An introduction to the use of control charts

Turning Data into Information for Improvement
Preface

This guide is designed to support Primary Care Trusts (PCTs) in using control charts. It can be read as a standalone document or as an adjunct to other resources such as those supplied as part of the NHS Institute programme ‘Turning Data into Information for Improvement’. Details are available at www.institute.nhs.uk/commissioning.

In this document we cover the basic concepts that PCT commissioners need to understand. The focus is on the key principles of the science of improvement and the application of control charts in supporting improvement initiatives. It is written in the context of the need for PCTs to deliver large scale, sustainable change where the impact of commissioning decisions has to be backed up by evidence of success.

This guide was created jointly by the Quality Observatory at NHS South East Coast and the NHS Institute. It draws upon an extensive history of work in the area of quality improvement by pioneers like Walter Shewhart, W Edwards Deming and Joseph Juran amongst countless others who have contributed to this field. We are particularly indebted to Robert C Lloyd PhD (Executive Director Improvement, Institute for Healthcare Improvement) for his support in introducing us to the concepts outlined in this guide. Many of the figures used draw directly from his work and more in-depth explanations of the concepts and other useful supporting ideas can be found in his book Quality Health Care: A Guide to Developing and Using Indicators.

Comments and ideas for further development are welcome. Please send comments to commissioning@institute.nhs.uk.

Contents

1 The Science of Improvement .................. 03
2 An Overview of Control Charts .................. 06
   The role of control charts .................. 06
   Definitions of variation .................. 07
   Using control charts in service improvement .................. 08
   Selecting an improvement strategy .................. 09
   Improving processes showing common cause variation .................. 10
   PDSA cycles and variation .................. 10
3 Demonstrating improvements .................. 14
   Introducing a case study .................. 14
   Proving that a change led to an improvement .................. 15
4 Conclusion .................. 17
5 Appendix: Frequently asked questions How do I select a control chart? .................. 18
   Where do the control limits come from? .................. 21
   Why 3-sigma and not 2-standard deviations? .................. 22
   What are the limitations of control charts? .................. 23
Glossary .................. 24
References and resources .................. 27
Quality improvement pioneers such as Deming, Shewhart and Juran introduced the world to what is now known in health settings as the ‘science of improvement’. They had realised that traditional ways of ensuring high quality products or services were inadequate.

Methods like ‘quality assurance’ tended to focus on identifying who had made a mistake or correcting faults when they were identified. This potentially addressed one end of the performance curve but did little to improve the overall quality of the products or services (often at the expense of waste and rework).

In contrast the science of improvement method seeks to look at the underpinning processes that create the products or services. This means that improvements can be made across the board as shown in Figure 1.

![Quality Assurance vs Quality Improvement](image)

**Figure 1:**
A comparison of Quality Assurance and Quality Improvement

These contrasting approaches can be recognised in the context of how different NHS providers have attempted to tackle the problem of long waiting times in A&E departments. Providers taking an assurance approach have tended to focus on actively managing patients who are nearing the four hour waiting time threshold. This has often consumed considerable manpower resources or led to inappropriate solutions (e.g. short term admissions). However, using the science of improvement principles, other providers have instead focused on the processes that introduce unnecessary delays thus reducing waiting times across the board whilst achieving the desired performance standard.

The afore-mentioned improvement pioneers also realised that the methods of data analysis used for improvement purposes needed to change. Common approaches that focus on describing and summarising performance distributions fail to recognise the wealth of information that is contained in the order of production (i.e. the performance of a process over time). They believed that to make improvements it is necessary to apply new forms of statistical analysis both to focus on the underlying processes and explore variations in time based performance. Only by truly understanding this variation can the right improvement strategy be selected.
Figure 2 reinforces this argument and shows how summary data can hide a story that is revealed when variation is explored over time. In this example, based upon attempts to improve immunisation rates, the aggregate data before and after a change seem to indicate that the change was an improvement. However, by exploring the variation in performance over time it can be seen that performance was actually improving up until the change and then seemed to decline as a result of the change. The story is therefore very different. The message is clear however, understanding variation is an important aspect of improving quality.

This example shows yearly figures for immunisation rates before and after a new system was introduced.

The aggregated data seems to indicate the change was a success.

However, viewing how the rates have changed within the two periods tells a very different story. Here we see that immunisation rates were actually improving until the new system was introduced. They then became worse.

Seeing this more detailed time based picture prompts a different response.

Wow! A “significant improvement” from 79% to 86%

Conclusion - The change was a success!

Now what do you conclude about the impact of the new system?

Figure 2: An example showing how summary data can be misleading
‘Run charts’ and ‘control charts’ are vital tools that allow improvement practitioners to understand the nature of variation in a process. They are part of a branch of statistics called ‘Statistical Process Control’ (SPC) which has a long and well researched academic pedigree. When coupled with the ‘Model for Improvement’ (shown in Figure 3) which underpins the science of improvement, these charts become a powerful force for helping commissioners achieve better healthcare services and health outcomes.

Figure 3: The Model for Improvement based on Langley et al. (1996)

For commissioners, run and control charts serve two vital purposes:

• Assisting commissioners to achieve a greater understanding of the variation inherent in current service delivery methods so as to select the appropriate improvement strategy

• Helping to prove when commissioning decisions have had an impact.
The role of control charts

Run and control charts usually show how a given measure (or ‘metric’) varies over time. They allow analysis of that variation through the application of simple analysis rules based upon patterns in the data. The two charts look broadly similar (i.e. showing the measure on the y-axis and time on the x-axis) with the main difference that a control chart includes additional features (i.e. horizontal lines representing the ‘upper control limit’ and ‘lower control limit’) based upon further statistical analysis of the data. These features will be described later in this section. Figure 4 shows an example of a control chart.

This guide focuses on using and understanding control charts since it is recognised that for some, the leap from a run chart to a control chart can be slightly bewildering especially given the existence of a range of different types of control chart. As we shall see, control charts are actually easy to create (especially with appropriate software) and just as easy as run charts to interpret. They also bring the added benefit of greater power to explore and understand process variation.

