Understanding variation for clinical governance: an illustration using the diagnosis and treatment of sore throat

Tom Marshall, Mohammed A Mohammed and Hei Toon Lim

SUMMARY

Background: *The aim of clinical governance is to improve clinical care. An understanding of the information contained in variation is central to any improvement effort. We must distinguish between variation intrinsic to a process (common cause variation) and variation caused by extrinsic factors (special cause variation). The control chart is a method of distinguishing between these two kinds of variation: it is used in industry to effect improvement and may be useful in primary care.*

Aim: *To illustrate the use of control charts to distinguish between common cause and special cause variation and to guide appropriate action.*

Design of study: *Analysis of diagnostic and treatment decisions for sore throat.*

Setting: *Single practice in the West Midlands.*

Methods: *We identified each general practitioner's (GP's) consultations for sore throat over a two-year period. We grouped these into two diagnostic categories (tonsillitis and non-tonsillar throat infection) and two treatment categories (antibiotics and no antibiotics). These data were illustrated graphically as XY control charts.*

Results: *In this practice, a special cause affects one GP's diagnosis — he is less likely to use the term 'tonsillitis'. A special cause also affects his treatment decisions — he is more likely to prescribe antibiotics. Diagnostic and treatment differences between the remaining GPs are consistent with common cause variation.*

Conclusion: *In this practice, action to improve the quality of diagnosis and treatment of sore throat should focus on investigating why one practitioner's diagnosis and treatment differs from that of his colleagues. Control chart analysis is valuable because it enables users to obtain practical guidance for action.*

Keywords: *variation; clinical governance; treatment; diagnosis; control chart; outlier.*

T Marshall, MRCGP, MFPHM, Lecturer in Public Health Medicine; M A Mohammed, PhD, research fellow; H T Lim, medical student, Department of Public Health and Epidemiology, University of Birmingham.

Address for correspondence

Tom Marshall, Lecturer in Public Health Medicine, Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, Birmingham, B15 2TT. E-mail: T.P.Marshall@bham.ac.uk.

Submitted: 23 February, 2001; Editor's response: 22 August 2001; final acceptance: 5 October 2001.

Introduction

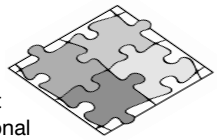
THE Royal College of General Practitioners identifies clinical governance as a method that enables health professionals to improve the quality of their clinical care.[1] It also refers to the problem of identifying and addressing unacceptable variations. However, the notion of acceptable and unacceptable variation may not be helpful to clinical governance. There is another way of understanding variation. This way sees that there is information contained in variation and that this information is a guide to action.

Variation in a clinical setting

We can think of consultation, symptom elicitation, diagnosis, and treatment as a single process. This process produces two outcomes: a diagnosis and a treatment. We would not expect the proportion of patients assigned a particular diagnosis or treated in a particular way, to be the same for every clinician. In other words, the outcome of the process will show variation between clinicians. However, if this is a single process then the variation will not be unlimited — it will be consistent with the variation intrinsic to a single process. If the variation we observe is greater than would be consistent with a single process, then we can conclude that a special circumstance is acting on this process to cause greater variation.

Statistical analysis of processes was pioneered in the 1920s by a physicist called Shewhart. He categorised variation as common cause (intrinsic to a process) or special cause (produced by something external acting on the process). Why does this distinction matter? Shewhart was concerned with quality improvement. The action required to improve quality is critically dependent on whether we observe common cause or special cause variation. If we observe common cause variation we take action on the process; if we observe special cause variaition we investigate the special causes. To provide guidance for action, we therefore need to distinguish between common cause and special cause variation. Shewhart developed a number of graphical tools to do this, called control charts. Control charts are simple to produce and straightforward to interpret. He recommended using a cut-off of three-sigma limits on either side of the mean.[2] If a value lies outside of this three-sigma limit (or there are unusual patterns in the data) then it is very likely to indicate a special cause. Shewhart's methodology has been used in clinical medicine, for infection control,[3,4] and the management of chronic diseases such as asthma.[5] It has recently been argued that it is particularly applicable to clinical governance.[6]

HOW THIS FITS IN



What do we know?

Medical care produces variation that research cannot explain. The traditional categorisation of this variation as acceptable or unacceptable fails to help us improve quality of care.

What does this paper add?

Control charts distinguish between common cause variation (due to a stable process) and special cause variation (due to extrinsic factors). Distinguishing between these kinds of variation guides us towards the most efficient action to improve medical care.

