Instead, we will focus on the acute care trusts from the first year of trainings through March 2019. We will select ten trusts that received the training during different months in order to maximise temporal heterogeneity. These ten trusts are the training intervention sample.

We will also identify ten acute trusts that have not received the training intervention to be external matched controls. The ten trusts in the intervention group will be matched to ten other trusts using the NHS Digital (2020) Peer Finder tool. This tool identifies trust peers based on variables such as attendances, deprivation, and patient profiles, and proposes ten peers with the smallest Euclidean distance to the selected trust. We will view the ten closest matches using the default tool weightings. From these ten closely matched trusts, we will select (without replacement) trusts that meet the following criteria (in order):

• did not receive a training intervention,
• similar number of attendances,
• similar degree of specialisation, and
• similar level of deprivation.

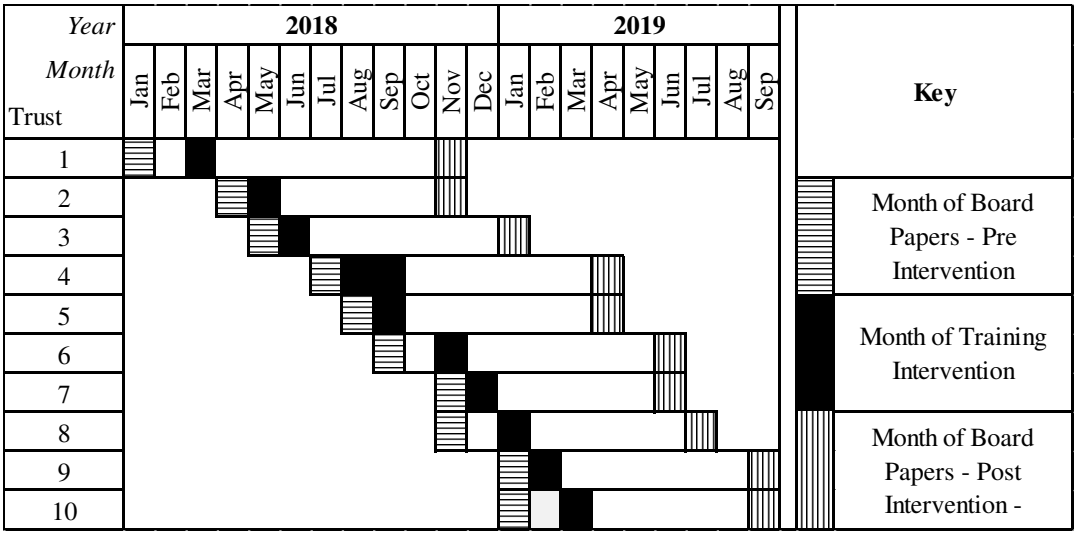
Occasionally, as tie breakers, other factors such as the number of FTE (full time equivalent) staff, urban location, and foundation status may be used as additional criteria.

Selection of board papers from acute care trusts

For the intervention group, we will identify board papers published in the month before the intervention was delivered (pre intervention observation) and approximately six months after delivery in each trust (post intervention observation). Boards do not publish their papers every month. In some cases, therefore, it is not possible to sample board papers exactly immediately prior to the training or six months immediately after the training. When it is not possible to select a board paper from the assigned month pre training, the first board paper published at least one month before intervention delivery will be selected; when it is not possible to select a board paper from the assigned month post-training, the first board paper published at least six months after the intervention delivery will be selected. This approach is shown in Figure 4, which represents the realised design in the intervention sample accounting for the fact that not all trusts have board papers available in the first month before the intervention or six months post intervention roll-out.

Supplementary 1 - Protocol

Figure 4: Selected board papers from for pre intervention and post intervention periods, and month of training intervention, for ten acute trusts that received 'Making Data Count' training sessions



Notes: The black boxes show the month of the training intervention. The ten cells with horizontal lines are those months where we will sample the pre intervention board papers from before the training. The ten cells with vertical lines are those months where we will sample the post-intervention board papers from after the training, which occur at least six months after the training. Trust 4 received trainings in August and September.

For the external matched control group (not shown in Figure 4), we will identify board papers published in the months closest to the pre and post intervention observations from the matched trust in the intervention group. Overall, this equates to 20 different intervention and matched control trusts in total, each contributing two papers, for a total sample of 40 board papers.

Quantitative measures

We will create three quantitative measures from data in the board papers. The main outcome measure will be the proportion of charts that were SPC charts out of all charts presented. The other two outcomes will be the proportion of charts that were SPC charts out of all time series charts, and the proportion of charts that were SPC charts out of all time series *and* between group charts.

The rationale behind selecting the first outcome is that increasing the use of SPC charts is a main focus of the training intervention, and it can be created from information that is publicly available in board papers. This outcome may be considered a broad level at which the effects of training on control chart usage may be evidenced. Not all charts, however, can be easily transformed into SPC charts. The rationale behind selecting the other two outcomes is that time series and between group charts can be more directly transformed into SPC charts than can other types of charts, such as pie charts. Time series and between groups charts are, therefore, the types of charts that we most expect the training sessions to influence. We focus on time series charts separately because time sequences "in order" were the types of charts that Shewart's original SPC methodology encouraged (Shewhart, 1939/1986, p.12). Some additional descriptive information about all of the charts in the board papers, as well as about the SPC charts specifically, will be recorded (this is discussed in the section further below on 'data extraction').

Data extraction from board papers

We will extract information from the board papers to populate the quantitative outcomes discussed above: number of SPC charts, total number of charts, number of time series charts, and number of between group charts. We will also extract additional information about the charts to illustrate the specific contexts where the training may be effective. The charts will be classified as 'quality and safety' charts or not, following Schmidtke et al (2017), which may be interpreted in various ways. One definition of quality and safety is whether care "conforms to established treatment goals and care processes" (quality) and "avoids injuries to patients" (safety), as discussed by the Institute of Medicine (2002, p.92). Guided by this definition, our approach will use multiple raters to assess whether a particular chart depicts quality and safety information.

Additional information about the nature and content of SPC charts identified will be recorded (see Appendix E, all Appendices here: https://bit.ly/3j0N4Iu): control limits (Appendix A, p. 38; Appendix E, item 10), recalculation of control limits (Appendix B, p. 44-46; Appendix E, item 11), run/trend points (Appendix A, p. 14; Appendix E, item 12), and whether there are comments about reasons for variation, or suggestions about intervening (Appendix B, p. 51; Appendix E, items 13-14). Whether the control limits are labelled is of interest because labels provide decision-makers with information that may engage reflective (vs. automatic)

cognitive processes. We will also assess whether r-a-g is still present in charts identified from the board papers (Appendix A, p. 4-7; Appendix E, item 1), and whether there are any icon summaries[3] (Appendix B, p. 47; Appendix E, item 8), which were also covered in training.

Blinding and agreement

One reviewer will download the board papers from the web and four independent reviewers will examine the board papers (reviewers R1, R2, R3, R4). Reviewers examining the board papers for the presence and nature of SPC charts will be blind as to whether the board paper is from the control or post intervention period. To ensure agreement and blinding, the below four steps will be taken. Steps one and three ensure agreement between raters, and steps two and four ensure blinding:

(1) *Identification and sampling of charts.* R1 will download the board papers. R1 and R2 will independently identify the total number of charts, and independently identify whether the chart is a quality and safety chart. R1 and R2 will discuss any disagreements to reach a consensus and inter-rater reliability will be calculated (prior to the consensus). Any unresolved disagreements will be referred to the chief project investigator.

(2) *Assessment of sample charts.* R1 screenshots the charts and removes any information about name of trust and/or date of board paper, randomises the order of trusts, and sends them to R3 and R4.

(3) *Examination of charts.* R3 and R4 will examine the charts and decide if the charts are SPC charts, time series charts, between group charts, or other types of charts. R3 and R4 will also give descriptions of the SPC charts according to the measures in Appendix E, described above.[4] Inter-rater reliability will be calculated and R3 and R4 will subsequently discuss to reach a consensus. Any unresolved disagreements will be referred to the chief project investigator.

(4) *De-blinding report.* R3 and R4 note if they have been de-blinded at any point.

[3] See Appendix E for example of SPC icon summary (here: https://bit.ly/3j0N4Iu).

Supplementary 1 - Protocol

Sample size calculation

We are looking for a substantial effect size because, in contrast to a clinical intervention which affects patients directly, this service intervention affects patients indirectly (Lilford et al., 2010). It is, therefore, doubtful whether service managers would want to replicate the training intervention unless they could achieve a substantial improvement in uptake. Our sample size is based on the detection of a 30 percentage point improvement in the proportion of charts that are SPC from 10% to 40% between pre and post intervention measures. Sample size is calculated with an alpha of 0.05 and power of 0.80. Due to the study design, adjustment for the correlation between pre and post intervention measures is made, which is estimated at r=0.90 (Frison & Pocock, 1992). A minimum of 16 hospitals with pre and post intervention measures is required.

Quantitative analysis

Information on the 20 hospitals will be summarised, including key characteristics used for the matching (attendances, specialisation, level of deprivation). Details about the identified SPC charts (control limits, recalculation of control limits, run/trend points, and whether there are comments about reasons for variation, or suggestions for intervention – see above and Appendix E) will be summarised as counts and proportional measures.

For each hospital, we will have information on the number of charts depicted as a SPC chart (the outcome), the total number of charts (an offset), the month of the observation, whether the observation was from the intervention or control group and whether the observation was from a pre or post intervention period. Other analyses will have an offset in two different ways; (1) time series charts only and (2) time series and between group charts.

A Poisson regression model will be fitted with an offset for the total number of charts and the outcome as the number of charts presented as an SPC control chart. We will adjust for group (intervention or control group), for period (pre or post intervention exposure) and an interaction between group and period (treatment effect) using a difference in difference approach. The offset in the model will be changed dependent on the outcome. Results will be reported on the rate ratio scale with 95% confidence intervals. Subgroup analysis will be conducted using quality and safety charts only.

Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

Supplementary 1 - Protocol

Inter-rater reliability will be calculated using Kappa statistics and percentage agreement to quantity the level of agreement between reviewers for deciding on whether they were SPC charts, time series charts, between group charts, and quality and safety charts.

Qualitative evaluation

In addition to quantitative outcomes and analyses, we will conduct a qualitative evaluation to better understand barriers and facilitators to the uptake of SPC charts. Our qualitative process outcomes come from feedback forms that were filled out by training session participants during the board sessions (see Appendix F). These forms were designed by NHS-I/E and shared with the research team. We will analyse responses to the following four items:

1. What went well today?
2. What could have been done differently?
3. What are your key takeaways?
4. Any other comments about today.

We will conduct thematic analysis of written responses to these questions to identify barriers to and facilitators of using SPC charts (Braun & Clark, 2014).

Ethical considerations

This research has been approved by the University of Warwick Biomedical and Scientific Research Ethics committee (BSREC 116/18-19).

Discussion

Summary

Overall, this research will provide evidence about the impact of training sessions on the use of SPC charts among acute care hospital trusts in England. In addition, qualitative reactions to the training will also be provided. The findings will provide new empirical evidence about

whether these training sessions are effective and may inform the design of any future work to increase the use of SPC charts. To the research team's knowledge, this is the first project to directly evaluate the effectiveness of such training using a controlled before and after analysis of the documents the training should influence.

There are some limitations to our approach that stem from our time and resource constraints, as well as the nature of the retrospective evaluation. One is about the validity of our outcome measures. Although our use of publicly available board papers does overcome potential errors resulting from self-reported data, such as social desirability bias and recall errors (Groves et al, 2011), it may not capture all of the ways that trusts use SPC charts. For example, trusts may increase their use of SPC charts in other routine monitoring reports. This would decrease the validity of our findings. However, it is not possible to assess the impact of this issue without further investigation with more time and resources, and we leave it for future research. Further, as the board papers comprise many sub-reports, and are monitored by top-level decision-makers, they serve as the best publically available documents for the present evaluation.

Another limitation relates to the precision of our estimates. It may be that having more pre and post intervention time period measurements would increase our precision. Given resource constraints, a decision was taken to include external matched controls rather than additional time series data. We may, therefore, sacrifice some precision for more plausible causal inference. Trusts who receive the training later on may get swept up in a 'rising tide' of greater use of SPC charts by trusts in general – and so the training could appear to be effective, even if it was not relatively effective within the context of greater usage overall (Chen et al., 2016). The external controls approach allows us to evaluate the rising tide phenomenon, although it is not a perfect solution. Control trusts were selected to be similar to intervention trusts on observable characteristics, and it is possible that control trusts will differ according to unobservable characteristics, such as motivation or openness to change, which could bias the results.

Finally, generalisability is an issue. We study a sample of self-selecting trusts that elected to take part in a training intervention. As such, our results may not apply to any mandated training initiatives if these become a requirement. To put this another way, trusts that elect to be part of the training may be more susceptible to change than other trusts that may not come

willingly. Moreover, because we limit our sample to acute trusts, our results may not hold when extended to other forms of trusts – such as mental health or community care trusts. That said, it seems unlikely that other types of healthcare institutions or that hospitals elsewhere would be 'immune' from the influence of training. While there may be quantitative differences, we consider it unlikely that there will be qualitative differences. Similarly, because we limited our investigation to trusts in England, the generalisability of our results may not hold in other geographic areas.

References

Anhøj J, Hellesøe AM. The problem with red, amber, green: the need to avoid distraction by random variation in organisational performance measures. BMJ Qual Saf. 2017 Jan 1;26(1):81-4.

Benning A, Dixon-Woods M, Nwulu U, Ghaleb M, Dawson J, Barber N, Franklin BD, Girling A, Hemming K, Carmalt M, Rudge G. Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. BMJ. 2011 Feb 3;342:d199. https://doi.org/10.1136/bmj.d199

Bird, P. Understanding the right and wrong time for intervention. Use of charts by hospital boards can help decision makers set control limits to performance indicating where intervention may be required. Health Service Journal; 2017 14 Feb.

Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers?. International Journal of Qualitative Studies on Health and Well-being. 2014;9. https://doi.org/10.3402/qhw.v9.26152

Britto MT, Fuller SC, Kaplan HC, Kotagal U, Lannon C, Margolis PA, Muething SE, Schoettker PJ, Seid M. Using a network organisational architecture to support the development of Learning Healthcare Systems. BMJ Quality & Safety. 2018 Nov 1;27(11):937-46. doi: 10.1136/bmjqs-2017-007219.

Chen YF, Hemming K, Stevens AJ, Lilford RJ. Secular trends and evaluation of complex interventions: the rising tide phenomenon. BMJ Quality & Safety. 2016 May 1;25(5):303-10. https://doi.org/10.1136/bmjqs-2015-004372

Deblois S, Lepanto L. Lean and Six Sigma in acute care: a systematic review of reviews. International Journal of Health Care Quality Assurance. 2016 Mar 14. https://doi.org/10.1108/IJHCQA-05-2014-0058

Deeny SR, Steventon A. Making sense of the shadows: priorities for creating a learning healthcare system based on routinely collected data. BMJ Qual Saf. 2015 Aug 1;24(8):505-15. doi:10.1136/bmjqs-2015-004278.

Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. Statistics in Medicine. 1992;11(13):1685-704. https://doi.org/10.1002/sim.4780111304

Goodman D, Ogrinc G, Davies L, Baker GR, Barnsteiner J, Foster TC, Gali K, Hilden J, Horwitz L, Kaplan HC, Leis J. Explanation and elaboration of the SQUIRE (Standards for Quality Improvement Reporting Excellence) Guidelines, V. 2.0: examples of SQUIRE elements in the healthcare improvement literature. BMJ Qual Saf. 2016 Dec 1;25(12):e7. https://doi.org/10.1136/bmjqs-2015-004411

Groves RM, Fowler Jr FJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R. Survey methodology. John Wiley & Sons; 2011 Sep 20.

Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ. 2015 Feb 6;350:h391. https://doi.org/10.1136/bmj.h391

Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, Altman DG, Barbour V, Macdonald H, Johnston M, Lamb SE. Better reporting of interventions: template for

intervention description and replication (TIDieR) checklist and guide. BMJ. 2014 Mar 7;348:g1687. https://doi.org/10.1136/bmj.g1687

Institute of Medicine. Fostering rapid advances in healthcare: Learning from system demonstrations. Washington, DC: National Academies Press; 2002. https://doi.org/10.17226/10565

Kaplan HC, Lannon C, Walsh MC, Donovan EF, Ohio Perinatal Quality Collaborative. Ohio statewide quality-improvement collaborative to reduce late-onset sepsis in preterm infants. Pediatrics. 2011 Mar 1;127(3):427-35. doi: 10.1542/peds.2010-2141.

Larsson J, Landstad BJ, Wiklund H, Vinberg S. Control charts as an early-warning system for workplace health outcomes. Work. 2011 Jan 1;39(4):409-25. doi: 10.3233/WOR-2011-1191.

Lilford RJ, Chilton PJ, Hemming K, Girling AJ, Taylor CA, Barach P. Evaluating policy and service interventions: framework to guide selection and interpretation of study end points. BMJ. 2010 Aug 27;341:c4413. https://doi.org/10.1136/bmj.c4413

Mazzocato P, Holden RJ, Brommels M, Aronsson H, Bäckman U, Elg M, Thor J. How does lean work in emergency care? A case study of a lean-inspired intervention at the Astrid Lindgren Children's hospital, Stockholm, Sweden. BMC Health Services Research. 2012 Dec;12(1):28. doi: 10.1186/1472-6963-12-28.

Michie S, Van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implementation Science. 2011 Dec 1;6(1):42. https://doi.org/10.1186/1748-5908-6-42

Mohammed MA, Cheng KK, Rouse A, Marshall T. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. The Lancet. 2001 Feb 10;357(9254):463-7. https://doi.org/10.1016/s0140-6736(00)04019-8

NHS Digital (2020). Peer Finder Too. Retrieved 18 May 2020 from https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/innovative-uses-of-data/multi-dataset-analysis/nhs-trust-peer-finder-tool

NHS Improvement. Making Data Count website. Retrieved 14 August 2019 from https://improvement.nhs.uk/resources/making-data-count/

Schmidtke KA, Poots AJ, Carpio J, Vlaev I, Kandala NB, Lilford RJ. Considering chance in quality and safety performance measures: an analysis of performance reports by boards in English NHS trusts. BMJ Qual Saf. 2017 Jan 1;26(1):61-9. https://doi.org/10.1136/bmjqs-2015-004967

Schmidtke KA, Watson DG, Vlaev I. The use of control charts by laypeople and hospital decision-makers for guiding decision making. The Quarterly Journal of Experimental Psychology. 2017 Jul 3;70(7):1114-28. https://doi.org/10.1080/17470218.2016.1172096

Shewhart, W. Statistical method from the viewpoint of quality control. Dover 1986 edition with foreword by W. Edwards Deming, originally published 1939. Mineola, N.W: Dover Publications.

Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science. 1974 Sep 27;185(4157):1124-31. https://doi.org/10.1126/science.185.4157.1124

Wheeler, W. Contra two sigma: the consequences of using the wrong limits. Quarterly Digest Daily, 2013 1 May, Manuscript 255. Retrieved from: https://www.spcpress.com/pdf/DJW255.pdf

Supplementary 2 – SQUIRE checklist

Reporting checklist for quality improvement in health care.

Based on the SQUIRE guidelines.

		Reporting Item	Page Number *[refers to submission document, not published document]*
Title			
	#1	Indicate that the manuscript concerns an initiative to improve healthcare (broadly defined to include the quality, safety, effectiveness, patientcenteredness, timeliness, cost, efficiency, and equity of healthcare)	Title page
Abstract			
	#02a	Provide adequate information to aid in searching and indexing	Abstract (0)
	#02b	Summarize all key information from various sections of the text using the abstract format of the intended publication or a structured summary such as: background, local problem, methods, interventions, results, conclusions	Abstract (0)
Introduction			
Problem description	#3	Nature and significance of the local problem	1=
Available knowledge	#4	Summary of what is currently known about the problem, including relevant previous studies	2-3
Rationale	#5	Informal or formal frameworks, models, concepts, and / or theories used to explain the problem, any	2-3

Supplementary 2 – SQUIRE checklist

		reasons or assumptions that were used to develop the intervention(s), and reasons why the intervention(s) was expected to work	
Specific aims	#6	Purpose of the project and of this report	3
Methods			
Context	#7	Contextual elements considered important at the outset of introducing the intervention(s)	3
Intervention(s)	#08a	Description of the intervention(s) in sufficient detail that others could reproduce it	4
Intervention(s)	#08b	Specifics of the team involved in the work	3
Study of the Intervention(s)	#09a	Approach chosen for assessing the impact of the intervention(s)	5-6
Study of the Intervention(s)	#09b	Approach used to establish whether the observed outcomes were due to the intervention(s)	5-6
Measures	#10a	Measures chosen for studying processes and outcomes of the intervention(s), including rationale for choosing them, their operational definitions, and their validity and reliability	6-8
Measures	#10b	Description of the approach to the ongoing assessment of contextual elements that contributed to the success, failure, efficiency, and cost	7-8
Measures	#10c	Methods employed for assessing completeness and accuracy of data	7
Analysis	#11a	Qualitative and quantitative methods used to draw inferences from the data	8-10
Analysis	#11b	Methods for understanding variation within the data, including the effects of time as a variable	8
Ethical considerations	#12	Ethical aspects of implementing and studying the intervention(s) and how they were addressed, including, but not limited to, formal ethics review and potential conflict(s) of interest	9
Results			

Supplementary 2 – SQUIRE checklist

	#13a	Initial steps of the intervention(s) and their evolution over time (e.g., time-line diagram, flow chart, or table), including modifications made to the intervention during the project	Fig 4, S4, S5
	#13b	Details of the process measures and outcome	17-18
	#13c	Contextual elements that interacted with the intervention(s)	N/A
	#13d	Observed associations between outcomes, interventions, and relevant contextual elements	9-16
	#13e	Unintended consequences such as unexpected benefits, problems, failures, or costs associated with the intervention(s).	N/A
	#13f	Details about missing data	N/A
Discussion			
Summary	#14a	Key findings, including relevance to the rationale and specific aims	18
Summary	#14b	Particular strengths of the project	18
Interpretation	#15a	Nature of the association between the intervention(s) and the outcomes	18
Interpretation	#15b	Comparison of results with findings from other publications	18-19
Interpretation	#15c	Impact of the project on people and systems	18-21
Interpretation	#15d	Reasons for any differences between observed and anticipated outcomes, including the influence of context	21
Interpretation	#15e	Costs and strategic trade-offs, including opportunity costs	N/A
Limitations	#16a	Limits to the generalizability of the work	23
Limitations	#16b	Factors that might have limited internal validity such as confounding, bias, or imprecision in the design, methods, measurement, or analysis	19-21

Supplementary 2 – SQUIRE checklist

Limitations	#16c	Efforts made to minimize and adjust for limitations	19-21
Conclusion	#17a	Usefulness of the work	21
Conclusion	#17b	Sustainability	21
Conclusion	#17c	Potential for spread to other contexts	21
Conclusion	#17d	Implications for practice and for further study in the field	21
Conclusion	#17e	Suggested next steps	21
Other information			
Funding	#18	Sources of funding that supported this work. Role, if any, of the funding organization in the design, implementation, interpretation, and reporting	29-30

Supplementary 3 – TIDieR checklist

Item No	Item	Current Research
Brief name		
1	Provide the name or a phrase that describes the intervention	Making Data Count SPC Training Sessions
Why		
2	Describe any rationale, theory, or goal of the elements essential to the intervention	Trainings are conducted because improved knowledge about statistical process control charts (SPCs) may increase their uptake and then prevent unnecessary interventions in the NHS
What		
3	Materials: Describe any physical or informational materials used in the intervention, including those provided to participants or used in intervention delivery or in training of intervention providers. Provide information on where the materials can be accessed (such as online appendix, URL)	Two examples of PowerPoints used in the training are shown in Supplementary Materials 4 and 5. The two Making Data Count guidebooks are available online (https://improvement.nhs.uk/resources/making-data-count/). The training events cover the strengths and weaknesses of presenting data in different ways, and include background on what SPCs are, when and how to use them, why they should be used, and how they can improve decision making. Topics include identifying trends (e.g. seven points in one direction), special versus common cause variation, and summarising data using icons (see Supplementary Materials 4, Slide 47). The limitations of r-a-g systems are discussed, and, importantly, each trust's data is presented to them using control charts to demonstrate the value of using SPCs.

Supplementary 3 – TIDieR checklist

4	Procedures: Describe each of the procedures, activities, and/or processes used in the intervention, including any enabling or support activities	Two examples of PowerPoints used in the training are shown in Supplementary Materials 4 and 5. The two Making Data Count guidebooks are available online (https://england.nhs.uk/resources/making-data-count/)
Who provided		
5	For each category of intervention provider (such as psychologist, nursing assistant), describe their expertise, background, and any specific training given	Trainings were delivered by two experienced trainers from NHS Improvement with higher educational background in statistics and work experience in data analytics.
How		
6	Describe the modes of delivery (such as face to face or by some other mechanism, such as internet or telephone) of the intervention and whether it was provided individually or in a group	One trainer visited each trust to deliver the training face-to-face to board members and, separately, to teams of analysts and ambassadors.
Where		
7	Describe the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features	The trainings were delivered in meeting rooms at each Trust, which varied in the available facilities
When and How Much		

Supplementary 3 – TIDieR checklist

8	Describe the number of times the intervention was delivered and over what period of time including the number of sessions, their schedule, and their duration, intensity, or dose	The trainings were delivered to board members and, separately, to teams of analysts and ambassadors. Board and analyst trainings were not necessarily given on the same day and could be separated by around a month. Board training sessions are delivered over about 90 minutes while training for analysts teams is delivered over one work day.
Tailoring		
9	If the intervention was planned to be personalised, titrated or adapted, then describe what, why, when, and how	Trusts' data were presented to them using control charts to demonstrate the value of using SPCs
Modifications		
10*	If the intervention was modified during the course of the study, describe the changes (what, why, when, and how)	N/A
How well		
11	Planned: If intervention adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them	N/A

Supplementary 3 – TIDieR checklist

12*	Actual: If intervention adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned	N/A
*If checklist is completed for a protocol, these items are not relevant to protocol and cannot be described until study is complete.		

Supplementary 4 - Making Data Count Powerpoint (1)

NHS Improvement

Making data count

[redacted]

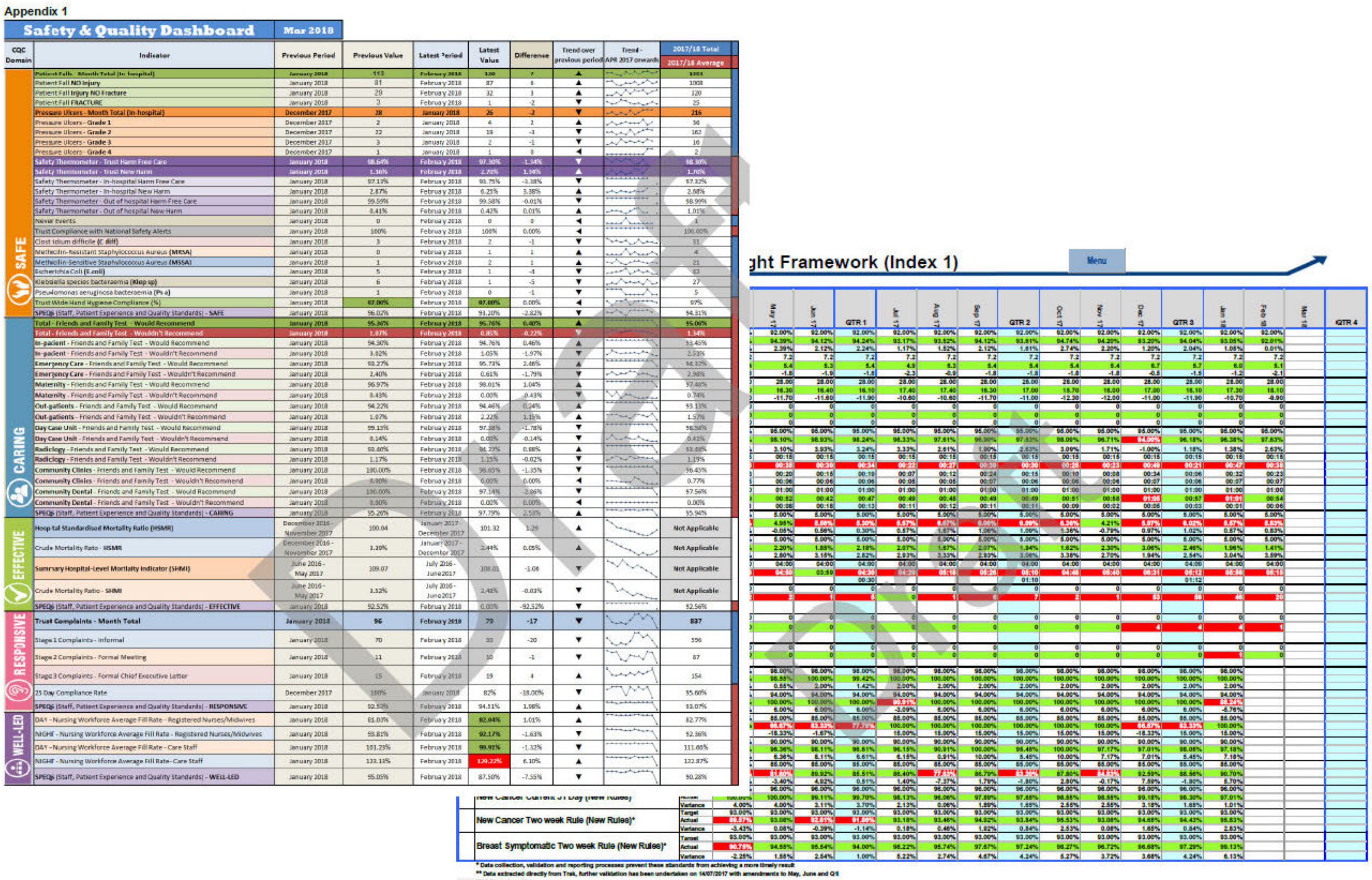
Samantha Riley, Head of Improvement Analytics

[redacted]

collaboration trust respect innovation courage compassion

Where we are now............

