Misuse of the *P* Value

Using Quality Improvement Analyses to Identify Clinically Significant Improvements

Staci S. Reynolds, PhD, RN, ACNS-BC, CCRN, CNRN, SCRN, CPHQ;
Julee Waldrop, DNP, PNP, FNP, SANE-A

Introduction: Quantitative research and quality improvement (QI) both seek to improve care provided to patients. However, clinicians often blur the lines between how to appropriately analyze data from these methodologies. Clinicians may inappropriately use statistical analyses for QI initiatives, rather than using run and statistical process control (SPC) charts to analyze improvements in outcomes.

Objective: The purpose of this article was to address the analytic methods used for QI initiatives in the clinical setting in an effort to show clinicians how to identify meaningful improvements in clinical practice.

Methods: In this article, we provide an example comparing the same evidence-based practice/QI initiative (chlorhexidine gluconate bathing in a medical intensive care unit) using a quasi-experimental pretest/posttest research design with statistical analyses completed with t tests with analyses using run and SPC charts to show the data trended over time. Using a pretest/posttest design, chlorhexidine gluconate bathing compliance improved from 63% to 65%, a nonsignificant change, P = .075. These same data plotted on run and SPC charts, however, show a shift and a trend, indicating clinically significant improvements per QI methodologies.

Conclusion: The example in this article highlights the pitfall of relying only on statistical analyses and P values to determine the importance of a clinical project, and provides a practical example for how run or SPC charts can be used to identify improvements over time.

Keywords: clinical significance, data visualization, quality improvement

[Dimens Crit Care Nurs. 2024;43(2):96-101]

Quality improvement (QI) teams may be missing evidence that their interventions are actually useful and may be using analytic methods inappropriately. For example, many teams use research methodologies to analyze data rather than standard QI methodologies to guide QI initiatives. Quantitative research methodologies allow clinicians and researchers to answer questions that will help generate new knowledge. These methodologies require rigorous research designs and statistical analyses that are appropriately powered to answer the research question(s). Experimental research designs include quasi-experimental and randomized controlled trials.[1] For quantitative research, *P* values are helpful in determining statistical, as well as clinical, significance of the research question(s). In health care, researchers generally want to see a *P* value of less than .05 or, in some instances, less than .01. However, experts agree

 DOI: 10.1097/DCC.0000000000000623

that there is more to statistical analyses than the arbitrary assignment of the *P* value level.[2-4] Indeed, Halsey[5] and Andrade[2] stated that the *P* value provides limited data and can be easily misinterpreted. Research data contain much more meaningful findings than can be demonstrated in a *P* value and statistical significance.[2]

Quality improvement, on the other hand, uses iterative methods, such as plan-do-study-act (PDSA) cycles, to help answer clinical inquiry questions and often seeks to improve processes in the health care environment.[6] Whereas QI is often considered less rigorous than research methodologies, this process can help improve care provided to patients by clinicians. Different from research methodologies, QI methodologies use tools such as run charts and statistical process control (SPC) charts to analyze data and identify signals of improvement or special cause variation. Data are plotted over time (either weeks or months) to identify whether the iterative cycles are indeed improving care based on run and SPC chart "rules."[7]

CHALLENGES WITH USING RESEARCH METHODOLOGIES FOR QI INITIATIVES

Whereas the intent and process of research and QI methodologies differ, they both seek to improve care provided to patients. However, clinicians often blur the lines between the 2 methodologies and inappropriately use statistical analyses for QI initiatives, and refer to them as pre/post-design studies, when iterative QI cycles would be more useful.[8,9] Often QI analyses are conducted with smaller sample sizes, for example, testing changes in 1 clinical unit, and therefore are not powered appropriately to identify statistical significance. Although clinicians may obtain a statistically significant difference (ie, a *P* value less than .05) after a QI practice change, this result is at risk of type I error; this occurs when a team concludes that the results are statistically significant when in reality the results were due to chance. More importantly, it may not be clinically significant, especially if the sample size was small. Anecdotally, many clinicians have an incorrect assumption that, to publish results of their QI initiative, they must (1) include statistical analyses and (2) have a statistically significant *P* value. The purpose of this article was to address the analytic methods used for QI initiatives in the clinical setting in an effort to show clinicians that use of run and SPC charts is sufficient and appropriate in identifying meaningful improvements in clinical practice.