Sore throat

Sore throat is a common reason for consulting a GP.[7] But not every person who experiences the symptom of sore throat consults. The decision to consult is influenced by the nature and severity of the symptoms and by a variety of psychological, social, and situational factors. It follows that the frequency with which patients consult their doctors and the severity of sore throat symptoms with which they consult (the case-mix), are intrinsically variable. The variation does not end with consultation. No two doctor–patient interactions are identical; the pattern of clinical symptoms and signs elicited differs from one occasion to another. This is also true of the patient concerns and expectations. The effect of this is that, even among GPs using the same diagnostic and treatment algorithms, the outcomes of the consultation process will show variation. This variation is well recognised.[8] Since this variation is intrinsic to the process, it is categorised as common cause variation.

As we said earlier, the consultation process for sore throat has two outcomes: a diagnosis and a treatment. The diagnostic outcomes can be grouped into two categories: tonsillitis or non-tonsillar throat infection (sore throat, pharyngitis or throat infection). The treatment outcomes can also be grouped into two categories: antibiotics or no antibiotics. Both these outcomes have important implications. Tonsillitis is regarded as a more serious diagnosis and is more likely to result in antibiotic treatment.[9] Diagnosis of repeated episodes of tonsillitis are held to be an indication for tonsillectomy — one of the most common surgical procedures. The appropriateness of referral for tonsillectomy therefore depends on a reliable and consistent diagnostic process. Antibiotic prescribing affects patient beliefs and influences future consultation.[10] In addition, prescriptions of antibiotics for sore throat account for a significant proportion of antibiotic prescribing in primary care. Government policy is to promote optimal prescribing through, among other measures, clinical governance.[11] The diagnosis and treatment of sore throat are therefore appropriate areas for clinical governance.

We applied the Shewhart methodology to the diagnostic and treatment decisions for sore throat undertaken by a single practice. By doing so we should be able to test the underlying reliability and consistency of the diagnostic process and the consistency of prescribing. The two out-

Commentary

Marshall and his colleagues present an interesting application of industrial methods for assessing variation in clinical general practice. It is more sophisticated than many measures to which we are currently subjected, such as the extent to which we are 'overspent' on our prescribing budget, or the proportion of our diabetic patients who have had recent ophthalmologic assessment. The method's strength is in its acceptance of the fact that volumes of research have been unable to explain large variations in practice across a large number of fields, such as prescribing, referral, and adherence to guidelines. We all have different approaches to different problems, with different patients in different circumstances, and these differences do not necessarily reflect the quality of our practice.

Variation is taken as an indicator of quality. However, rather than using the Orwellian classification of variation (acceptable or unacceptable), they hope to use Shewart's classification (intrinsic or extrinsic). While the former requires attention to the process the latter requires investigation of the special cause — in this case the GP. The problem is that the distinction between these two types of variation is an arbitrary statistical one; what we are really assessing, therefore, is how far the practitioner deviates from the mean. The authors sensibly emphasise that deviation does not necessarily mean fault, but it is tempting to conclude that Doctor C is 'wrong' — in other words, that extrinsic variation is equivalent to unacceptable variation.

The language of the paper presents another problem. It is the language of industry, including raw materials and tools; in particular I am worried about being assessed by control charts. In industry, processes and outcomes are generally well defined, usually relating to the generation of profit. In health care, on the other hand (particularly primary care), processes and outcomes are often more nebulous and more numerous. It is simplistic, I think, to view the consultation as a single process resulting only in diagnosis and treatment. Rather, the consultation comprises multiple processes (if we choose to use this language) occurring over many years, in which an episode such as sore throat plays a tiny part. Outcomes include advice, education, *other* diagnoses, and lifelong health. *Not* having a stroke is a valid outcome. The condition selected for this study — sore throat — is one that conforms to binary measures with arguable ease: tonsillitis/non-tonsillitis and antibiotics/no antibiotics. This is often true for industrial processes, but not true for the majority of primary care conditions. Depression, for example, is a spectrum whose classification includes many terms and euphemisms, and the treatment ranges from nothing to emergency admission, including drugs, counselling, and judicious review. It is difficult to see how this would submit to Shewart's graphic analysis, and yet easy to envisage a health manager's attempt. Equally, most chronic conditions would not conform. This difficulty is not with the value of Shewart's methods, but with their application.

There are also statistical considerations. This study used a small sample. It is feasible that a larger sample might display greater variation, such that Doctor C came within three standard deviations of the mean. Alternatively, a larger sample might produce smaller standard deviations, with the result that other doctors (certainly Doctor E) would be subject to the same investigation as his or her colleague. It is essential, if this is to form a basis for clinical governance, that the limits of intrinsic variation are based on sound evidence, including valid measures and large samples, otherwise there is a danger of mis-classification. Let calibration precede interpretation.