The control chart will always be at its most powerful if used as part of an overarching improvement programme or within ‘PDSA’ (Plan, Do Study Act) cycles where the impact of changes to a process are explored. It is of paramount importance that the control chart is not seen as a project in itself; it is a tool for understanding. It has to be used in conjunction with a good knowledge of the process or system to be improved. Control charts therefore need to be used by SMEs (Subject Matter Experts) not just analysts.

Control charts are used in any setting where variation occurs. Figure 4 shows a control chart looking at systolic blood pressure over time.

This simple example typifies the way in which a control chart is presented and used. Discussion of a ‘mean’ systolic blood pressure over a 28 day period is almost meaningless without exploring the detailed variation on a daily basis. By exploring this variation we can ask questions such as whether the blood pressure is predictable (i.e. statistically ‘in control’) or whether special events are causing unusual patterns of variation (termed ‘special cause’ variation). We can further explore the question of whether the variation in blood pressure is clinically acceptable and desirable. The answers to all these questions help us to formulate an improvement (or treatment) strategy that would not be possible from just a monthly average.
Figure 4: An extract from a XMR-control chart showing changes in systolic blood pressure over 26 consecutive days. Source: Based upon Mohammed et al. (2008)

Definitions of variation
The variation seen in the data shown in a control chart is split into two types.

‘Common cause variation’ is just that; common. Every normal process will exhibit natural (or common cause) variation. In healthcare, every patient is different, so every process will have inbuilt variation. Even simple processes exhibit variation. Just think about signing your name. Is your signature identical every time or just similar?

‘Special cause’ variation is something abnormal, the result of a specific change that has affected the process in a way beyond what would be expected under normal conditions. The analogy would be signing your name 20 times but inserting one signature made with your less dominant hand. This special cause change would probably show up as a glaring exception in the list of signatures. Similarly, if asked about your typical travel time to work you may describe a (common cause) variation in your journey times of between 45 and 60 minutes. However, you might also comment upon the day when a motorway accident led to a (special cause) trip time of 120 minutes. Common cause variation and special cause variation therefore refer to the ordinary and the extraordinary respectively.

By plotting data on a control chart over time, special cause variation can be clearly seen and differentiated from the common cause variation.
Using control charts in service improvement

Control charts can be used as part of an initial diagnostic process to understand the performance of a system. They can also be part of the related improvement journey where they can be used to show whether an intervention has had an impact.

In order to fully understand what a control chart is telling you, you should be familiar with your system or process, understand the drivers that can impact on it and be aware of any limitations of the metrics chosen to represent it. For example, is breastfeeding initiation a reasonable indicator (driver) of infant mortality and will improvement in this metric truly evidence an improvement in the system as a whole? Asking these kinds of questions to explore the drivers in your system is an important step towards using data effectively.

A helpful mechanism for representing a model for the drivers in a system is the ‘driver diagram’. Figure 5 shows part of a driver diagram created to illustrate the local factors that could lead to a reduction in teenage pregnancy rates. Here, two of the major primary drivers affecting pregnancy rates have been identified for a locality and metrics have been defined to represent them. The diagram then goes to the next level to define secondary drivers that impact upon the primary drivers, again with supporting metrics. These secondary drivers then become the improvement goals for specific improvement projects (not shown in the figure). Throughout this model, control charts can be used to understand the performance changes resulting from the individual projects. In turn, other control charts can show their resulting impact upon the secondary drivers and their subsequent impact upon the primary drivers. Ultimately the driver diagram allows a range of improvement projects to be connected to the overall goal of reduced pregnancy rates through a series of measurements that are interpreted through control charts.

In Figure 5 we show both the drivers and the data that might be used for the associated control charts.

**Figure 5**: A partial driver diagram exploring some of the drivers and associated measures linked to reducing teenage pregnancy rates
For PCTs, driver diagrams represent a useful tool for meaningfully connecting commissioning decisions (i.e. service changes) to the PCT’s overarching improvement goals (e.g. reducing inequalities). This ‘joining the dots’ between actions and strategic intentions is a key part of PCT governance arrangements as recognised in the world class commissioning assurance process.

As the driver diagram and its metrics illustrate, control charts are applicable to any scale of improvement project. The earlier clinical example relates to a single patient where the ultimate aim is the effective management of their blood pressure. At a larger scale, control charts can play a key part in complex system wide change programmes such as efforts to reduce alcohol related admissions or initiatives to reduce teenage pregnancy rates (as shown in figure 5).

**Selecting an improvement strategy**

When attempting to make improvements to a process, the improvement strategy you select should depend upon the type of variation you see in your data.

Where special cause variation is identified you are observing one-off factors affecting your system which are different from the usual (common cause) fluctuations that you see. Recall the earlier examples of signing your name with your less dominant hand or being delayed on the motorway on your way to work. In this situation, the appropriate change strategy begins with locating the origin of the specific special cause variation (e.g. the motorway accident). You then need to consider whether the resulting variation is desirable or not. If the variation is undesirable (i.e. it leads to a poorer performance of the process) then you may try to put in place mechanisms to guard against its effects in the future (i.e. manage the risk). You do not however change the underlying process since, by definition, this event was exceptional and out of the ordinary and unlikely to be due to a problem with your current process. The fact that I cannot sign my name very well with my less dominant hand does not mean I should alter the way I usually sign my name with my other hand. Similarly, I do not change my route to work because of a single unusual motorway delay.

However, if the special cause variation is desirable (i.e. representing improved performance) then you may have stumbled across a great improvement idea and could look for ways to make the unusual performance the norm. Again you would need to isolate the reason for the special cause variation and then make a considered judgement about whether to attempt to build it into the process. For example, if chlamydia screening rates temporarily (and dramatically) increase after a one-off talk by sexual health staff at the local college, you may want to try and see if these talks can become a regular feature at local colleges and universities.