NHS Improvement



The importance of focus

NHS Improvement

Safety & Quality Dashboard									Mar 2018
CQC Domain	Indicator	Previous Period	Previous Value	Latest Period	Latest Value	Difference	Trend over previous period	Trend - APR 2017 onwards	2017/18 Total / 2017/18 Average
	Emergency Care - Friends and Family Test - Would Recommend	January 2018	93.27%	February 2018	95.73%	2.46%	▲		94.32%

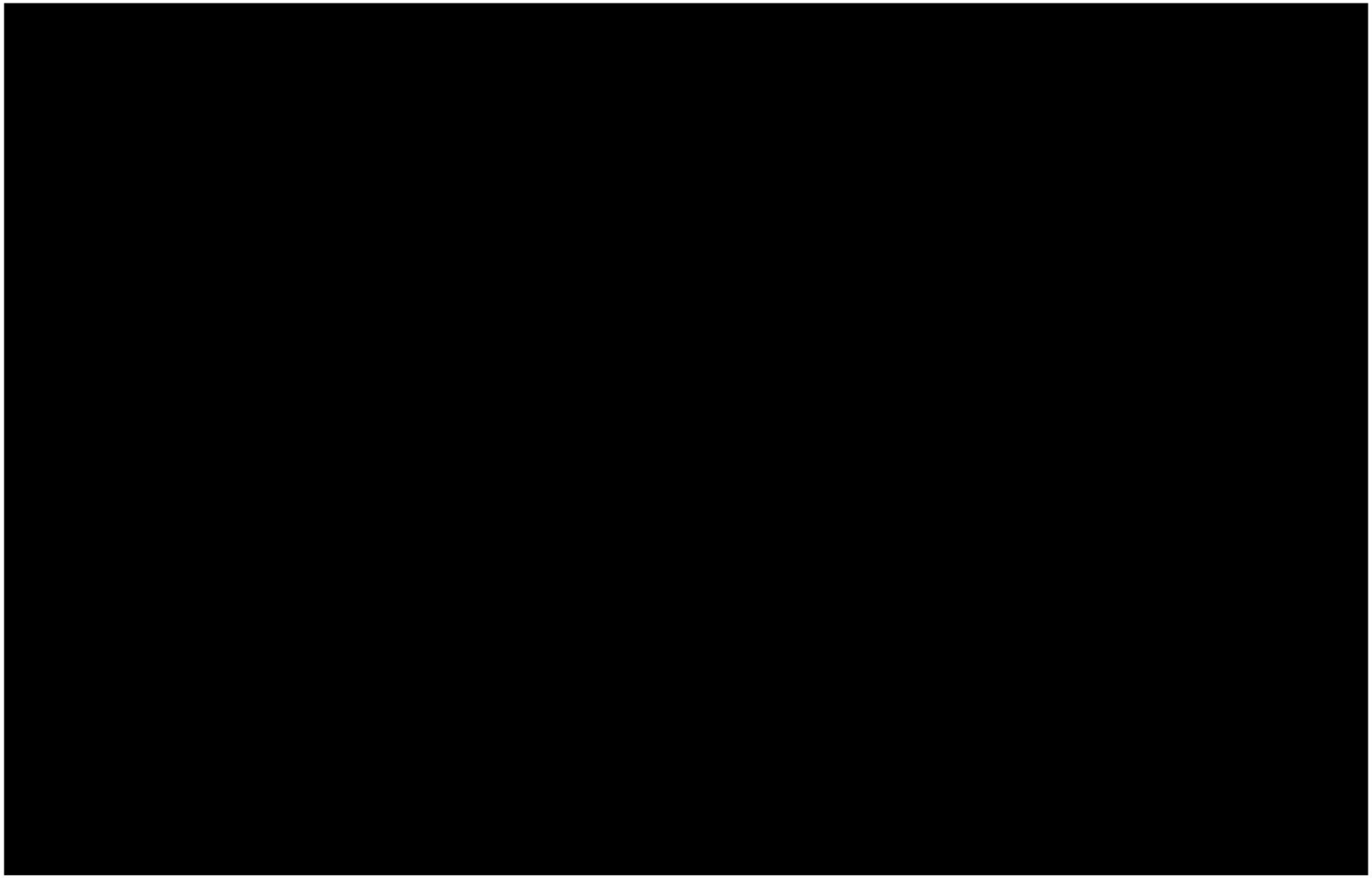
One month trend.....

Is an increase from 95.36% to 95.76% important or distracting narrative?

Caring

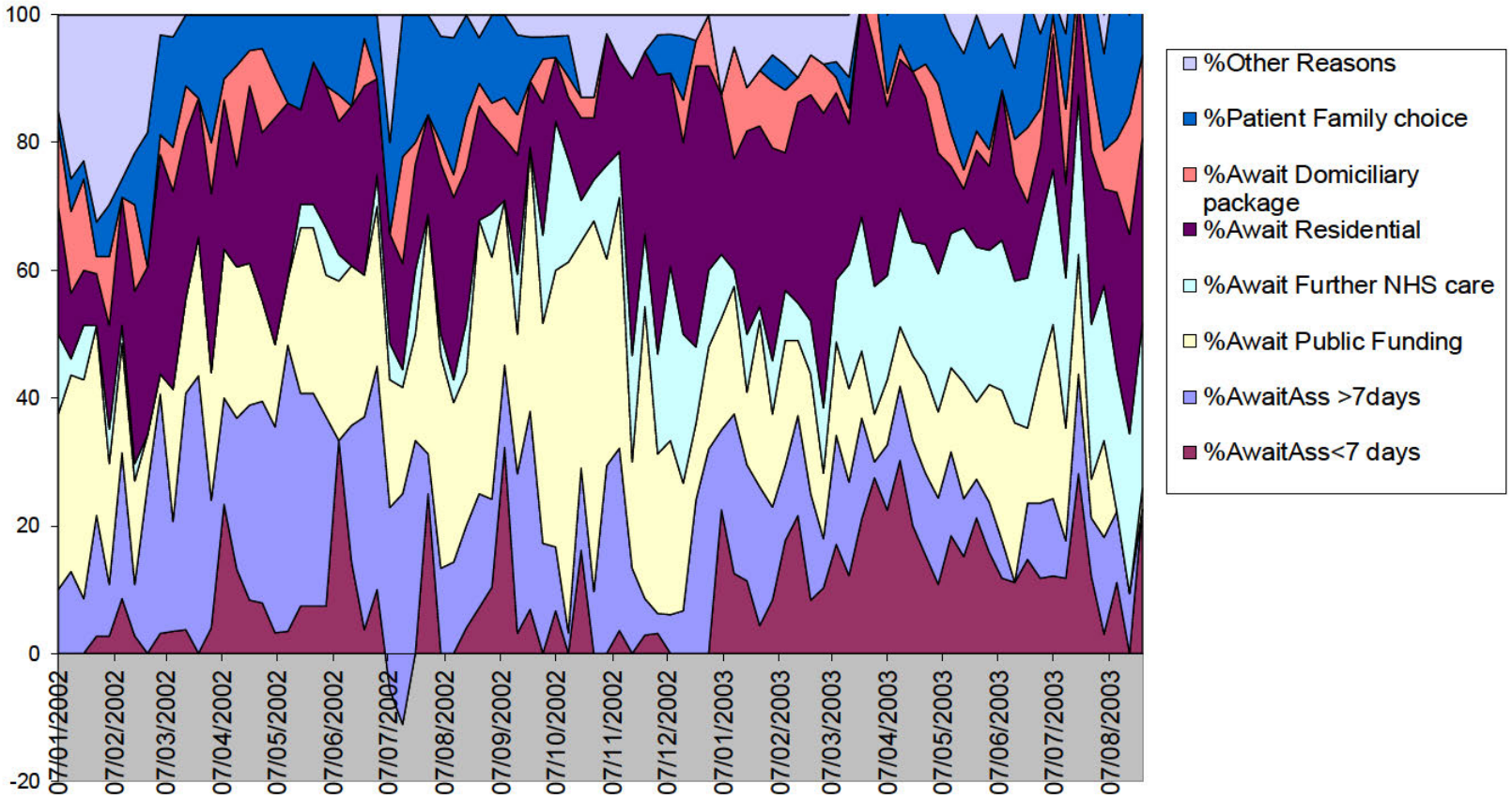
7 Family and Friends Test (FFT) (data up to February 2018)

7.2 The Trusts 'Would Recommend' for Friends and Family returns *increased* to **95.76%** for February 2018 from 95.36% in January 2018. The percentage of patients who stated they 'Wouldn't Recommend' *decreased* to **0.85%** in February 2018 from 1.07% in January 2018.



NHS Improvement

% Delayed transfers of Care by Type - source SITREPS 7/1/02-31/08/03



5 | Making data count

NHS Improvement

Activities summary from the monthly measures : Mar'02-

	Patients within the scope of the project treated this month		Patients in scope of the project treated this month under an agreed Care Pathway		Time* from referral to first definitive treatment (days)		Maximum wait* for 1st specialist appointment		Average wait* for 1st specialist appointment		Booked appointments through new clinics		Booked admissions through new clinics		Number of Patient Discovery Interviews		Team self-assessment
Pilot sites	Mar '02	change from last month	Mar '02	change from last month	Mar '02	change from last month	Mar '02	change from last month	Mar '02	change from last month	Mar '02	change from last month	Mar '02	change from last month	Mar '02	change from last month	Mar '02
	-	0	-	0	184	-17	-	-140	45	-2	278	54	14	4	-	0	-
	8	-1	8	8	741	434	175	-7	123	-4	0	0	0	0	0	-1	2
	-	0	-	0	-	-8	-	-70	-	-30	-	0	-	0	-	0	-
	97	22	17	11	-	-16	84	0	57	-4	17	2	7	7	0	0	4
	37	-19	12	6	41	-65	84	0	82	22	0	0	0	0	0	0	-
	-	-38	-	0	-	-669	-	-182	-	-123	-	0	-	-48	-	0	-
	15	0	15	0	294	-34	245	14	84	0	0	0	0	-87	0	0	3
	366	24	291	3	395	48	55	10	39	-8	0	-379	0	0	0	0	4
	-	-46	0	-46	282	-52	235	64	119	0	1217	46	0	0	0	0	4
	15	2	3	3	-	0	97	-8	23	1	699	238	6	0	-	0	3
	29	-4	9	3	369	53	226	-49	186	37	0	0	0	0	8	8	2
	18	10	14	9	200	-40	32	1	30	2	0	0	39	2	0	0	-
	175	-77	38	-88	172	-1	137	11	45	5	0	0	18	3	0	0	-
	25	8	25	8	1	-1	98	-14	77	14	123	-3	62	-6	14	4	3
	-	0	-	0	-	0	-	-618	-	-106	-	0	-	0	-	0	-
	30	12	15	5	318	0	89	-155	209	43	0	0	0	0	0	0	3
	11	-3	0	0	400	90	214	-47	210	13	0	0	0	0	0	0	-
	-	0	-	0	-	-416	-	-297	-	-91	-	0	0	-29	9	-9	4
	12	2	12	2	210	113	306	-250	114	43	0	0	0	0	0	0	-
	-	-4	-	-4	204	57	222	-445	192	99	0	0	0	0	0	0	-
	3	0	3	0	226	-81	322	-195	195	104	0	0	0	0	0	0	-

NHS Improvement

Safer Staffing Report

	Assessment			Medical			Stroke			Surgical		
	Current month	Last month	Year to date	Current month	Last month	Year to date	Current month	Last month	Year to date	Current month	Last month	Year to date
Day fill rate	104	80	99	101	79	104	96	86	87	94	101	104
Night fill rate	94	70	101	105	104	93	72	97	100	85	94	71
Sickness	20	39	24	30	36	32	39	29	38	27	37	28
Vacancy	23	21	35	39	37	37	26	39	21	39	30	21

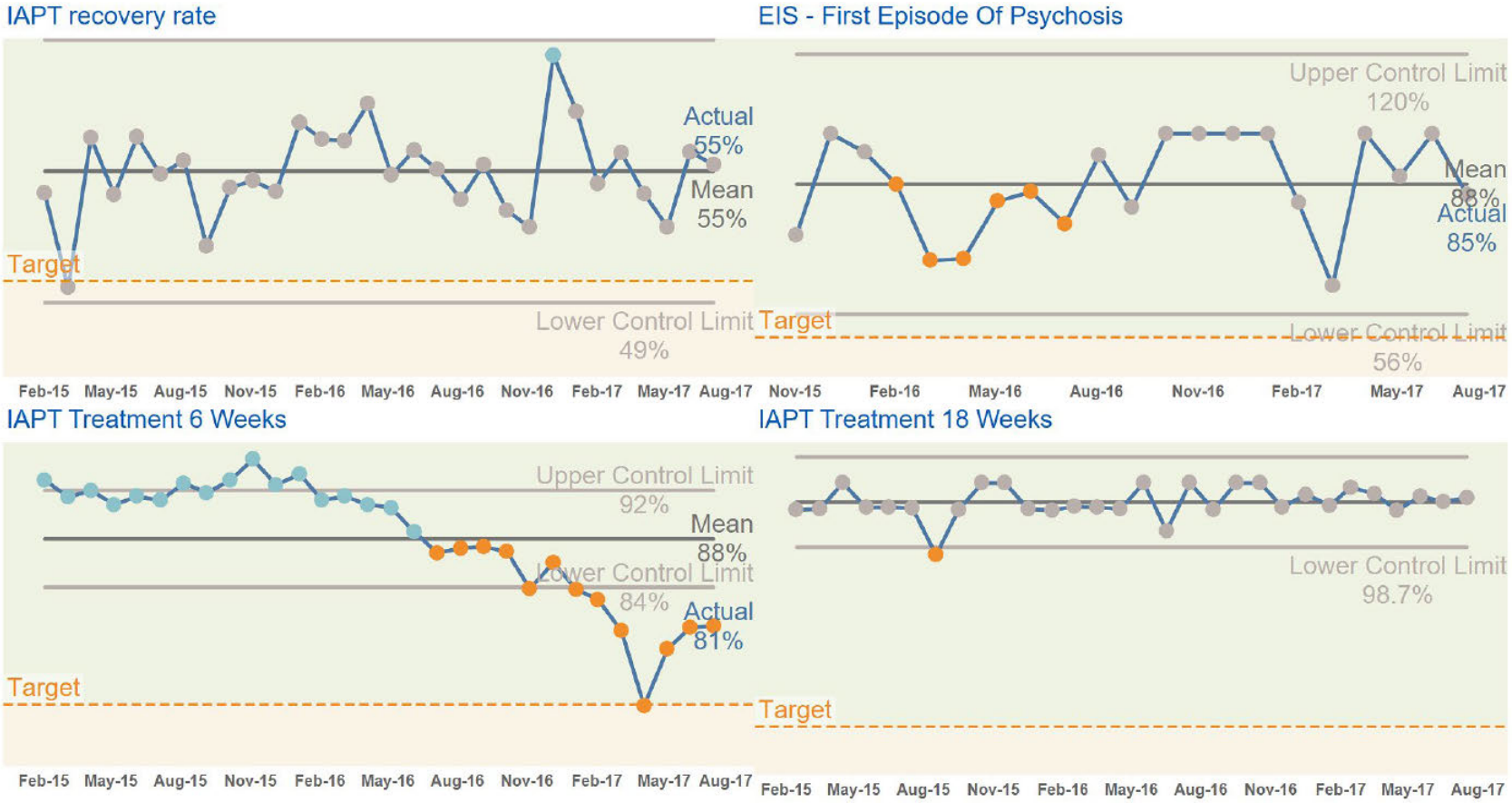
7 | Making data count

NHS Improvement

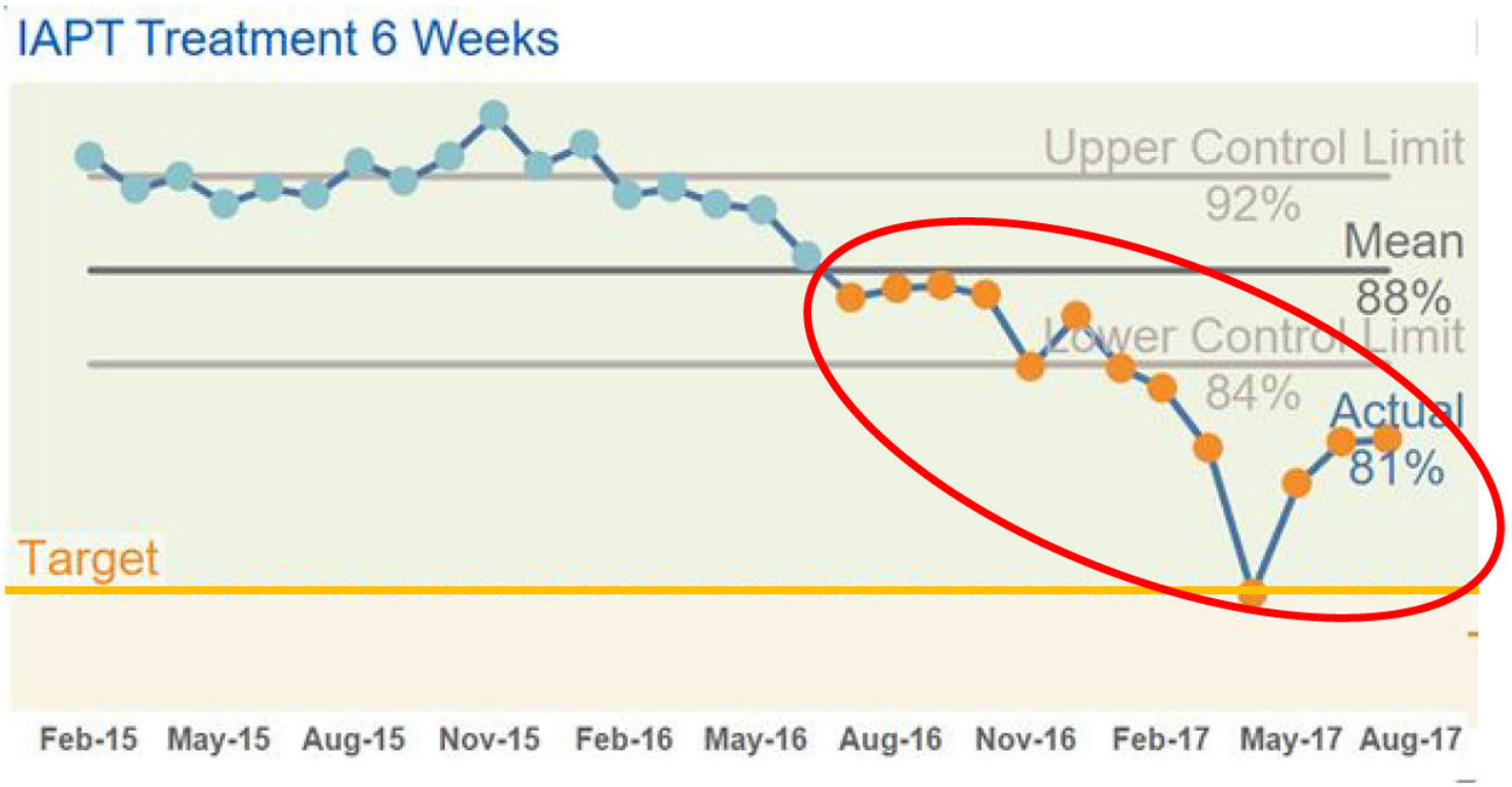
Improving Access to Psychological Therapies – performance against target

Metric	Target	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17
IAPT Treatment 18 Weeks	95%	100.0%	99.5%	99.9%	99.8%	99.4%	99.7%	99.6%	99.7%
IAPT Treatment 6 Weeks	75%	86%	84%	83%	81%	75%	80%	81%	81%
IAPT Recovery Rate	50%	59%	57%	54%	55%	54%	52%	55%	55%
EIS First Episode Psychosis	50%	100%	100%	83%	63%	100%	89%	100%	85%

NHS Improvement



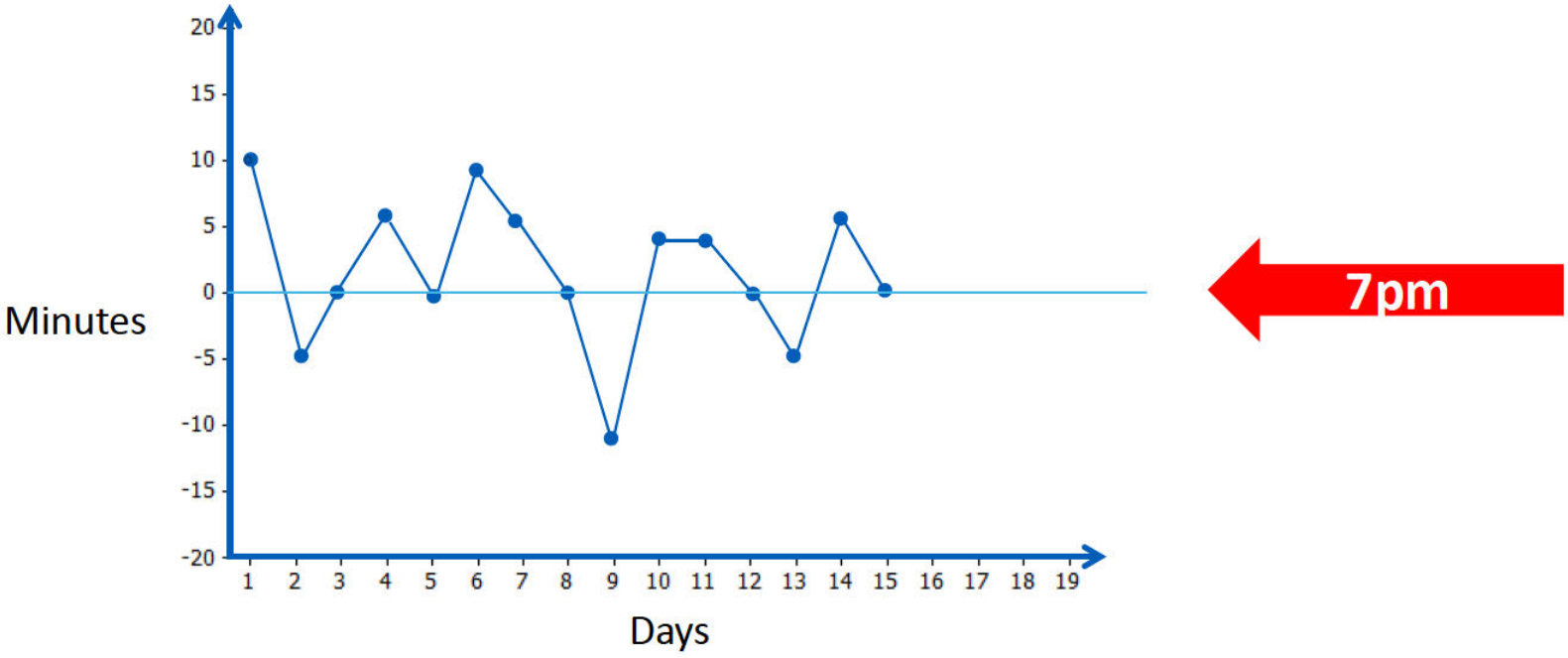
Did green provide true assurance?



Introducing John and Mary

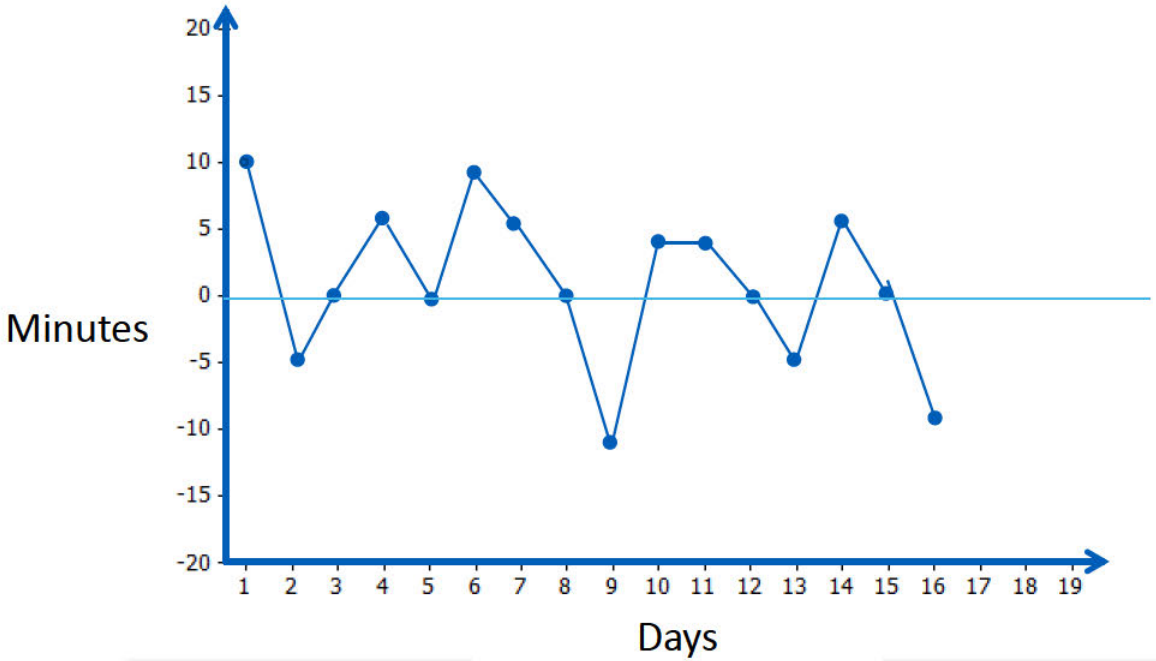
NHS Improvement

NHS Improvement



Now John comes back...

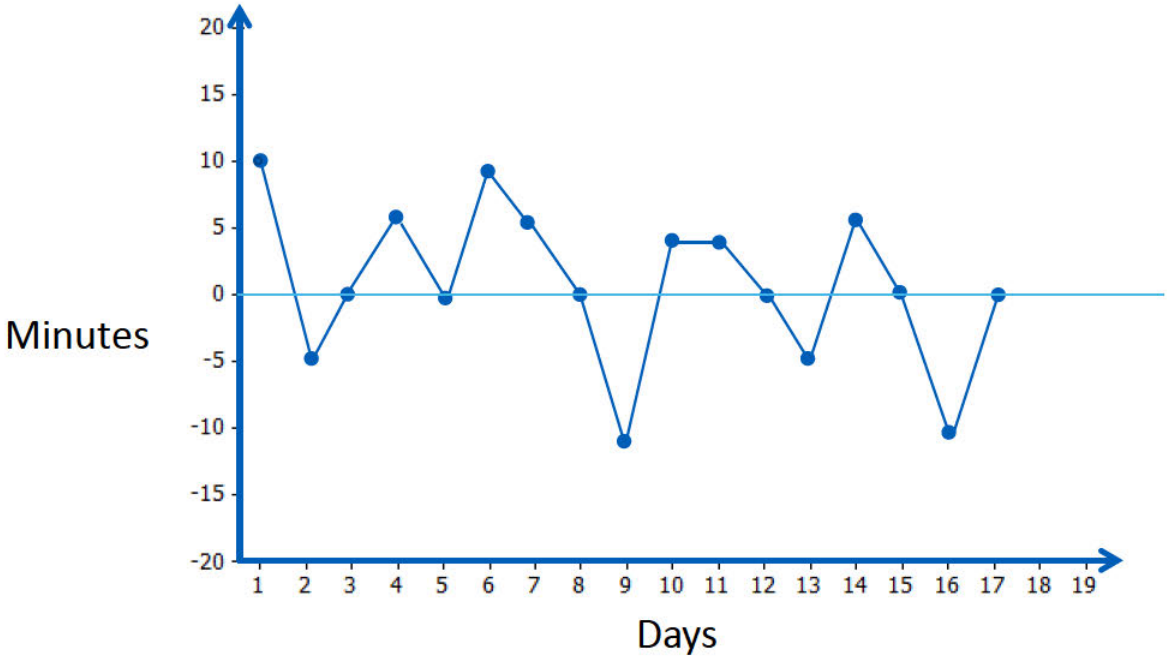
NHS Improvement



Mary arrives at 18:50

John asks, why have you arrived 10 minutes early?

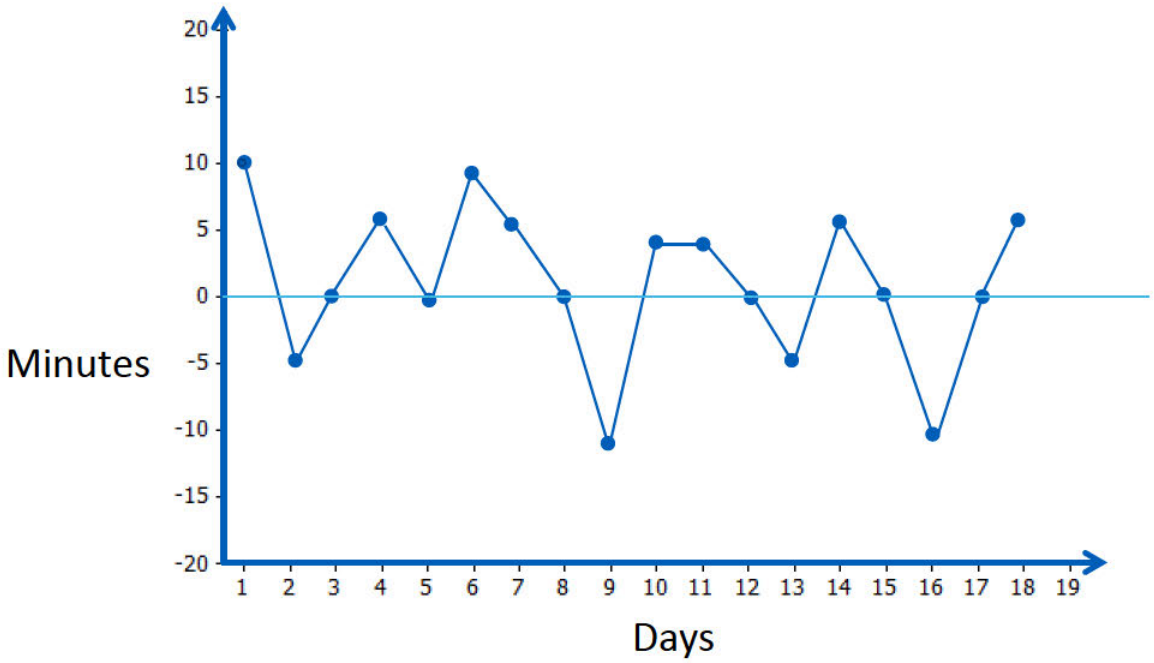
NHS Improvement



Mary arrives at 19:00.

John asks: yesterday you arrived at 18.50 – why have you arrived at 19:00 today?

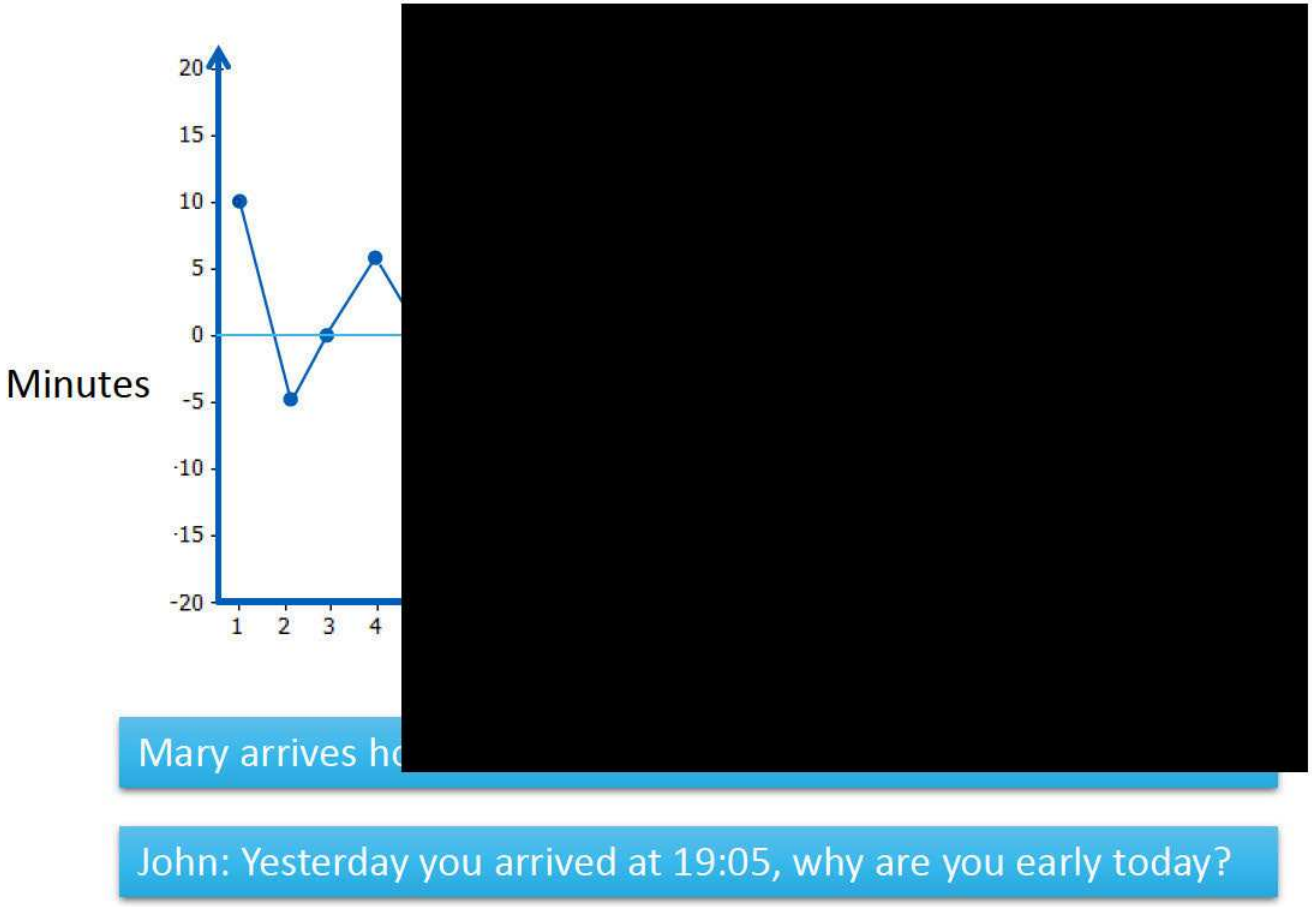
NHS Improvement



Mary arrives at 19:05

John asks: yesterday you arrived at 7pm – why are you late?