It is also essential that clinicians and managers distinguish between variation and quality. I am worried that a superficial understanding of the clinical issues might lead political and managerial reviewers of our control charts to (choose to) assume synonymity. We can only require conformity if the evidence against non-conformity is sound. It may be the orchestra that is out of tune and the singer who is correct!

BLAIR H SMITH

Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen

comes — tonsillitis or non-tonsillar throat infection, and antibiotics or no antibiotics — are presented graphically as a binomial probability plot. Illustrated by this method, the control limits are parallel lines on either side of a trend line representing the mean. Such a chart is called an XY control chart. Visual inspection indicates whether data points lie within the control limits (indicating common cause variation) or outside the control limits (indicating special cause variation).

Method

We searched the electronic records of a single metropolitan practice for consultations between 1998 and 2000 with the diagnosis of sore throat, pharyngitis, throat infection or tonsillitis. For each consultation we identified the GP and whether or not the patient was prescribed antibiotics. From this we calculated the total number of throat-infection consultations for each GP and the numbers recorded as tonsillitis or as non tonsillar throat infection. We also recorded the numbers of occasions on which antibiotics were prescribed for tonsillitis and for non-tonsillar throat infection, respectively. We present these data graphically as XY control charts.

Results

Table 1 shows the data which forms the basis of the analysis. Figure 1 illustrates the outcomes of the diagnostic decision-making process. All but one of the GPs' diagnostic decisions lie within the control limits, indicating common cause variation. In other words, despite a fivefold variation (from 10% to 50%) in the proportion diagnosed as tonsillitis their diagnostic decisions are consistent with a single process. GP C's diagnostic decisions indicate special cause variation. They are not consistent with the same diagnostic process.

Figure 2 shows the treatment decisions for consultations where the diagnosis was tonsillitis. All of the GPs' treatment decisions lie within the control limits, indicating common cause variation. In other words, their treatment decisions for tonsillitis are consistent with a single process.

Figure 3 shows the treatment decisions for consultations where the diagnosis was non-tonsillar throat infection. All but two of the GPs' diagnostic decisions lie within the control limits, indicating common cause variation. However, simple inspection of the distribution suggests that there is one outlier and that it is exerting an undue influence on the control limits. We can exclude this data point and recalculate the control limits. Ten GPs now lie within the control limits, indicating common cause variation. This indicates that, despite a seven-fold variation (from 11% to 78%) in the proportion of prescribed antibiotics for non-tonsillar throat infection, the treatment decisions of these ten practitioners are consistent with a single process. GP C's treatment decisions indicate special cause variation (Figure 4).

Figure 5 shows the treatment decisions of the 11 GPs, all for sore throat consultations. Again, the treatment decisions of ten GPs indicate common cause variation. General practitioner C's treatment decisions indicate special cause variation, prescribing more antibiotics than his colleagues.

Discussion

Shewhart was concerned with improving quality. His categorisation of variation indicates the most efficient action required to improve quality. His methodology is widely used in industry for monitoring and improving manufacturing processes.[12,13] If there is special cause variation, this should be investigated and the special causes identified. The aim of investigation is to learn lessons rather than to single out individuals for blame or praise. If the special causes result in lower quality then they should be eliminated. If the special causes result in higher quality then they should be investigated to identify lessons or good practice which can be applied elsewhere. What kind of special causes should we look for? We first look at the data collection process itself, including operational definitions. In a clinical setting this means looking at data recording and at case definitions. We then look at the raw materials. In a clinical setting this means the case-mix, including the age, sex, ethnicity, and social background of patients. We then look at the worker's tools, equipment and facilities. In a clinical setting this means the diagnostic equipment; it could include the amount of time available for consultation and perhaps skills or training in eliciting particular symptoms or signs. Finally, we look at the processes. In this case we could inquire about the clinician's own diagnostic and treatment algorithms (the way in which he reaches decisions).