USING QI METHODOLOGIES

To begin a QI initiative, clinicians should focus on a clinical problem, often identified through local data or a clinical concern. Next, the multidisciplinary team should identify a formal purpose statement that will guide the initiative and keep the team focused. Several QI tools and processes are available to identify barriers or contributors to the ideal state of practice, such as the fishbone diagram, driver diagram, and brainstorming, among others.[10] From these data, measurable aims can be identified, including outcome, process, and balancing aims. *Outcome* aims are those focused on the ultimate outcome the team is hoping to achieve (eg, improvement in functional ability through early intensive care unit mobility). *Process* aims measure the steps that will help the team achieve their ultimate goal (eg, compliance with ambulating a patient per the mobility protocol). Process aims are logically connected to the outcome and typically show improvements before the outcome measure does. *Balancing* aims help detect unintended consequences of a project or a complication that may arise from the project (eg, a rise in fall rates due to increased mobility).[7] The QI project design should include rigorous attention to choosing good measures that will help determine whether changes are leading to improvements.[11]

For an example of a QI initiative, a team is focused on reducing central line–associated bloodstream infection in the medical intensive care unit. Using a fishbone diagram activity with clinical staff, they identify that clinicians are not compliant with performing daily chlorhexidine gluconate (CHG) baths, a well-supported evidence-based intervention,[12] per the hospital's policy. The *outcome* metric would be reduction in central line–associated bloodstream infections (what the team is ultimately trying to improve); the *process* metric would be compliance with daily CHG bathing. A potential *balancing* metric may be patient and nurse satisfaction with using CHG cloths for bathing.

For a QI initiative, the methods would entail iterative QI cycles (such as PDSA) including small tests of change implemented to improve compliance with the process metric (CHG bathing). Unlike research, where the intervention must be implemented based on a somewhat rigid structure as outlined in the research protocol, QI methodology allows for modification of tests of change over time. The team can implement small tests of changes over several weeks and months; if one test of change failed to result in improvements, they could stop implementation of that change. Generally, process metrics will show improvement before outcome metrics. Outcome, process, and balancing metrics would be measured over time to identify any improvements, based on run and SPC chart rules.[7,13]

Conversely, if a clinician used a quasi-experimental preimplementation/postimplementation research design for the initiative, they would collect data preimplementation (eg, over the course of 3 months), implement the intervention(s), and then collect postimplementation data (again, over the course of 3 months). Data would then be analyzed for differences between preimplementation and postimplementation, using a statistical test such as an independent samples *t* test.

IDENTIFYING IMPROVEMENTS USING QI METHODOLOGIES

For the CHG bathing example, Table 1 provides a chart of weekly CHG bathing compliance results. However, it is challenging to see whether there have been any improvements over time just from viewing the chart. Weeks 1 to 20 represent baseline data; weeks 21 to 40 represent data collected after implementing changes to improve CHG bathing compliance. If a clinician completed a typical quasi-experimental pre/post research study design using an independent samples *t* test to find differences between time point 1 (weeks 1-20) and time point 2 (weeks 21-40), they would receive the results visible in Figure 1. The preimplementation mean was 63%, with a slight increase in the postimplementation mean to 65%. However, this change is not statistically significant, with a *P* value of .075. After this pre/post research design study, clinicians may be discouraged and abandon the changes that they implemented, as a nonsignificant *P* value was identified.