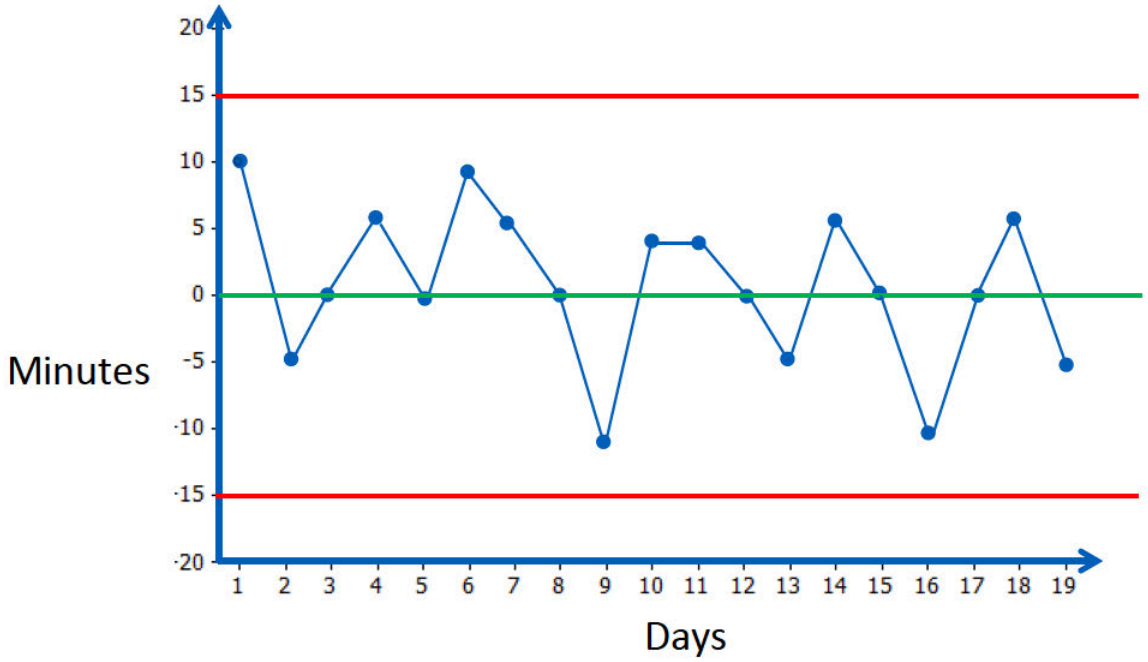
NHS Improvement



16 | Making data count

Thoughts on John & Mary ?

NHS Improvement



Frequently seen in the NHS

NHS Improvement

Spuddling

To make a lot of fuss about trivial things, as if they were important

Tampering

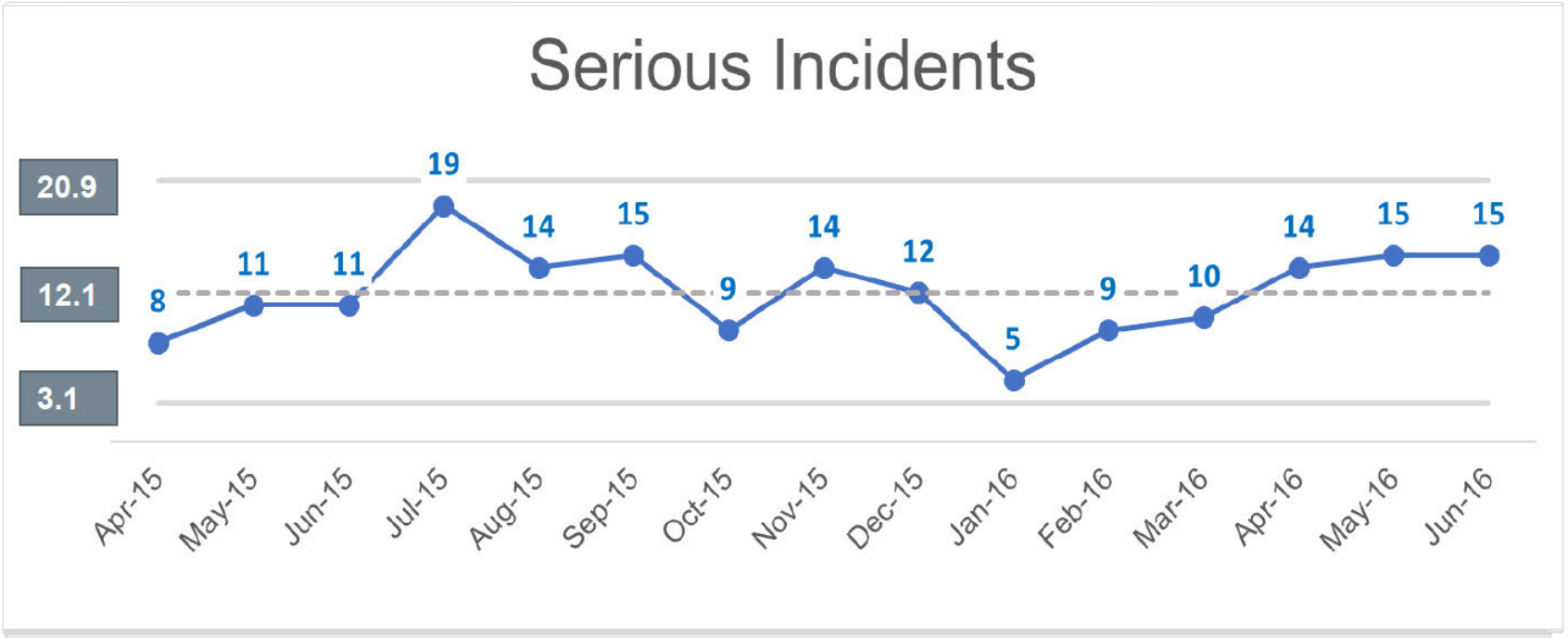
Scenario

We're going to simulate some **real data** in a healthcare setting

We'll be thinking about **how people react to patterns and trends** in data

Can you spot an **improvement or decline** when it occurs? We'll begin plotting our data in a **run chart**

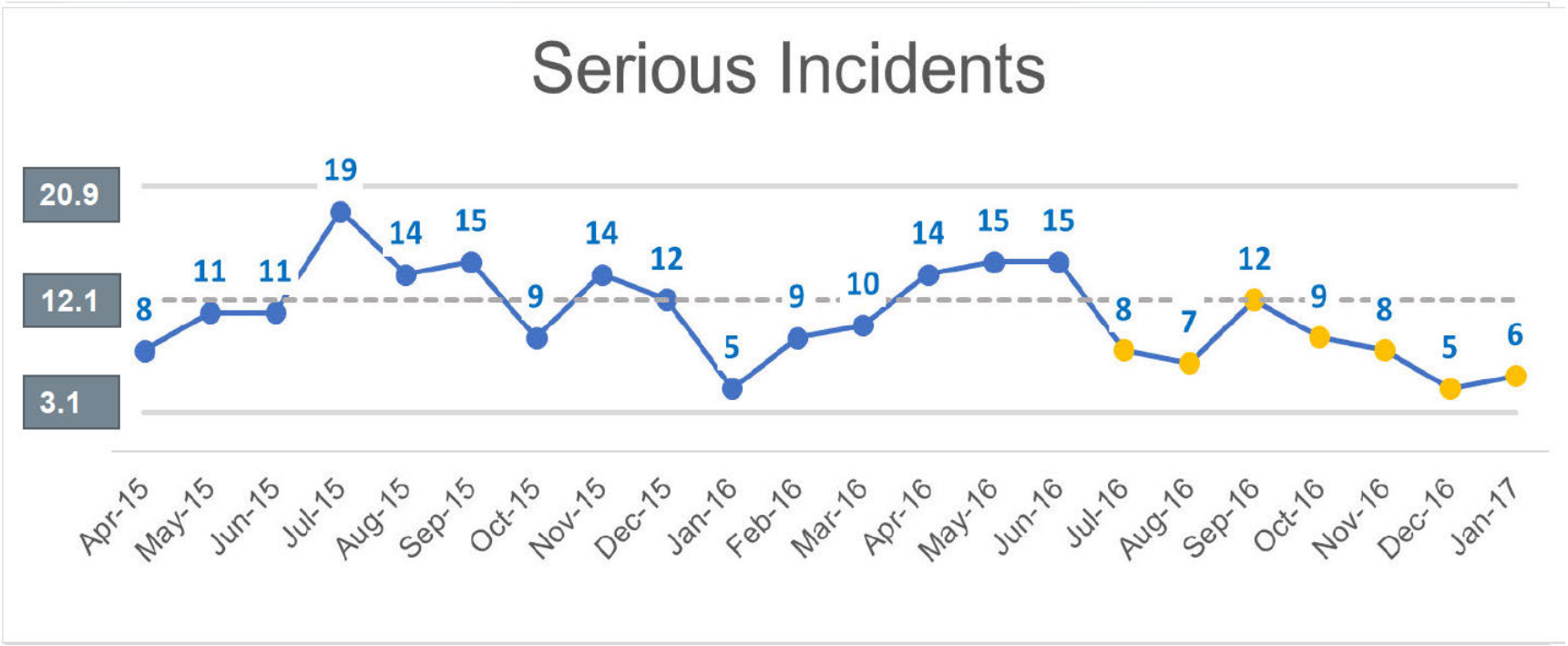
NHS Improvement



We now have enough data for robust process limits, lets change our run chart to an SPC chart

20 | Presentation title

NHS Improvement

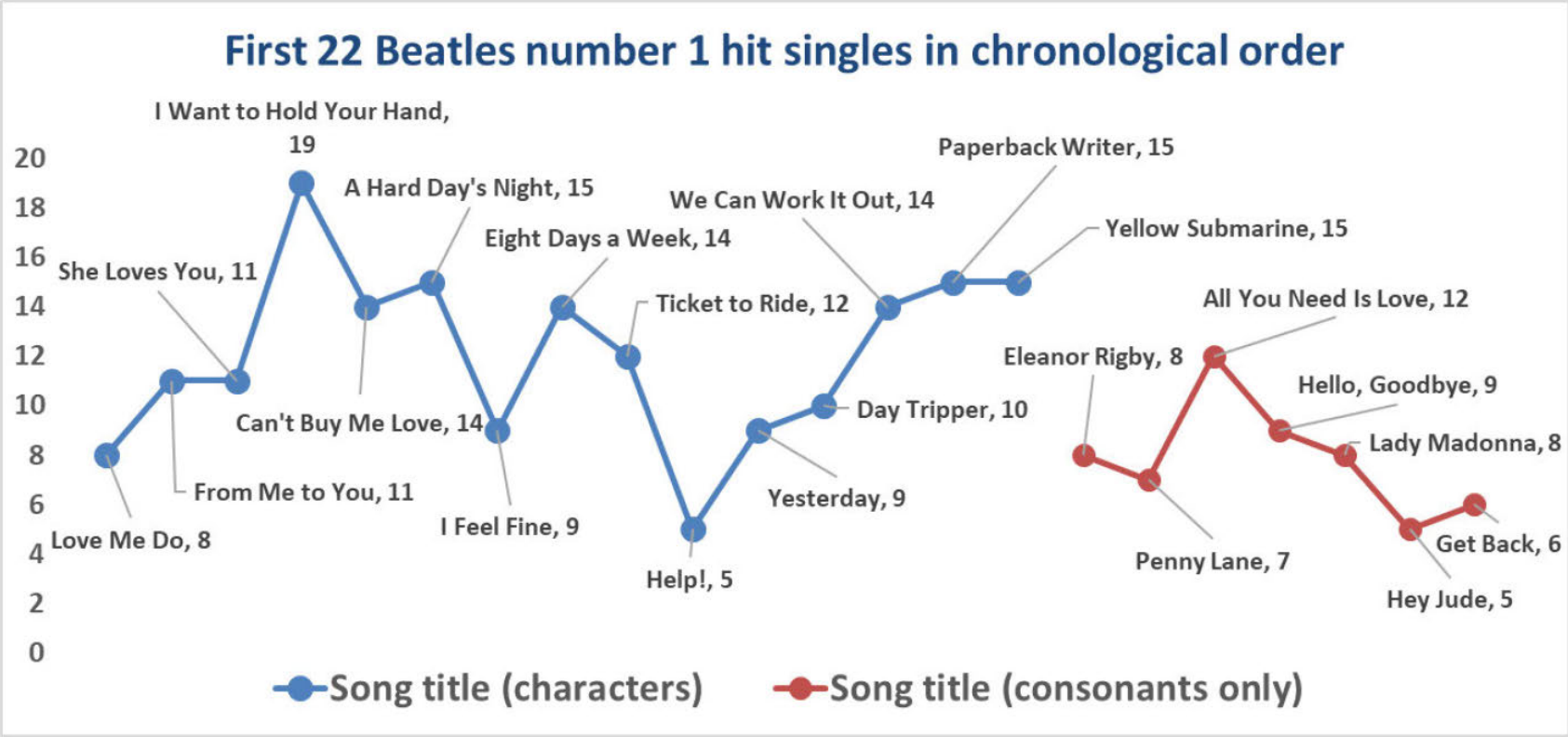


7 points below mean line put your hand if you think the improvement is successful

21 | Presentation title

NHS Improvement

This data set was randomly generated using the number of letters and consonants in Beatles number 1 singles



Strong evidence base

NHS Improvement

THE PROBLEM WITH...

The problem with red, amber, green: the need to avoid distraction by random variation in organisational performance measures

Jacob Anhøj, Anne-Marie Blok Hellesøe

Centre for Diagnostic Investigation, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

Correspondence to
Dr Jacob Anhøj, Centre for Diagnostic Investigation, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, Copenhagen 2100, Denmark; jacob@anhoej.net

Accepted 18 January 2016
Published Online First 31 March 2016

Linked

► http://dx.doi.org/10.1136/bmjqs-2015-004967
► http://dx.doi.org/10.1136/bmjqs-2016-005503

CrossMark

To cite: Anhøj J, Hellesøe A-MB. *BMJ Qual Saf* 2017;**26**:81–84.

'The Problem with...' series covers controversial topics related to efforts to improve healthcare quality, including widely recommended but deceptively difficult strategies for improvement and pervasive problems that seem to resist solution.

INTRODUCTION

Many healthcare organisations now track a number of performance measures like infection and complication rates, waiting times, staff adherence to guidelines, etc. Our own organisation, The Capital Region of Denmark, provides healthcare for 1.7 million people and runs 6 hospitals and 11 mental health centres. Measures of clinical quality have been widely used in our region locally at hospitals and departments for many years. Recently, our region started to systematically define and track strategical key performance measures also at the top management level. Approximately 25 measures on a wide range of subjects from hospital infections to public transportation are being tracked by the top management and the Regional Council.

The measurement strategy for hospitals involves a bottom-up approach allowing each hospital and department to, if needed, define its own performance measures that feed into one or more of the overall measures. For example, bacteraemia is one of the overall measures, and some acute-care departments, who rarely see hospital-acquired bacteraemia, have started to work on reducing the use of bladder catheters in order to reduce the risk of bacteraemia from catheter-related urinary tract infections diagnosed after their patients have been transferred to other departments. To support their work, they have developed a handful of measures that track the use of catheters and staff compliance with standard procedures related to catheter use.

We welcome this development very much. The choice of relatively few overall measures combined with the bottom-up approach is a helpful strategy that focuses and aligns improvement work and stimulates the use of data at all levels of the organisation while leaving room for meaningful local adaptations of performance measures.

However, we do not at all welcome the widespread use of red, amber, green approaches to data analysis that is everywhere in our organisation.

By 'red, amber, green', we are referring to graphical data displays that use colour coding of individual data values based on whether this value is on the right (green) or wrong (red) side of a target value. Often amber or yellow is used to indicate data values that are somewhere between 'right' and 'wrong'.

The problem with red, amber, green management is that at best is it useless, at worst it is harmful.

THE PROBLEM WITH RED, AMBER, GREEN

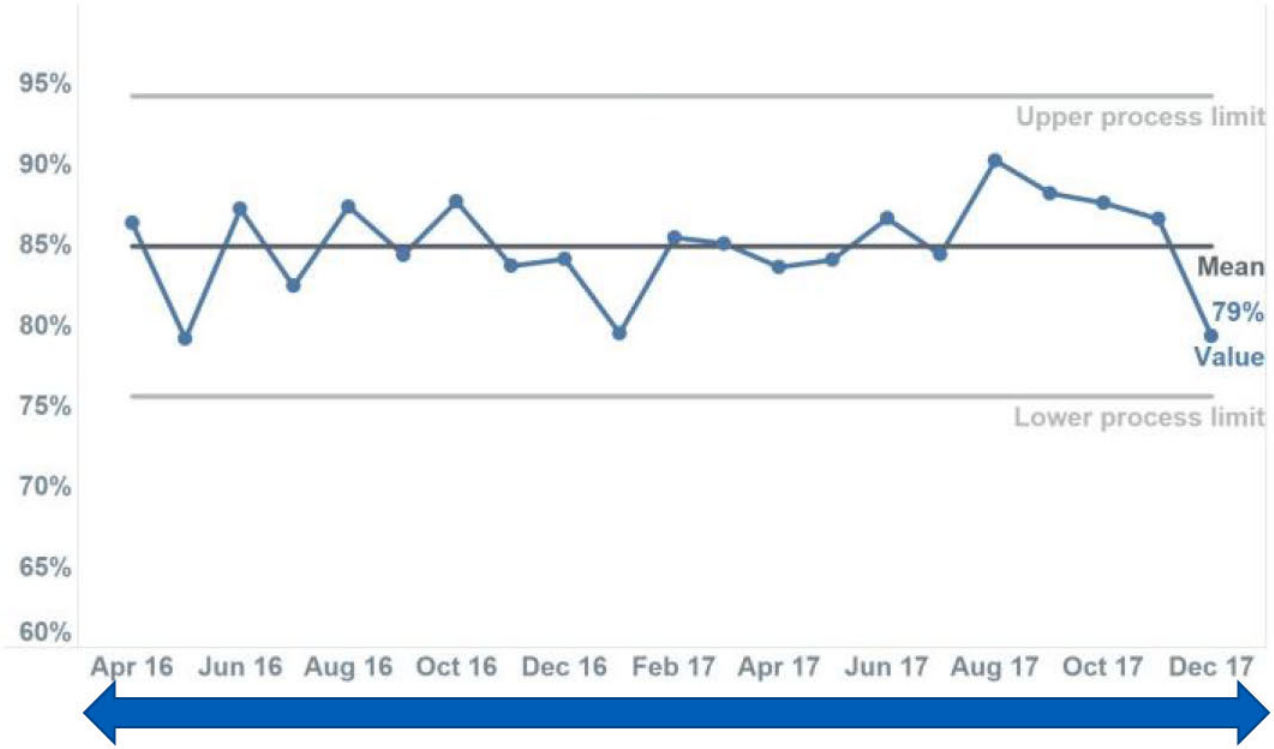
Figure 1 was captured from the February 2015 report on regional performance measures. It shows the monthly count of a certain type of unwanted incident in mental healthcare. The horizontal line represents the target value of 10.5. That is, we do not want more than 10 incidents per month. Red bars show months above target. Green bars show months below target.

The data display in figure 1 is formally correct (green is better than red). However, it fails to convey a very

The anatomy of a SPC chart

NHS Improvement

Time series line chart with 3 reference lines



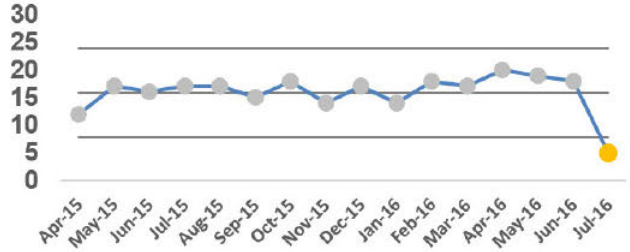
≈ 99% of data

20 plus data points for a robust analysis

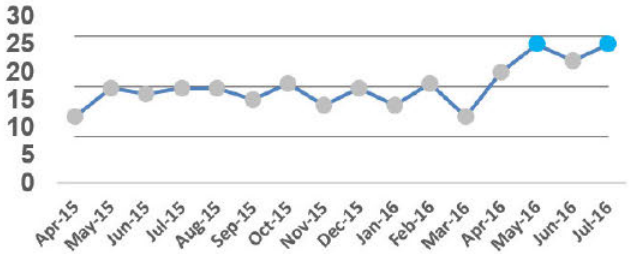
SPC rules : special cause variation

NHS Improvement

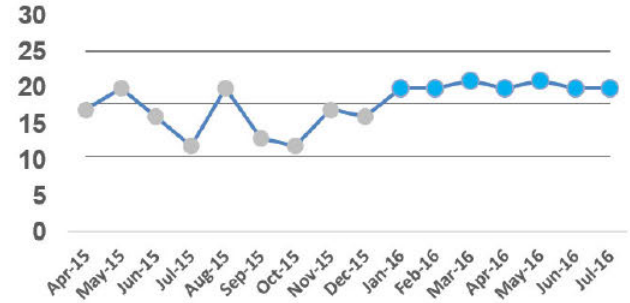
A single data point outside the process limits



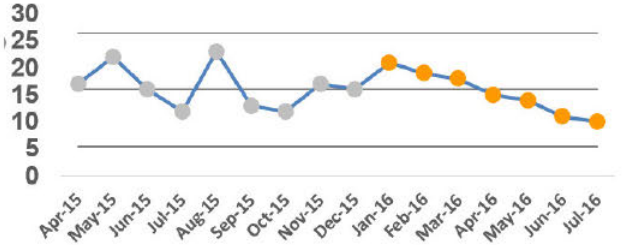
Two out of three points close to the process limits



Shift of points above / below mean line



Run of points in consecutive ascending / descending order



Why is 7 points significant?

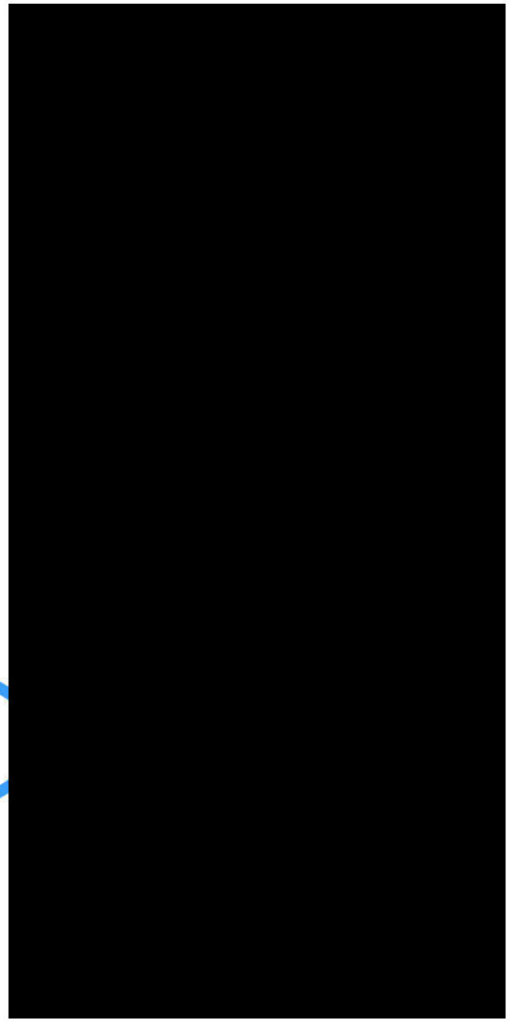
NHS Improvement

A trend of 2 has the probability of 25% occurrence (**one in four**)

A trend of 4 has the probability of 6.25% occurrence (**one in sixteen**)

A trend of 6 has the probability of 1.56% occurrence (**one in sixty-four**)

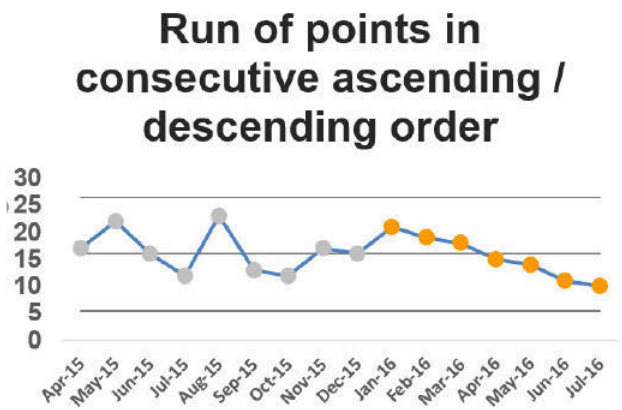
A trend of 7 has the probability of 0.8% occurrence (**one in one hundred and twenty-eight**)



If there is special cause…..

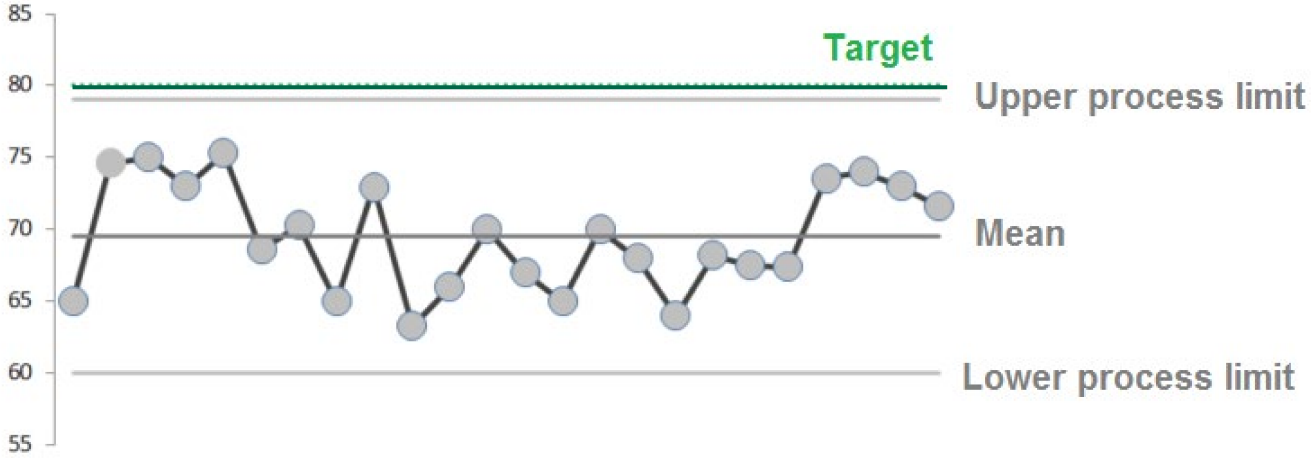
NHS Improvement

[redacted]



NHS Improvement

In control but unacceptable variation (common cause variation)

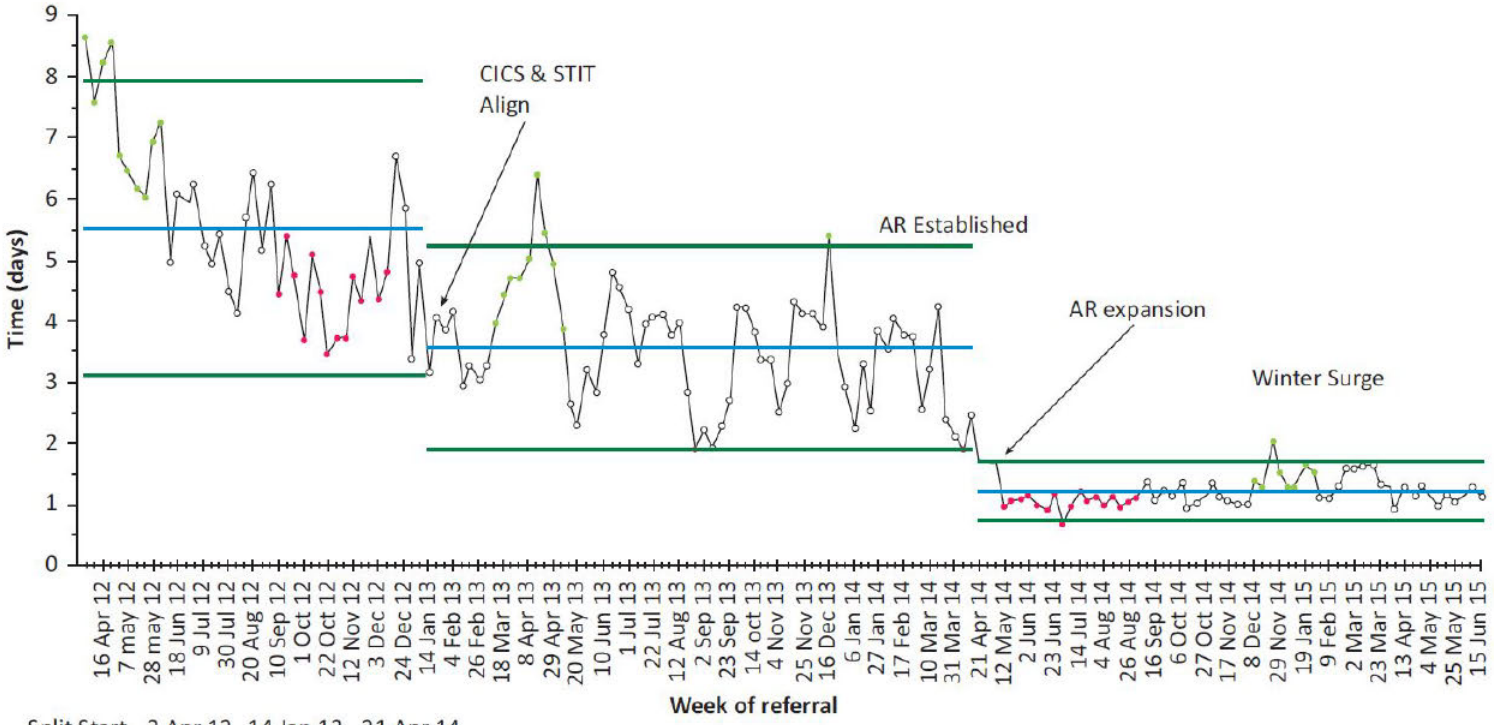


Redesign the system

Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

Has the change worked?