If there is only common cause variation we can regard all our GPs as part of a single underlying process. Any variation between them is a result of the complex interplay of all those factors which are intrinsic to the process itself. These may include case-mix, mode of presentation, differences between individual consultations, differences in symptom elicitation, and so on. There is little to be gained in trying to disentangle the different influences of these factors. No single factor is exerting an undue influence on outcomes. It would not be meaningful to urge one to improve or to hold up another as exemplary. The most efficient way to improve the quality of care is to make fundamental changes to the process as a whole. This is likely to be informed by detailed mapping of the process (plan), making changes (design) and auditing the effects of these changes (study) before implementation (act).[12]

How does this apply to our data? In the first place we have looked at process measures (prescribing) rather than clinical or patient-related outcomes. This is not ideal. However, important clinical outcomes are often infrequent events and audits must therefore be very large to detect differences in outcome. It has been argued that a more practical alternative is to measure aspects of the process of care that have been shown by randomised controlled trials to influence outcome.[14]

A special cause affects GP C's diagnostic decisions: he diagnoses tonsillitis less frequently than his colleagues. What is the special cause? It could be different data recording or different case definitions. This is possible. He may see a different case-mix of patients. This is unlikely: emergency patients are seen on a first-come-first-served basis in this practice. The observation of variation in the diagnosis of tonsillitis is far from new.[15] For diagnostic data to be useful, there needs to be consistency between clinicians.[16]

Table 1: Numbers of consultations, diagnoses, and antibiotic prescriptions for sore throat among eleven GPs in a single practice.

GP	Whole time equivalents	Total number of sore-throat consultations			Number prescribed antibiotics	
		Non-tonsillar throat infection	Tonsillitis	All sore-throat consultations	Non-tonsillar throat infection	Tonsillitis
A	1	31	29	60	19	23
B	0.5	29	11	40	4	9
C	1	100	3	103	96	3
D	1	57	35	92	24	34
E	0.85	38	15	53	5	10
F	1	17	7	24	8	7
G	0.75	39	18	57	11	15
H	0.45	18	5	23	2	5
I	1	20	20	40	7	15
J	0.5	14	8	22	3	3
K	0.33	9	1	10	7	1

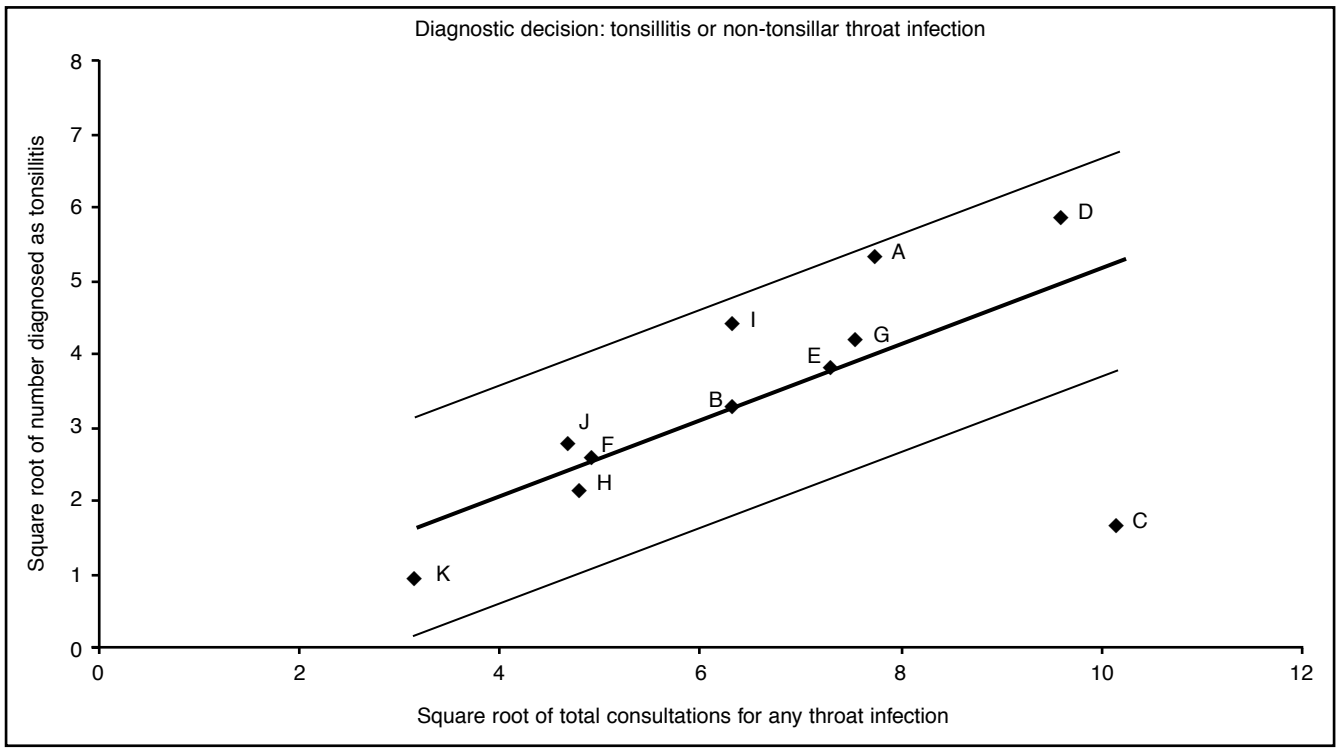


Figure 1. Control chart illustrating the diagnostic decisions of eleven GPs.

