Assist Prof Gaj Vidmar, PhD

University Rehabilitation Institute, Ljubljana, Slovenia
University of Ljubljana, Faculty of Medicine, Institute for Biostatistics and Medical Informatics
Univ. of Primorska, Fac. of Mathematics, Natural Sciences and Information technologies (Koper)
University of Primorska, Faculty of Health Sciences (Izola)

STATISTICAL QUALITY CONTROL IN HEALTH CARE – a general overview and the specific problem of outlier detection among overdispersed proportions

Univerza v Ljubljani

ibmi

Institute for Biostatistics and Medical Informatics
Ljubljana, 10 December, 2013

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI SCIENZE BIOMEDICHE PER LA SALUTE
PhD SCHOOL IN INTEGRATIVE BIOMEDICAL SCIENCES

ISTITUTO AUXOLOGICO ITALIANO
Istituto di ricovero e cura a carattere scientifico

Dipartimento di Scienze Biomediche per la Salute
Milano, 13 December, 2013

Part I: General overview

- the Bristol heart scandal and the Harold Shipman case
- key concepts and methods of SPC/SQC in health care
    - league tables and caterpillar plots
    - Shewhart charts and funnel plots
    - adjusted risk
    - exponentially weighted moving average (EWMA)
    - cumulative sum charts (CUSUM)
    - sequential probability ratio test (SPRT)
    - variable life adjusted display (VLAD)
- recommended literature

Part II: Our research on outlier detection among overdispersed proportions

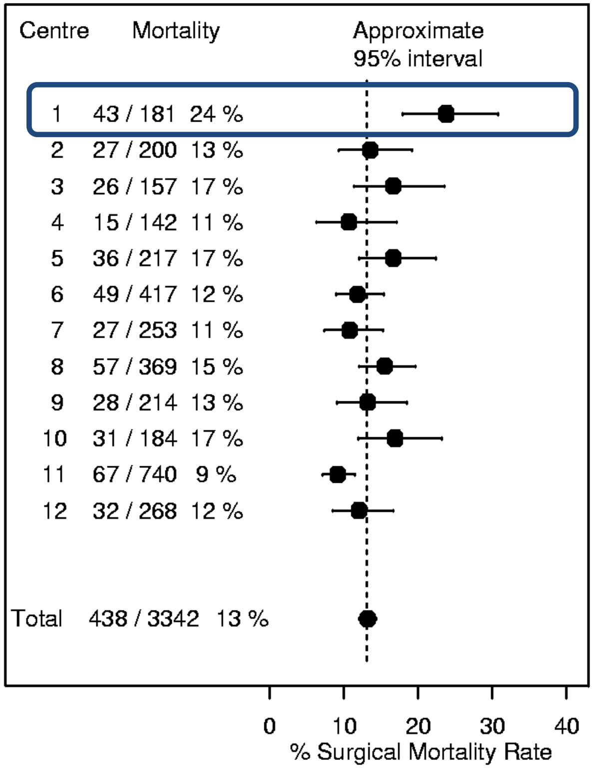
- double-square-root chart
    - control limits based on prediction intervals from regression-through-origin
    - confidence level adjusted to sample size
    - transformed into asymmetric funnel plot
- comparison with established methods
    - general-purpose outlier-detection tests and rules
    - Laney's p' control chart for cross-sectional data
    - Spiegelhalter's regression approach (multiplicative and additive)
    - large-scale simulations and real national-level data
- some general thoughts on applying statistics for quality improvement in health care and elsewhere in the public service sector