NHS Improvement

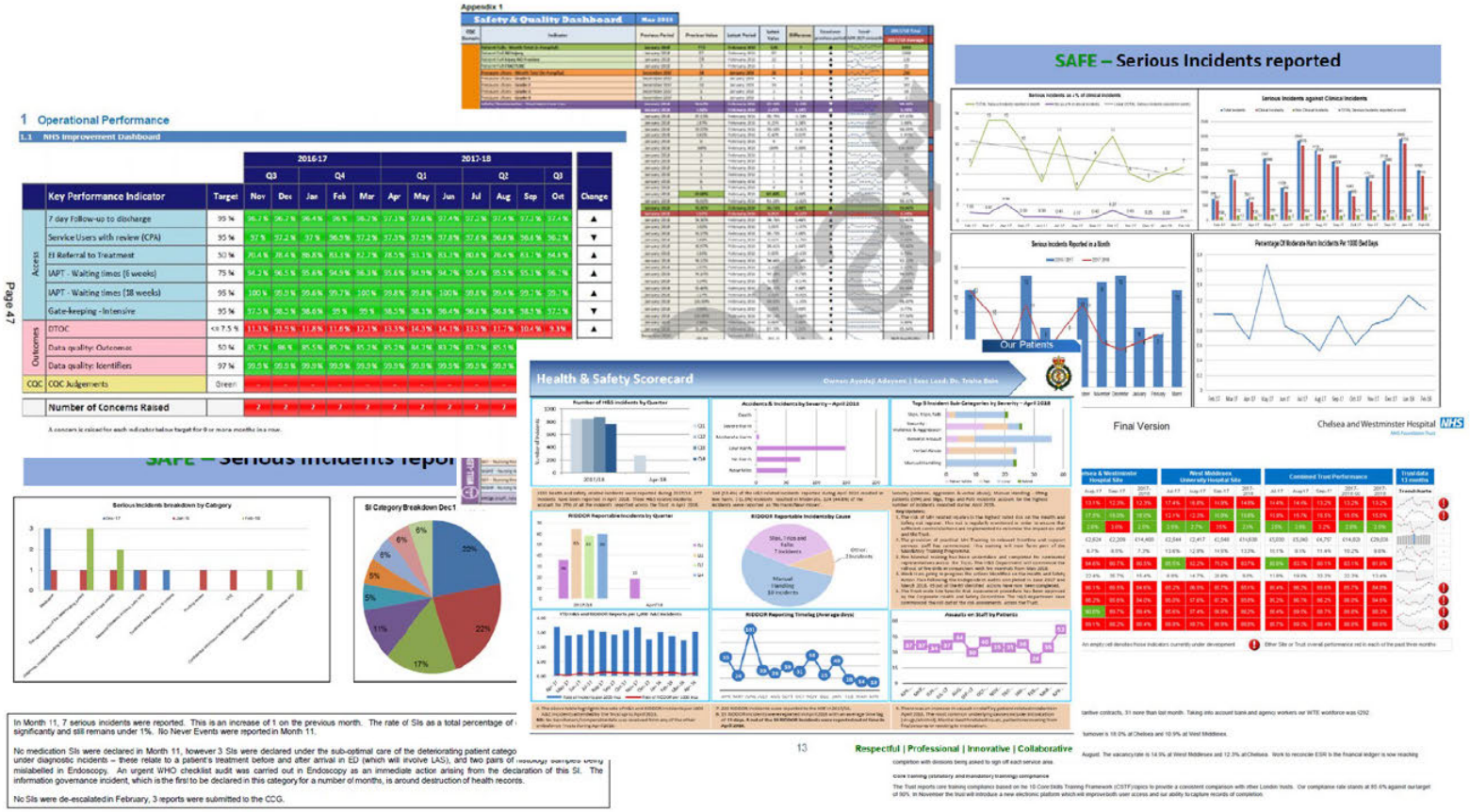


Split Start	2 Apr 12	14 Jan 13	21 Apr 14
U.C.L.	−7.9	5.2	1.7
Mean	−5.5	6.6	1.2

Fig 2. Reducing patient wait for active recovery from a hospital bed. AR = Active Recovery; CICS = Community Intermediate Care Service; STIT = Short Term Intervention Team

What extra insight could SPC provide?

NHS Improvement



Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

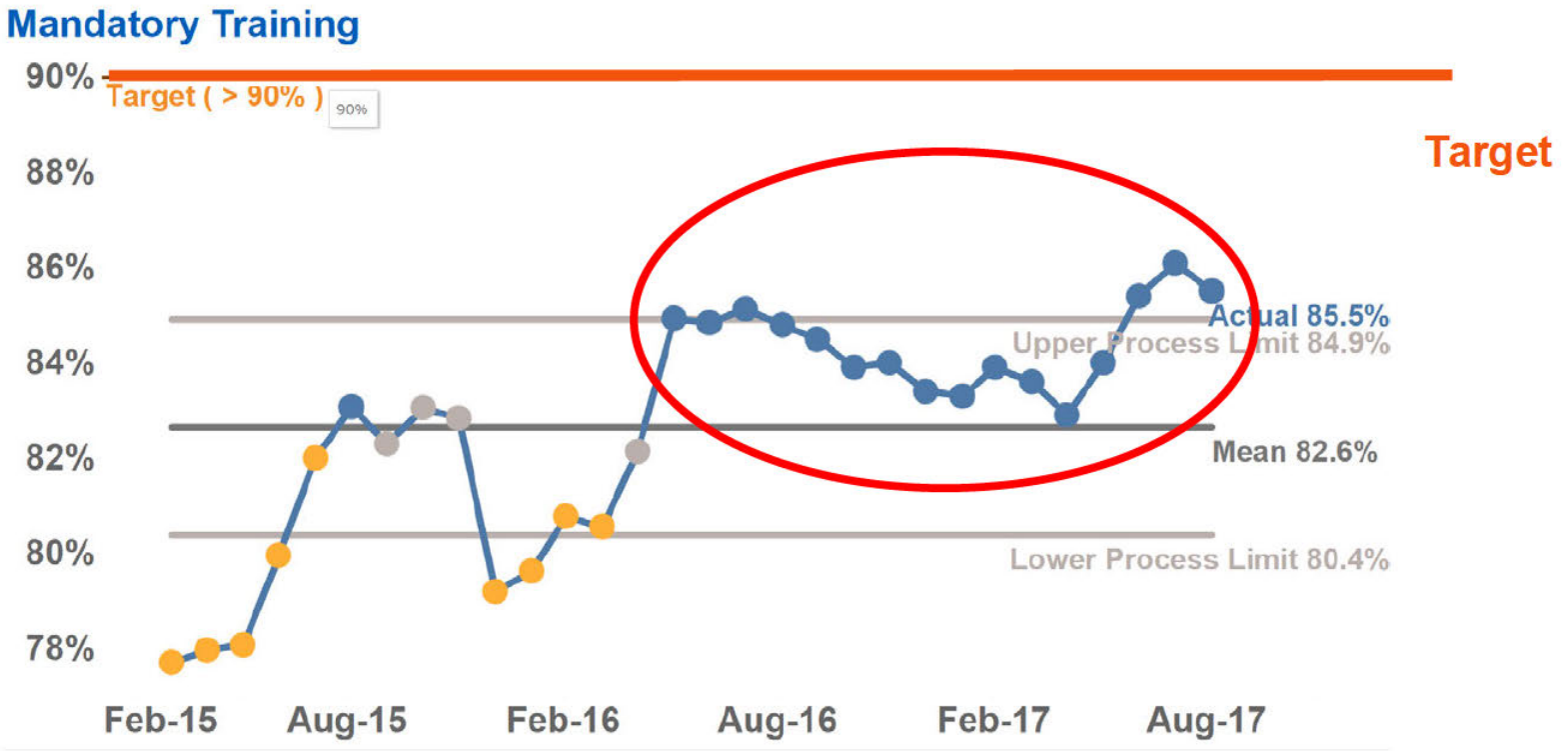
What do you think when you see this?

NHS Improvement

Domain	Indicator	Jul-17	Aug-17	Sep-17	2017-2018	Jul-17	Aug-17	Sep-17	2017-2018	Jul-17	Aug-17	Sep-17	2017-2018 Q2	2017-2018	Trust data 13 months Trend charts
Training	Mandatory training compliance (Target: >90%)	85.4%	86.1%	85.5%	84.6%	85.2%	86.5%	85.7%	85.1%	85.4%	86.2%	85.6%	85.7%	84.8%	

Presentation influences discussion

NHS Improvement



Can you spot improvement?

NHS Improvement

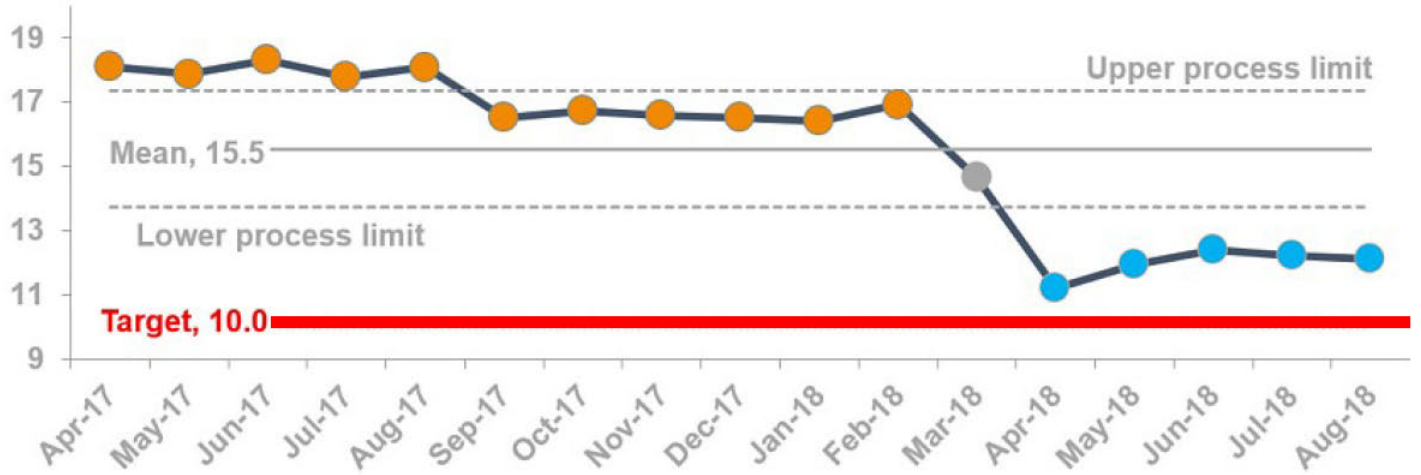
Turnover trust wide (target 10%) *source ESR*

This remains high for a number of factors, which includes service decommissioning and termination of a number of fixed term contract worker across numerous operational and corporate services.

Quarter 1						Quarter 2			
Apr-17	Apr-18	May-17	May-18	Jun-17	Jun-18	Jul-17	Jul-18	Aug-17	Aug-18
18.08%	11.19%	17.86%	11.95%	18.31%	12.40%	17.91%	12.20%	18.15%	12.10%

Improvement through the red

NHS Improvement



Turnover trust wide (target 10%) ***source ESR***

This remains high for a number of factors, which includes service decommissioning and termination of a number of fixed term contract worker across numerous operational and corporate services.

Quarter 1						Quarter 2			
Apr-17	Apr-18	May-17	May-18	Jun-17	Jun-18	Jul-17	Jul-18	Aug-17	Aug-18
18.08%	11.19%	17.86%	11.95%	18.31%	12.40%	17.91%	12.20%	18.15%	12.10%

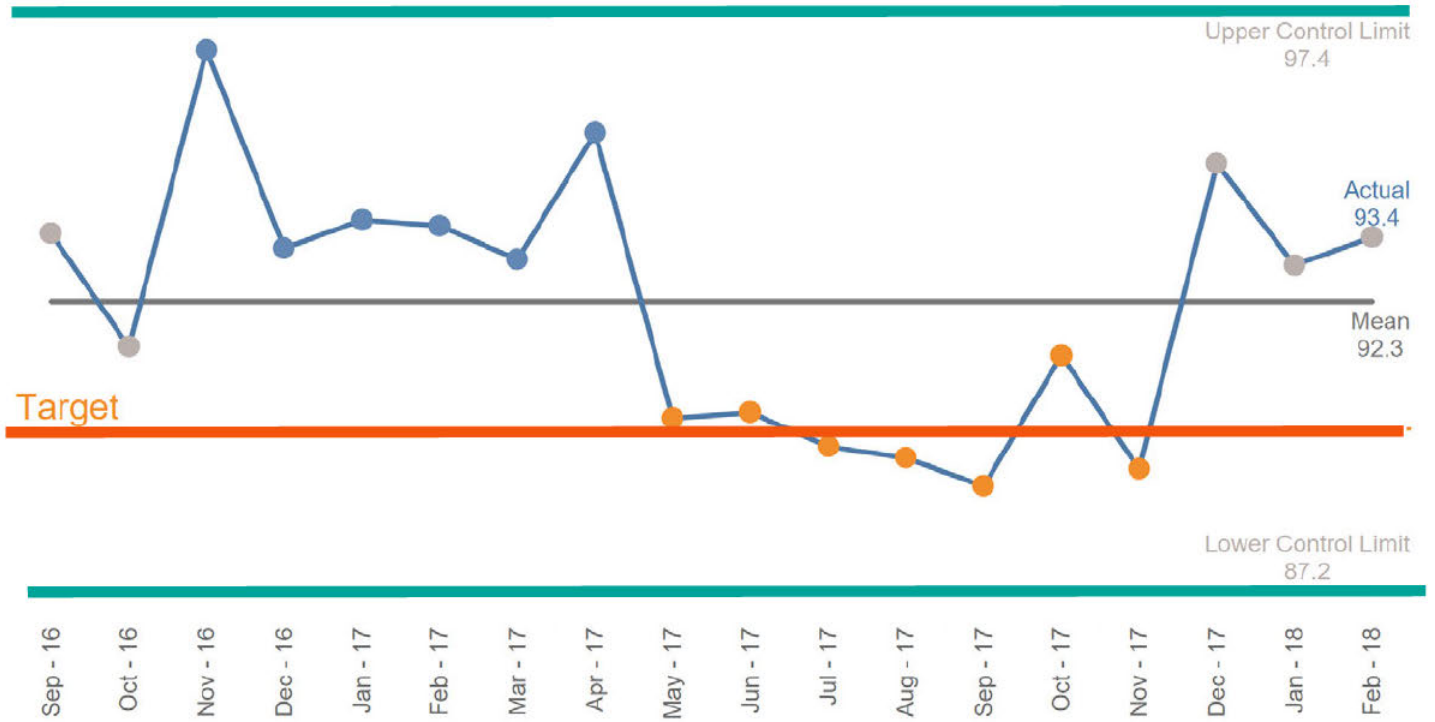
Encourages knee jerk reactions?

NHS Improvement

Caring Standards	Month 10	Month 11	Month 12	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	FYTD Actual	YTD Target	Trend on Month
Friends and Family Test - % Likely to Recommend A&E	93.7	93.6	93.%	95.24	90.2	90.3	89.7	89.5	89.0	91.31	89.8	94.7	92.9	93.4	90.49	90.00	↑

System not capable

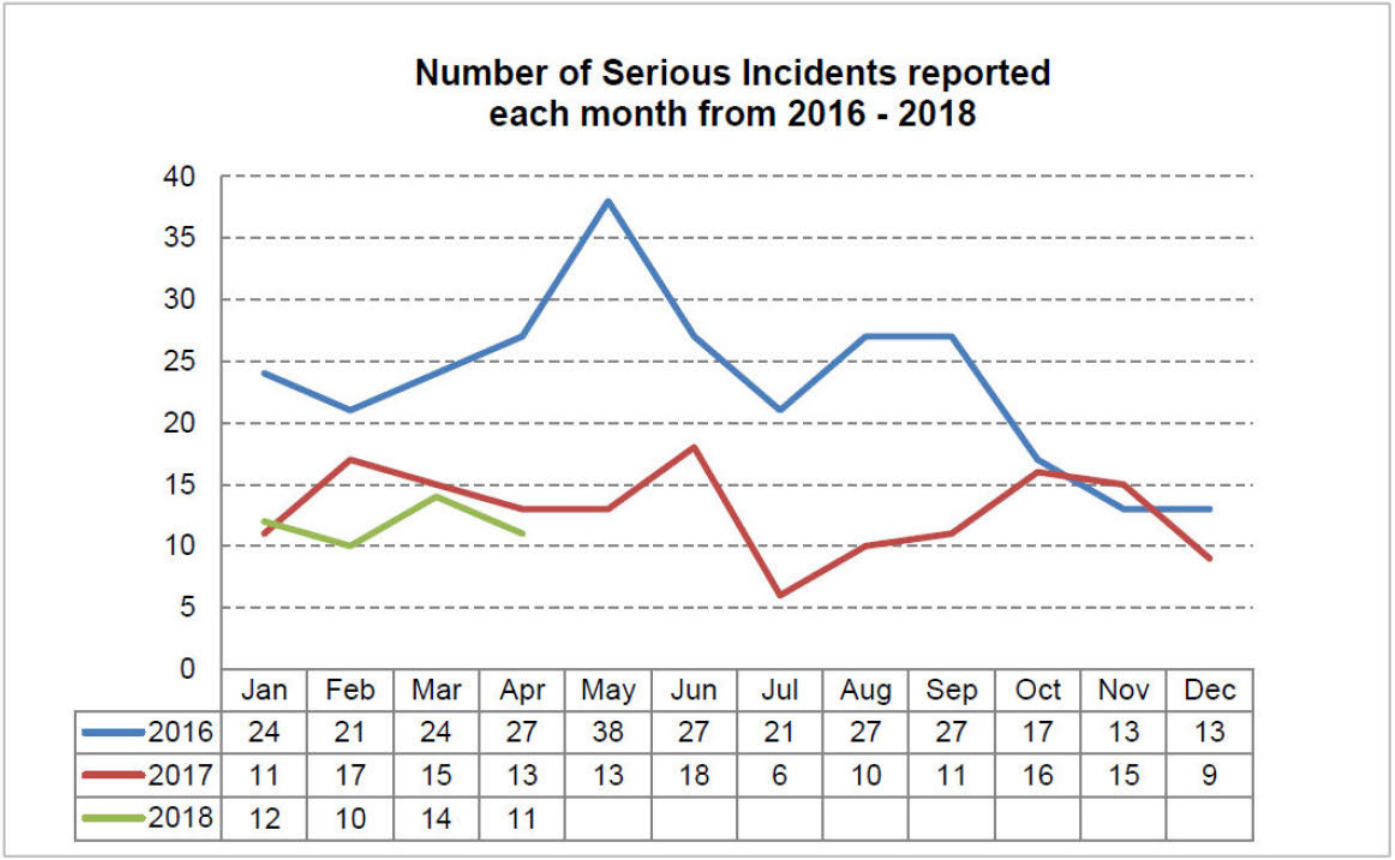
NHS Improvement



Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

Serious incidents : 3 years

NHS Improvement

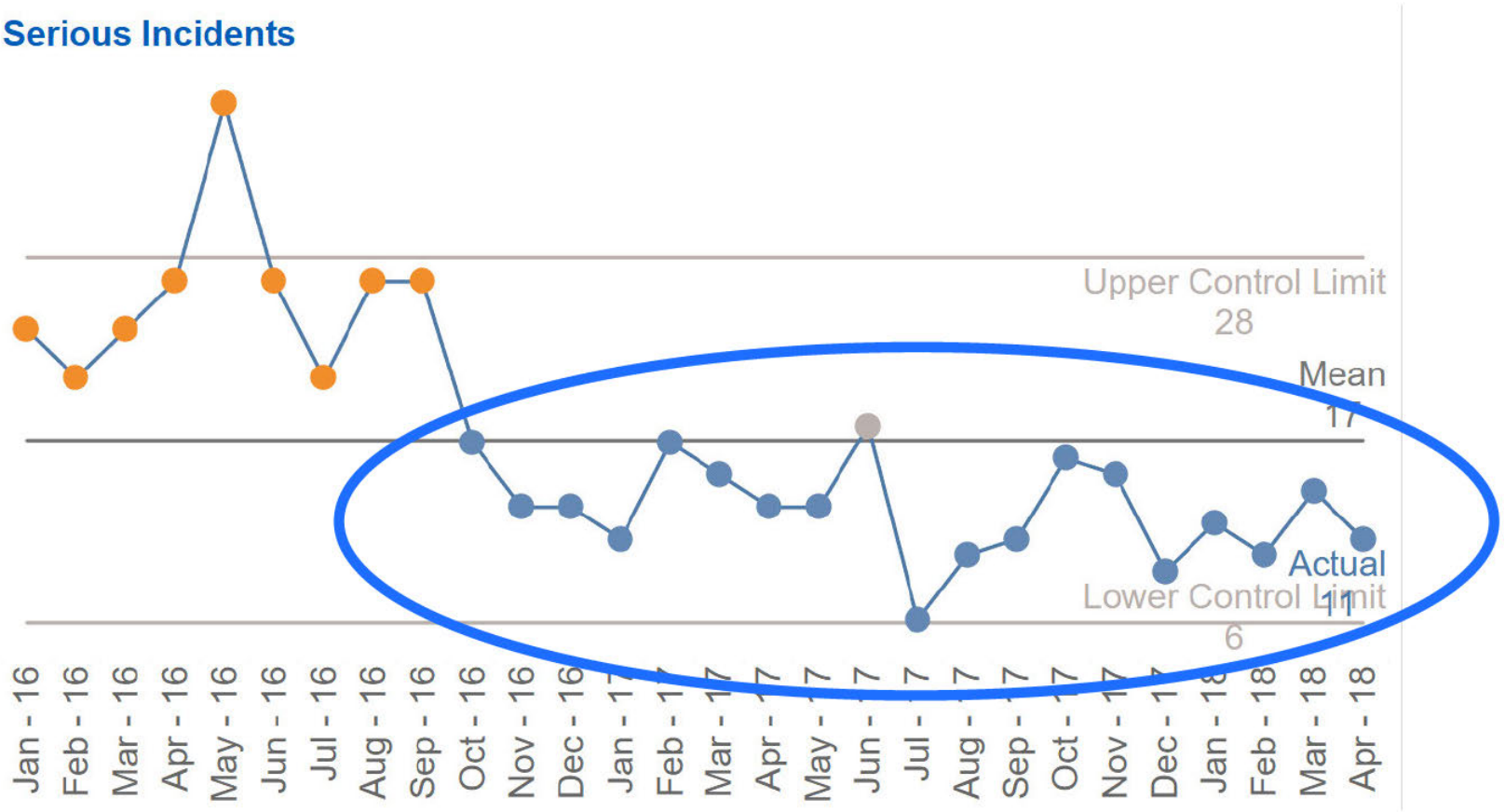


	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2016	24	21	24	27	38	27	21	27	27	17	13	13
2017	11	17	15	13	13	18	6	10	11	16	15	9
2018	12	10	14	11								

Improvement (?)

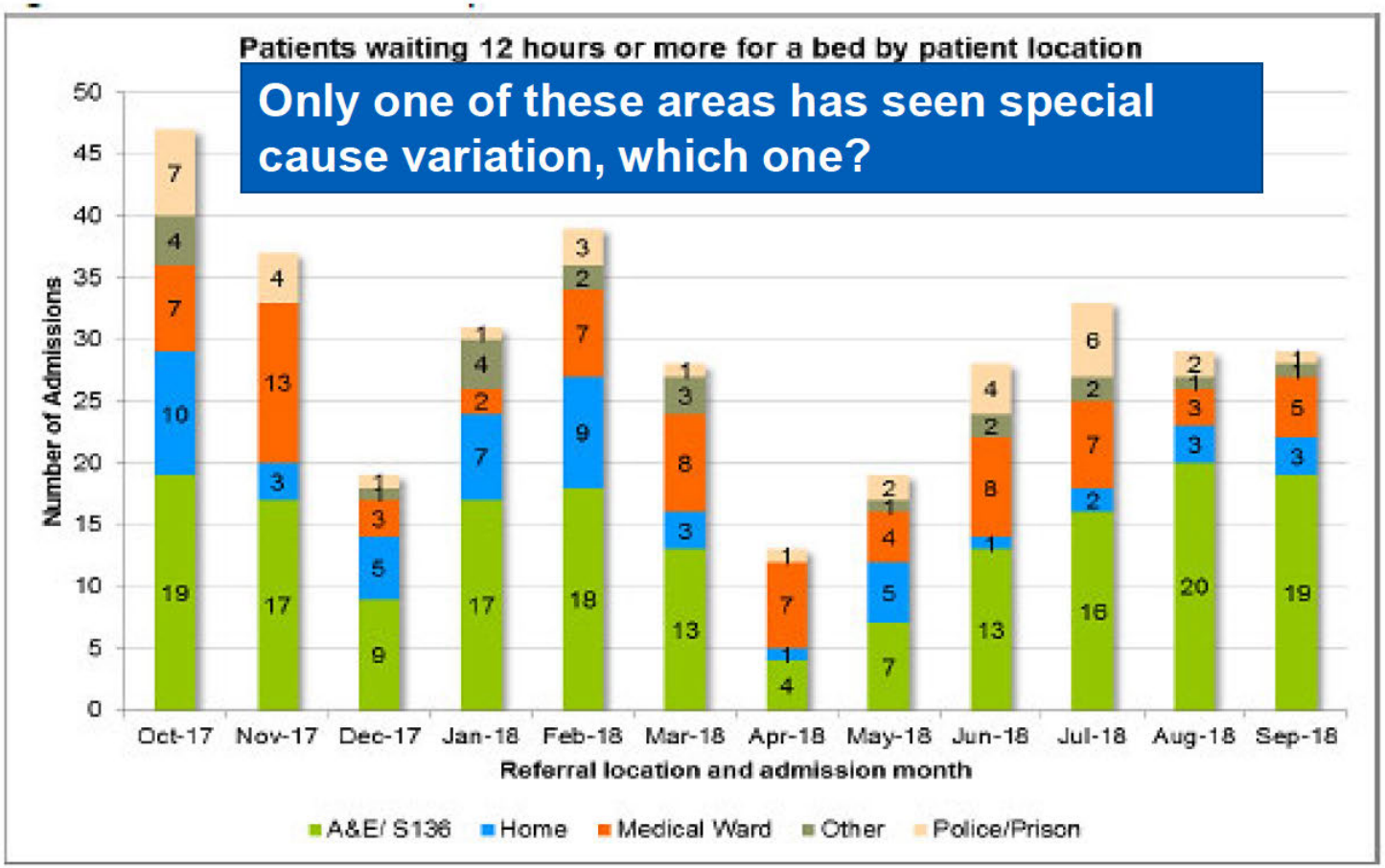
NHS Improvement

Serious Incidents



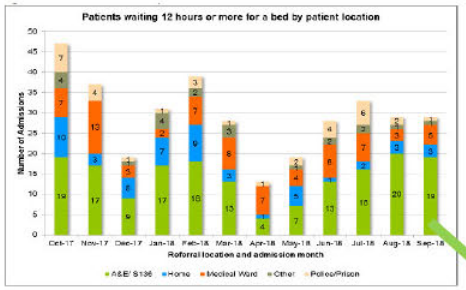
Spotting improvement and decline

NHS Improvement

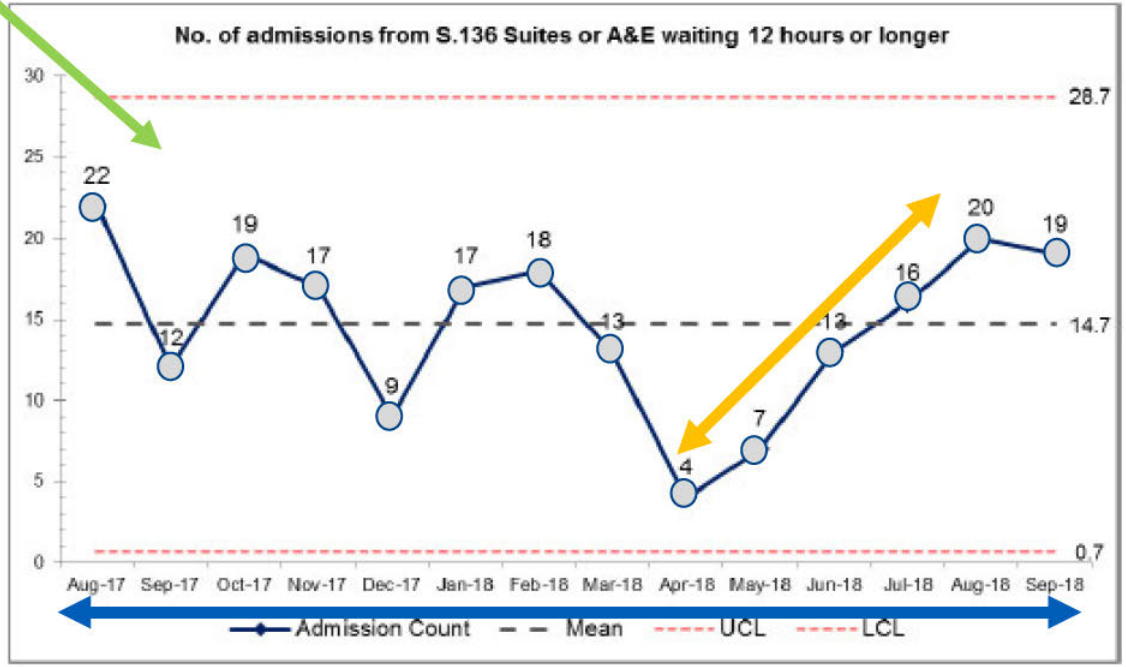


Was it green?

NHS Improvement



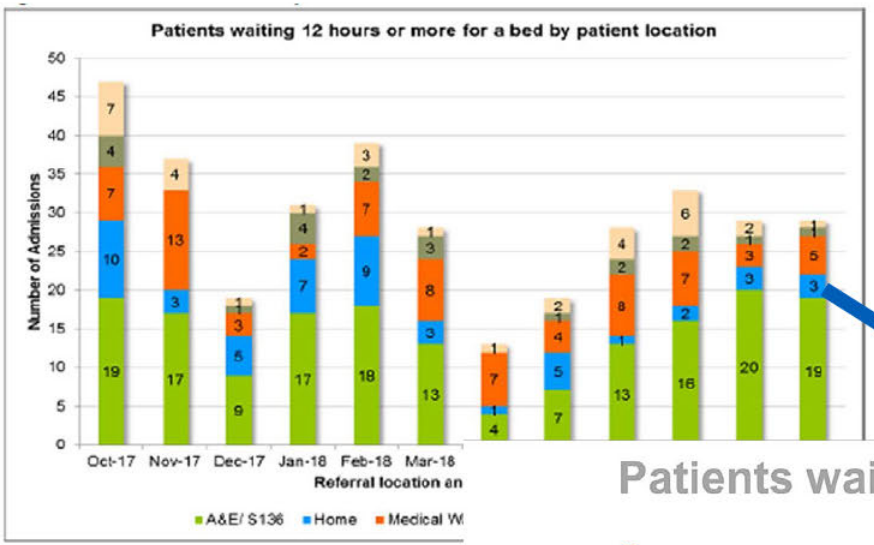
Is this significant? Count the dots….



Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

What was significant?