Improving the quality of data, perhaps by agreeing diagnostic criteria and data coding, may be a necessary prerequisite to improving quality of care.

There is evidence that doctors choose diagnostic labels to fit their preferred prescribing decisions.[17] The prescribing decisions for tonsillitis reported here seem to be consistent, but a special cause affects GP C's prescribing decisions for non-tonsillar throat infection. Here we must make a decision. If antibiotic prescribing matters, both approaches (C's and that of the remaining practitioners) cannot be optimum. Is the appropriate level of prescribing high or low? Evidence suggests that the absolute benefits of antibiotic prescribing in sore throat are modest.[18] Government policy is to reduce antibiotic prescribing.[11] There is also evidence that prescribing has a medicalising effect.[10,19] Arguably, we can regard low rates of prescribing as preferable. Nevertheless, whichever we decide, we should investigate and try to understand the special cause. As almost all prescriptions are issued electronically, differences in data recording are unlikely. Differences in case-mix could explain C's infrequent use of the term tonsillitis, but cannot also explain his high rates of antibiotic prescription. Given the fact that his rate of consultation for sore throat is higher than his colleagues, a more plausible explanation is that he lacks consultation time. Indeed this may be partly a consequence of high prescribing.[10] (Thus creating a vicious circle from which it may be difficult to escape). As he is the senior partner in the practice it is highly implausible that skills in eliciting symptoms or clinical signs are the problem. The remaining interesting possibility is that his diagnostic and treatment algorithms differ from those of his colleagues. If so then these should be the focus of attention.