Part I

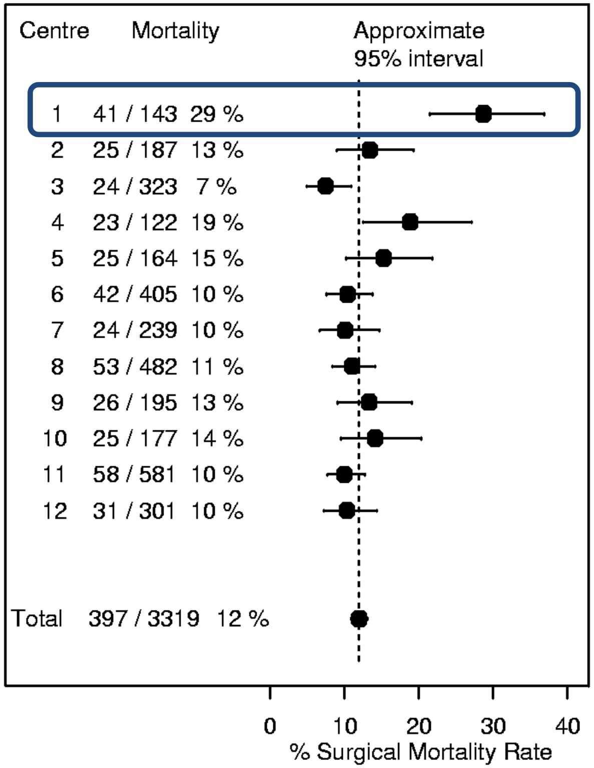
General overview

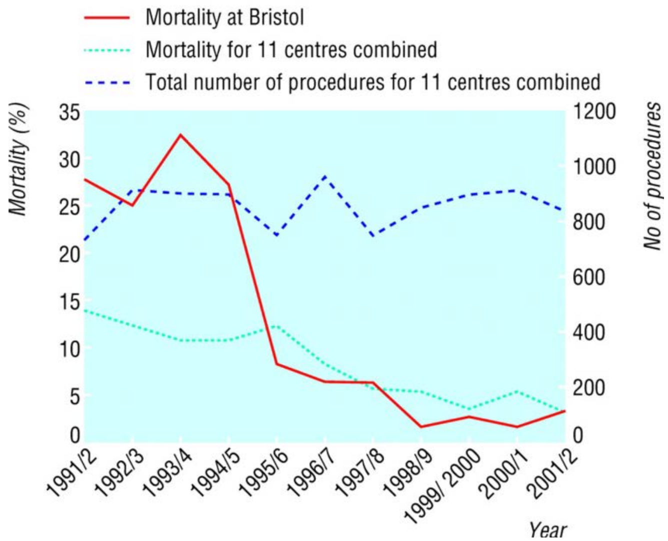
Bristol heart scandal

CSR data :
1991 - Mar95: under 1 yr



HES data :
1991 - Mar95: under 1 yr





The Bristol Royal Infirmary Inquiry

- media coverage
  - The Guardian
  - BBC
- the report
- Learning from Bristol
- the RCN response

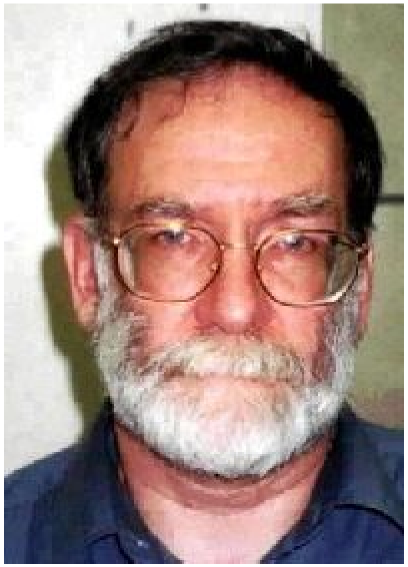
Key scientific articles

- Aylin *et al.*, *Lancet*, 2001
- Spiegelhalter *et al.*, *JRSSA*, 2002
- Spiegelhalter, *BMJ*, 2002
- Aylin *et al.*, *BMJ*, 2004

Data sources issues

- Response by Aylin *et al.*, *Lancet*, 2001
- Responses to Spiegelhalter, *BMJ*, 2002
- Westaby *et al.*, *BMJ*, 2007
- Responses to Westaby *et al.*, *BMJ*, 2007