NHS Improvement



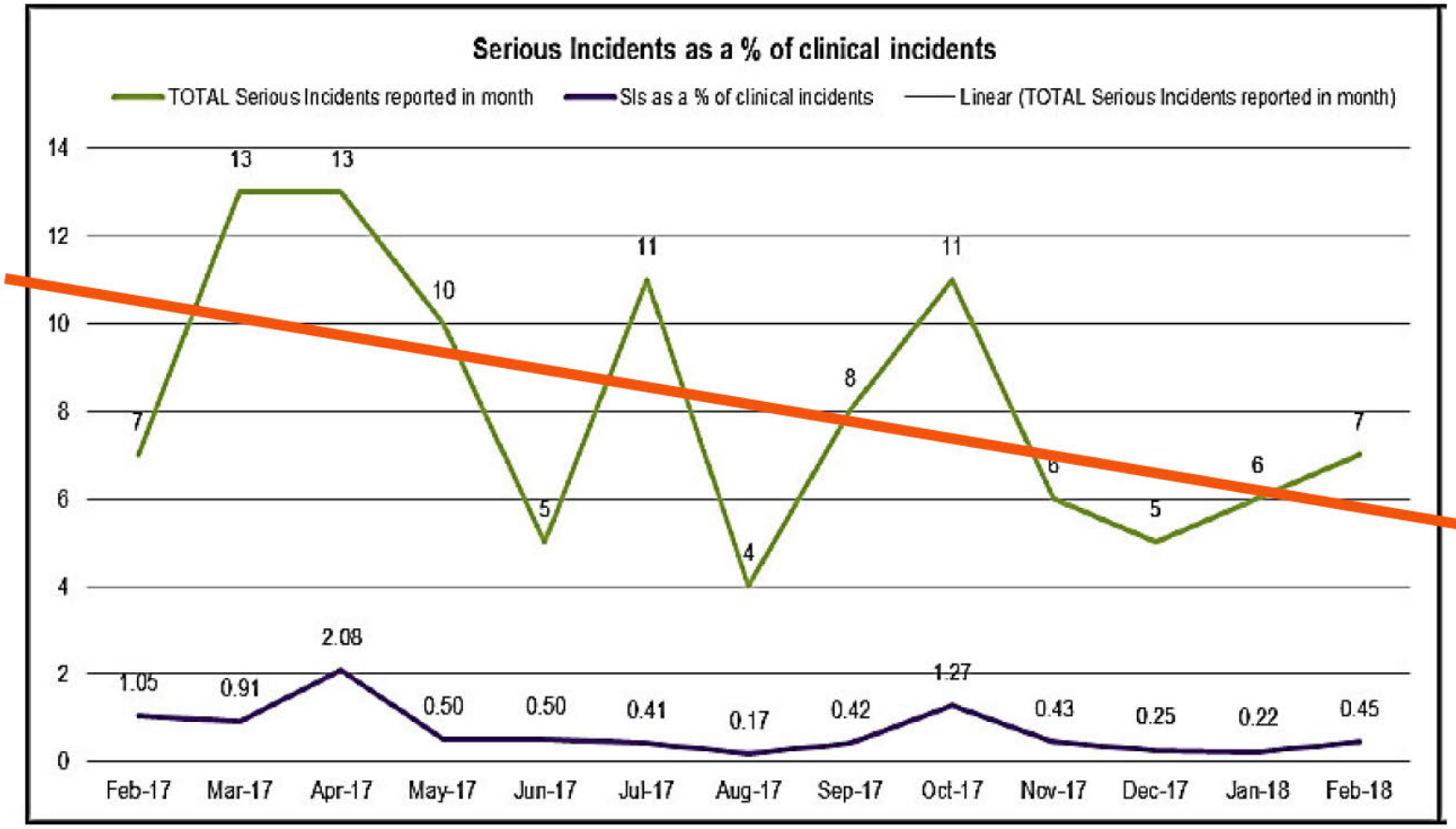
It wasn't Green

It was Blue

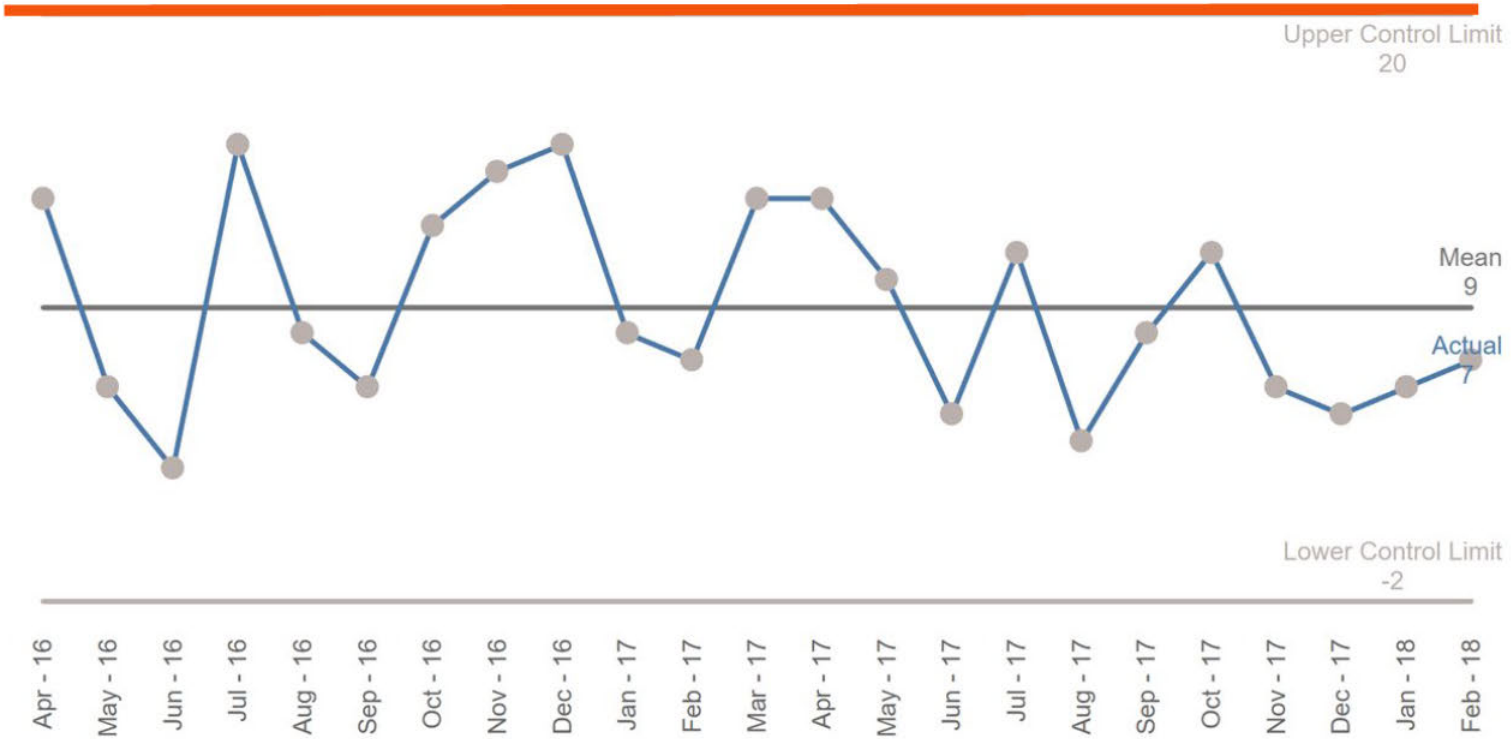
7 months of improvement

Are things improving?

NHS Improvement

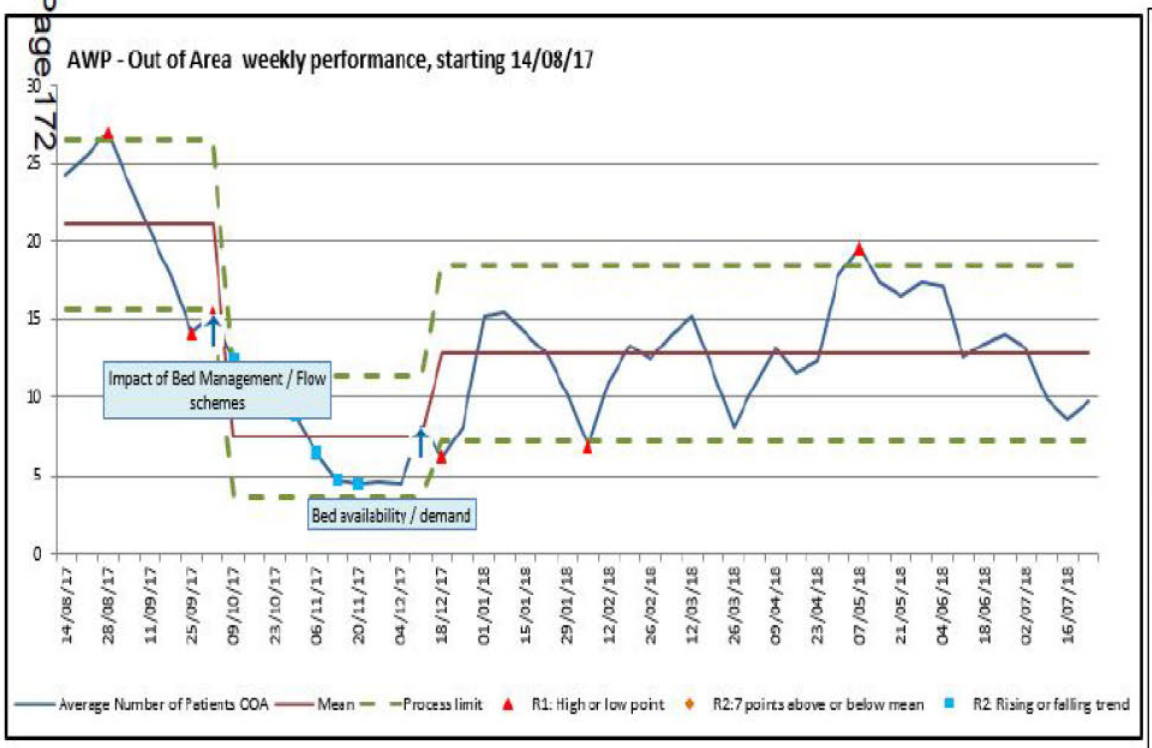


20 serious incidents a month acceptable?



Changes being made at Avon

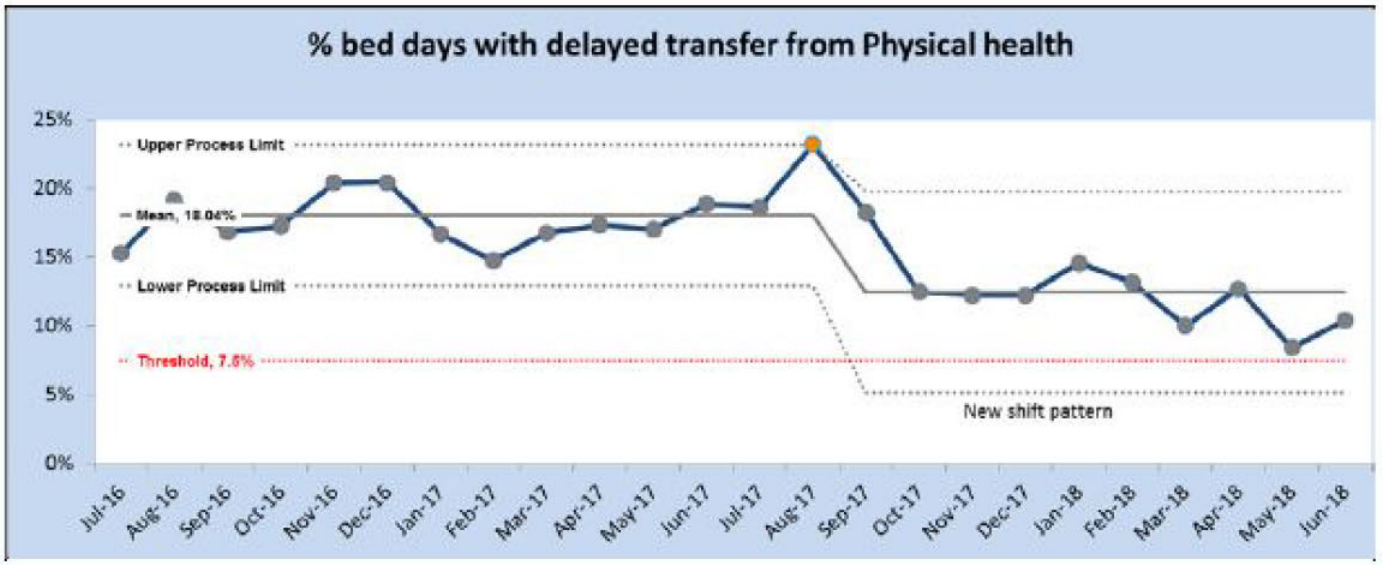
NHS Improvement



The table to the left highlights the levels of OOA placements on a weekly basis over the last 50 weeks (updated) and highlights the initial drop from 20+ down to a level of circa 7 for a 2 month period before the advent of winter and bed closures shifted the average back up to 13. Although numbers have dropped over the last three weeks of the period (below the 10 beds originally budgeted for), this doesn't yet represent a sustained shift in activity levels.

Dorset Healthcare

NHS Improvement

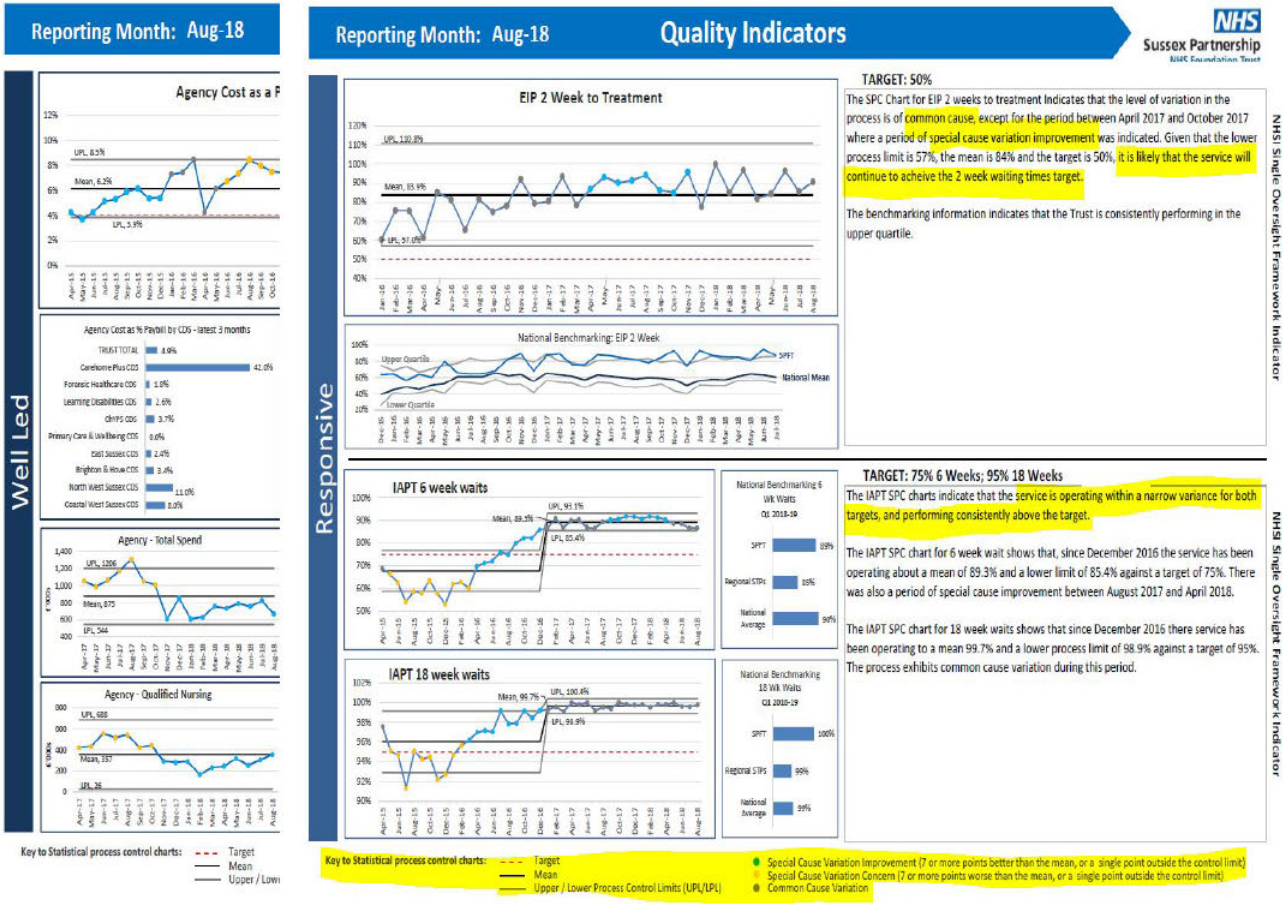


A new shift pattern was introduced in September 2017 and this improved the average DToC performance. However, SPC analysis shows that as the mean is 12.4% and the data is predicted to vary between 5.1% and 19.8% the Trust is unlikely to consistently achieve the threshold. Progress sheet 2.2.2 details improvement actions being taken.

https://www.youtube.com/watch?v=tHUMLtlJxGw

Sussex Partnership Trust

NHS Improvement

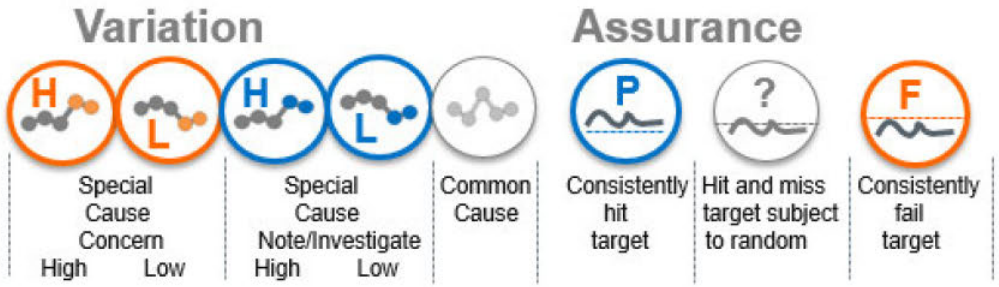


https://www.sussexpartnership.nhs.uk/sites/default/files/documents/v4_final_papers_-_public_board_of_directors_-_26_september_2018.pdf

Alternative summary report

NHS Improvement

Single line indicator w automated decision

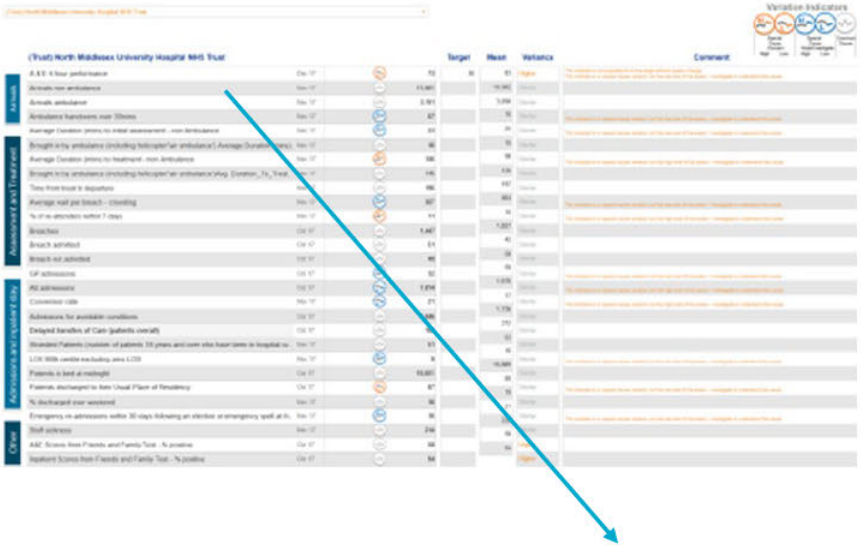


	Jun-18	Target	Variation	Target Capability	Comment
Staff Sickness absence	4.4%	3.5%		F	Shift change in August 2017 showing increase in sickness - staff survey review indicated…..

47 | Making data count

What could good look like?

NHS Improvement



For those indicators that cause concern – ability to find out more and ask questions of the system....

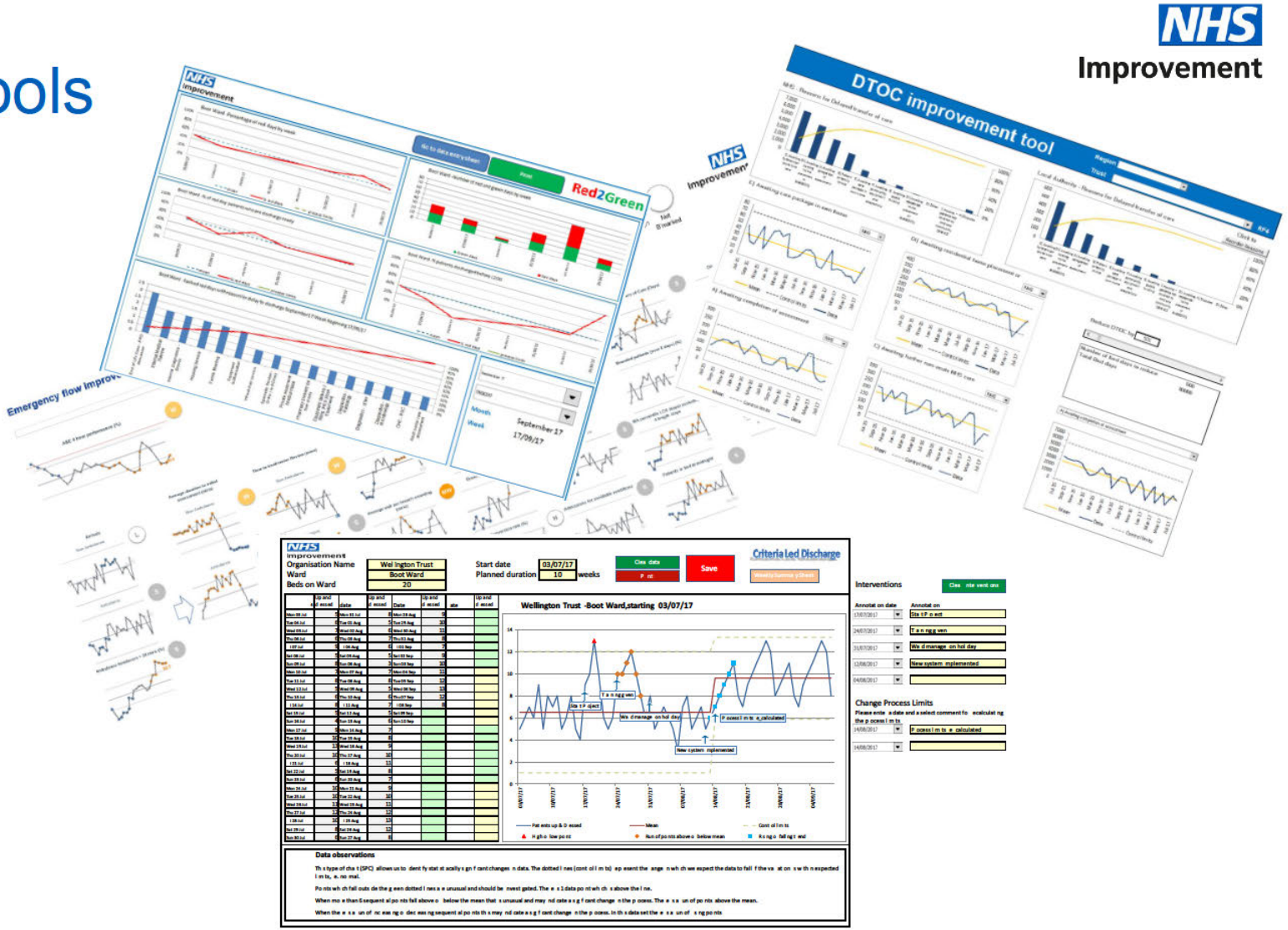
SPC SOF dashboard

NHS Improvement

Operational performance	A&E performance	July 2018		87.4		95.0	84.1	88.8	93.5	Common cause variation which is the type of variation expected
	A&E Quarter Performance	July 2018		87.4		95.0	85.3	88.1	91.0	Common cause variation which is the type of variation expected
	Cancer GP Performance	June 2018		78.9			73.9	83.7	93.6	Common cause variation which is the type of variation expected
	Cancer NHS Performance	June 2018		68.8			65.3	89.6	113.9	Common cause variation which is the type of variation expected
	Diagnostic Performance	June 2018	H	0.8			-0.5	0.3	1.0	Special cause variation (on the high side of the scale) - investigate to understand the cause
	RTT Performance	June 2018	L	90.2	?	92.0	88.9	90.6	92.2	Concerning special cause variation (on the low side of the scale) - investigate to understand the cause
	Cdiff - Infection Rate	June 2018	H	19.5	F	0.0	13.5	15.9	18.3	Concerning special cause variation (on the high side of the scale) - investigate to understand the cause
	Cdiff - Variance Plan	June 2018		-1.0			-4.2	1.2	6.6	Common cause variation which is the type of variation expected
Quality of care	FFT - A&E	2018		89.2			88.2	78.6	91.1	Common cause variation which is the type of variation expected
	FFT - Community	June 2018		95.9			89.1	96.3	103.5	Common cause variation which is the type of variation expected
	FFT - Inpatient	June 2018		94.9			89.3	95.0	100.6	Common cause variation which is the type of variation expected
	FFT - Maternity Qtr2	June 2018		97.6			92.9	97.3	101.7	Common cause variation which is the type of variation expected
	FFT - Staff	March 2018		70.8			68.6	76.1	83.6	Common cause variation which is the type of variation expected
	MRSA - InfectionRate	March 2018	L	0.7			0.6	0.9	1.1	Special cause variation (on the low side of the scale) - investigate to understand the cause
	MSSA	June 2018	L	7.5			7.1	8.5	9.9	Special cause variation (on the low side of the scale) - investigate to understand the cause
	NRLS - UnderReporting	May 2018	H	46.2			23.9	35.2	46.4	Special cause variation (on the high side of the scale) - investigate to understand the cause
	VTE - Risk	March 2018	H	1.0			0.9	1.0	1.0	Special cause variation (on the high side of the scale) - investigate to understand the cause
rship ement ility	Workforce - Staff Sickness	March 2018	H	3.8			3.1	3.6	4.2	Concerning special cause variation (on the high side of the scale) - investigate to understand the cause

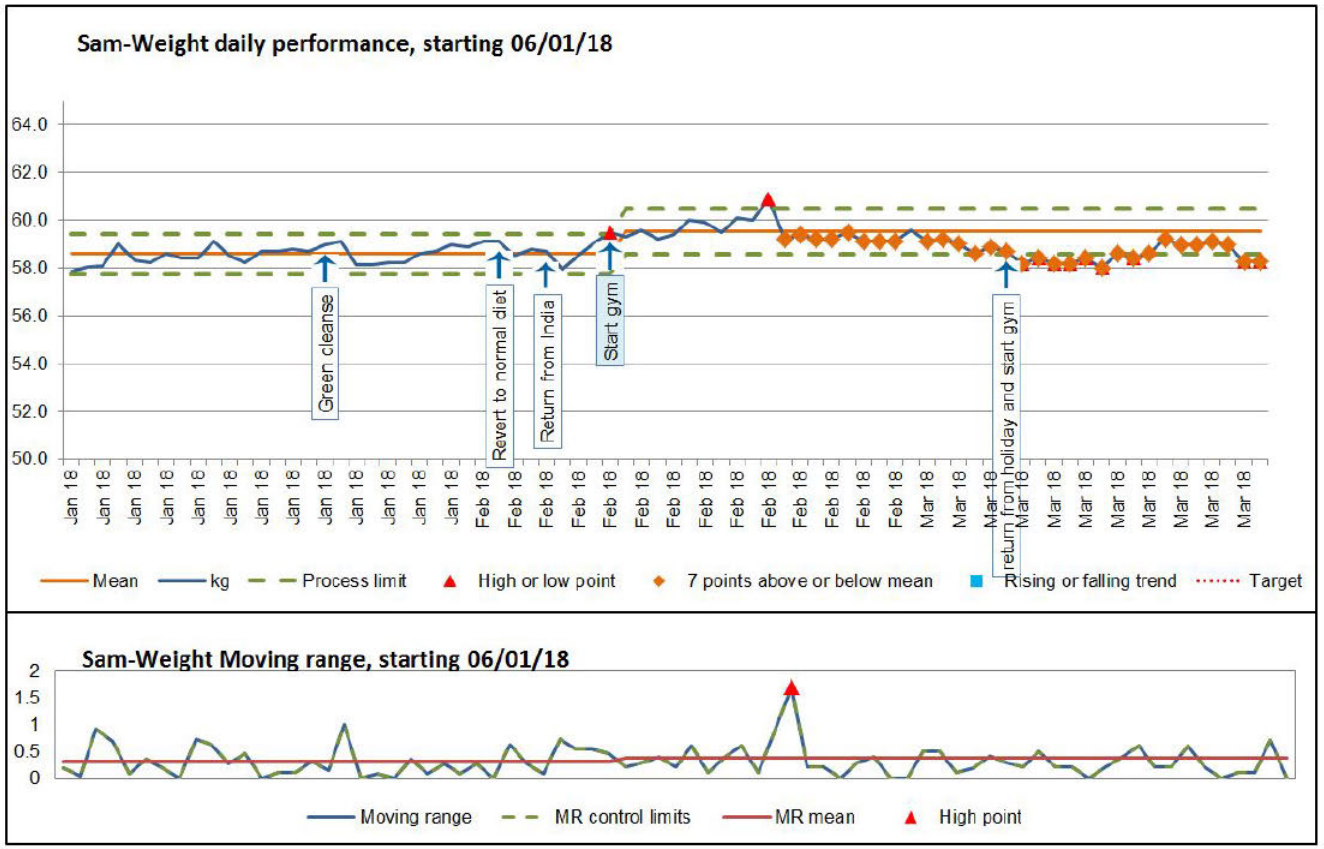
RTT Performance June 2018 90.2 92.0

Tools


NHS Improvement
DTOC improvement tool
Red2Green
Criteria Led Discharge
Wellington Trust -Boot Ward,starting 03/07/17
Data observations

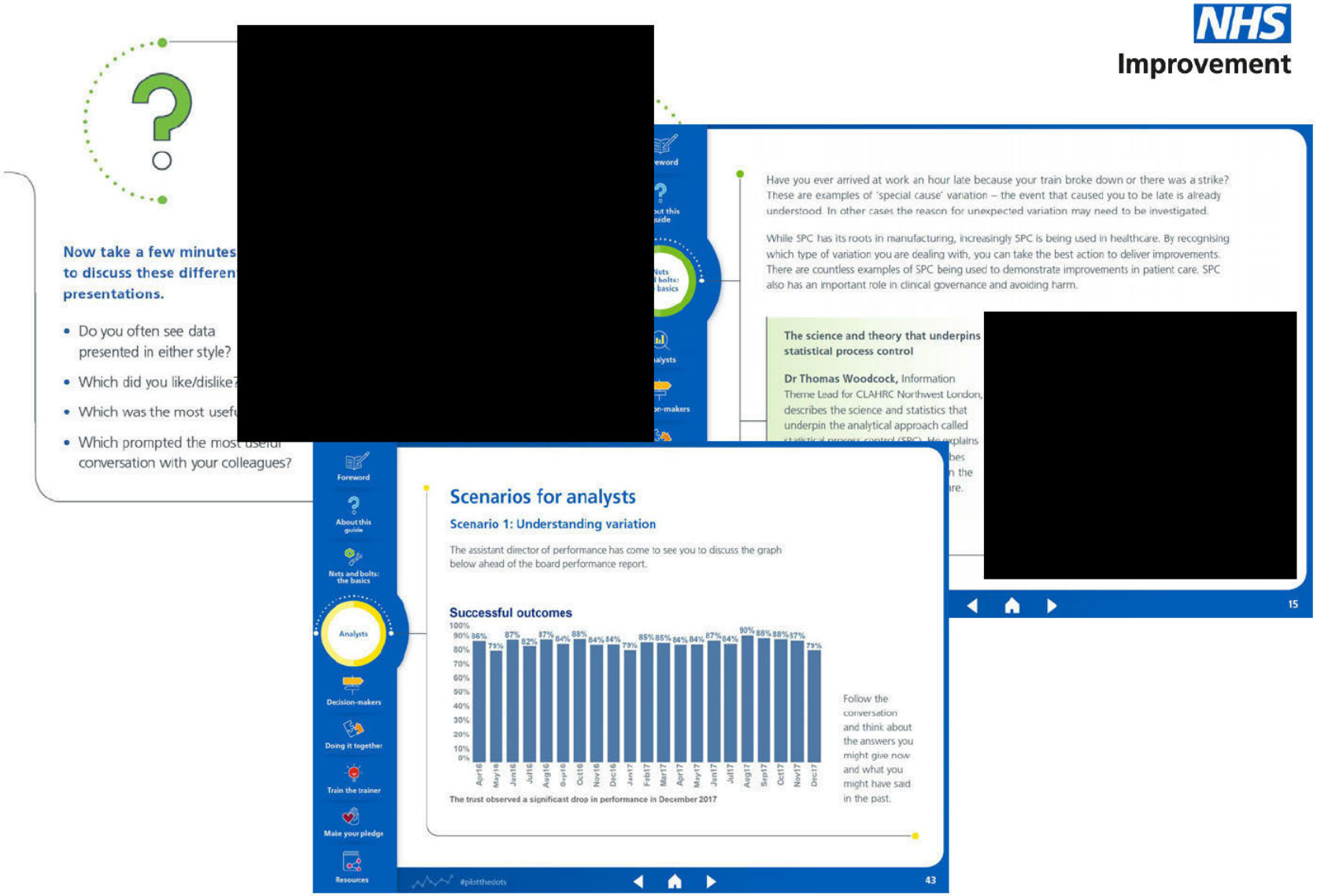
Free SPC tool

NHS Improvement



https://improvement.nhs.uk/resources/making-data-count/

NHS Improvement


NHS
Improvement
Now take a few minutes to discuss these differen presentations.
Do you often see data presented in either style?
Which did you like/dislike?
Which was the most usefu
Which prompted the most useful conversation with your colleagues?
Have you ever arrived at work an hour late because your train broke down or there was a strike? These are examples of 'special cause' variation – the event that caused you to be late is already understood. In other cases the reason for unexpected variation may need to be investigated.
While SPC has its roots in manufacturing, increasingly SPC is being used in healthcare. By recognising which type of variation you are dealing with, you can take the best action to deliver improvements. There are countless examples of SPC being used to demonstrate improvements in patient care. SPC also has an important role in clinical governance and avoiding harm.
The science and theory that underpins statistical process control
Dr Thomas Woodcock, Information
Theme Lead for CLAHRC Northwest London, describes the science and statistics that underpin the analytical approach called
15
Foreword
About this guide
Nuts and bolts: the basics
Analysts
Decision-makers
Doing it together
Train the trainer
Make your pledge
Resources
Scenarios for analysts
Scenario 1: Understanding variation
The assistant director of performance has come to see you to discuss the graph below ahead of the board performance report.
Successful outcomes
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
86% 79% 87% 82% 87% 84% 88% 84% 84% 79% 85% 85% 84% 84% 87% 84% 90% 88% 88% 87% 79%
Apr16 May16 Jun16 Jul16 Aug16 Sep16 Oct16 Nov16 Dec16 Jan17 Feb17 Mar17 Apr17 May17 Jun17 Jul17 Aug17 Sep17 Oct17 Nov17 Dec17
The trust observed a significant drop in performance in December 2017
Follow the conversation and think about the answers you might give now and what you might have said in the past.
#plotthedots
43