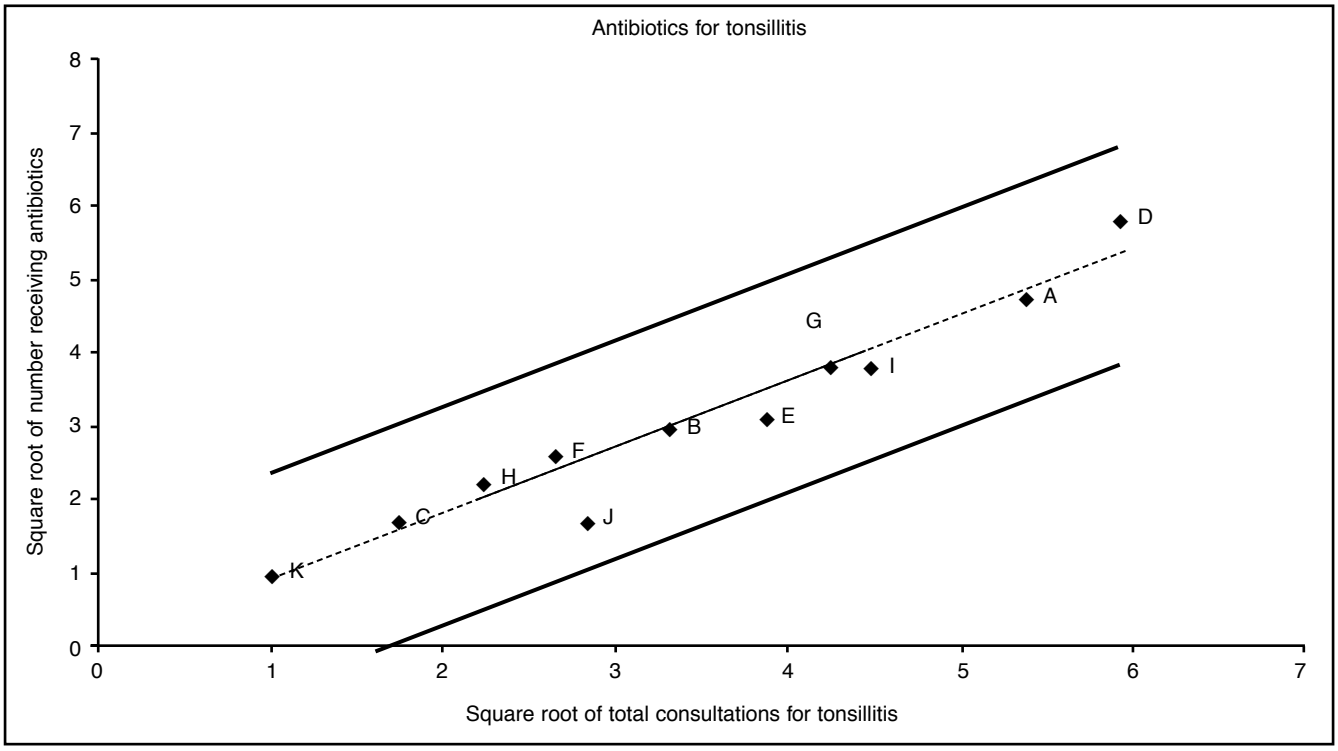


Figure 2. Control chart illustrating the treatment decisions for tonsillitis.

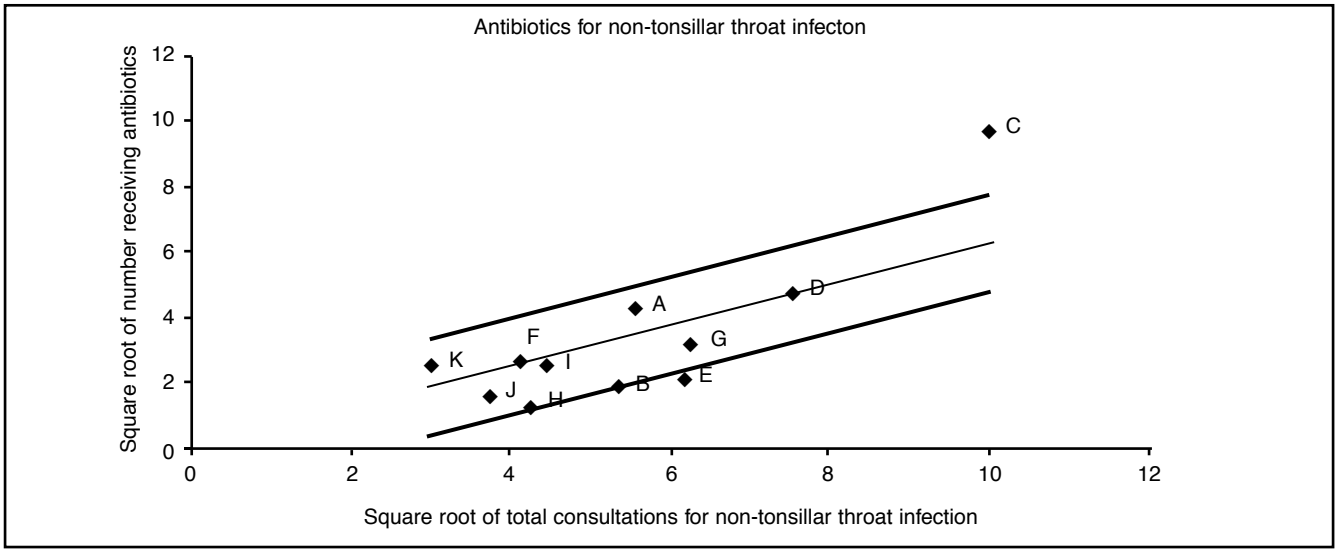


Figure 3. Control chart illustrating the treatment decisions for non-tonsillar throat infection.

It is axiomatic in this approach that individuals are not at fault, the faults lie in the system. We have no reason to doubt that all our GPs are motivated to do the best for their patients. If GP C's diagnostic and treatment algorithms are different then this is not a matter for praise or blame. In the first place he may not have been aware that his practice differs from his partners'. Since shared consultation and peer observation are rare after the trainee year, how could he have known until the data were collected? If we take the view that low prescribing is preferred, it is clear is that GP C needs help to improve his prescribing. Does he have access to the clinical evidence? Clinicians often interpret evidence differently, but training in critical appraisal can reduce this particular kind of variation.[20] Is he overworked? Senior partners often accumulate administrative responsibilities. Are there wider lessons to learn about data recording, accessing evidence, critical appraisal skills or workload? Were we to take the opposite view, that high prescribing is preferred. We might ask similar questions. What is special about GP C's practice? How can we apply the wider lessons of his work?