Had the QI team displayed the data in either run charts or SPC charts, clinicians would have been able to see changes more easily in data that are often hidden in tables or inferential analyses. This technique is known as data visualization, which helps promote transparency in QI initiatives through organizing data in an easy-to-interpret manner by using various types of charts.[14] In addition, run or SPC charts provide visualization of changes over the course of 40 weeks, rather than just at 2 time points, as shown in Figure 1 (preimplementation and postimplementation bar graphs). Run charts are simple visual displays of data on a graph that includes the data clinicians are collecting, as well as the median. Once tests of change (improvement cycles) begin, the median should be "frozen" or "locked" and extended into the future. The same data from Table 1 were plotted on a run chart in Figure 2. From this figure, we can see that there was more variability in CHG bathing compliance in the preimplementation period (weeks 1-20). After PDSA cycle 2, there was an identified shift in the data, defined as 6 or more data points all above the median. Likewise, after PDSA cycle 3, there was an improving trend in the data, defined as 5 or more data points all going in the same direction.[7] Per QI methodology, these small tests of changes do, in fact, signal an improvement in practice per standard run chart rules.[7] These data are also displayed in an SPC chart in Figure 3; a shift, defined as 8 or more data points all above the mean, indicates special cause variation (eg, improvement). Statistical process control charts have similar rules; however, an SPC chart includes the data line, mean, an upper control limit, and a lower control limit; control limits are typically set at 3 sigma above and below the mean. Statistical process control charts aid in describing the *stability* of the data and can identify special cause variation or when a process is "out of control" (ie, outside the upper or lower control limit).[7] See Table 2 for an overview of how to interpret run and SPC charts.

TABLE 1 Table of CHG Bathing Compliance Score

Week	CHG Bathing Compliance	Week	CHG Bathing Compliance
Week 1	58%	Week 21	63%
Week 2	67%	Week 22	64%
Week 3	59%	Week 23	60%
Week 4	65%	Week 24	63%
Week 5	58%	Week 25	62%
Week 6	65%	Week 26	64%
Week 7	67%	Week 27	63%
Week 8	55%	Week 28	67%
Week 9	67%	Week 29	66%
Week 10	65%	Week 30	66%
Week 11	58%	Week 31	67%
Week 12	67%	Week 32	66%
Week 13	59%	Week 33	67%
Week 14	65%	Week 34	66%
Week 15	58%	Week 35	66%
Week 16	65%	Week 36	61%
Week 17	67%	Week 37	62%
Week 18	56%	Week 38	64%
Week 19	67%	Week 39	68%
Week 20	65%	Week 40	69%

Abbreviation: CHG, chlorhexidine gluconate.

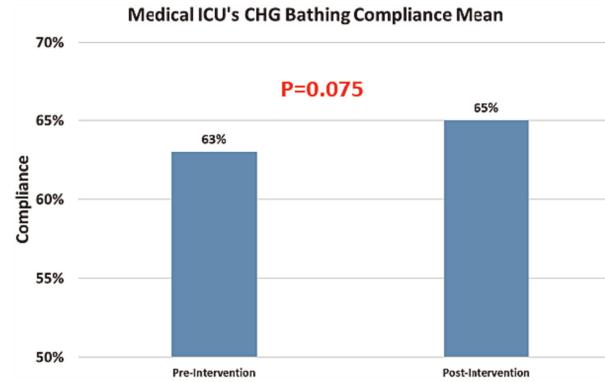

Figure 1. Preimplementation/postimplementation *t* test analysis of compliance data from Table 1.

DISCUSSION

In the provided example, a pre/post statistical analysis using *t* tests yielded nonsignificant improvements in CHG bathing

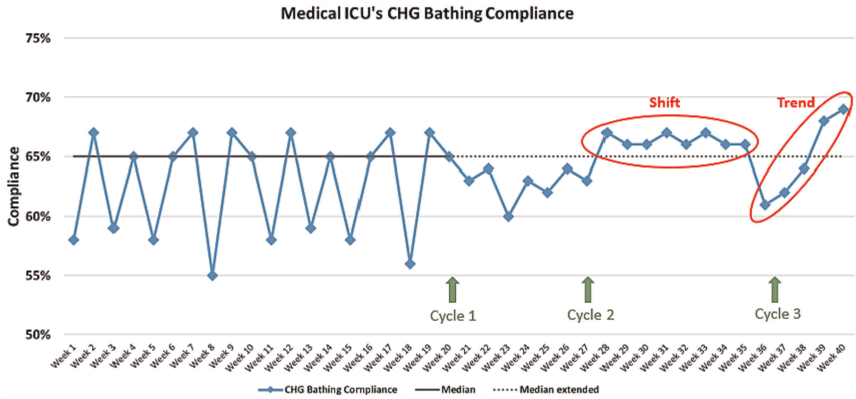

Figure 2. Run chart analysis of compliance data from Table 1.