Pledges for action

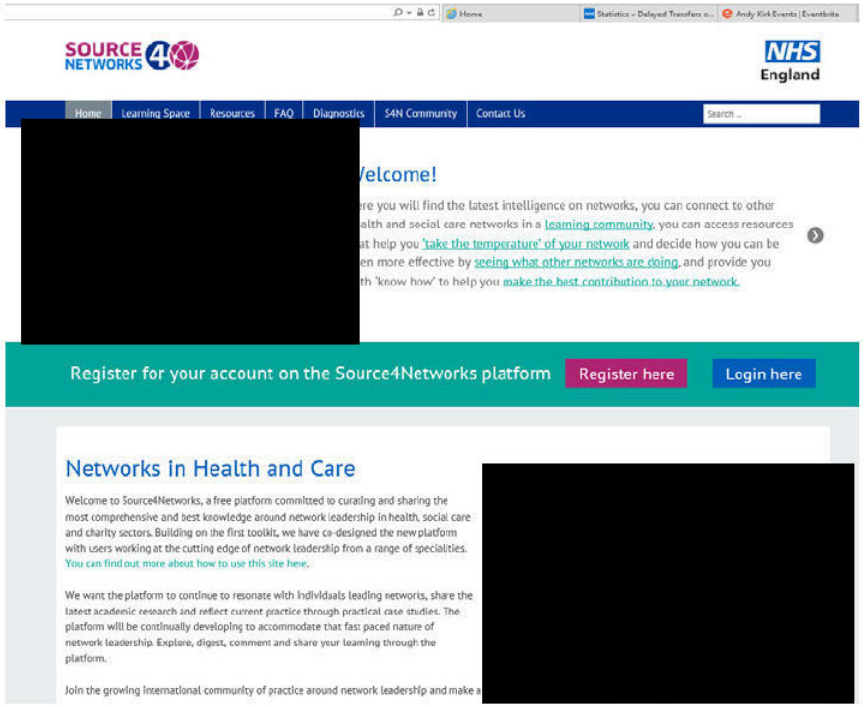
[redacted]

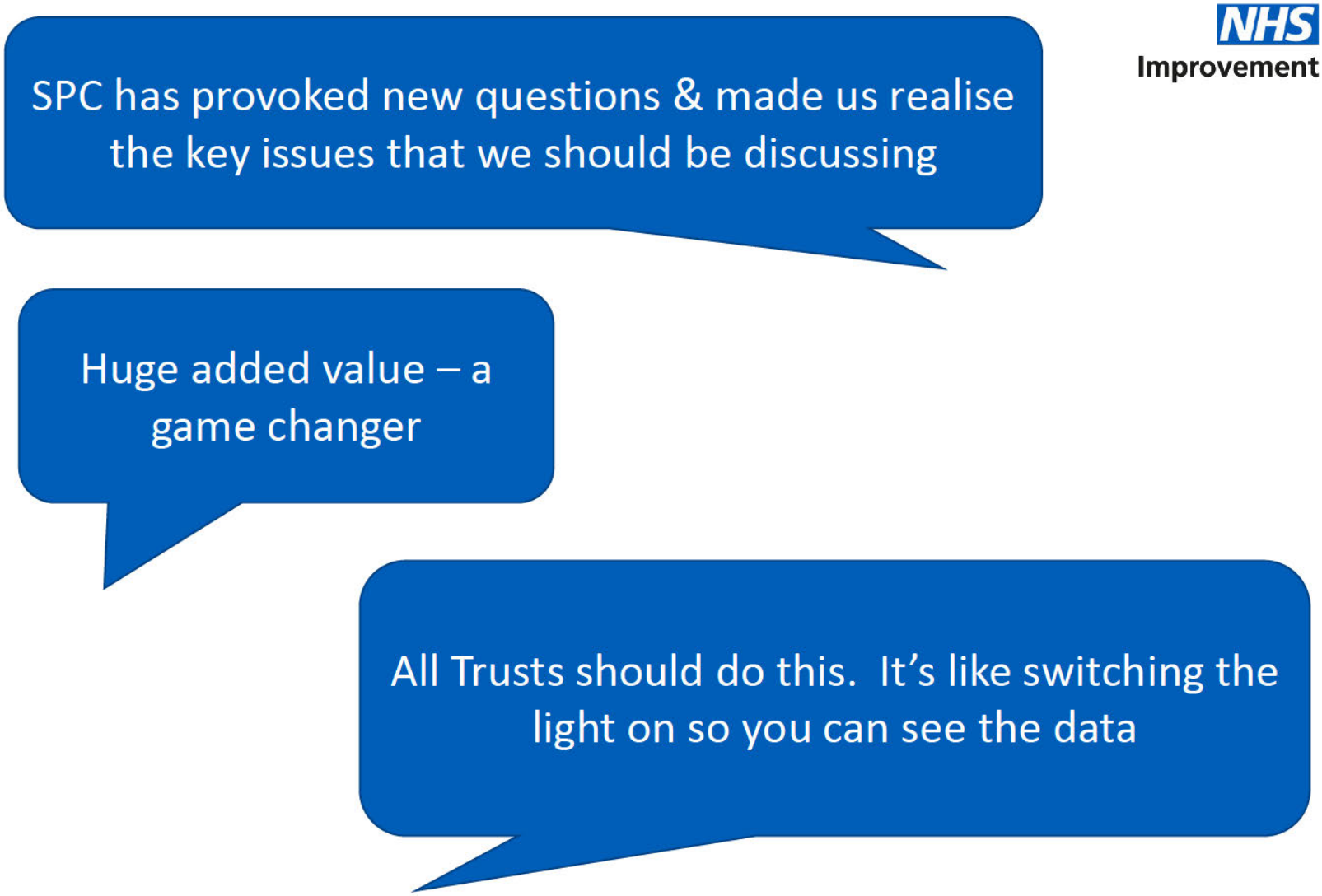
54 |

Making Data Count network

NHS Improvement

To register go to:
https://www.source4networks.org.uk




NHS
Improvement
SPC has provoked new questions & made us realise the key issues that we should be discussing
Huge added value – a game changer
All Trusts should do this. It's like switching the light on so you can see the data
56 | Making data count

NHS Improvement

NHS Improvement

Supplementary 5 - Making Data Count Powerpoint (2)

NHS Improvement

Making data count

– the why, the how and the experience so far

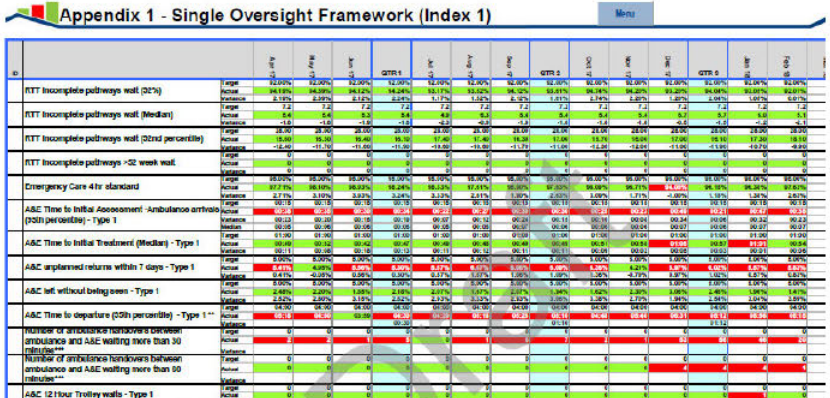
6th July 2018

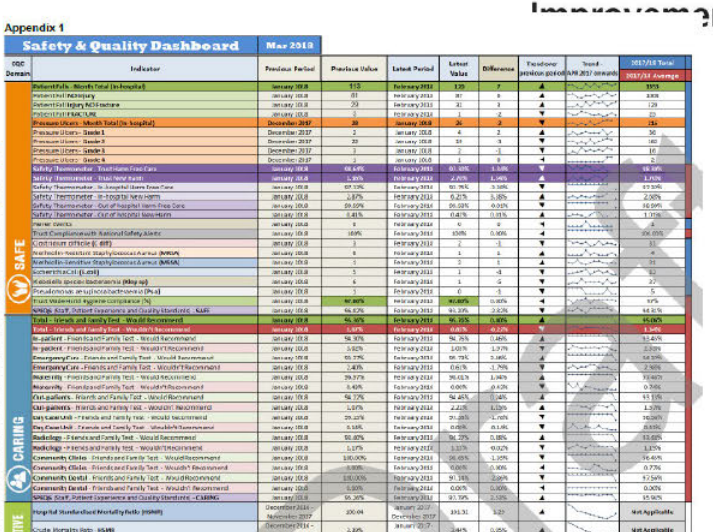
Samantha Riley, Head of Improvement Analytics, NHS Improvement

Mark Outhwaite, Non Exec Director, Avon & Wiltshire Mental Health Partnership NHS Trust

collaboration trust respect innovation courage compassion

Where are we now?





Where are we now?

NHS Improvement

Safety & Quality Dashboard									Mar 2018
CQC Domain	Indicator	Previous Period	Previous Value	Latest Period	Latest Value	Difference	Trend over previous period	Trend - APR 2017 onwards	2017/18 Total / 2017/18 Average
	Emergency Care - Friends and Family Test - Would Recommend	January 2018	93.27%	February 2018	95.73%	2.46%	▲		94.32%

One month trend…….

Is an increase from 95.36% to 95.76% important or distracting narrative?

Caring

7 Family and Friends Test (FFT) (data up to February 2018)

7.2 The Trusts 'Would Recommend' for Friends and Family returns *increased* to 95.76% for February 2018 from 95.36% in January 2018. The percentage of patients who stated they 'Wouldn't Recommend' *decreased* to 0.85% in February 2018 from 1.07% in January 2018.

NHS Improvement

Poll 1

What best describes your current integrated performance for the board :

- Mainly RAG charts
- A mixture of RAG and time series data/spark lines
- Presence of SPC charts

collaboration trust respect innovation courage compassion

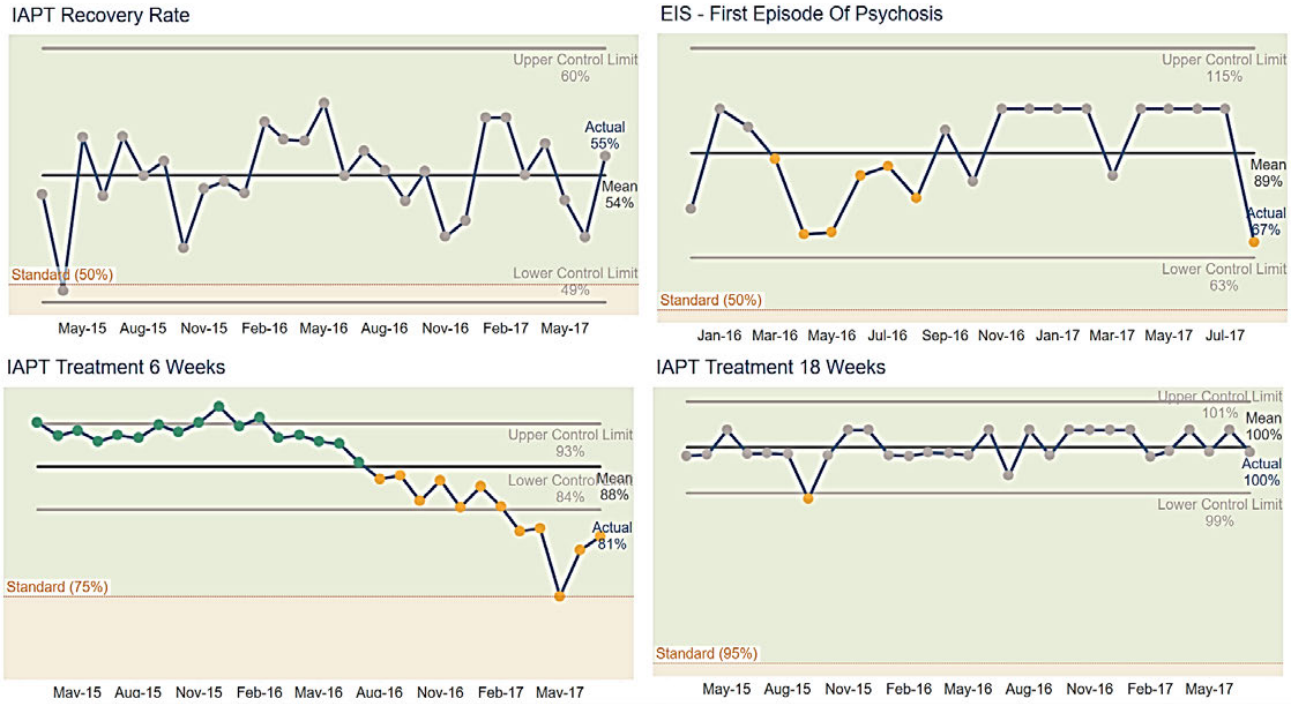
NHS Improvement

Improving Access to Psychological Therapies – performance against target

Metric	Target	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17
IAPT Treatment 18 weeks	95%	99.8%	99.5%	99.9%	99.8%	99.4%	99.7%	99.6%	99.7%
IAPT Treatment 6 weeks	75%	86.3%	84.1%	83.3%	80.9%	74.9%	79.5%	81.1%	81.2%
IAPT Recovery Rate	50%	59.3%	57.0%	54.0%	55.3%	53.6%	52.2%	55.3%	54.8%
EIS First Episode Psychosis	50%	100.0%	100.0%	83.0%	62.5%	100.0%	89.5%	100.0%	85.0%

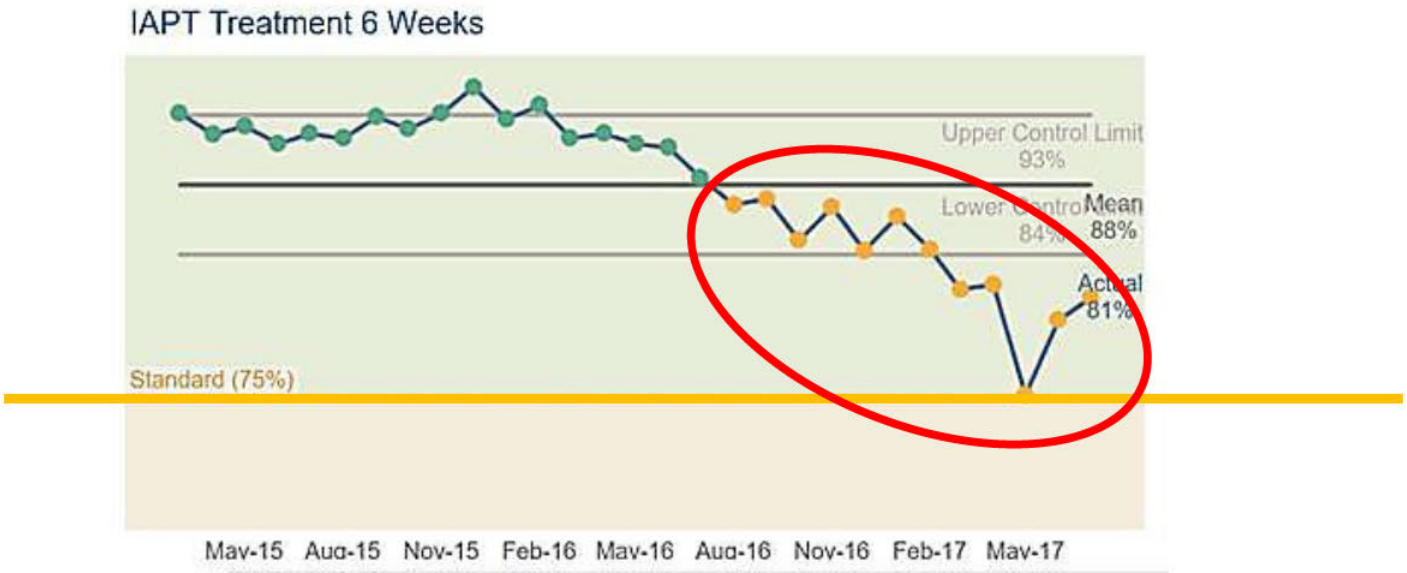
Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

NHS Improvement



Did green provide true assurance?

NHS Improvement

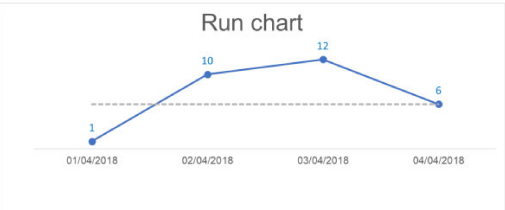


Scenario

We're going to simulate some **real data** in a healthcare setting

We'll be thinking about **how people react to patterns and trends** in data.

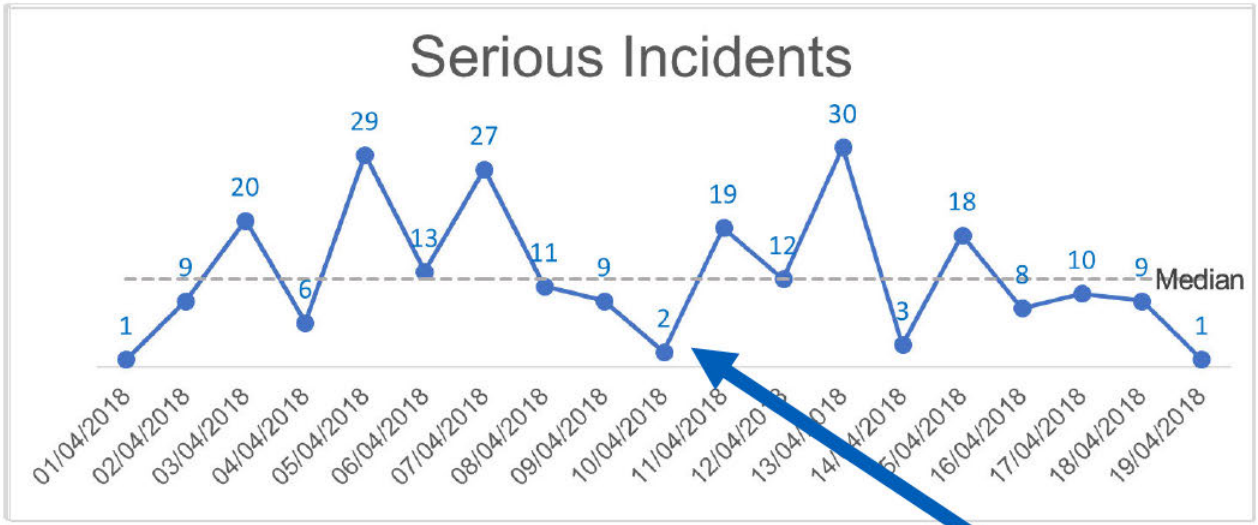
Can you spot an **improvement or decline** when it occurs? We'll begin plotting our data in a **run chart.**



Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

Reducing serious incidents

NHS Improvement

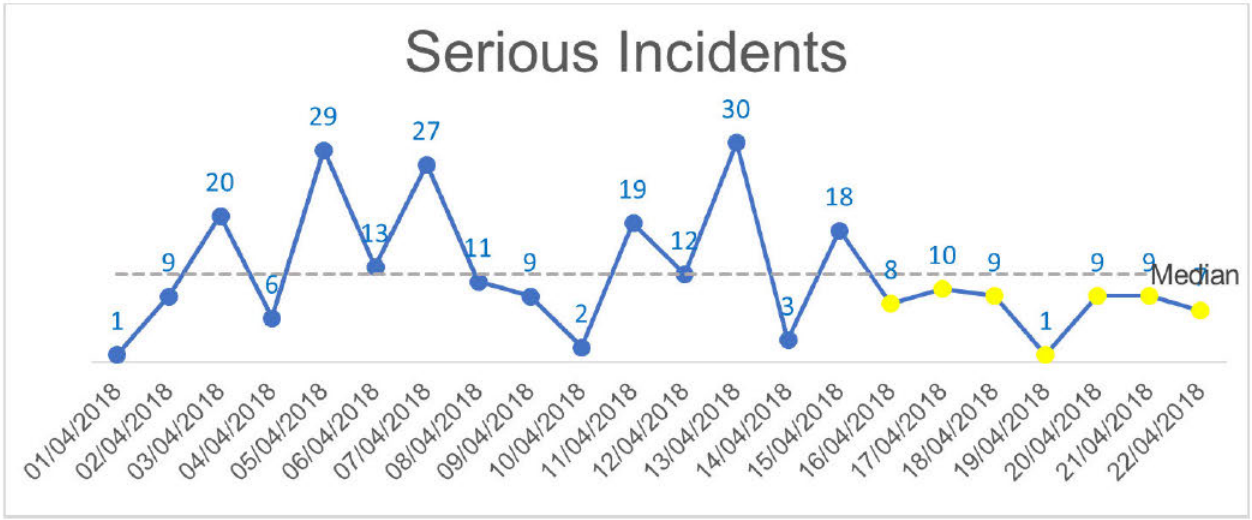


Has the improvement idea been successful?

Are you worried you might have seen this pattern before?

Improvement idea

NHS Improvement

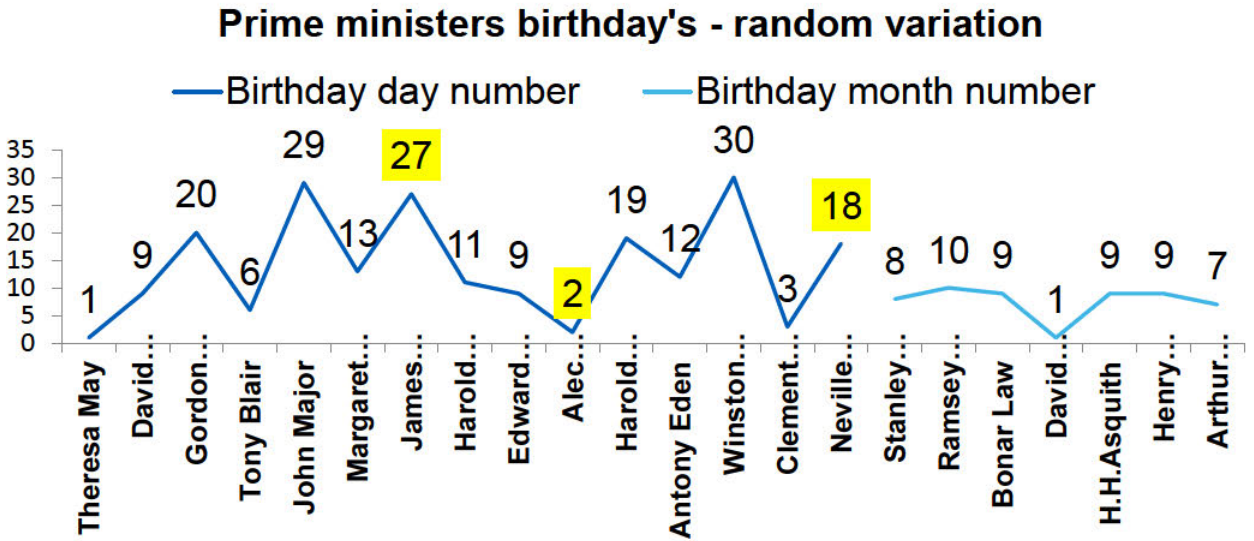


Now seven days below the baseline median…

We could go on… when should we recognise a trend?

The data that created this scenario

NHS Improvement



Any patterns at these points were randomly generated, then I changed the rules of the scenario….

Anatomy of a SPC chart

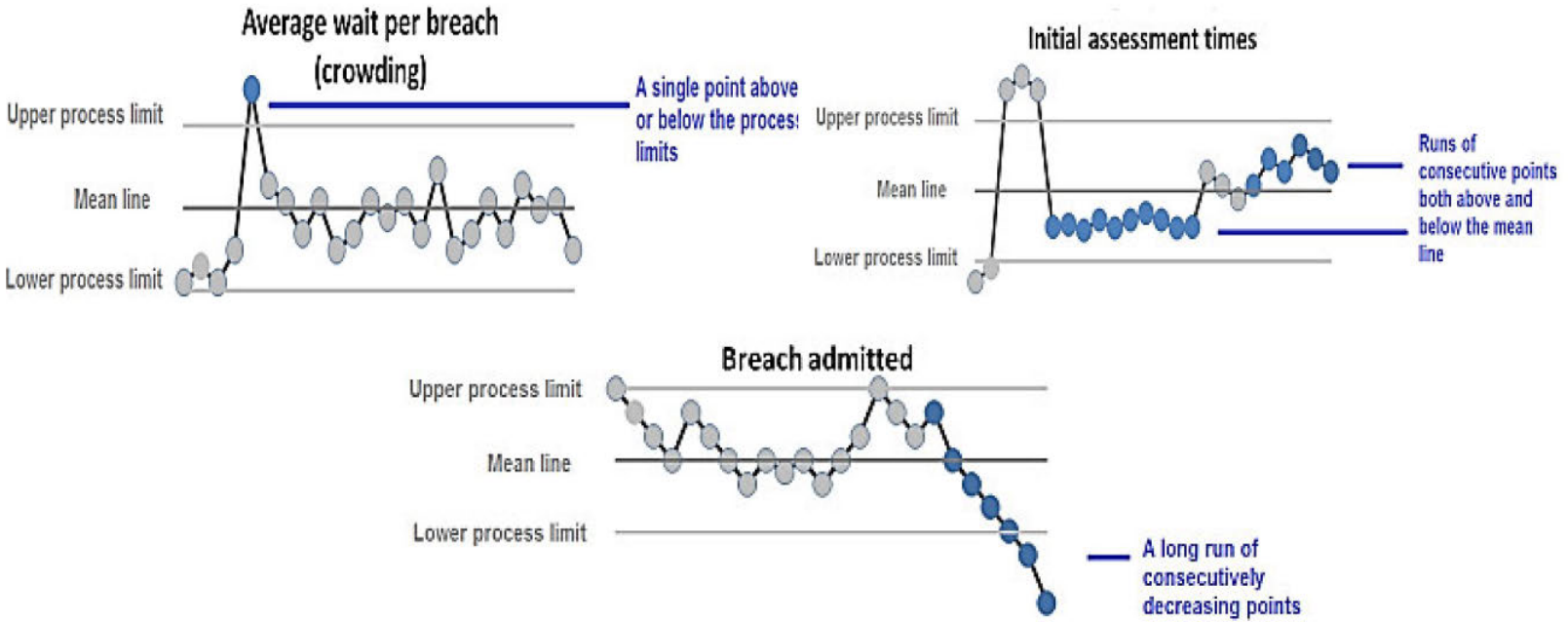
NHS Improvement

Time series line chart with 3 reference lines

20 plus data points for a robust analysis

SPC rules

NHS Improvement



Why is 7 significant?

A trend of 2 has the probability of 25% occurrence (**one in four**)

A trend of 4 has the probability of 6.25% occurrence (**one in sixteen**)

A trend of 7 has the probability of 0.8% occurrence (**one in one hundred and twenty-eight**)

Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

Evidence base

Public health

Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons

Mohammed A Mohammed, K K Cheng, Andrew Rouse, Tom Marshall

During the past century, manufacturing industry has achieved great success in improving the quality of its products. An essential factor in this success has been the use of Walter A Shewhart's pioneering work in the economic control of variation, which culminated in the development of a simple yet powerful graphical method known as the control chart. This chart classifies variation as having a common cause or special cause and thus guides the user to the most appropriate action to effect improvement. Using six case studies, including the excess deaths after paediatric cardiac surgery seen in Bristol, UK, and the activities of general practitioner turned murderer Harold Shipman, we show a central role for Shewhart's approach in turning the rhetoric of clinical governance into a reality.

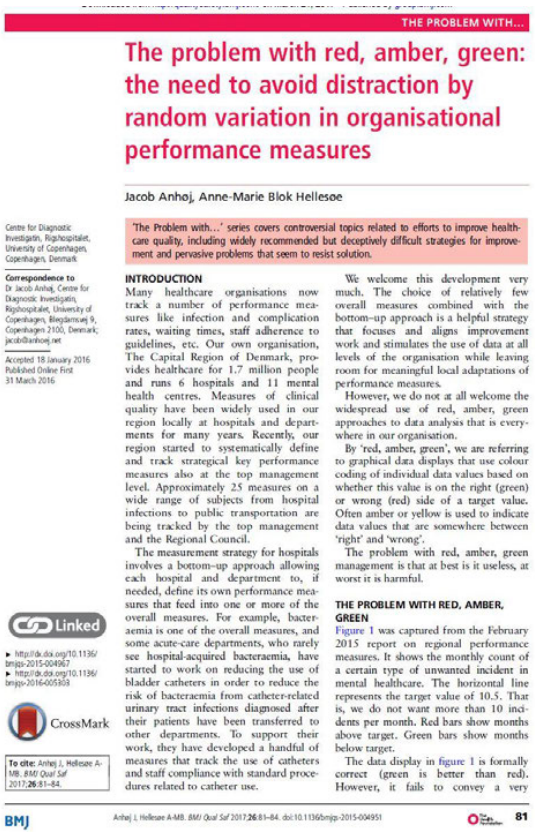
During the past century, manufacturing industry has achieved great success in improving the quality of its products. In industry, the definition of quality is "on target with minimum variation".[1] Reduction of variation is also a core concern in clinical governance;[2] however, there are fundamental and profound differences between the ways in which health services and industry make sense of variation. We begin with an illustration of the industrial approach to understanding and controlling variation, followed by application of this approach to health care, using six clinical governance case studies: mortality rates after paediatric cardiac surgery in Bristol, UK; mortality rates in older women treated by the general practitioner and convicted serial killer Harold Shipman; success rates of in-vitro fertilisation (IVF) treatment; neonatal deaths; prevalence of coronary heart disease in primary care; and mortality after fractured neck of femur.

Common-cause and special-cause variation

Consider a process such as writing a signature. Five of MAM's signatures are shown in the left of figure 1. Although these signatures were produced under the same conditions and by the same process, they are not identical. However, although they show variation, the signatures on the left are identical. No signature is better or worse than the others. If we want to reduce the variation between signatures, we must change the way we write all signatures, not just the ones that fail an adequate test. Thus, conventional approaches to understanding variation from a stable system can misguide us to act on individual failures rather than acting on the underlying process.