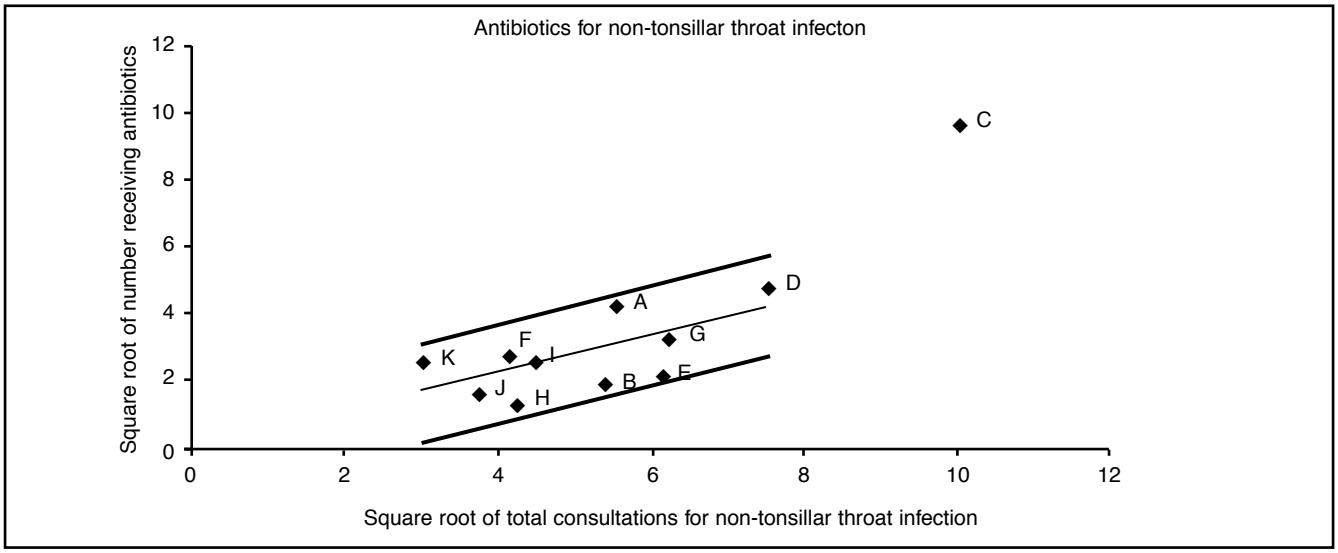


Figure 4. Control chart of treatment decisions for non-tonsillar throat infection. (Control limits have been recalculated for all GPs except C.)

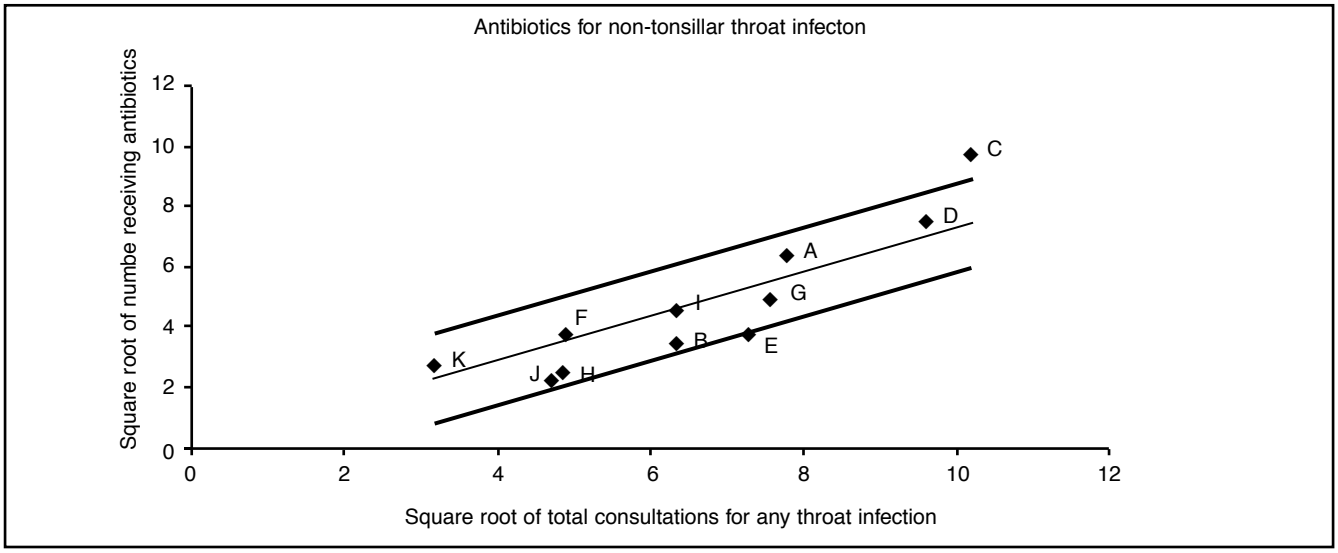


Figure 5. Control chart of treatment decisions for any sore throat (tonsillitis or non-tonsillitis).