In contrast, where a system or process shows common cause variation the appropriate change strategy is to look for ways to improve the underlying process (i.e. make changes that affect the majority of patients rather than just the exceptions). This requires a detailed understanding of the process (or system) and complimentary techniques like ‘process mapping’ and data ‘stratification’ (where tools like the ‘Pareto chart’ can be useful). Using the chlamydia screening example, you may choose to map out the booking process if you find you routinely have high DNA rates for screening and then experiment with changes in that process (e.g. trial using text reminders). The focus is clearly on the underpinning process.

These different change strategies are often confused. It is not uncommon to see staff hunting for the cause behind a month on month change in a particular metric, which if viewed over a longer time period would just be seen as within the natural (common cause) variation in the process. How often do we hear someone say “Explain why this figure has gone up 1% this month?” whilst they ignore the fact that the figure routinely changes by...
a percent or more between months. Here, treating a common cause variation as if it was a special cause is pointless as it is the underlying process (which creates all the monthly figures) that needs to be understood and changed to create an improvement. Reacting to every small (common cause) change in a system just leads to wasted efforts and even less predictable results.

The reverse can also apply where staff overreact to a genuine, unique (special cause) event by changing the whole process. Does it really seem sensible to alter the entire process for emergency care based on what happens on New Year’s Eve each year or during a major incident? Instead you put in place contingency plans to manage these rare events as and when they occur.

Figure 6 summarises the decisions you need to make when determining your change strategy.

Figure 6: The decision tree for determining a change strategy
Improving processes showing common cause variation

Hunting for special causes in a control chart is quite simple using the ‘rules’ that we will explore later and it is often not too difficult to attribute reasons for the special causes that are identified. However, improvement teams sometimes struggle to apply an improvement strategy when their process just shows common cause variation (i.e. the right hand branch of Figure 6).

As noted earlier, the basic technique is to focus upon the underlying process. You might start by mapping the process and then choose to test changes to the process to see if they lead to better outcomes. In the Model for Improvement this is where you would use the PDSA cycles.

Sometimes you will need to unpick your problem to a greater level of detail. ‘Stratification’ (i.e. dividing up data into categories) is a simple and powerful technique for providing a focus to improvement efforts. It can help where the outputs you see are actually due to multiple processes or where a single process may have different results for different patient groups. Let us take the example of smoking cessation clinics and see how stratification could be applied.

As a PCT imagine you have four clinics operating along broadly similar lines in different parts of the PCT area. A control chart showing the total weekly ‘quit’ rate for the whole PCT displays just common cause variation. Unfortunately the mean (average) quit rate is much lower than you hope for. What could you do? This is where the subject matter knowledge of your SMEs comes into play as you need to select ways to stratify your data that will reveal different aspects of the underlying processes. The first (and obvious) way to stratify your data is by clinic since it is not unreasonable to expect that processes may differ slightly in different areas and thus yield different outcomes.

A second way to stratify the data might be by intervention type (e.g. advice courses, hypnotherapy etc.). You could perhaps consider different groups of people (e.g. young smokers versus long term smokers).

For the clinic based stratification you could plot individual control charts for each clinic. Sometimes when we do this type of stratification we find that although the aggregated ‘all clinics’ data is in control (i.e. just showing common cause variation) the more detailed charts for ‘individual clinics’ show some special causes that get hidden in the ‘all clinics’ results. These newly revealed special causes may yield useful learning.

Also, in creating the control charts you might identify one clinic that has comparatively poor performance in relation to the others and so choose to focus your efforts there. Similarly, if you stratify by intervention type you may find that one specific type of intervention accounts for many of the failures to ‘quit’. In both cases, this method of targeting our improvement efforts can pay dividends if used appropriately.

Discussions of stratification approaches and ways to identify improvement opportunities are given in many of the books on process improvement. Helpful healthcare related examples described in the context of control charts are given by Balestracci et al. (1998) and Lloyd et al. (1995).

PDSA cycles and variation

The preceding sections have talked about how we analyse the variation in a process to determine the appropriate change strategy. But what about when we deliberately introduce a change to a process as part of a PDSA cycle? What role do control charts play then?

This is where the control chart changes its role from being a diagnostic tool to become a tool for identifying whether a change is an improvement. When we use control charts diagnostically we identify special causes through analysis of the data, determine their origin and then decide whether to build the cause into our usual process (or manage the risks). The PDSA cycle reverses this to actively seek special cause variation as a result of a pre-determined intervention (i.e. a pre-determined test of a change in the process).
We hope to see a special cause variation in performance since we have deliberately altered the process so that it is different from its usual (common cause) way of operating.

We will work through an example of how PDSA cycles are analysed through control charts in Section 3.

**Rules for interpreting control charts**

There are several simple rules, based on statistical theory and empirical observation, that can be applied in order to assess whether or not data exhibits special cause variation. These are born from the work and experience of many experts around the world. These ‘rules’ can sometimes vary slightly depending on the exact nature of the process or how sensitive one wishes to be in identifying special causes, but for the vast majority of healthcare processes a simple set of these rules are sufficiently rigorous to be of great use.

Some of the rules rely on defining what are known as ‘control limits’. The calculation of these will be discussed in more detail later but they can be thought of as lines (‘upper’ and ‘lower’ control limits) representing a natural boundary for identifying extremely high or extremely low data. They are based upon the variation observed in the data and thus they are a statistically derived calculation of the point at which it becomes very unlikely that data points will be observed. They may sound complicated but they actually make intuitive sense. My journey to work tends to take 45-60 minutes and I would know something very different from the usual has happened if I make it there in 20 minutes or it takes as long as 85 minutes. If I recorded my regular journey times I could calculate my exact lower and upper control limits and check out my intuitive figures of 20 and 85 minutes.