Now consider the sixth signature, on the right. It is clearly different from the others. A casual look suggests that there must be a special reason why this is so. If we want to address this kind of variation, we need to identify this special cause and prevent it from interacting with an otherwise stable process. (In this case, the signature is a forgery, attempted by TM under the same essential conditions!)

This approach categorises variation according to the action needed to reduce it. Common-cause variation is intrinsic to the process. To decrease common-cause variation, we need to act on the process. Special-cause variation is the result of factors extrinsic to the process, and its reduction therefore requires identification of and action on the special causes. The originator of these fundamental concepts was a physicist and engineer—Walter A Shewhart.[3] His pioneering work at



THE PROBLEM WITH...

The problem with red, amber, green: the need to avoid distraction by random variation in organisational performance measures

Jacob Anhøj, Anne-Marie Blok Hellesøe

Centre for Diagnostic Investigation, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

Correspondence to
Dr Jacob Anhøj, Centre for Diagnostic Investigation, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, Copenhagen 2100, Denmark; jacob@anhoej.net

Accepted 18 January 2016
Published Online First 31 March 2016

'The Problem with…' series covers controversial topics related to efforts to improve healthcare quality, including widely recommended but deceptively difficult strategies for improvement and pervasive problems that seem to resist solution.

INTRODUCTION

Many healthcare organisations now track a number of performance measures like infection and complication rates, waiting times, staff adherence to guidelines, etc. Our own organisation, The Capital Region of Denmark, provides healthcare for 1.7 million people and runs 6 hospitals and 11 mental health centres. Measures of clinical quality have been widely used in our region locally at hospitals and departments for many years. Recently, our region started to systematically define and track strategical key performance measures also at the top management level. Approximately 25 measures on a wide range of subjects from hospital infections to public transportation are being tracked by the top management and the Regional Council.

The measurement strategy for hospitals involves a bottom–up approach allowing each hospital and department to, if needed, define its own performance measures that feed into one or more of the overall measures. For example, bacteraemia is one of the overall measures, and some acute-care departments, who rarely see hospital-acquired bacteraemia, have started to work on reducing the use of bladder catheters in order to reduce the risk of bacteraemia from catheter-related urinary tract infections diagnosed after their patients have been transferred to other departments. To support their work, they have developed a handful of measures that track the use of catheters and staff compliance with standard procedures related to catheter use.

We welcome this development very much. The choice of relatively few overall measures combined with the bottom–up approach is a helpful strategy that focuses and aligns improvement work and stimulates the use of data at all levels of the organisation while leaving room for meaningful local adaptations of performance measures.

However, we do not at all welcome the widespread use of red, amber, green approaches to data analysis that is everywhere in our organisation.

By 'red, amber, green', we are referring to graphical data displays that use colour coding of individual data values based on whether this value is on the right (green) or wrong (red) side of a target value. Often amber or yellow is used to indicate data values that are somewhere between 'right' and 'wrong'.

The problem with red, amber, green management is that at best is it useless, at worst it is harmful.

THE PROBLEM WITH RED, AMBER, GREEN

Figure 1 was captured from the February 2015 report on regional performance measures. It shows the monthly count of a certain type of unwanted incident in mental healthcare. The horizontal line represents the target value of 10.5. That is, we do not want more than 10 incidents per month. Red bars show months above target. Green bars show months below target.

The data display in figure 1 is formally correct (green is better than red). However, it fails to convey a very

Linked
► http://dx.doi.org/10.1136/bmjqs-2015-004967
► http://dx.doi.org/10.1136/bmjqs-2016-005303

CrossMark

To cite: Anhøj J, Hellesøe A-MB. *BMJ Qual Saf* 2017;**26**:81–84.

CQC – signs of a mature QI approach

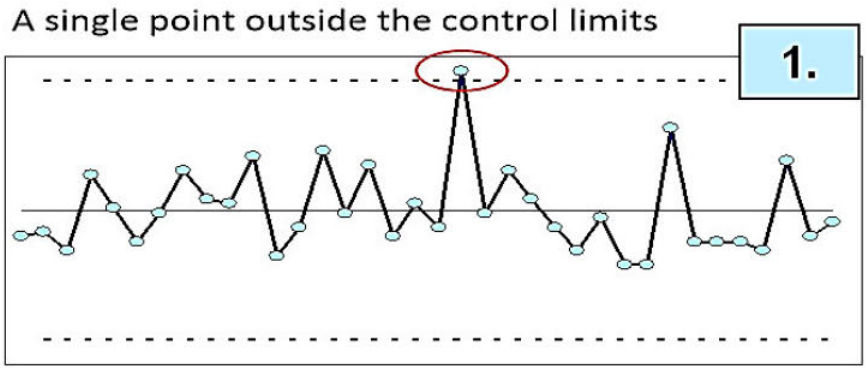
3. The Board looks at data as time series analysis, and makes decisions based on an understanding of variation.[1]

[1] data are presented as run or control charts, instead of bar graphs, pie charts or RAG rated. Narrative analysis describes system quality and performance using terminology of common cause and special cause variation.

https://www.cqc.org.uk/sites/default/files/20180404_9001395_briefguide-quality_improvement_healthcare_provider%20v1.pdf

If there is 'special cause'

NHS Improvement



Unacceptable variation

NHS Improvement

Redesign the system

Everything is failing?

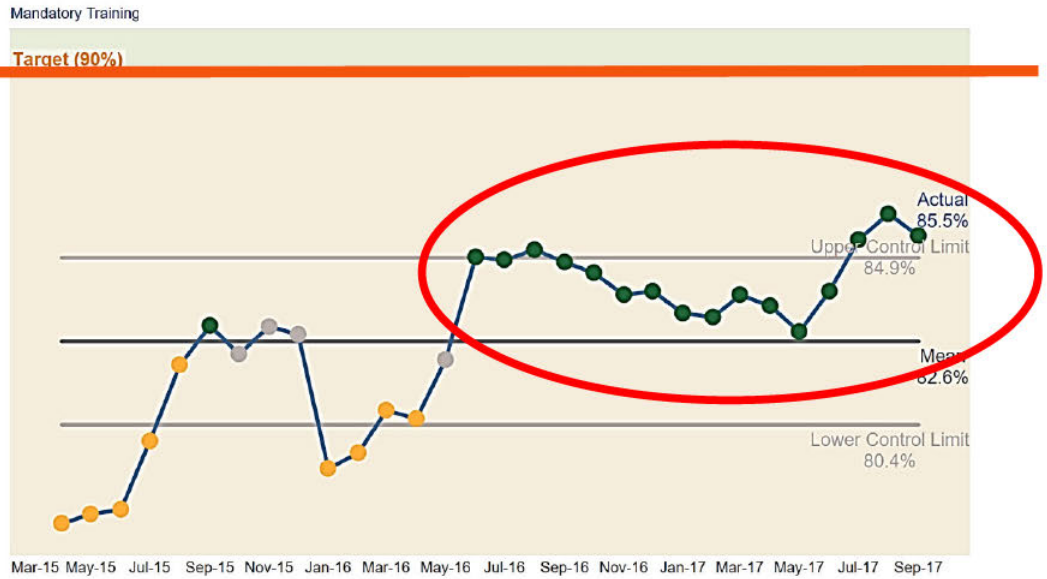
NHS Improvement

									Combined Trust Performance					Trust data 13 months	
Domain	Indicator	Jul-17	Aug-17	Sep-17	2017-2018	Jul-17	Aug-17	Sep-17	2017-2018	Jul-17	Aug-17	Sep-17	2017-2018 Q2	2017-2018	Trendcharts
Training	Mandatory training compliance (Target: >90%)	85.4%	86.1%	85.5%	84.8%	85.2%	86.5%	85.7%	85.1%	85.4%	86.2%	85.6%	85.7%	84.8%	

Presentation influences discussion

NHS Improvement

Target



Are things improving?

NHS Improvement

Patient Experience Dashboard

Domain	Indicator
Friends and Family	FFT: A&E recommend % (Target: >90%)

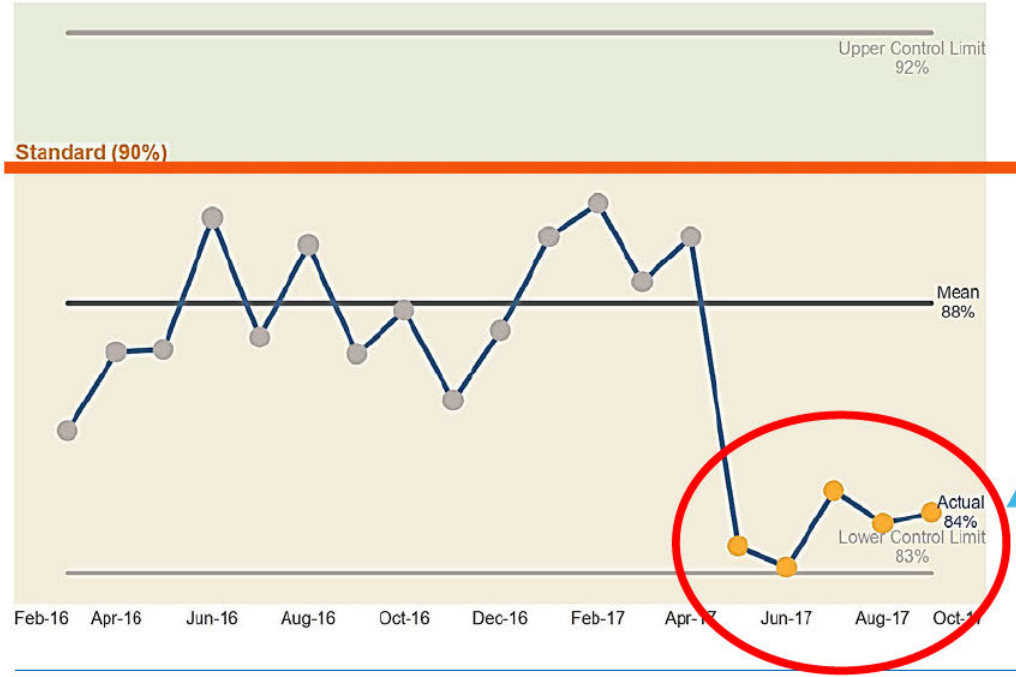
Combined Trust Performance					Trust data 13 months
Jul-17	Aug-17	Sep-17	2017-2018 Q2	2017-2018	Trend charts
84.7%	83.5%	84.4%	84.2%	84.8%	

Friends and Family Test - A&E recommend %

The recommend rate improved from the previous month however remains below the 90%.

SPC changes the narrative

NHS Improvement

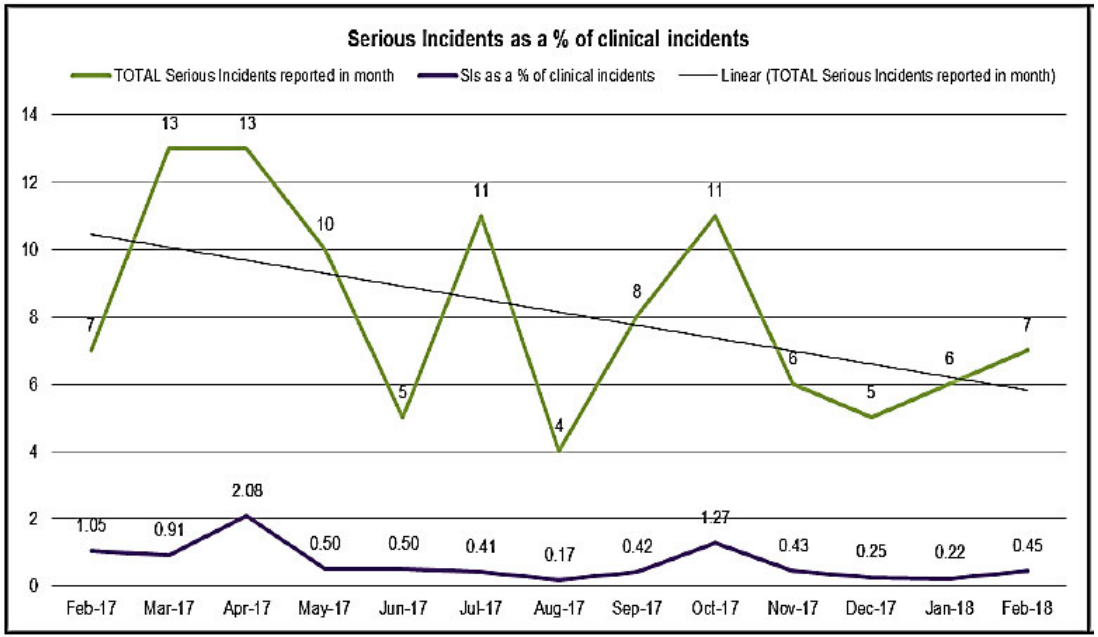


Target

No signs of improvement - signs of decline

22 | Making data count

Serious incidents



Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

NHS Improvement

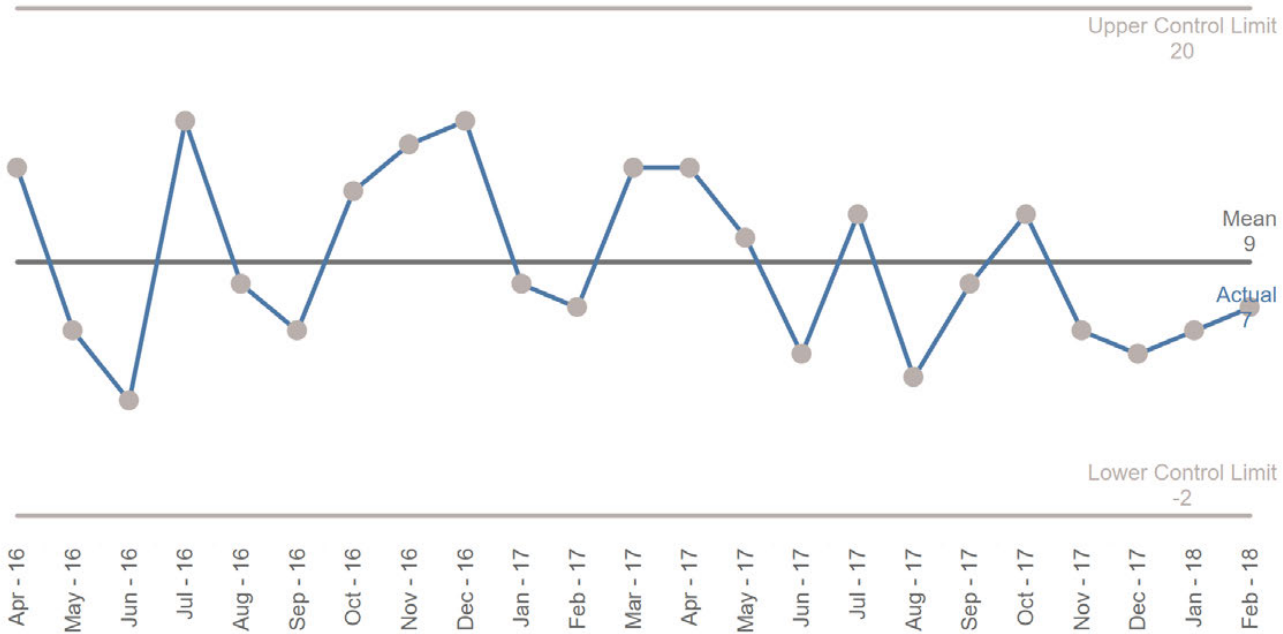
Poll 2

The number of serious incidents occurring is :

- Improving
- Declining
- Staying the same

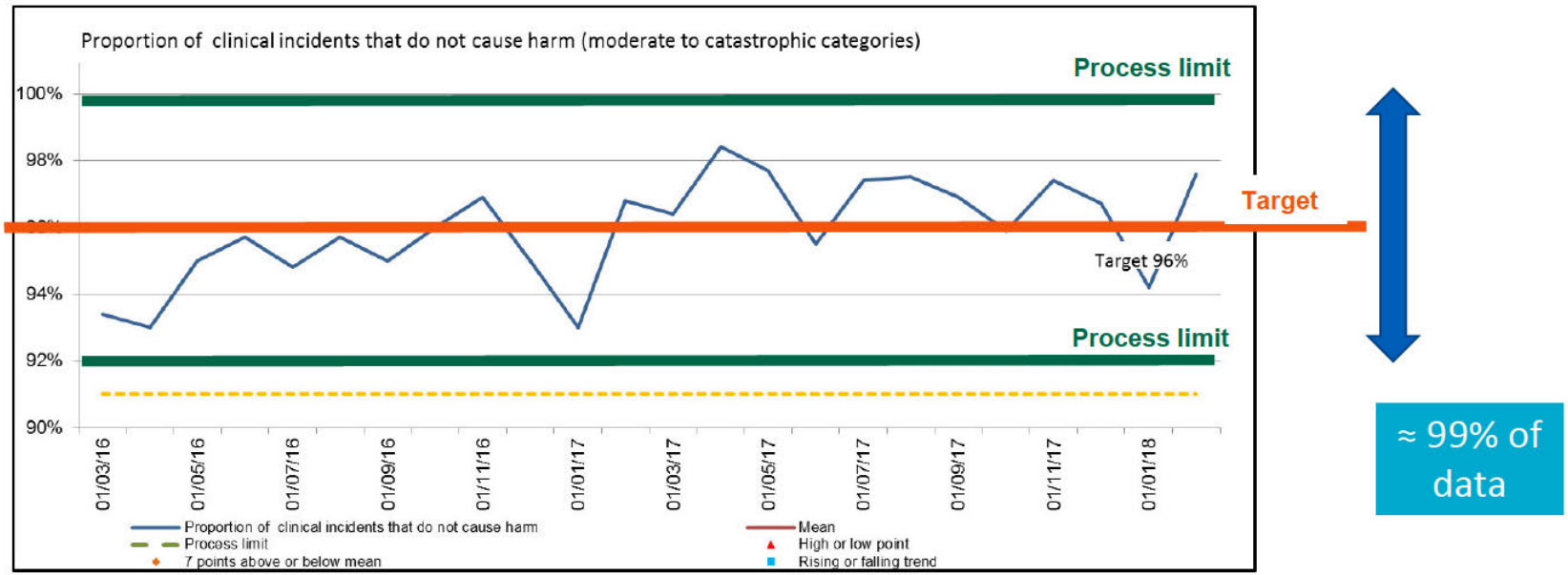
collaboration trust respect innovation courage compassion

Level of variation acceptable?



Will the target always be achieved?

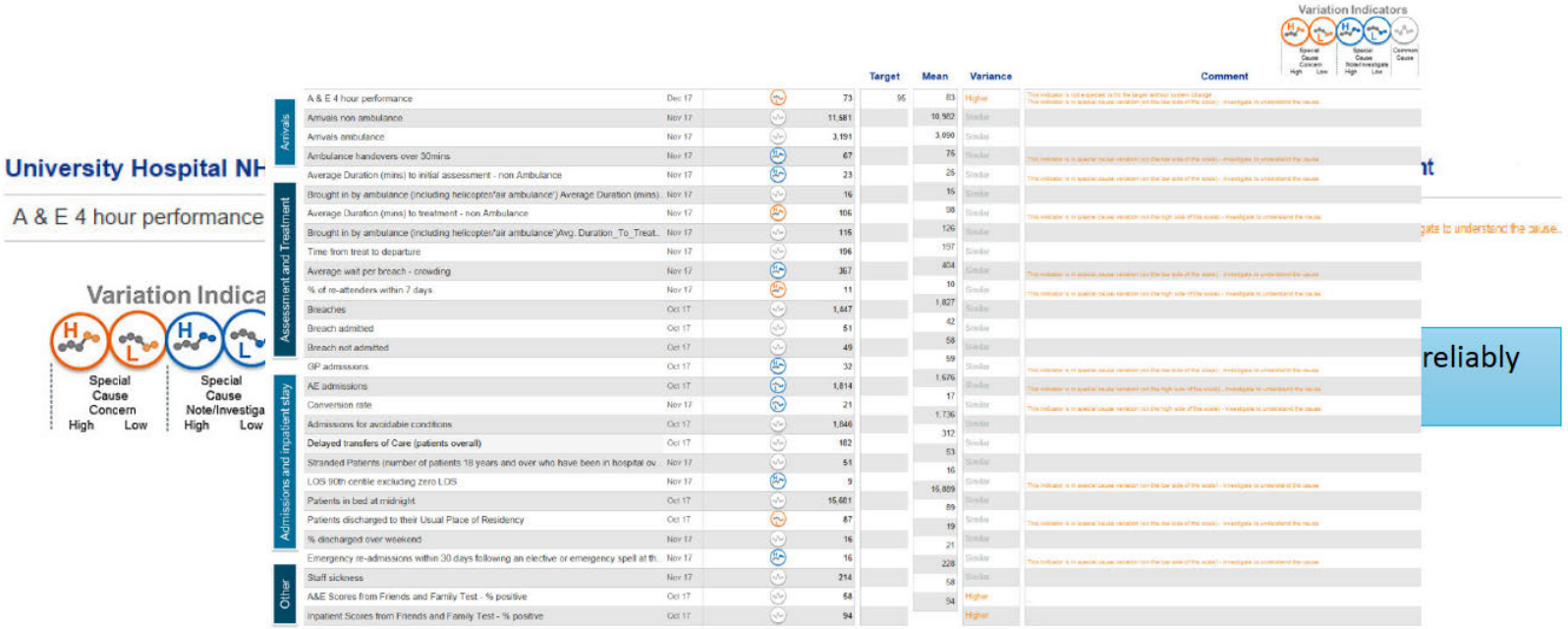
NHS Improvement



≈ 99% of data

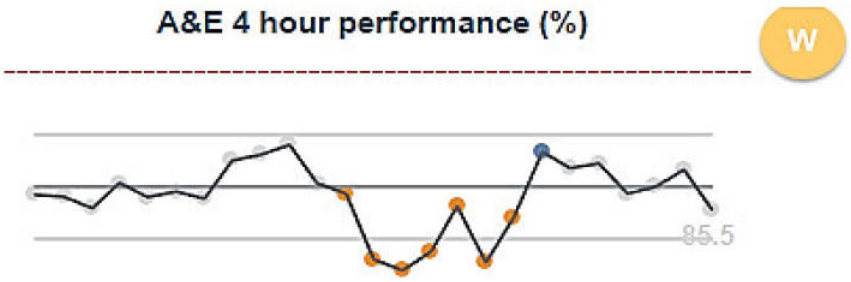
Thinking outside the box

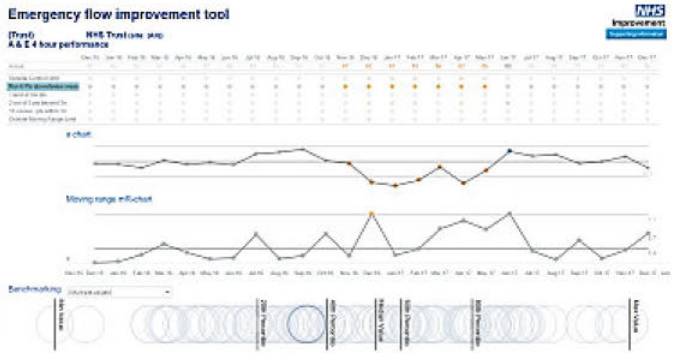
NHS Improvement



SPC Appendix

NHS Improvement

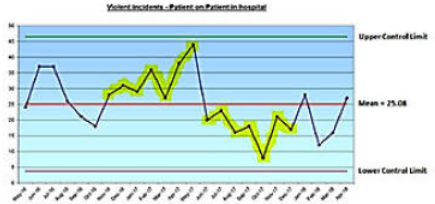




Supporting contextual commentary

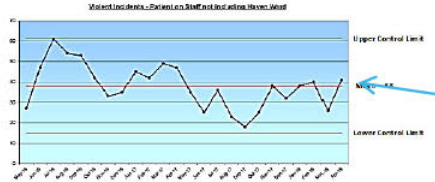
What is changing?

NHS Improvement



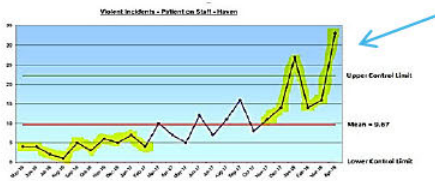
There were 74 patient on staff violent incidents reported trust wide. SPC analysis shows that this is a special cause variation as it is outside of the predicated range for the number of monthly incidents.

Further analysis shows that both the number of incidents on Haven Ward and Haven incidents as a proportion of total incidents is increasing. When Haven Ward is excluded from the total figures for the Trust the number of incidents is stable and predictable; it will range between 14.87 and 61.13 with a mean of 38.



During April one violent patient on staff incident was a moderate incident on Haven Ward and will be subject to a Serious Incident Investigation. A staff member was assaulted by a patient following a Section 136 from Police custody. The staff member has since returned to work, well supported by wider team/management. The rest were low or no harm incidents. Haven and St Brelades were the top two reporters with Haven reporting 33 incidents and St Brelades 20. SPC analysis for St Brelades shows that April's figure is within predicted range of variation (0 to 21.95). On St Brelades Ward one patient was involved in seven incidents; the same patient was involved in three of the patient on patient incidents as well.

SPC analysis for patient on staff violent incidents on Haven shows that this is special cause variation (part run of 6 above mean).



During April two patients on Haven were involved in 26 incidents. The majority of incidents were caused by three patients who were/are acutely unwell. The rest of the violence towards staff was one-off incidents by individual patients. Following the weekly incident reviews in place on Haven, the team identified the individuals concerned, conducted RCA's and organised risk huddles with all team members to create behaviour support plans in order to

SPC charts

Highlighting special cause

Supporting narrative

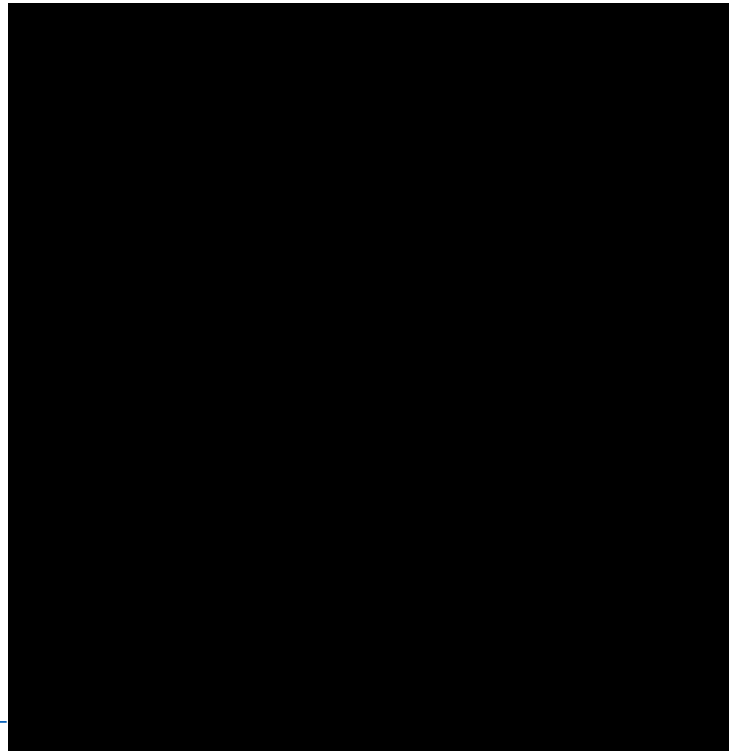
Dorset Healthcare's SPC Journey

Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

Advice to others

Don’t forget the PORC

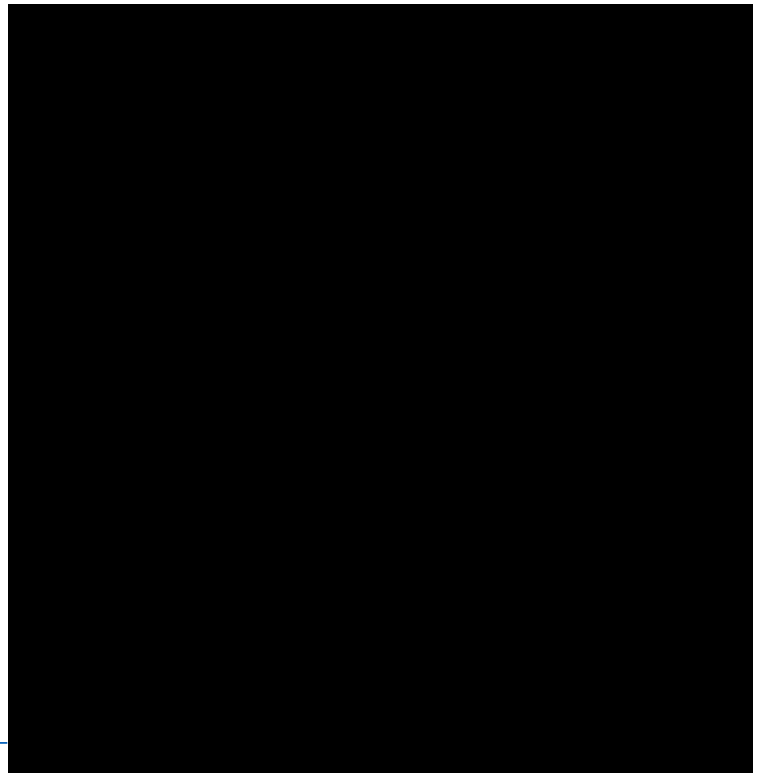
In the excitement of introducing SPC and putting control limits on your charts don’t lose sight of the utility and accessibility of the ‘Plain Ole Run Chart’ (PORC)



Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

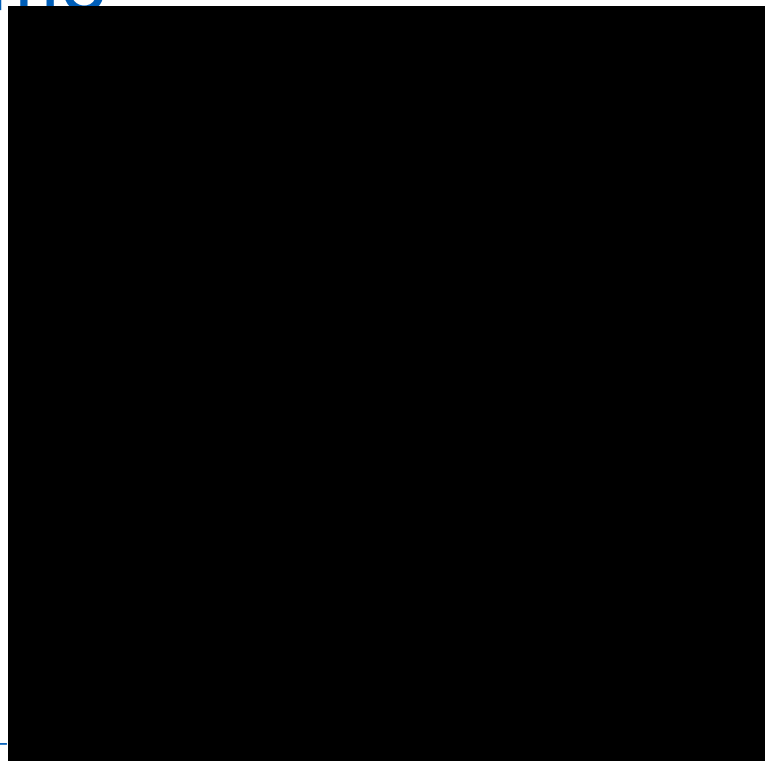
Top table exclusive

The top table at the feast always used to get the best food. Are SPC and Run Charts seen as rich fare only for the nobs on the top table? Are they routinely used in the front-line?