Conclusions

The control chart method of illustrating variation is easy to understand and intuitive. Even analysis of this small data set provides a basis for action. When we view clinical practice as a process we can use control charts to form some practical conclusions about the diagnosis and treatment of sore throat. First, within this practice the diagnostic term tonsillitis is not used in the same way by all GPs. This is an important point, seriously affecting the basis of referral for tonsillectomy. It is difficult to attach any special significance to recurrent episodes of tonsillitis as opposed to recurrent episodes of non-tonsillar throat infection if they simply reflect GP preferences for one or another diagnostic category. Secondly, GPs in this practice are consistent in their treatment of tonsillitis. Thirdly, the GPs are not consistent in their treatment of non-tonsillar throat infection. Fourthly, the most efficient action to improve the quality of sore throat diagnosis care in this practice is to identify and eliminate the special causes affecting the diagnostic decisions of one GP. Finally the most efficient action to improve the quality of prescribing for sore throat is to identify and act on the special causes affecting the prescribing decisions of this GP. Only when this has been addressed, should action focus on improving the process as a whole, such as the introduction of an evidence-based protocol.

Acknowledgements

The authors are grateful to Dr Brendan Delaney of the Department of Primary Care and General Practice, University of Birmingham for comments made on an earlier draft of this paper.

References

1. Royal College of General Practitioners. *Clinical governance: practical advice for primary care in England and Wales.* London: RCGP. http://www.rcgp.org.uk/rcgp/webmaster/quality_and_standards.asp.
2. Shewhart WA. *Economic control of quality of manufactured prod-*

uct. (1931.) Republished 1980 by American Society for Quality Control Quality Press, Michigan, USA.
3. Benneyan JC. Statistical quality control methods in infection control and hospital epidemiology. Part 1: introduction and basic theory. *Infect Control Hosp Epidemiol* 1998; **19(3):** 194-214.
4. Benneyan JC. Statistical quality control methods in infection control and hospital. Part 2: chart use, statistical properties, and research issues. *Infect Control Hosp Epidemiol* 1998; **19(4):** 265-277.
5. Boggs PB, Wheeler D, Washburne WF, Hayati F. Peak expiratory flow rate control chart in asthma care: chart construction and use in asthma care. *Ann Allergy Asthma Immunol* 1998; **81(6):** 552-562.
6. Mohammed MA, Cheng KK, Rouse A, Marshall T. Bristol Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001; **357:** 463-467.
7. Office of Population Censuses and Surveys. *Morbidity Statistics from General Practice. Fourth National Study 1991*. London: HMSO, 1994.
8. Howie JGR, Richardson IM, Gill G, Durno D. Respiratory illness and antibiotic use in general practice. *J R Coll Gen Pract* 1971; **21:** 657-661.
9. Pitts N, Vincent SH. Diagnostic labels, treatment and outcome in sore throat. *Practitioner* 1988; **232:** 343-346.
10. Little P, Gould C, Williamson I, *et al*. Reattendance and complications in a randomised trial of prescribing strategies for sore throat: the medicalising effect of prescribing antibiotics. *BMJ* 1997: **315:** 350-352.
11. Department of Health. *UK antimicrobial resistance strategy and action plan.* London: DoH, June 2000,.
12. Deming WE. *Out of the crisis.* (1986). Cambridge, MA: Massachusetts Institute of Technology, 1986.
13. Deming WE. *The New Economics* (1994). Cambridge, MA: Massachusetts Institute of Technology, 1994.
14. Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995; **311:** 793-796.
15. Howie JGR, Richardson IM, Gill G, Durno D. Respiratory illness and antibiotic use in general practice. *J R Coll Gen Pract* 1971; **21:** 657-663.
16. Howie JGR. Diagnosis — the Achilles heel? *J R Coll Gen Pract* 1972; **22:** 310-315.
17. Howie JGR. Department of General Practice, University of Edinburgh. Unpublished data. Presented at Scottish Intercollegiate Guidelines Network Workshop on 3 October 1997.
18. Del Mar CB, Glasziou PP, Spinks AB. Antibiotics for sore throat (Cochrane Review). In: *The Cochrane Library,* Issue 3, 2001. Oxford: Update Software.
19. Herz MJ. Antibiotics and the adult sore throat — an unnecessary ceremony. *Fam Pract* 1988; **5(3):** 196-199.
20. Doust JA. Silagy CA. Applying the results of a systematic review in general practice. *Med J Aust*, 2000; **172(4):** 153-156.