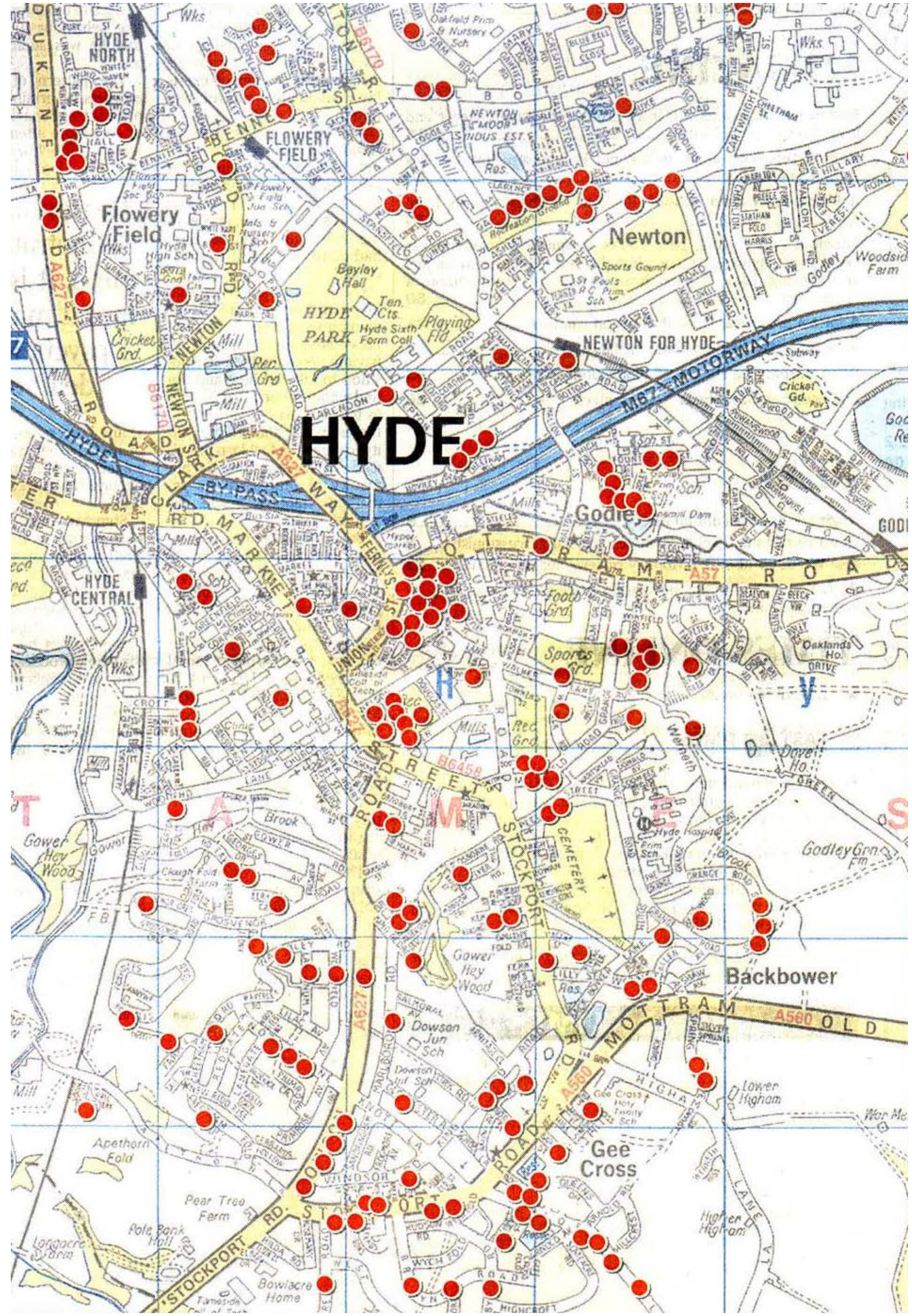
Harold Shipman



The Shipman Inquiry: 215 definite victims, 45 probable

Key scientific articles

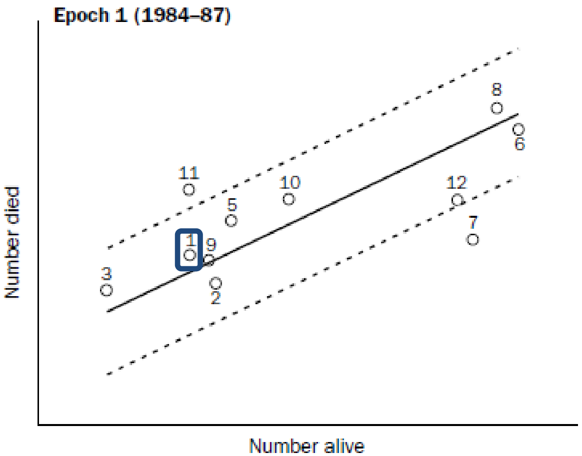
- [Baker *et al.*, *BMJ*, 2002]
- [Aylin *et al.*, *Lancet*, 2003]

