HHS Public Access

Author manuscript

J Registry Manag. Author manuscript; available in PMC 2015 July 10.

Published in final edited form as:

J Registry Manag. 2011 ; 38(3): 162–165.

Using a Statistical Process Control Chart during the Quality Assessment of Cancer Registry Data

Zachary M. Myles, MPH[a], **Robert R. German, DrPH, MPH**[a], **Reda J. Wilson, MPH, RHIT, CTR**[a], and **Manxia Wu, MD, MPH**[a]

[a]Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion, CDC.

Abstract

Statistical process control (SPC) charts may be used to detect acute variations in the data while simultaneously evaluating unforeseen aberrations that may warrant further investigation by the data user. Using cancer stage data captured by the Summary Stage 2000 (SS2000) variable, we sought to present a brief report highlighting the utility of the SPC chart during the quality assessment of cancer registry data. Using a county-level caseload for the diagnosis period of 2001–2004 (n=25,648), we found the overall variation of the SS2000 variable to be in control during diagnosis years of 2001 and 2002, exceeded the lower control limit (LCL) in 2003, and exceeded the upper control limit (UCL) in 2004; in situ/localized stages were in control throughout the diagnosis period, regional stage exceeded UCL in 2004, and distant stage exceeded the LCL in 2001 and the UCL in 2004. Our application of the SPC chart with cancer registry data illustrates that the SPC chart may serve as a readily available and timely tool for identifying areas of concern during the data collection and quality assessment of central cancer registry data.

Keywords

process control; quality control; statistical methods

Introduction

Historically, the quality assurance of cancer registry data has relied on the interpretation of counts, proportions and/or rates for the identification of irregularities within data collected.[1–3] However, frequent quality control analyses using these methods with cancer registry data may be impractical and can show statistically significant but not clinically relevant changes in data. To maintain the high quality of data that are collected, consolidated, and edited by cancer registrars at the central cancer registry, methods—such as a statistical process control (SPC) chart— that use quantifiable numeric indicators to readily detect and identify variations in registry data may be more preferable to descriptive analyses that are often subjective and qualitative.

Address correspondence to Zachary Myles, 4770 Buford Highway NE, Mailstop K-53, Atlanta, GA 30341. Telephone: (770) 488-6521. Fax: (770) 488-4759.

Originally created as an industrial tool for measuring the performance of routine manufacturing processes, the SPC chart's strength is that it allows the data user to draw logical decisions, rather than sudden judgments based on "last month" regarding the need for investigation, change in procedures, etc.[4] The underlying principle of a SPC chart is that all processes vary inherently and the end result can be described in statistical terms. The assumption of a SPC chart is that if the occurrence of a particular event is examined over time, the number of events in a set period of time (eg, 1 month) will follow a statistical distribution generally approximating a bell-shaped curve.[4] If the distribution of the occurrence over time is consistent, then the occurrences in the current time period will be statistically consistent with the historical experience of the event. In regards to the assessment of cancer registry data, a SPC chart has the capacity to identify changes in the reporting of data elements, while discriminating when and where during the data collection the change had occurred and prompt the registry's quality assurance staff to investigate the issue further.

Structurally, SPC charts are chronological graphs of data with control limits (typically set at 2 or 3 standard deviations from the expected value) represented by horizontal lines on the graph (see sample chart in Figure 1). These lines, encompassing the center line (CL) (usually the mean of the statistic of interest), upper control limit (UCL) and the lower control limit (LCL), define the central tendency and range of natural variation of plotted values assuming that change within the data will not occur without being due to chance. Values plotted outside the UCL and LCL are considered to be statistically significant indications that a process is not producing outcomes from one consistent and homogenous process, while values plotted within these limits are considered to be a result of natural variation and "in control."

As iterated earlier, the SPC chart may be used to detect acute variations in the data while simultaneously evaluating unforeseen aberrations that may warrant further investigation by the data user. In this current example, we will present a brief report highlighting the utility of the SPC chart during the quality assessment of cancer registry data using the Summary Stage 2000 (SS2000) variable as collected by Centers for Disease Control and Prevention's National Program of Cancer Registries (CDC-NPCR). Considering the well-documented change in the reporting of cancer stage by central cancer registries (CCRs) from 1998–2004, we believe the analysis of the SS2000 variable provides an opportunity to present the strengths of the SPC chart using a period of time where known changes in data collection had occurred and were due to factors outside of the statistical realm of chance. Since 1996, changes in the reporting to the SS variable have included: (1) cancer staging based on direct codes from Summary Stage 1977, (2) cancer staging based on direct codes from SS2000, and (3) the implementation of collaborative staging (deriving stage based upon TNM staging and SEER SS) that began in diagnosis year 2004. If effective, the SPC chart should detect the implementation of these changes and produce plots outside the LCL or UCL for diagnosis years occurring prior to 2001 and during the diagnosis year 2004.