The rules introduce another term that can sometime create confusion. The upper and lower control limits are also called the ‘3-sigma limits’. This term is less common but it does have the advantage that it becomes easier to talk about where data points are positioned in relation to the control limits. As we shall see, two of the rules identify regions based upon 1-sigma and 2-sigma limits. To understand what this means, let’s take the earlier systolic blood pressure example (figure 4) where the upper control limit (3-sigma limit) was 202.5 mmHg (the topmost dashed line in the chart) and the mean was 173.2 mmHg. The difference between the mean and the 3-sigma limit (upper control limit) is 29.3 mmHg. We would therefore place the 2-sigma limit at two thirds of this difference (i.e. 192.7 mmHg) and the 1-sigma limit at one third of the difference (i.e. 183.0 mmHg). This creates convenient ‘zones’ which can be used to describe where data points lie (e.g. the zone between the 2-sigma and 3-sigma limits).

The rules for identifying special cause variation are very straightforward:

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule 1</td>
<td>One or more points outside of the control limits</td>
</tr>
<tr>
<td>Rule 2</td>
<td>Eight successive consecutive points above (or below) the ‘centre line’ (ignoring any points exactly on the centre line)</td>
</tr>
<tr>
<td>Rule 3</td>
<td>Six or more consecutive points steadily increasing or decreasing (ignoring any point that is identical to its predecessor). This is defined as a ‘trend’.</td>
</tr>
<tr>
<td>Rule 4</td>
<td>Two out of three successive points beyond +/- 2-sigma</td>
</tr>
<tr>
<td>Rule 5</td>
<td>Fifteen consecutive points within +/- 1-sigma of the centre line (i.e. ‘hugging’ the centre line)</td>
</tr>
</tbody>
</table>

Examples of these rules are shown in Figure 7
The rules should make intuitive sense since they are simply looking for patterns that are unlikely to occur by chance. For example, consider the ‘drive to work’ situation already described. We have already considered the exceptionally long or short journeys (Rule #1). How about if I experience a run of longer than usual journeys due to a local diversion? This might give a pattern like Rule #2 where the special cause is the diversion. How about if it is the school holidays and the roads are quieter than usual? Might these lead to a Rule #2 example (i.e. lots of shorter than usual journeys) or perhaps even Rule #5 (i.e. less variation than usual in my journey times)? Hopefully you can see that the rules are simply reflecting our common sense interpretation of ‘unlikely patterns’ and applying a few probability calculations to add more rigour.

**Figure 7:** Examples of the five control chart rules

1. A single point outside the control limits

2. A run of 8 or more points in a row above (or below) the centre line

3. Six consecutive points increasing (trend up) or decreasing (trend down)

4. Two out of three consecutive points near a control limit (in the outer one-third)

5. Fifteen consecutive points close to the centre line (in the inner one-third)
Introducing a case study

We will now explore an example of how a control chart has been used in a real improvement project. The example, from the work of Saturno et al (2000), is based on door-to-needle times for provision of thrombolytic drugs in acute myocardial infarction. In this example, we follow the use of control charts throughout a programme to improve the efficacy of treatment, where it was known that the sooner thrombolysis is administered, the greater the medical benefit.

At the start of this particular programme, it was decided to measure the time from arrival at hospital to administration of thrombolytic therapy. Other measures were also taken for each stage of the patient journey but we will concentrate on the time to administration. The first recording phase was used to capture baseline data (i.e. create a measure of how the current process was performing). This was plotted on an x-type control chart, which is shown in the first section of Figure 8. The ‘door-to-needle’ time is shown in minutes on the vertical (y) axis, with the measurement number shown on the horizontal (x) axis (i.e. the ‘door-to-needle’ time was recorded consecutively for each patient).

Figure 8

Three phases of monitoring in a project to improve door-to-needle times for thrombolytic treatment.

Source: Saturno et al. (2000)

To interpret the chart it is necessary to apply the simple interpretation rules mentioned earlier to identify whether special cause variation is present. It can be seen from the figure that the variation is quite considerable and that one data point matches a special cause rule (Rule #1). It is above the dotted line which is marking the upper control limit. We can see that this patient waited almost four hours for thrombolytic treatment. It is important to keep in mind that this “special cause” is a statistically derived figure (based on Rule #1). It is not by itself evidence of clinical unacceptability as the ‘unacceptability’ is a judgment rather than a statistical measure. This special cause shows instead that something special or one-off happened to delay this patient’s treatment that was not inherent to the usual process which is experienced by the other patients. Other than this point, the pre-intervention phase shows no other indication of special cause variation, and the ‘door-to-needle’ time would be considered as being subject to common cause variation only.
As noted earlier, the change strategy appropriate for data with special cause variation is to firstly identify the reason for the special causes. In this instance the improvement team investigated the long wait and identified the reason behind it. As it was clearly an undesirable variation the team did not alter the underlying process on the basis of this one-off event. The project team then moved on to focus on the remaining common cause variation in the process (i.e. the pattern of variation affecting all of the other patients). Having patients waiting an average of 70 minutes for administration of the therapy, with some waiting up to three hours, is clearly unacceptable and potentially life threatening. The project team therefore decided that an improvement was needed.

By reviewing the process for giving thrombolysis, assessing clinical notes and undertaking surveys of staff, the improvement team created an improvement programme which was largely focussed on communication between A&E and ICU as well as patient transport issues. These improvements focused on the underlying process common to all patients receiving thrombolysis.

After these changes were implemented (the post intervention phase, shown in the second section of Figure 8), the data was again plotted. This second phase data shows much less variation between patients, and the mean has dropped considerably to 30 minutes. It might be tempting to just leave it there, and indeed the improvement appears to have been dramatic. But is it sustainable? A close inspection of the second phase data shows an indication of special cause variation (eight consecutive points above the mean line – Rule #2) which is worth investigating as it may have implications for continuing good practice. The improvement team in this example investigated this special cause and then continued to a third phase to ensure continuity and a stable (predictable) process that met their clinical expectations. This phase shown in the third section of Figure 8 shows no special cause variation and thus represents a stable (‘in control’) process which, for the team, also showed clinically acceptable waiting times.