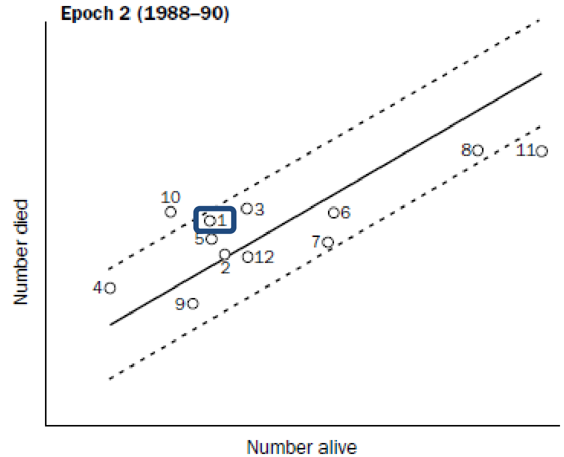


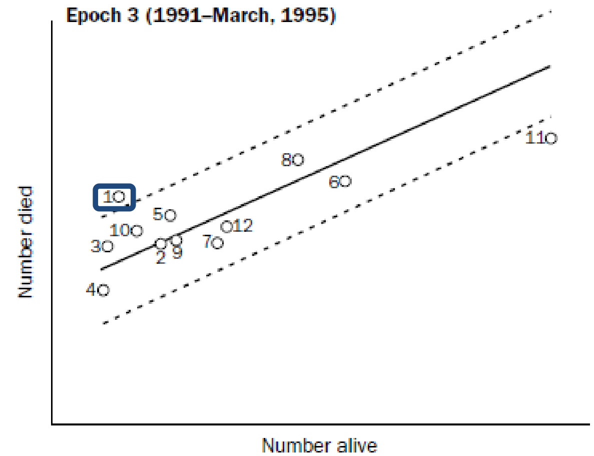
Double square root chart

(Shewhart chart for binomial data drawn on double square-root paper – both axes on all charts are in square-root scale)

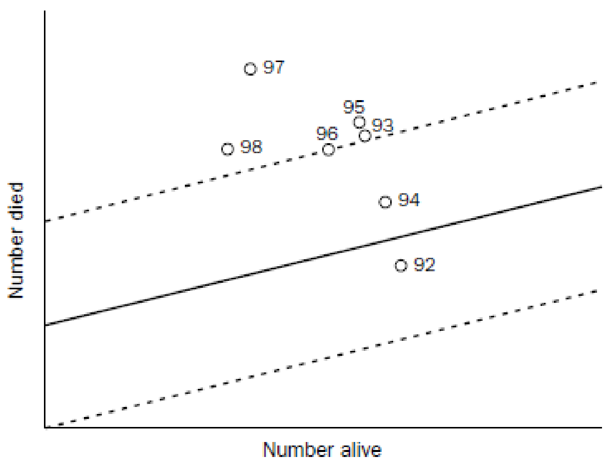
Bristol cardiac surgery (hospital 1 is Bristol Royal Infirmary).