Methods

Data for the SS2000 variable were collected from 46 population-based cancer registries participating in the CDC-NPCR. The CDC-NPCR data used were submitted in the NPCR-CSS Call for Data and reported to CDC as of January 31, 2009. These registries met established criteria for high-quality data. Information on states that met established data quality criteria for United States Cancer Statistics (USCS) is available at the CDC-NPCR's website, http://www.cdc.gov/cancer/npcr/uscs/data/00_data_quality.html.

In testing our null hypothesis, we created an analytic dataset consisting of all cancer diagnoses from 1 US county (n=25,648) during the years of 1998–2004. Using a local registry-level caseload allowed us a true assessment of the utility of the SPC chart during a real-life scenario—in an ideal setting, an SPC chart will be performed by staff at the local or central cancer registries at the time of data collection and that will ensure potential irregularities in data have been assessed prior to the data submission.

Using the *proc shewart* function packaged with SASv9.2 statistical software (SAS Institute Inc., Cary, North Carolina) we then plotted the SS2000 variable (for each reported diagnosis) against the year of diagnosis on a SPC chart for 2 separate diagnosis periods including 1998–2004 and 2001–2004 with the alpha (α) for each chart equaling 0.01. We repeated this step to create stage-specific SPC charts for stages in situ/localized, regional and distant for both diagnosis periods. The comparison of 2 separate diagnosis periods allowed us to determine the SPC chart's ability to detect changes in the cancer registry data using known time parameters where reporting practices had discretely changed over time.

Results

For the diagnosis period of 1998–2004, the overall variation (reflecting registry reports of in situ, localized, regional, distant, "blank(s)" and unstaged) of the SS2000 variable was never within the control limits (Figures 2a, 2b, 2c, and 2d, respectively). During this diagnosis period, stage-specific SPC charts presented plots for each cancer stage (*in situ*/localized, regional, and distant) consistently beyond the UCL following the diagnosis year of 2000. In comparison, during the diagnosis period of 2001–2004, we found the overall variation of the SS2000 variable to be in control during diagnosis years of 2001 and 2002, (unexpectedly) exceeded the LCL in 2003, and exceeded the UCL in 2004 (Figure 3a). Stage-specific charts found the staging of in situ/localized to be the only SS category to be in control throughout the diagnosis years of 2001–2004 (Figure 3b). During this same diagnosis period, the staging of regional exceeded UCL in 2004 (Figure 3c), while the staging of distant exceeded the LCL in 2001 and the UCL in 2004 (Figure 3d).

Discussion

Using the SPC chart here, we identified variation in CDC-NPCR data using a scenario in which we knew changes in data collection had occurred and were due to factors outside of the statistical realm of chance. Our example illustrated that the SPC chart may serve as an invaluable tool for prospectively identifying areas of concern during the data collection and quality assessment of cancer registry data. However, the generalizability of our charts is

challenged because of our consolidation of stage data from diagnoses of all cancer sites. For real-time application of the SPC chart, site-specific charts may be more appropriate to control for site-specific factors affecting cancer staging. German et al[5] found the data quality analysis from the Patterns of Care Study (POC1) underscores the importance of examining the quality of specific data elements by cancer site.[5] Further validation for the use of the SPC chart with cancer registry data should incorporate varying caseloads (ie state, regional and/or national levels) and utilize additional registry data elements beyond variables associated with cancer staging. Nonetheless, SPC charts presented here displayed their usefulness in identifying changes in cancer stage reporting, and potential for being a tool used during the ongoing monitoring of cancer registry data.

Acknowledgments

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention and the National Cancer Institute.