Proving that a change led to an improvement

However, in telling this improvement story we ignored a crucial stage in the improvement process. Although visually it is clear that a dramatic change occurred between Phase 1 and Phase 2, we neglected to say how we know statistically that a change (in this case an improvement) occurred. There is a risk that the apparent improvement is no more than an artefact of chance (common cause) variation rather than a genuine improvement.

To prove that this change was more than just a random change in performance the team followed three simple steps:

**Proofing that a change led to an improvement**

**Step 1** They calculated the mean waiting time and control limits for their baseline period (the first section of Figure 8). The calculation of these dotted line limits is discussed shortly.

**Step 2** They then froze these values at the time of the change and continued plotting more data, looking for evidence of a special cause using the frozen mean and control limits. Note that the special cause they found in the Phase 2 data is not the one circled in the figure as that is based on the recalculated control limits. The actual special cause is the first eight points which are all below the original mean (Rule #2).

**Step 3** When they found a special cause they realised they did indeed have evidence of a change in the waiting times so they could celebrate success and calculate a new mean waiting time and control limits using the data from after the change (as shown in the second section of Figure 8).
These steps are not complicated but it is important to understand the logic embedded with them.

In Step (1) the team were applying a statistical analysis diagnostically to differentiate between special cause and common cause variation. They did this because their change strategy depended on the type of variation present. Existence of a special cause required a strategy of isolating its source. Once isolated, the special cause case could be safely ignored allowing the team to focus on the underlying common cause variation. Here the appropriate change strategy was to alter the process which is common to all the patients receiving thrombolytic therapy. The team did this by reviewing notes etc. and identifying changes relating primarily to communication and transport.

In Step (2) they wanted to prove that their attempts to change the process had indeed had a measurable, lasting effect on the waiting times. This shows up (statistically) through the introduction of new special causes into the data identified through the ‘rules’ as we noted earlier. In this case, the special cause is a marker that something ‘out of the ordinary’ for the original process has occurred (i.e. we have successfully changed the underlying performance of the process). Ideally in this case we would hope to see Rule #2 applied (as the waiting times settle on a new low value) or Rule #3 (where the waiting times steadily decrease down to a new low). Since they saw new special causes (all in the right direction!) they realised they must have indeed changed the underlying process successfully.

Visually it is simple to follow this step yourself. Just extend the dotted lines shown in Phase 1 into the Phase 2 area. It is pretty clear that all the points in Phase 2 are below the old mean. This is a very obvious example of a Rule #2 special cause.

Finally in Step (3) they re-analysed the new waiting times generated by the process (ignoring the old data) to understand how predictably the new process was now operating.

In fact if one looks carefully at the Phase 2 analysis one can see that the new process data showed the team that (after recalculating the control limits) other new special causes were now present in the data after the change. Once these had been investigated (much like the special cause in Phase 1) and the team was confident the process was now stable (i.e. predictable) they decided to identify a new ‘business as usual’ phase. For this Phase 3 they therefore calculated a new mean and control limits where they could confidently describe the average waiting times as having been reduced to an impressive 26 minutes with a predictable (in control) performance that ensured virtually all patients would be treated within the upper control limit time period of 46 minutes.

This thrombolysis improvement project is an excellent example of the application of control charts as an ongoing monitoring method within an improvement initiative and shows how control charts can be used as an organic tool, growing and evolving to meet the team’s needs. The chart clearly shows the changes in performance and the labelling makes it clear how these changes relate to actions taken to improve the process. It should also be clear that the chart is an integral part of the improvement story but it is not the whole story. The local SMEs had to know where to look for improvement opportunities and be capable of doing the detective work required to understand the special causes and redesign the underlying processes.
We have seen that control charts play a vital role in healthcare improvement and are a key element of the science of improvement developed by pioneers like W Edwards Deming, Walter Shewhart and Joseph Juran. Fundamentally, control charts help us maintain a focus on the underlying processes that deliver healthcare and support us in understanding the variation that occurs in the outputs of a process over time.

Our understanding of special cause and common cause variation allows us to move from a process of diagnosing problems to selecting appropriate change strategies and ultimately to seeking evidence of change after we have altered a process. Control charts are therefore part of the entire commissioning journey as we assess the value and success of current patterns of services and move on to intervene in the healthcare system by altering existing services or commissioning new ones.

At the end of the day control charts are just a tool, albeit a powerful one. They have to be accessible and understandable to those who need them and only have value if wielded by teams and individuals who have local knowledge of the systems and processes the control charts describe. Fortunately, control charts successfully combine statistical rigour with simplicity. All commissioners should be capable of determining when to use a control chart and how to interpret one. Commonly available software makes it easy to produce these charts whilst the rules outlined here make the interpretation of process variation simple.

In the era of world class commissioning, commissioners increasingly have to be capable of understanding, exploring and changing the health systems they work with. Control charts are a vital aid in this process and can also help to answer the key question “Have we made a difference?”.
Frequently asked questions
The following sections provide more technical detail for those who will be required to create control charts or who simply want a deeper level of understanding.

How do I select a control chart?
Control charts are easy to create using SPC software and with a little effort can also be created in basic spreadsheet programmes like MS Excel. The key issue is selecting the right sort of chart for the data that you have. Experts vary in their views on whether it is necessary to use more than one control chart type. One view is that the XmR chart (described later) is appropriate for most data types. Here we follow an alternative view that specialist chart types (e.g. the p-chart) offer a more sensitive way to detect special cause variation.

To determine the specialist chart type, the most crucial distinction is whether you have ‘continuous’ data (also known as variable or measurement data) or ‘attribute’ data (also known as discrete or count data). Continuous data can be measured along a continuous scale (e.g. length, weight etc.) and examples include waiting times, blood pressure readings and the number of procedures performed in a day. In contrast, attribute data is based on categories. A patient is either alive or dead, pregnant or not pregnant. They either fell or did not fall.

For attribute data we also need to draw a distinction between ‘defectives’ and ‘defects’. Essentially this relates to whether or not we can count up the opportunities for an event of interest to occur.