compliance after the intervention. However, when these same data were plotted on a run or SPC chart, clinically meaningful signals of improvement (shift and trend) were identified. For QI initiatives, especially those with a small sample size and perhaps confined to 1 unit or other local contexts, clinicians should use QI methodologies including run or SPC charts to show data over time, rather than pre/post statistical analysis. Only conducting pre/post statistical analyses can result in statistically nonsignificant findings, which may overshadow improvements that are clinically significant.[15] Indeed, Gallagher-Ford and Melnyk[16] noted that clinicians (including those completing doctor of nursing practice projects in the clinical setting) should not be educated in rigorous research methods and designs; rather, they should be experts in leading evidence-based practice and QI initiatives in hospitals and health systems. A part of this QI expertise is understanding the role and use of QI methodologies and analyses such as run and SPC charts to identify signals of improvement in data. In addition, many clinicians do not have access to or the skill set to complete statistical analyses using a statistical software package system, such as SAS or SPSS. R and Microsoft Excel are 2 software packages that are readily available and can conduct statistical analyses; however, R statistical software requires knowledge of how to properly input and code data, which could be a deterrent from practicing clinicians. Furthermore, all statistical packages require clinicians to know which statistical tests to use and how to interpret the data; many practicing clinicians may not have this skill set. Thankfully, run and SPC charts can be easily developed in Microsoft Excel or other widely available spreadsheet software packages.

Run and SPC charts also represent more frequent collection and monitoring of the project aims and, as such, can be shared with all stakeholders on a regular basis, which increases continued engagement in the improvement process. When measures are only taken and shared months after a practice change has taken place, an opportunity is lost not only from a continued improvement perspective but also from an empowerment perspective for

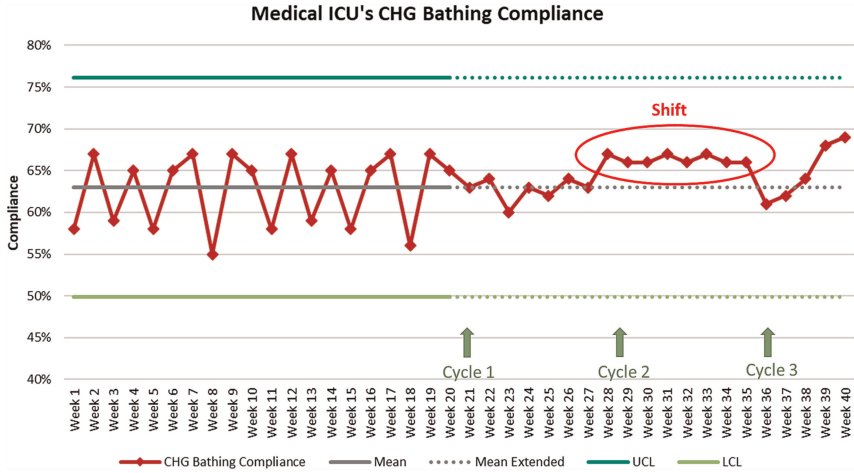

Figure 3. Statistical process control chart analysis of compliance data from Table 1.