References

1. Centers for Disease Control and Prevention. Guidelines for investigating clusters of health events. MMWR. 1990; 39:1–22.
2. Jacquez GM, Grimson R, Waller LA, Watenberg D. The analysis of disease clusters, part II: introduction to techniques. Infect Control Hosp Epidemiol. 1996; 17:385–397. [PubMed: 8805074]
3. Arantes A, Carvalho Eda S, Medeiros EA, Farhat CK, Mantese OC. Use of statistical process control charts in the epidemiological surveillance of nosocomial infections. Rev Saude Publica. 2003; 37(6):768–774. [PubMed: 14666307]
4. Sellick JA. The use of statistical process control charts in hospital epidemiology. Infect Control Hosp Epidemiol. 1993; 14(11):649–656. [PubMed: 8132986]
5. German RR, Wike JM, Wolf HJ, et al. Quality of cancer registry data: findings from CDC-NPCR's Breast, Colon, and Prostate Cancer Data Quality and Patterns of Care study. J Registry Manage. 2008; 35(2):67–74.

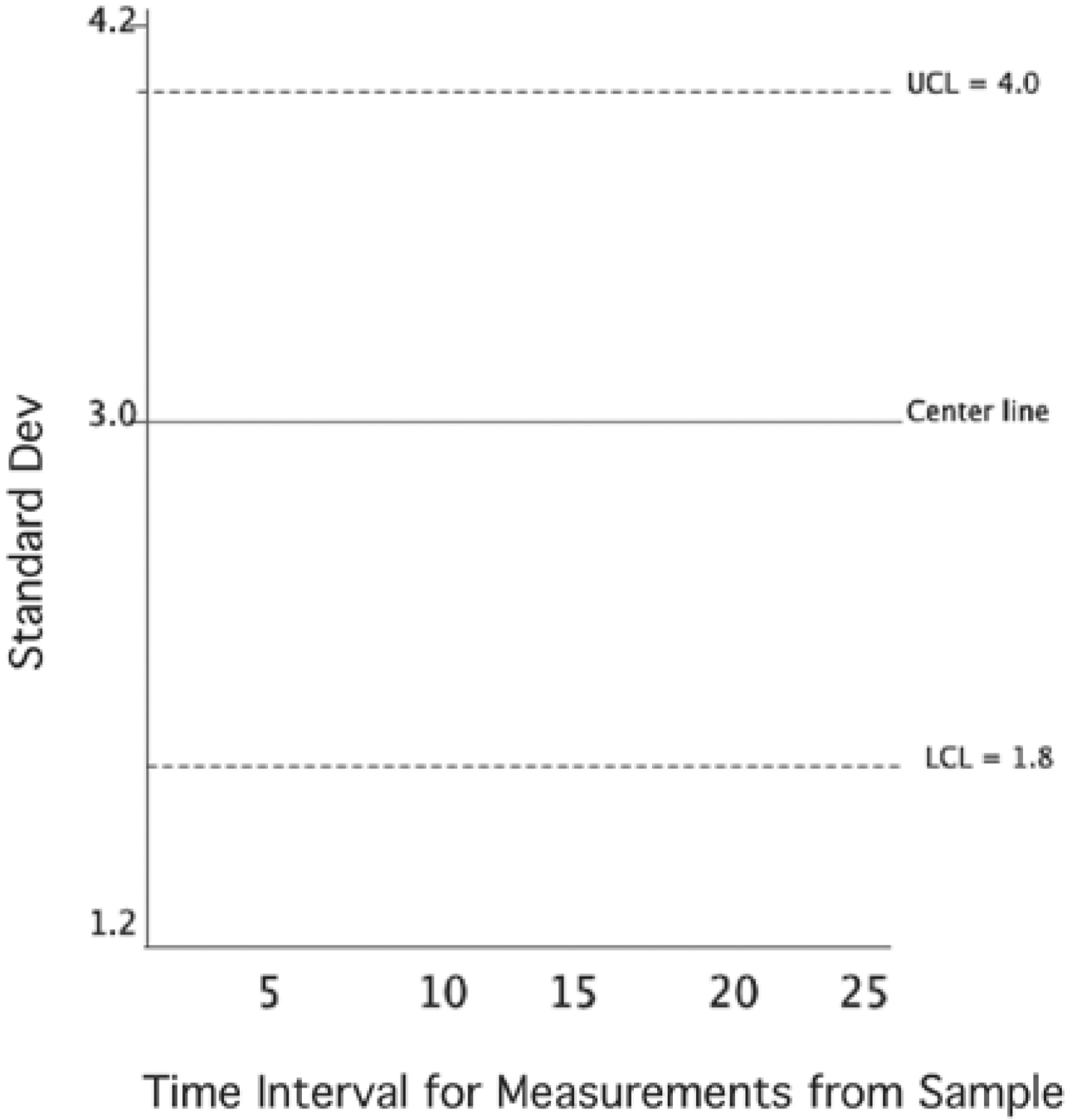

Figure 1.
Sample Statistical Process Control (SPC) Chart

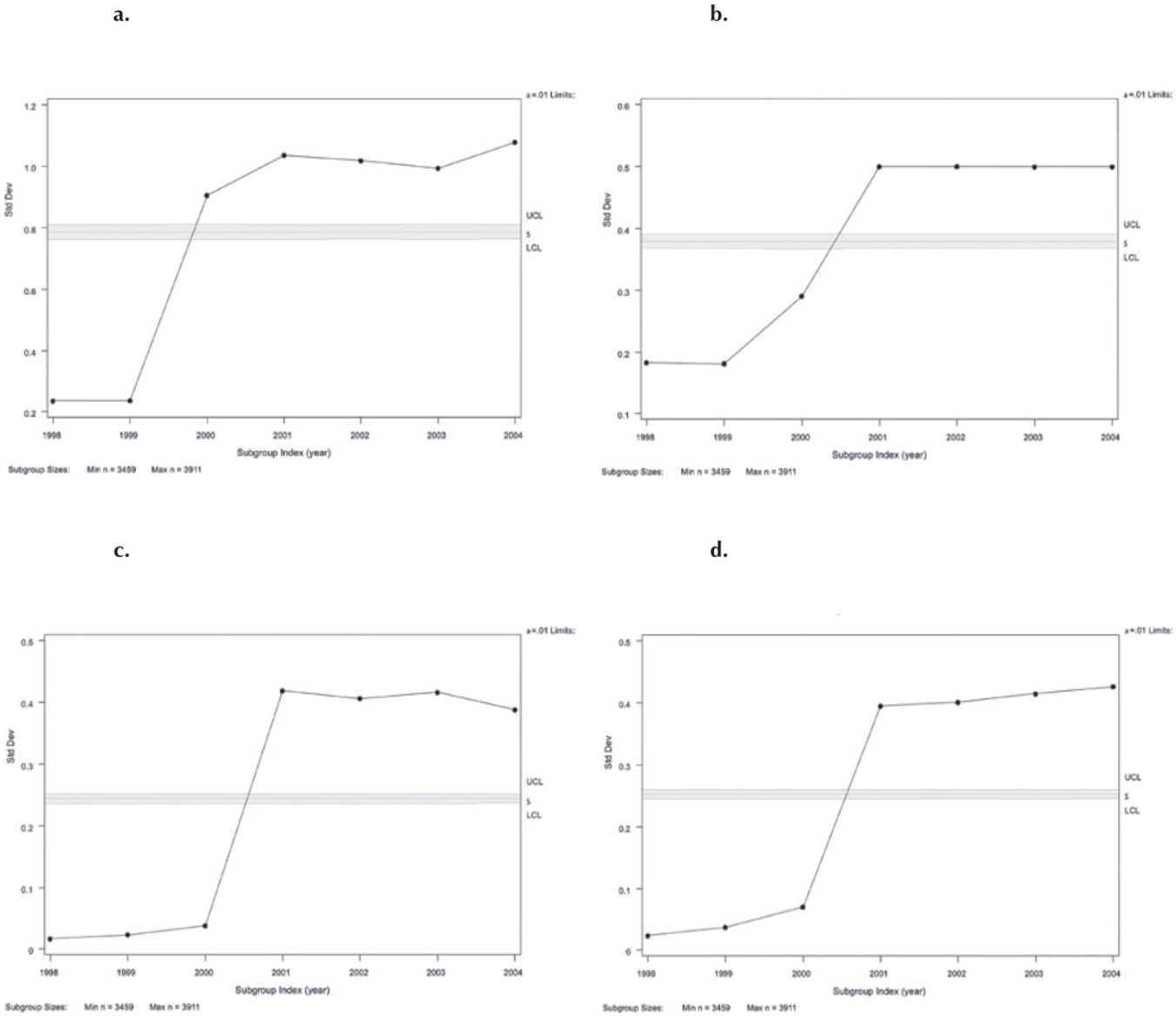

Figure 2.

a. Statistical Process Control Chart Indicating Variation in Summary Stage 2000 Variable, Diagnosis Years 1998–2004

b. Statistical Process Control Chart Indicating Variation in Summary Stage 2000 Variable Reported as In Situ/Localized, Diagnosis Years 1998–2004

c. Statistical Process Control Chart Indicating Variation in Summary Stage 2000 Variable Reported as Regional, Diagnosis Years 1998–2004

d. Statistical Process Control Chart Indicating Variation in Summary Stage 2000 Variable Reported as Distant, Diagnosis Years 1998–2004