For defectives we know the actual opportunities. If we take the example of a particular surgical procedure we can measure the number of patients who died and the total number who had the procedure. In this case we know the defectives (i.e. number of deaths) and the opportunity (i.e. the total number of patients who had the procedure). We can thus calculate a percentage.

For defects we do not know the actual opportunity and only know an ‘area of opportunity’. Falls are a good example. We can measure whether someone fell. We cannot measure the number of times someone did not fall. All we can calculate is the area of opportunity which in this case could be the length of time the patient was in hospital. So instead of a percentage we calculate a fall rate (i.e. the number of falls per inpatient day). Thus another way of recognising defectives from defects is to see whether you choose to calculate a percentage (i.e. numerator and denominator are in the same units) or a rate (i.e. numerator and denominator are different units).

The terms defect and defective just refer to the measure of interest such as deaths, falls or pregnancies and they do not imply a good or bad judgment in this context.

To determine your control chart type you need to answer a number of questions about the measure you are going to plot. Figure 9 shows how you can choose the correct control chart for your dataset.
Figure 9: Choosing the correct control chart for the dataset
Source: Based upon Lloyd et al. (1995)
We can walk through Figure 9 by applying a number of questions at the different levels shown:

At Level 1 the first simple question is “What is your data type?” (i.e. variable or attribute data).

If you have variable data, you follow the left hand branch of figure 9 and your next question (at Level 2) is “Do I have more than one observation (or measurements) for each subgroup?” In simple terms, a ‘subgroup’ is a grouping which is represented by just a single point on a control chart. For example, you could measure and plot the waiting time in A&E for each patient in which case you have one observation per patient. Here the single patient is the subgroup. Similarly you could choose to look at the average time it takes ambulances to respond to category A calls by taking a sample of five response times each hour. In this case your observation size is five as you have taken five measurements for each average point you plot on your control chart (where the subgroup is now the “hour of the day”).

If you have just one observation per subgroup then you use an ‘XmR chart’. The XmR chart is actually made up of two charts, an ‘x-chart’ that plots the actual measurements and the ‘mR-chart’ that plots the moving range (i.e. the change between successive points). Both types of chart use the same rules mentioned earlier to identify special causes. In the earlier thrombolysis example we had variable data (waiting times) and plotted measurements for each patient (i.e. one observation per subgroup). Figure 6 thus showed the X-chart component of the XmR-chart. The X-chart component is also known as an ‘individuals chart’.

If you have variable data and more than one observation per subgroup, you need to ask (Level 3) “Do I have less than ten observations per subgroup?”.

When you are using small samples (i.e. less than ten observations) then the appropriate chart is an ‘X-bar & R chart’. If your sample is larger than ten you instead use an ‘X-bar and S chart’. The ‘R’ refers to ‘range’ whilst the ‘S’ refers to ‘standard deviation’.

If you have attribute data instead of variable data you take the alternate right hand branch in Figure 7. Here your Level 2 question is “Am I measuring ‘defectives’ or ‘defects’?”. This question can be answered by referring to the earlier descriptions.

To determine the type of control chart you need to ask a further (Level 3) question. For defectives this is “Do I have a constant subgroup size?”. In this context the ‘subgroup’ is the opportunity described above. Thus, if we were looking at caesarean section percentages and we always looked at 150 births for each period (and thus each point on our control chart was based on 150 births) the subgroup size would be constant (at 150). In healthcare the more likely scenario is that you will not have constant subgroup sizes. For example we would tend to look at caesarean sections performed in a week where the number of births will vary between weeks. In cases where we have constant subgroup sizes we use an ‘np-chart’ whereas for unequal subgroup sizes we use a ‘p-chart’.

A similar question is used if we are now considering ‘defects’. In this case the question is “Do I have an equal area of opportunity?” We discussed areas of opportunity earlier. In healthcare it is actually much more common to have an unequal area of opportunity. Medication errors in primary care are a good example. For a particular GP practice the area of opportunity per week will vary according to the numbers of patients seen. We would therefore choose to calculate something like ‘errors per 1000 patients’ and plot this on a control chart. For this type of unequal opportunity we would use a ‘u-chart’. If however we presumed that the numbers of patients did not actually change each week (which might be the case for some specific services with dedicated appointment slots) we would just choose to count the number of medication errors. In this instance we have decided we have an equal area of opportunity and thus the appropriate chart is a ‘c-chart’.
Where do the control limits come from?

The control limits (or 3-sigma limits) on a control chart are usually based on the assumed data distribution type for each version of chart (e.g. normal, poisson, binomial etc.). The exception is for the XmR chart which uses the actual variation in the data by considering the moving ranges between points. In each case the control limit represents a boundary beyond which we will only rarely expect to see data points (i.e. the extreme tails of each distribution or a multiple of the average moving range for XmR charts).

Calculating where to place the centre line and control limits for each type of chart is relatively straightforward. For XmR charts the calculation of the control limits is very easy as it only requires that you work out the average moving range between points (and then multiply this by 2.66). For the more adventurous it is possible to also calculate the control limits for other types of charts with a bit more effort. The formulae are given in Table 1 for some of the more simple charts. However, many SPC software applications exist that will calculate control limits for you and most will even apply the special cause rules for you.

As can be seen from Figure 10, for some types of control charts the control limits take on a ‘stair-stepped’ appearance. For example, in a p-chart the upper and lower control limits depend on the denominator of each percentage plotted on the chart (and thus vary from point to point). This does not affect how you use the rules described earlier. In some instances when you calculate upper or lower control limits you may find that they are beyond natural boundaries (e.g. > 100% or < 0 for measures that cannot become negative such as waiting times). In these instances it is conventional to reset the control limits to match the natural boundaries. Clearly, where this is done Rule #1 will never apply. The 1-sigma and 2-sigma boundaries are maintained at their calculated values so that Rules #4 and #5 can still be applied.

Where observations happen very infrequently (e.g. MRSA infections, or deaths from certain rare diseases), a rare event chart can be used. To plot this, the time between events is calculated and used as your measure on the chart. This can sometimes require a transformation of the data to the normal distribution to allow the calculation of control limits. This type of chart is known as a ‘g-chart’. Detail on this chart and other rarely used types can be found in specialist SPC texts. See for example Benneyan (2001).
Why 3-sigma and not 2-standard deviations?