TABLE 2 Interpreting Run Charts and Control Charts

Rule	Interpretation
Run charts	
• Median is used as the centerline.	
• A minimum of 10 baseline data points are needed to use the below probability-based rules.	
• Any of the below rules indicate a signal of improvement.	
Shift	Six or more consecutive points either all above or all below the median
Trend	Five or more consecutive points are all going up or down.
Statistical process control charts	
• Mean is used as the centerline.	
• A minimum of 20 baseline data points are needed to use the below probability-based rules.	
• Any of the below rules indicate special cause variation.	
Data point is out of the control limits	A single point outside the upper or lower control limits. Upper Control Limit = Mean + (3 × Standard Deviation) Lower Control Limit = Mean − (3 × Standard Deviation)
Shift	Eight or more consecutive points either all above or all below the mean
Trend	Six or more consecutive points are all going up or down.

those involved in the daily work of improvement. Ongoing monitoring and feedback of data via run or SPC charts can also aid in sustainability efforts.[17]

Contrary to popular belief, QI initiatives without inferential statistical analyses can be published and provide clinically meaningful information. For example, Reynolds and colleagues[18] used SPC charts to determine whether there was a reduction in urinary catheter device days after implementing a nurse-driven urinary catheter removal protocol. They found no signals of improvement, but rather, special cause variation was identified with a shift of 8 data points above the extended mean, indicating there was an increase in urinary catheter device days after the intervention. Findings demonstrated a continued need to implement further tests of change to improve the catheter utilization rate.[18] In another QI initiative, Williams and colleagues[19] implemented small tests of change through PDSA cycles to improve oral care compliance. Data were plotted using SPC charts, finding a shift in oral care compliance, with data points above the upper control limit, indicating clinical improvements.[19] These articles provide great examples of how data analysis using QI methodologies (and not including statistical analyses) can be implemented and published, and provide clinically meaningful findings.

CONCLUSIONS

This article provides a brief overview of how research and QI analyses differ, and how clinicians should focus more on clinically meaningful improvements from QI initiatives, rather than dwelling on statistical analyses and *P* values. The example in this article highlights the pitfall of relying only on statistical analyses and *P* values to determine the impact of a clinical project. Displaying data on run or SPC charts not only provides great data visualization for how metrics are changing over time; clinicians can also use QI "rules" for quickly interpreting run and SPC charts to identify signals of improvement and clinically meaningful changes that can be used in practice.

IMPLICATIONS

This article provides an introduction of how to interpret run charts and SPC charts for QI initiatives, and can be used in conjunction with implementation science methodologies to create an effective, sustainable practice change.[20] This practical knowledge and skill set may reduce the concern from clinicians about not having access to statistical software packages or statistician assistance, and increase their confidence with publishing projects that have clinical significance—even without a *P* value. Quality improvement teams may also fail to recognize the meaningful results of their initiative and may be at risk for wasting time and more resources searching for better interventions. In addition, QI initiatives focus on clinically significant improvements rather than statistically significant *P* values, which may reduce discouragement that clinicians may feel if they get a *P* value greater than .05.

References

1. Bloomfield J, Fisher MJ. Quantitative research design. *J Australas Rehabil Nurs Assoc*. 2019;22(2):27-30. doi:10.3316/informit.738299924514584.
2. Andrade C. The *P* value and statistical significance: misunderstandings, explanations, challenges, and alternatives. *Indian J Psychol Med*. 2019;41(3):210-215. doi:10.4103/IJPSYM.IJPSYM_193_19.