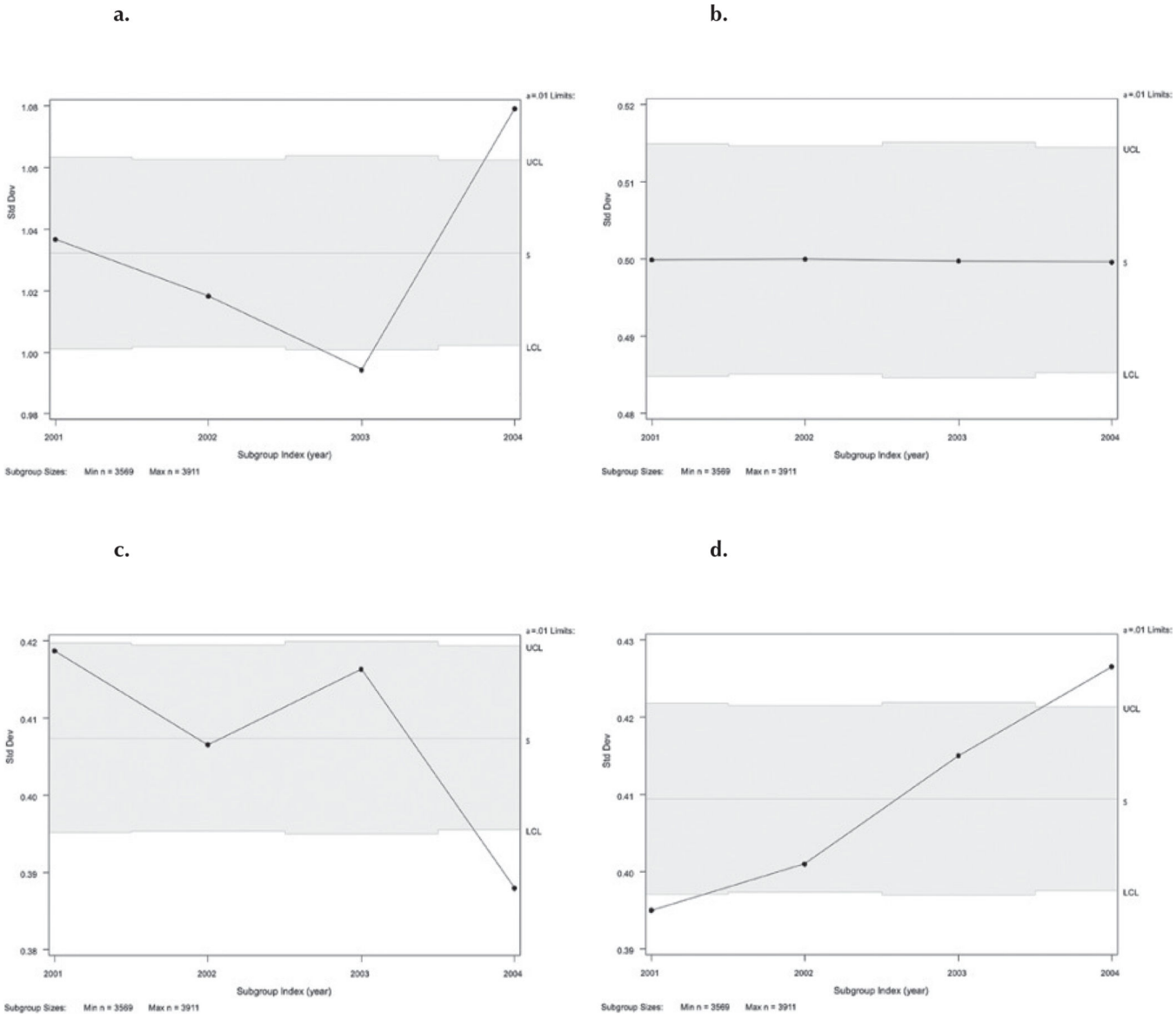

Figure 3.

a. Statistical Process Control Chart Indicating Variation in Summary Stage 2000 Variable, Diagnosis Years 2001–2004

b. Statistical Process Control Chart Indicating Variation in Summary Stage 2000 Variable Reported as In Situ/Localized, Diagnosis Years 2001–2004

c. Statistical Process Control Chart Indicating Variation in Summary Stage 2000 Variable Reported as Regional, Diagnosis Years 2001–2004

d. Statistical Process Control Chart Indicating Variation in Summary