People sometimes incorrectly use the terms ‘sigma’ and ‘standard deviation’ interchangeably. Although based on similar concepts relating to the distribution of data around the mean, these two measures are calculated differently. The calculations used for 3-sigma for some of the basic control charts are shown in Table 1 and one can see that the upper and lower control limits depend upon the underlying data type.

The question of why 3-sigma is used instead of 2-standard deviations to identify outlier (special) points arises from the common use of 2-standard deviations in research practice as a threshold for significance. One can see that the concept of a ‘significant’ research result bears a striking similarity to Rule #1 which also hunts for individual results considered beyond normal chance. In research, the choice of two standard deviations is essentially a rounding of the 1.96 standard deviations that encompass 95% of observations in a normal distribution (i.e. a result is exceptional if chance would make it occur just one time in 20). Whilst this might be appropriate in a research environment, where variables are tightly controlled and conditions are conducive to a ‘fair test’, in the applied field where we are considering multiple time-base measurements the chance of a false positive is too high. For example, a typical control chart consists of 20-25 data points. With this many data points we would generally expect one point to be outside 2-standard deviations purely by chance rather than due to a special cause. Thus to be reasonably confident of a special cause we need to be more restrictive on our control limits hence the 3-sigma limit. The choice of 3-sigma has actually been made through a combination of statistical theory and empirical observations. The intention is to avoid false positives (which is called a ‘Type I error’) whilst also avoid missing actual special causes (which is called a ‘Type II error’).

Although this paper has focused on the control limits and their relationship to Rule #1, the concept of looking for statistically unusual patterns is what underpins all of the five rules mentioned earlier. For example, consider Rule #2 (8 consecutive points above or below the centre line). In effect we can picture this rule like tossing a coin. If it lands as ‘heads’ we plot a point above the centre line. If it is ‘tails’ goes below the centre line. The chance of getting 8 ‘heads’ (or ‘tails’) in a row is rare enough for us to deem any occurrence as likely to be due to a special cause event.
Variable X Normal \[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \]
\[ \bar{X} \pm 2.66 (mR) \]

mR Normal \[ \bar{mR} = \frac{\sum_{i=2}^{n} mR_i}{n-1} \]
\[ 3.27 (mR) \]

Attribute p (or np) Binomial
\[ \bar{p} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} n_i} \]
\[ \bar{p} \pm 3 \sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}} \]

U Poisson
\[ \bar{u} = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} n_i} \]
\[ \bar{u} \pm 3 \sqrt{\bar{u}/n_i} \]

C Poisson
\[ \bar{c} = \frac{\sum_{i=1}^{n} c_i}{n} \]
\[ \bar{c} \pm 3 \sqrt{\bar{c}} \]

* The calculations for the X bar and R chart and the X bar and S chart are not shown here as these are slightly more complex. Readers are recommended to use one of the references at the end of this guide if you want the formulae for these charts.

**Table 1: Calculating 3-sigma limits for different data types**

<table>
<thead>
<tr>
<th>Data type*</th>
<th>Type of chart</th>
<th>Distribution</th>
<th>Centre line</th>
<th>3-sigma control limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>X</td>
<td>Normal</td>
<td>[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} ]</td>
<td>[ \bar{x} \pm 2.66 (mR) ]</td>
</tr>
<tr>
<td>mR</td>
<td>Normal</td>
<td>[ \bar{mR} = \frac{\sum_{i=2}^{n} mR_i}{n-1} ]</td>
<td>[ 3.27 (mR) ] (upper control limit only)</td>
<td></td>
</tr>
<tr>
<td>Attribute</td>
<td>p (or np)</td>
<td>Binomial</td>
<td>[ \bar{p} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} n_i} ]</td>
<td>[ \bar{p} \pm 3 \sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}} ]</td>
</tr>
<tr>
<td>U</td>
<td>Poisson</td>
<td>[ \bar{u} = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} n_i} ]</td>
<td>[ \bar{u} \pm 3 \sqrt{\bar{u}/n_i} ]</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Poisson</td>
<td>[ \bar{c} = \frac{\sum_{i=1}^{n} c_i}{n} ]</td>
<td>[ \bar{c} \pm 3 \sqrt{\bar{c}} ]</td>
<td></td>
</tr>
</tbody>
</table>

What are the limitations of control charts?

There are very few limitations of control charts themselves; they are a simple, effective and reasonably intuitive tool. However, there are a number of barriers that, whilst not insurmountable, may affect the efficacy of applying control charts:

- **Aggregation of data:** The more data is aggregated the harder it becomes to make meaningful interpretations (as it masks some of the useful variability that underpins how control charts are constructed).

- **Lack of clear and consistent data collection standards and criteria:** If these are not defined early on, changes in data values can be due to data artefacts or a change in definition rather than actual interventions.

- **Lack of skills and knowledge to construct the most appropriate control chart:** There are a number of different types of chart that can be used and there is some mathematical knowledge required to construct these properly. However, by using this guide and appropriate SPC software this issue can be easily addressed.
3-sigma limits: See control limits

Area of opportunity: The background against which the count of defects must be interpreted. For example in measuring falls per inpatient day the area of opportunity is based upon the inpatient days. The concept is used in creating c-charts and u-charts.

Attribute data: Data represented as counts of events that can be aggregated into discrete categories (e.g. infected versus not infected, alive versus dead).

c-chart: A control chart that is used for attribute data where it is impossible to count non-occurrences of defects (e.g. the number of non-falls) and where there is an equal area of opportunity for each measurement point.

Centre line: The horizontal line used in run charts and control charts to represent the average of the measurement points. For run charts the median is used. For control charts the mean is usually used.

Common cause variation: Variation in a process due to regular, natural or ordinary causes affecting all outcomes of the process. This is distinct from special cause variation.