Harold Shipman



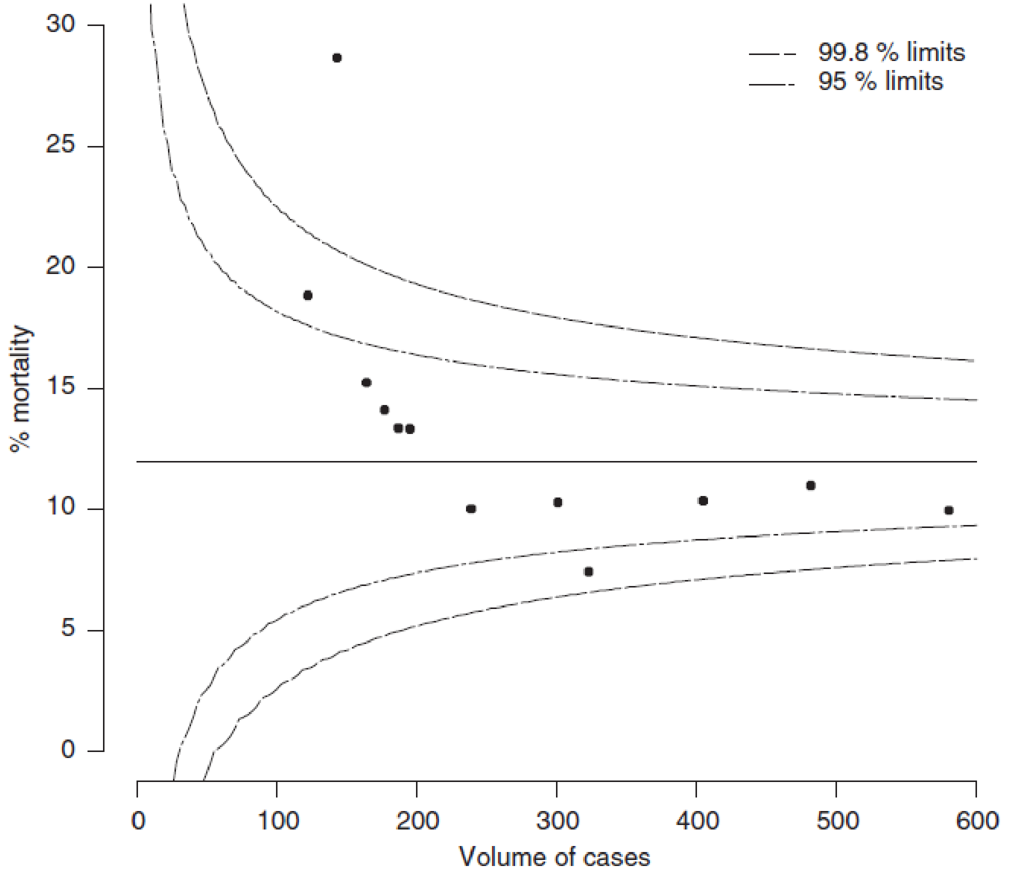
The three lines indicate the background mortality for women aged 65 years or more in Thameside and Glossop (inclusive of Shipman's patients). Shipman's annual mortality rates are imposed on this.

Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons
Mohammed MA, Cheng KK, Rouse A, Marshall T. *The Lancet*, 357(9254), 463-467, 10 February 2001

Funnel plots

- introduced in meta-analysis (usually vertical and drawn with straight sides – log scale on x-axis)
- in health care quality monitoring, p is plotted against n (horizontal curved funnels)
- basic property: the larger the subsample size, the closer to the overall mean any proportion is expected to be (i.e., increased variability is expected from smaller centres)
- observed event-rates are displayed (more intuitive than Shewhart charts)
- easy to calculate exact control limits (spreadsheet software suffices)

Mortality rates following paediatric cardiac surgery in under-1's in 12 English specialist centres (HES data) 1991-1995.
The target is the overall average rate of 12%.
The outlier is Bristol Royal Infirmary.