New hammer syndrome

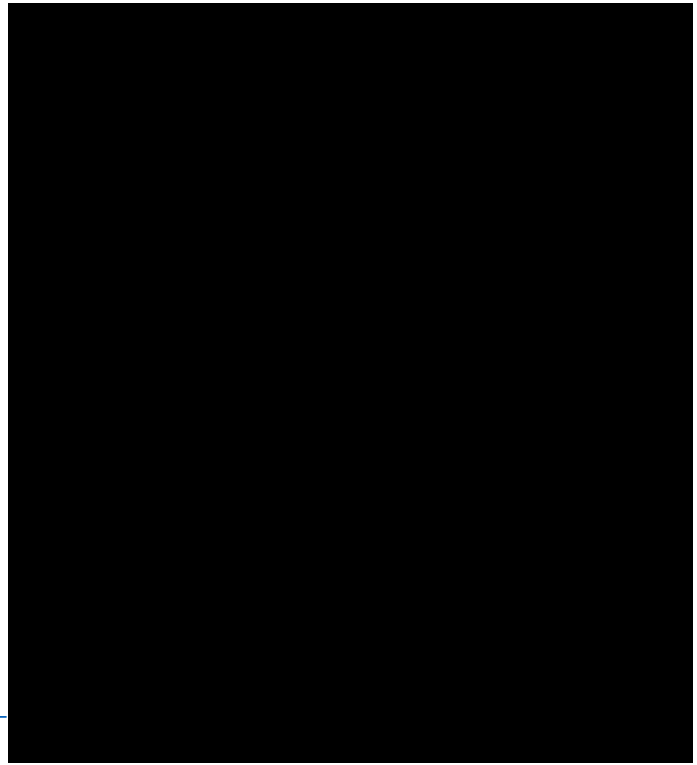
To someone with a new hammer everything looks like a nail! Not everything is appropriate for SPC or a Run Chart



Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

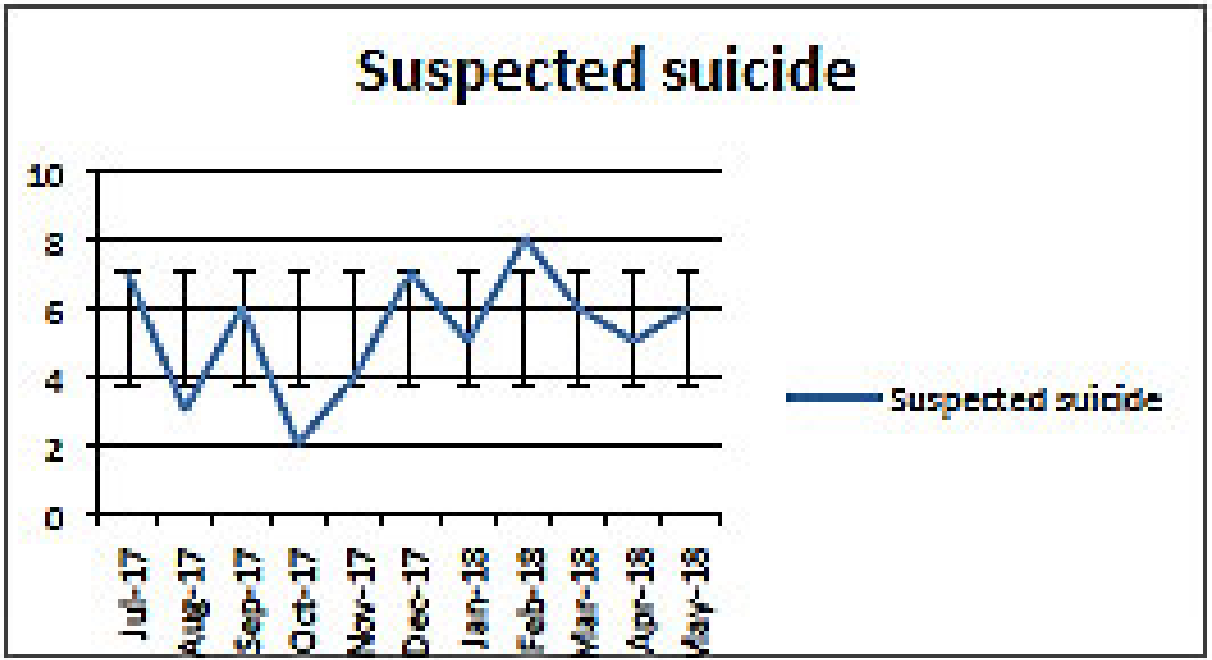
Cargo Cults

Measure it and something will happen. More about Cargo Cults here: https://en.wikipedia.org/wiki/Cargo_cult



Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

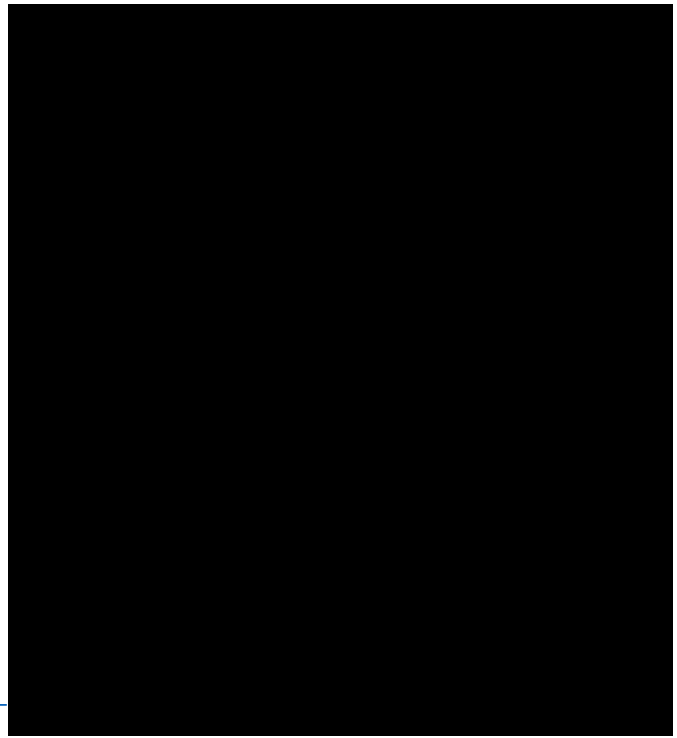
Cargo cults – an example



Kudrna L, *et al. BMJ Qual Saf* 2022;0:1–9. doi: 10.1136/bmjqs-2021-013514

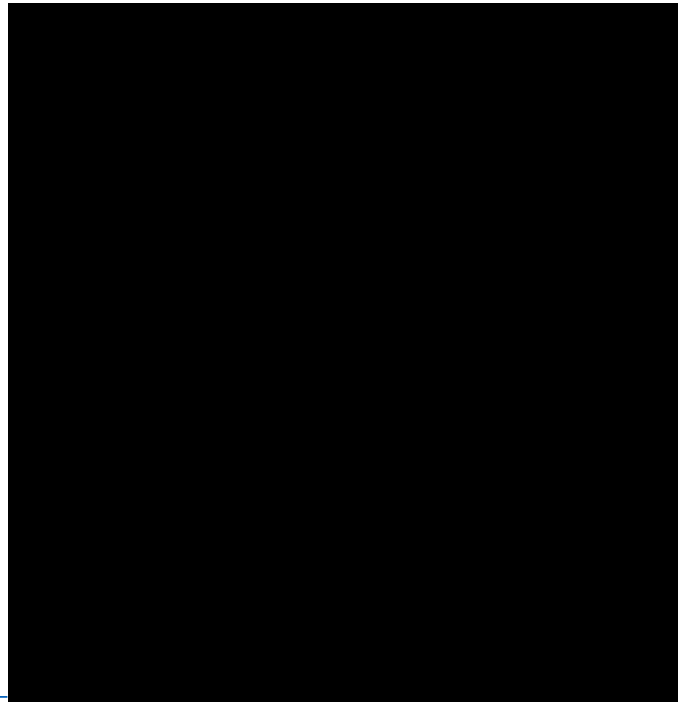
Where's Wally?

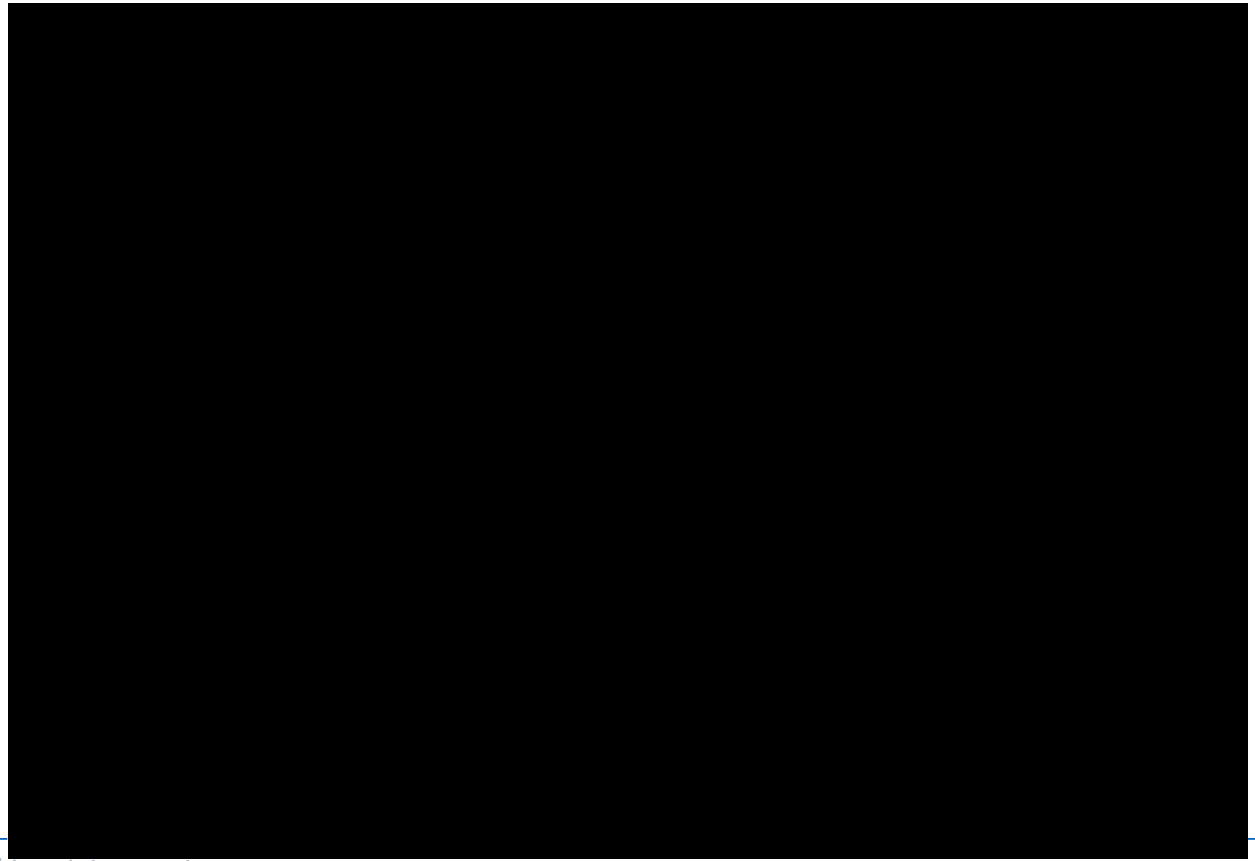
Just how many charts can you cram onto an A4 page? If you cannot even read the legend without a magnifying glass then what is the point? How do you identify the chart(s) that indicate significant change in that crowd?



How many angels on the head of the SPC pin?

Watch the newly minted SPC experts start to argue about how many points constitute a shift, a trend, a run – how many points to calculate control limits, sampling etc





NHS Improvement

https://improvement.nhs.uk/documents/2748/NHS_MAKING_DATA_COUNT_FINAL.pdf

ACT Academy

NHS Improvement

https://improvement.nhs.uk/documents/1241/QSIR-A5-4pp.pdf

NHS Improvement

Poll 3

Which statement best describes how you feel about your performance report:

- I am confident that my report supports effective decision making
- I am concerned that my report may not focus discussion on the most important issues
- I need time to reflect on today's session

collaboration trust respect innovation courage compassion

NHS Improvement

Poll 4

Please rank the following in order of priority – which of these will be most helpful?

- Test a different approach to regulation
- Implement a regional train the trainer programme
- Establish regional networks
- Facilitate mechanisms to share learning
- Providing analytical products to aid decision making

Supplementary 6 – Coding Frame and Selected Results

Coding Frame

The coding frame is shown in Table S6-1. We identified if the charts were time series, between groups, time series and between groups, pie charts, other charts, and/or statistical process control charts (SPCs).

We identified if aspects of the training were reflected in the charts and board papers. One aspect of the training was R-A-G colouring anywhere on the board paper (see slide 7, Supplemental File 4). R-A-G charts consist of tables of coloured boxes that show whether data fail to meet a specific target (red), are in danger of not meeting that target (amber), or are achieving that target (green).[1] The colour coding is not informed by statistical variation. In contrast, SPCs have set statistical limits, typically at two or three standard deviations above and below the mean value.[2–4] The training discusses the limitations of R-A-G charts.

The training encourages hospitals to use SPC icons (slide 47, Supplemental File 4) that summarise statistical variation visually using colours and letters that indicate special or common cause variation and performance relative to a target. We identified if any SPC icons were present on the SPCs.

For the SPCs, we also identified if other factors included in the training were present (see Powerpoint slides in Supplemental File 4):

- Labelling of limits (slides 32, 34), including sigma versus standard deviation control limits. Sigma and standard deviation limits are different because sigma calculations do not assume homogeneity in the underlying data (for example, that they are

derived from an unchanged process), whereas standard deviation calculations assume a constant mean value [2–4];

- recalculation of control limits (slides 44-46), which occurs when the formula used to calculate the control limits is altered due to a change in the underlying process, such as a new bed management system;
- highlighted runs or trends (slides 25-27);
- and comments about reasons for variation or suggestions for improvement (slide 47).

Table S6-1 – Coding Frame

Chart Number
1. Is RAG present?
2. Type of chart: time series?
3. Type of chart: between groups?
4. Type of chart: time series and between groups?
5. Type of chart: pie chart?
6. Type of chart: other chart?
7. Type of chart: notes on other
8. Is this a statistical process control chart summary icon?
9. Is this a statistical process control chart?
IF YES statistical process control chart:
10a. Are the control limits labelled?
10b. If labelled are the control limits labelled as sigma?
10c. One sigma?
10d. Two sigma?

Supplementary 6 – Coding Frame and Selected Results

10e. Three+ sigma?
10f. If labelled are the control limits labelled as standard deviation?
10g. One standard deviation?
10h. Two standard deviations?
10i. Three+ standard deviations?
11. Control limits recalculated?
12a. Run/trend highlighted?
12b. Run/trend 7+ points?
12c. Run/trend <7 points?
13. Does the chart have comments about reasons for variation?
14. Does the chart have suggestions for intervention?
15. Notes:

Response options to all except items 7 and 15 are either 'yes' or 'no'. Items 1-9 are filled out for all charts. Items 10-14 are only filled out for SPCs. Items 10b and 10f are only answered if the response to 10a is 'yes'. Items 10c-e are only answered if the response to 10b is 'yes'. Items 10g-I are only answered if the response to 10f is 'yes'. Items 12b-c are only answered if the response to 12a is 'yes'. RAG is captured at the board paper rather than chart level an aspect of the training was focussed on reducing RAG colouring more generally.

Selected Results

The main results for time series, between group, and SPCs are reported in the main manuscript. Considering all of the charts identified, there were also 129/6,287 (2%) pie charts

and 46/6,287 (1%) 'other' charts. Of the 46 'other' charts, 27/46 were donuts (59%), 9/46 were heat maps (20%), 4/46 were population pyramids (9%), 2/46 were scatterplots (4%), 3/46 were spider diagrams (7%), and 1/46 was a people plot (2%). RAG colouring was used by 18 of the 20 hospitals (90% of hospitals) at least once in their board papers.

Around half of all charts were contained within dashboards (3,348/6,287, 53%). Dashboards are Tables of data, including Tables of charts (an example is in Supplemental File 4, slide 2).

Further information specific to the 449 SPCs identified is shown in Table S6-2. The control limits were labelled for 342/449 (76%) of the SPCs. Considering only those 342 SPCs with labelled limits, sigma limits (139/342, 41%) were more common than standard deviation limits (12/342, 4%). However, it was most common that the labelled limits were not labelled as either sigma or standard deviation (191/342, 43%), using text such as UCL ('upper confidence limit') or LCL ('lower confidence limit').

Considering all 449 SPCs, half of the SPCs highlighted if a run or trend was present (215/449, 48%). Around one quarter included comments about reasons for variation (123/449, 28%) or suggestions for improvement (109/449, 24%). A minority recalculated the control limits (59/449, 13.2%). Note that we do not inspect differences between intervention and control hospitals on the characteristics of SPCs due to the small overall number of SPCs identified.

Supplementary 6 – Coding Frame and Selected Results

Table S6-2

Chart characteristics (SPCs)

Characteristic	SPCs (n=449) n (%)
Control limit labelled	342 (76.2)
Sigma limit	139 (31.0)
One sigma	0
Two sigma	0
Three sigma	139 (40.6)
Standard deviation limit	12 (3.5)
One standard deviation	0
Two standard deviation	12 (3.5)
Three standard deviation	12 (3.5)
Not sigma or standard deviation	191 (42.5)
Control limits recalculated	59 (13.2)
Run/trend highlighted	215 (48.0)
Comments about variation	123 (27.5)
Suggestions for improvement	109 (24.4)

References

1 Anhøj J, Hellesøe AMB. The problem with red, amber, green: The need to avoid distraction by random variation in organisational performance measures. BMJ Qual. Saf. 2017;**26**:81–4. doi:10.1136/bmjqs-2015-004951

2 Wheeler DJ. The Right and Wrong Ways of Computing Limits. *Qual Dig* Published Online First: 2010.https://www.qualitydigest.com/inside/six-sigma-column/right-and-wrong-ways-computing-limits.html

3 Wheeler DJ. Contra Two Sigma. 2013. www.spcpress.com/pdf/DJW255.pdf1 (accessed 29 Jan 2021).

4 Banchs RJ, Pop MR. *The quality improvement challenge : a practical guide for physicians*. Oxford: Wiley Blackwell

Supplementary 7- Sensitivity Analyses

Sensitivity analyses

Firstly to allow for full information on the pre-intervention number of SPCs (as opposed to condensing this information into a proportion as in our primary analysis) a Poisson regression model was fitted with an offset for the total number of charts and the outcome as the number of charts presented as an SPC, and with adjustment for group (intervention or control group), for period (pre or post intervention exposure) and an interaction between group and period (treatment effect). To allow for the dependence between repeated measures on the same hospital a random cluster effect (without any small sample correction, which is not supported in Stata 16 for count or binary outcomes) is included. We additionally tried to allow for degree of clustering to depend on period of measurement, but models with a random interaction between cluster and period failed to converge. Results are reported on the rate ratio scale with 95% confidence intervals. This is Model 1 in supplementary tables. This model was our planned primary analysis, but to allow for the over dispersion evident in the counts and for the many zero counts, we additionally extended these models to zero-inflated Poisson (Model 2, without a random cluster effect as this is not supported in Stata 16) and negative Binomial (Model 3), with a random hospital effects – again models with random hospital by period effects failed to converge. Due to lack of convergence of the random hospital by period interaction, we additionally fitted these models using logistic regression (this model can have better convergence properties), but this model with the random interaction also failed to converge and so is included with a random hospital effect only (Model 4).

Finally, to avoid reliance on mixed models (which might not be stable with only 20 hospitals especially without a small sample corrections) we additionally modelled the data conditioning on the proportion of charts which were SPCs in the pre-intervention period (as a fixed categorical effect) using a Poisson distribution (Model 5), negative Binomial (Model 6),

zero-inflated Poisson (Model 7) and zero-inflated negative Binomial (Model 8, our primary model). All models suggest a positive and large impact of the intervention, with the exception of model 1 to 3 (these models are not expected to be very reliable, as they use a random effect with only 20 clusters and does not allow for the over dispersion and zero counts which are prominent features of the data).

Supplementary 7- Sensitivity Analyses

Table S7-1 Different model approaches for primary analysis

	Model 1 – Mixed effects Poisson model	**Model 2 – Zero inflated Poisson model**	**Model 3 – Mixed effects negative Binomial model**	**Model 4 – Mixed effects logistic model**
	Rate ratio (95% CI)	**Rate ratio (95% CI)**	**Rate ratio (95% CI)**	**Odd ratio (95% CI)**
Group – intervention	14.61 (2.03 to 104.86)	4.46 (2.07 to 9.60)	7.49 (1.19 to 47.18)	13.96 (1.62 to 120.50)
Period – post intervention	1.76 (0.81 to 3.85)	1.81 (0.77 to 4.21)	1.46 (0.22 to 9.50)	1.79 (0.81 to 3.96)
Treatment effect - group # period	1.84 (0.81 to 4.18)	1.19 (0.491 to 2.86)	2.81 (0.22 to 35.29)	3.19 (1.36 to 7.48)
Changes	Original analysis plan	Model used adjusts for zero cells in outcome	Model used adjusts for overdispersion in outcome	Model treats outcome as binomial not count data
Outcome	Number of SPCs	Number of SPCs	Number of SPCs	Number of SPCs
Covariates	Group (intervention or control), period (pre/post intervention) and treatment effect	Group (intervention or control), period (pre/post intervention) and treatment effect	Group (intervention or control), period (pre/post intervention) and treatment effect	Group (intervention or control), period (pre/post intervention) and treatment effect
Random effect	Yes, hospital	No	Yes, hospital	Yes, hospital
Comments	**Issue using random effects for small number of clusters (hospital) and doesn't adjust for zero cells in outcome**	**This model adjusts for zero cells in outcome but not clusters (hospital).**	**Issue using random effects for small number of clusters (hospital)**	**Issue using random effects for small number of clusters (hospital)**

Supplementary 7- Sensitivity Analyses

	Model 5 – Poisson model	**Model 6 – Negative Binomial model**	**Model 7 – Zero inflated Poisson model**	**Model 8 – Zero inflated negative Binomial model**
	Rate ratio (95% CI)	**Rate ratio (95% CI)**	**Rate ratio (95% CI)**	**Rate ratio (95% CI)**
Group – intervention	14.71 (9.58 to 22.58)	17.90 (3.63 to 88.3)	4.71 (3.03 to 7.31)	9.24 (2.68 to 31.87)
Pre-measurement	10.23 (4.42 to 23.65)	5.66 (0.01 to 6258.06)	4.96 (2.03 to 12.13)	3.24 (0.10 to 100.30)
Changes	Data is reshaped to account for repeated measurements and avoid reliance on mixed models	Data is reshaped to account for repeated measurements and avoid reliance on mixed models and model used adjusts for overdispersion	Data is reshaped to account for repeated measurements and avoid reliance on mixed models and model used adjusts for zero cells.	Data is reshaped to account for repeated measurements and avoid reliance on mixed models and model used adjusts for zero cells and overdispersion.
Outcome	Number of SPCs in post-intervention measurements	Number of SPCs in post-intervention measurements	Number of SPCs in post-intervention measurements	Number of SPCs in post-intervention measurements
Covariates	Group (intervention or control) and pre-intervention proportion of SPCs.	Group (intervention or control) and pre-intervention proportion of SPCs.	Group (intervention or control) and pre-intervention proportion of SPCs.	Group (intervention or control) and pre-intervention proportion of SPCs.
Comments	**No adjustment made for zero cells or overdispersion.**	**This model adjusts for overdispersion in outcome but not zero cells.**	**This model adjusts for zero cells in outcome but not overdispersion.**	**This model adjusts for zero cells and overdispersion in outcome, out primary model.**

Supplementary 7- Sensitivity Analyses

Table S7-2 SPC usage by group, hospital and period out of time series charts

Control group				**Intervention group**			
Hospital	**Pre-training** SPC/Chart (%)	**Post-training** SPC/Chart (%)	**Post– Pre** % difference	**Hospital**	**Pre-training** SPC/Chart (%)	**Post-training** SPC/Chart (%)	**Post– Pre** % difference
1	0/57 (0)	0/69 (0)	0	**11**	0/190 (0)	9/184 (5)	5
2	0/71 (0)	0/97 (0)	0	**12**	0/149 (0)	0/117 (0)	0
3	0/12 (0)	2/53 (4)	4	**13**	0/109 (0)	0/77 (0)	0
4	0/638 (0)	0/665 (0)	0	**14**	3/115 (3)	91/243 (37)	34
5	0/146 (0)	0/163 (0)	0	**15**	52/107 (49)	47/63 (75)	26
6	0/78 (0)	11/155 (7)	7	**16**	0/69 (0)	58/81 (72)	72
7	0/138 (0)	0/137 (0)	0	**17**	0/11 (0)	27/52 (52)	52
8	0/92 (0)	0/93 (0)	0	**18**	18/137 (13)	42/404 (10)	-3
9	2/148 (1)	6/178 (3)	2	**19**	0/80 (0)	25/80 (31)	31
10	0/99 (5)	0/85 (0)	0	**20**	8/112 (7)	20/93 (22)	15
Average difference in control group (95% CI)			0 (0 to 2)	**Average difference in intervention group (95% CI)**			19 (7 to 30)
				Average difference between intervention and control group* (95% CI)			18 (7 to 29)
				Average rate change between intervention and control group ** (95% CI)			9 (3 to 29)

For each hospital in pre and post intervention period, the number of SPCs, the number of all charts and percentage of SPCs out of time series charts are reported

* T-test comparing average difference in proportions between intervention and control group. Percentage difference and 95% confidence intervals are reported.

** Zero-inflated negative Binomial regression models. Outcome is number of SPCs in post-intervention period, adjusting for pre-intervention proportion of SPCs. Exposure is time series charts. Rate ratios and 95% confidence intervals are reported.

Supplementary 7- Sensitivity Analyses

Table S7-3 SPC usage by group, hospital and period out of time series and between group charts

Control group				Intervention group			
Hospital	**Pre-training** SPC/Chart (%)	**Post-training** SPC/Chart (%)	**Post–Pre** % difference	**Hospital**	**Pre-training** SPC/Chart (%)	**Post-training** SPC/Chart (%)	**Post– Pre** % difference
1	0/13 (0)	0/36 (0)	0	**11**	0/13 (0)	0/11 (0)	0
2	0/34 (0)	0/51 (0)	0	**12**	0/27 (0)	0/29 (0)	0
3	0/4 (0)	0/32 (0)	0	**13**	0/60 (0)	0/43 (0)	0
4	0/1 (0)	0/8 (0)	0	**14**	2/25 (8)	1/28 (4)	-4
5	0/6 (0)	0/16 (0)	0	**15**	20/29 (69)	17/26 (65)	-4
6	0/7 (0)	0/9 (0)	0	**16**	0/42 (0)	18/34 (53)	53
7	0/2 (0)	0/1 (0)	0	**17**	0/6 (0)	0/13 (0)	0
8	0/38 (0)	0/37 (0)	0	**18**	0/21 (0)	0/37 (0)	0
9	0/10 (0)	0/18 (0)	0	**19**	0/21 (0)	0/20 (0)	0
10	0/5 (0)	0/1 (0)	0	**20**	0/56 (0)	0/37 (0)	0
Average difference in control group (95% CI)			0 (0 to 0)	**Average difference in intervention group (95% CI)**			10 (0 to 21)
				Average difference between intervention and control group* (95% CI)			10 (0 to 20)
				Average rate change between intervention and control group ** (95% CI)			Non-convergence

For each hospital in pre and post intervention period, the number of SPCs, the number of all charts and percentage of SPCs out of time series and between group charts are reported

* T-test comparing average difference in proportions between intervention and control group. Percentage difference and 95% confidence intervals are reported.

** Zero-inflated negative Binomial regression models. Outcome is number of SPCs in post-intervention period, adjusting for pre-intervention proportion of SPCs. Exposure is time series and between group charts. Rate ratios and 95% confidence intervals are reported.

Supplementary 8 – Discussion of Proportional Changes

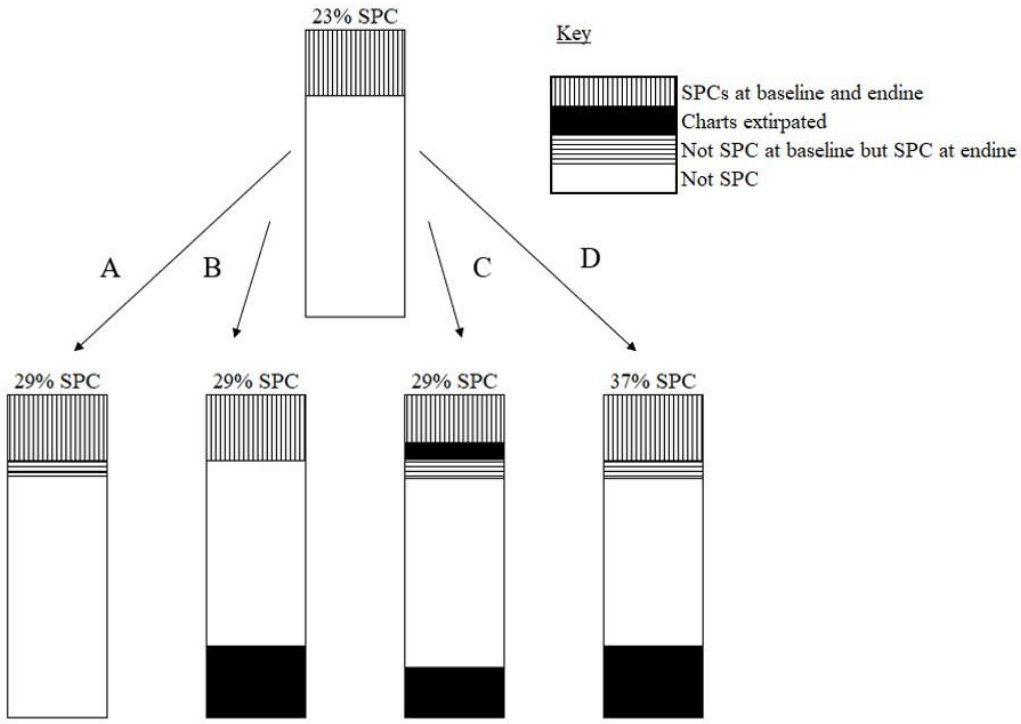
There are several mechanisms by which the proportional changes in this study could be brought about. Firstly, as intended, charts that were previously not produced using statistical process control methodology could be transformed into SPCs. This scenario is represented in Pathway A in Figure S8-1 below. However, we propose three additional ways that the proportion of SPCs could increase. First, non-SPCs are eliminated, and all SPCs maintained, with no new SPCs generated (Pathway B). Although there are no new SPCs in this scenario, the proportion of SPCs increases only because the denominator decreases. Second, SPCs and non-SPCs are eliminated in equal or similar proportions while some previously non-SPCs are now produced using statistical process control methodology (Pathway C). Increased SPC usage is more ambiguous in this scenario because SPCs have been both eliminated and created alongside a decrease in the denominator. Third, all SPCs are retained, new SPCs are produced, and some non-SPCs are eliminated (Pathway D). Here more SPCs have been produced, but the reduced total number of charts exaggerates the apparent effect of the intervention on converting one type of chart to another. Note that these mechanisms are all possible, at least in theory, even in a randomised controlled trial, since the intervention could have prompted changes in the number of charts presented to boards as well as causing non-SPCs to be converted to SPCs.

We can investigate some of these possibilities by visual inspection of the data in Table 3 (in the main manuscript). There is one case where a drop in the denominator was accompanied an improving SPC ratio with a no substantive increase in the numerator (Trust 15). Here, the proportion could have increased purely because of a decline in the denominator. But we would have to assume that the type of chart that was eliminated was selectively the type that was not already captured in statistical process control form (Figure S8-1, Pathway B). If the SPCs and the non-SPCs both declined, then the only way an overall increase in SPCs could

be observed is through increased adoption of statistical process control methodology (Figure S8-1, Pathway C). There are also two cases in Table 3 where the number of SPCs in the numerator increased, and the denominator also increased (Trust 14, Trust 17). In this case, at least part of the increase could have resulted from adopting the type of chart that lends itself to statistical process control methods.

In theory, it would be possible to evaluate these other mechanisms more precisely with a more time intensive approach tracking longitudinal changes at the level of individual charts rather than aggregating information about charts at the hospital level. However, the same individual charts do not always appear across board papers over time.

Figure S8-1. Four examples of mechanisms by which the proportional changes in this study could occur.



Supplementary 8 – Discussion of Proportional Changes