Continuous data: Data that can take on different values on a continuous scale (e.g. time, weight, blood sugar levels etc.)

Control chart: A tool used to determine whether a process shows special cause or common cause variation (also known as a Shewhart chart) based upon Statistical Process Control principles.

Control limits (upper and lower): Statistically determined limits used to distinguish those high and low measurements that represent ‘extraordinary’ (i.e. special cause) events. They are also known as 3-Sigma limits.

Defectives: An attribute data type where for each event it is possible to count occurrences and non-occurrences (e.g. alive / dead, pregnant / not pregnant). This is often presented as a percentage (e.g. the percentage of patients who died following a particular procedure).

Defects: An attribute data type where for each event it is possible to count occurrences but not possible to count non-occurrences (e.g. falls / non-falls). This is often presented as a rate using the area of opportunity as the denominator (e.g. falls per inpatient day).

Discrete data: An alternative name for attribute data.

Distribution type: A classification system for the shape of histograms based upon the underlying statistical processes that generate them (e.g. the normal distribution versus the binomial distribution).

Driver diagram: A tool used to map out the system drivers that contribute towards achievement of a chosen goal. Drivers are classified as primary or secondary drivers.

g-chart: A type of control chart used for rare events (e.g. healthcare acquired infections or rare death events).

In control: A description applied to a process that has been shown to exhibit no special cause variation implying future predictability of the process.

Individuals chart: An alternative name for the X component of an XmR chart.

Mean: The value resulting from summing a set of measurements and dividing the resulting figure by the number of measurements in the set. It is often used as the centre line in control charts.
**Median:** The value at the midpoint in a dataset when all the data is arranged in value order. It is used as the centre line in run charts.

**Metric:** A process or outcome measurement that is viewed as important in a quality improvement project, often used as the measure of interest in a control chart.

**Model for Improvement:** An approach to process improvement that helps teams accelerate the adoption of proven and effective changes. It is often associated with PDSA cycles.

**np-chart:** A control chart that is used for attribute data where it is possible to count defectives (e.g. deaths) and where there are subgroups of equal size.

**Pareto chart:** A graphical method for presenting categories of data (e.g. fall types) ordered according to the counts in each category. This chart is frequently used to identify priorities amongst categories and separate the ‘vital few from the useful many’.

**p-chart:** A control chart that is used for attribute data where it is possible to count defectives (e.g. deaths) and where there are subgroups of unequal size.

**PDSA:** A structured trial of a process change based upon the Shewhart cycle. It involves a planning phase (Plan), the implementation of a change (Do), analysis of the results (Study) and next steps based upon the results (Act).

**Process:** An ordered set of tasks (or steps) that achieve a purpose.

**Process mapping:** A technique used to define (and ultimately alter) the steps in a process in order to deliver an improvement in the process outcome.

**Quality assurance:** A set of systematic actions to provide confidence that a product (or service) is suitable for its purpose. It is often associated with reviewing suitability at the end of production with remedial actions focusing on the causes of defective products (or services).

**Quality improvement:** A systematic approach to ensuring that a product (or service) is suitable for its purpose involving a focus on the processes used to create all products (or services).

**Rules:** A set of tests used to determine whether a run or control chart exhibits special cause variation.

**Run chart:** A basic chart used to understand the types of variation present in a process.

**Science of improvement:** A set of concepts (based primarily upon the work of Deming, Shewhart and Juran) that provide a framework for quality improvement.

**Sigma:** A value calculated when constructing control charts which is used to determine upper and lower control limits (set at +/- three sigma). The calculation used depends upon the type of control chart selected.

**Special cause variation:** Variation in a process due to irregular or unnatural causes that are not inherent in the process. This is distinct from common cause variation.

**Standard deviation:** A measure of the variability or dispersion of a data set. Its formula is different from that used to calculate sigma.

**Statistical process control (SPC):** The use of a specific set of statistical techniques (such as control charts and the associated rules) usually to measure and analyse the variation in a process.
**Stratification**: The process used to place members of a population into relatively homogeneous groups. It is used in sampling methods. Typically in improvement projects, stratification will be used to look for variation between subgroups.

**Subgroup**: A sample of data pulled from the stream of data produced by a process, selected to maximise the chances of identifying a special cause between subgroups. Typical subgroups samples might be based upon the hour of the day, the shift or the machine used for a specific laboratory test type.

**System**: A group of interdependent processes that interact to produce an outcome. For example, obesity levels will be due to a complex system involving access to health advice, culture, food availability and various other interacting elements. Driver diagrams are often used to try and understand systems.

**Trend**: A series of five or more consecutive points all increasing or decreasing on a run chart. On a control chart this is amended to six or more consecutive points. Where two consecutive points are of equal value, the second point is ignored for the purpose of identifying a trend.

**Type I & II errors**: The misinterpretation of a common cause data point as being a special cause variation is referred to as a Type I error. A Type II error makes the opposite mistake (i.e. misinterprets special cause variation as common cause variation).

**u-chart**: A control chart that is used for attribute data where it is possible to count defectives (e.g. deaths) and where there are subgroups of equal size.

**X bar and R chart**: A control chart that is used for variable data where there is more than one observation per subgroup but less than ten observations per subgroup. The ‘X bar’ refers to the mean whilst the ‘R’ refers to the range.

**X bar and S chart**: A control chart that is used for variable data where there are ten or more observations per subgroup. The ‘X bar’ refers to the mean whilst the ‘S’ refers to the standard deviation.

**XmR chart**: A control chart that is used for variable data where there is one observation per subgroup. The ‘X bar’ refers to the mean whilst the ‘mR’ refers to the moving range.

**X-type**: The X component of an XmR chart, also referred to as an Individuals chart.
An excellent tutorial in the technicalities in plotting control charts is by Mohammed, Worthington and Woodall. It’s available online via Athens or via your local NHS library:


Other references in this paper, which serve as good examples, are:


And some books that serve as excellent practical references:


Additional tools used in the science of improvement are available on our website: www.institute.nhs.uk/commissioning