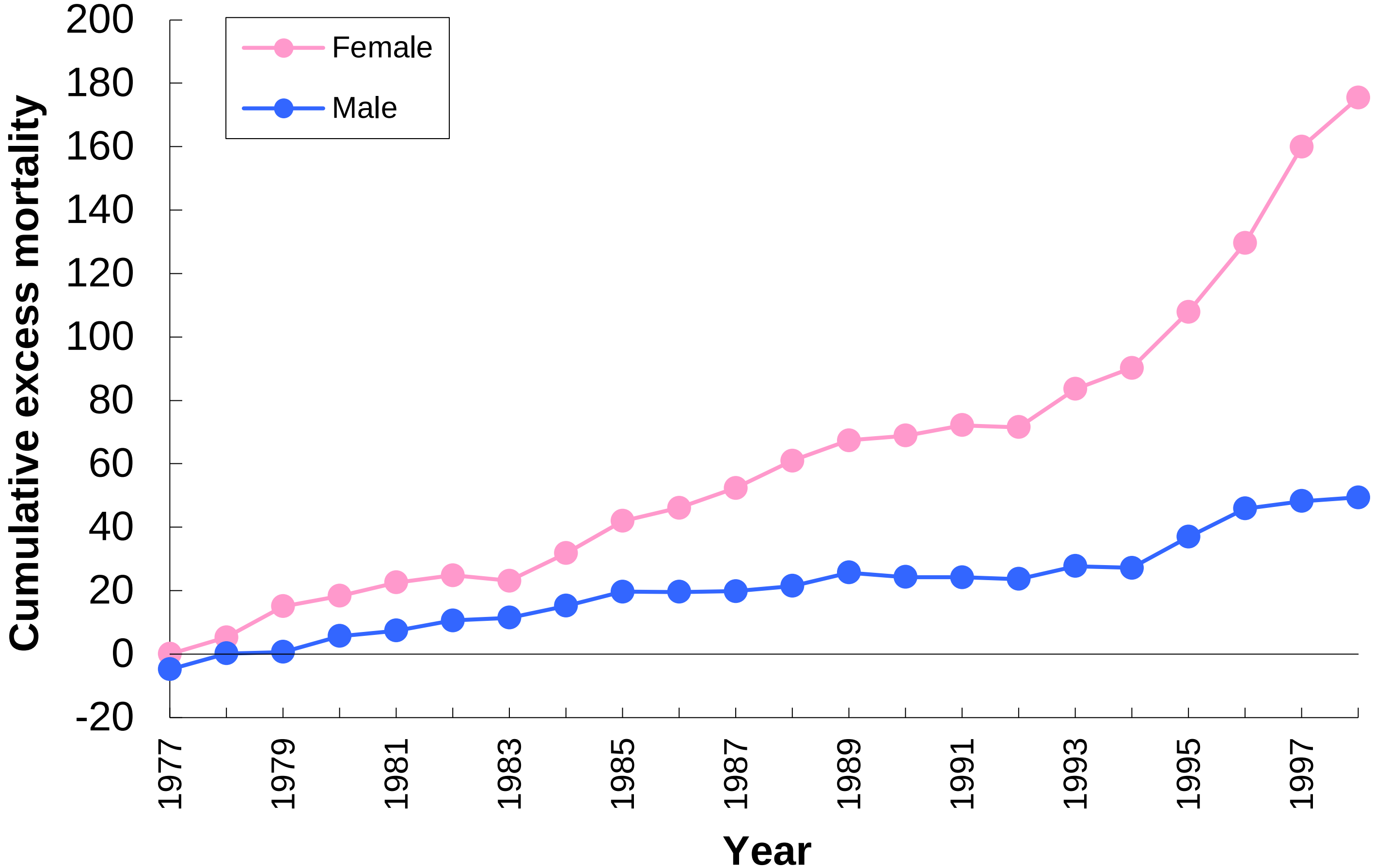
Funnel plots for institutional comparison
Spiegelhalter D. *Qual Saf Health Care*, 11(4), 390-391, 2002

Funnel plots for comparing institutional performance
Spiegelhalter D. *Stat Med*, 24(8), 1185-1202, 2005

CUSUM chart

(excess mortality = observed – expected)

Cumulative excess death certificates signed by Shipman: age >64 and death in home/practice



Can be risk-adjusted

- Steiner *et al*., *Biostatistics*, 2000
- Grigg & Farewell, *JRSSA*, 2004

CUSUM charting is **also applicable** for the purpose of **personal audit or credentialing of physicians** (Lim *et al*., Int J Qual Health Care, 14(3), 251-258, 2002)

Sequential probability ratio test

From Wikipedia, the free encyclopedia

Theory [edit]

As in classical hypothesis testing, SPRT starts with a pair of hypotheses, say H_0 and H_1 for the null hypothesis and alternative hypothesis respectively. They must be specified as follows:

$$H_0 : p = p_0$$
$$H_1 : p = p_1$$

The next step is calculate the cumulative sum of the log-likelihood ratio, $\log \Lambda_i$, as new data arrive:

$$S_i = S_{i-1} + \log \Lambda_i$$

The stopping rule is a simple thresholding scheme:

- $a < S_i < b$: continue monitoring (*critical inequality*)
- $S_i \geq b$: Accept H_1
- $S_i \leq a$: Accept H_0

where a and b ($0 < a < b < \infty$) depend on the desired type I and type II errors, α and β. They may be chosen as follows:

$a \approx \log \dfrac{\beta}{1-\alpha}$ and $b \approx \log \dfrac{1-\beta}{\alpha}$