3. Hayat MJ, Staggs VS, Schwartz TA, et al. Moving nursing beyond $P < 0.05$. *Res Nurs Health*. 2019;42(4):244-245. doi:10.1002/nur.21954.
4. Concato J, Hartigan JA. *P* values: from suggestion to superstition. *J Invest Med*. 2016;64(7):1166-1171. doi:10.1136/jim-2016-000206.
5. Halsey LG. The reign of the *P*-value is over: what alternative analyses could we employ to fill the power vacuum? *Biol Lett*. 2019;15(5):20190174. doi:10.1098/rsbl.2019.0174.
6. Picarillo AP. Introduction to quality improvement tools for the clinician. *J Perinatol*. 2018;38(7):929-935. doi:10.1038/s41372-018-0100-4.
7. Provost LP, Murray SK. *The Health Care Data Guide: Learning from Data for Improvement*. San Francisco, CA: Jossey-Bass; 2011.
8. Waldrop J, Reynolds S. Letter to the editor. *Nurs Outlook*. 2022;70(3):368. doi:10.1016/j.outlook.2022.01.006.
9. Hanners A, Melnyk B, Guo J. Considering evidence from studies with findings beyond the sacred cow of $P < .05$ for best practice. *Worldviews Evid Based Nurs*. 2020;17(6):409-411. doi:10.1111/wvn.12471.
10. Institute for Healthcare Improvement. Science of improvement. Published 2021. http://www.ihi.org:80/about/Pages/ScienceofImprovement.aspx. Accessed April 18, 2021
11. Chassin MR, Loeb JM, Schmaltz SP, Wachter RM. Accountability measures—using measurement to promote quality improvement. *N Engl J Med*. 2010;363(7):683-688. doi:10.1056/NEJMsb1002320.
12. Reynolds SS, Woltz P, Keating E, et al. Results of the chlorhexidine gluconate bathing implementation intervention to improve evidence-based nursing practices for prevention of central line associated bloodstream infections study (CHanGing BathS): a stepped wedge cluster randomized trial. *Implement Sci*. 2021;16:45. doi:https://doi.org/10.1186/s13012-021-01112-4.
13. Langley G, Moen R, Nolan K, Nolan T, Norman C, Provost L. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. 2nd ed. San Francisco, CA: Jossey-Bass; 2009.
14. Weissgerber TL, Winham SJ, Heinzen EP, et al. Reveal, don't conceal. *Circulation*. 2019;140(18):1506-1518. doi:10.1161/CIRCULATIONAHA.118.037777.
15. Carpenter R, Waldrop J, Carter-Templeton H. Statistical, practical and clinical significance and doctor of nursing practice projects. *Nurse Author Ed*. 2021;31(3–4):50-53. doi:10.1111/nae2.27.
16. Gallagher-Ford L, Melnyk BM. Evaluating outcomes of evidence-based practice initiatives versus research: clarifying the confusion with a call to action. *Worldviews Evid Based Nurs*. 2022;19(4):258-259. doi:10.1111/wvn.12604.
17. Granger BB. Six simple steps to sustainability: a checklist for ongoing monitoring of clinical practice improvements. *AACN Adv Crit Care*. 2020;31(2):203-209. doi:10.4037/aacnacc2020667.
18. Reynolds SS, Lozano H, Fleurant M, Bhandari K. Using statistical process control charts to measure changes from a nurse-driven protocol to remove urinary catheters. *Am J Infect Control*. 2022;50(12):1355-1359. doi:10.1016/j.ajic.2022.03.005.
19. Williams B, Shelley PD, Patel V, Prothro C, Reynolds SS. Using PDSA cycles to improve oral care compliance. *Am J Infect Control*. 2023;51(1):110-113. doi:10.1016/j.ajic.2022.05.006.
20. Reynolds SS, Granger BB. Implementation science toolkit for clinicians: improving adoption of evidence in practice. *Dimens Crit Care Nurs*. 2023;42(1):33-41. doi:10.1097/DCC.0000000000000556.

ABOUT THE AUTHORS

Staci S. Reynolds, PhD, RN, ACNS-BC, CCRN, CNRN, SCRN, CPHQ, is associate clinical professor, Duke University School of Nursing, Durham, NC. She is the current editor of the Journal of Nursing Care Quality.

Julee Waldrop, DNP, PNP, FNP, SANE-A, is clinical professor and assistant dean, Doctor of Nursing Practice Program, Duke University School of Nursing, Durham, NC. She is the current editor of the Journal for Nurse Practitioners.

The authors have disclosed that they have no significant relationships with, or financial interest in, any commercial companies pertaining to this article.

Address correspondence and reprint requests to: Staci S. Reynolds, PhD, RN, ACNS-BC, CCRN, CNRN, SCRN, CPHQ, Duke University School of Nursing, 307 Trent Dr, Durham, NC 27710 (Staci.reynolds@duke.edu).