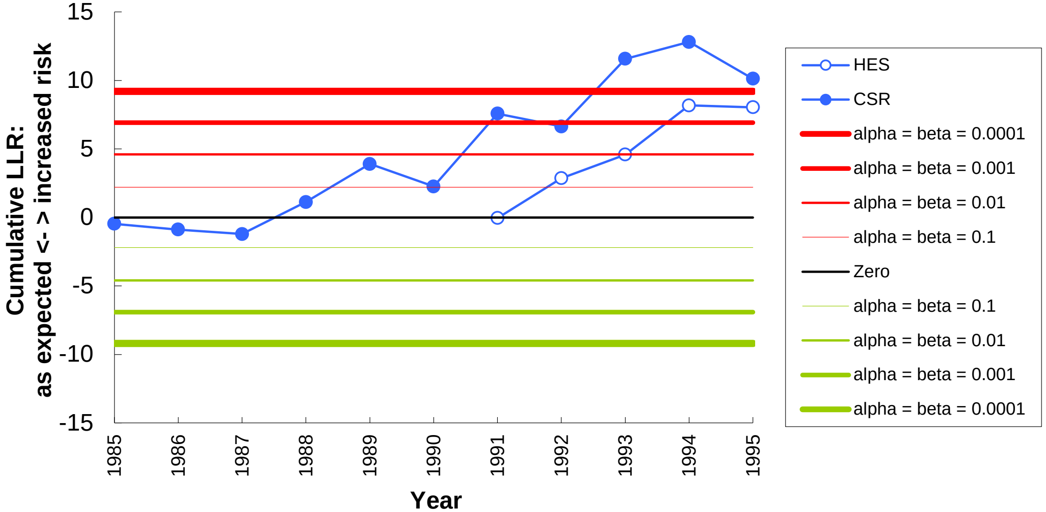
In other words, α and β must be decided beforehand in order to set the thresholds appropriately. The numerical value will depend on the application. The reason for using approximation signs is that, in the discrete case, the signal may cross the threshold between samples. Thus, depending on the penalty of making an error and the sampling frequency, one might set the thresholds more aggressively. Of course, the exact bounds may be used in the continuous case.

Risk-adjusted sequential probability ratio tests and longitudinal surveillance methods

Spiegelhalter D, Grigg O, Kinsman R, Treasure T. *Int J Qual Health Care*, 15(1), 7-13, 2003

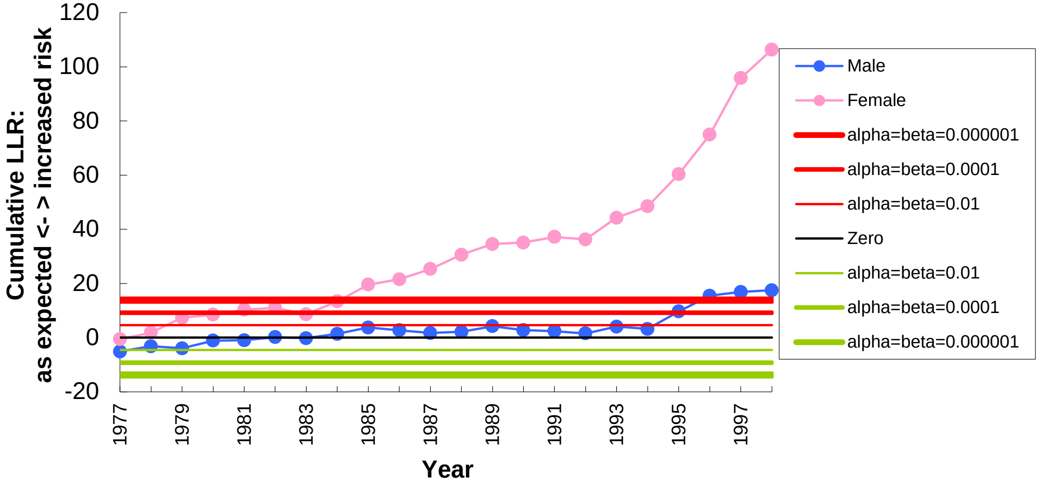
Bristol

SPRT for detection of a doubling in the odds on mortality



Shipman

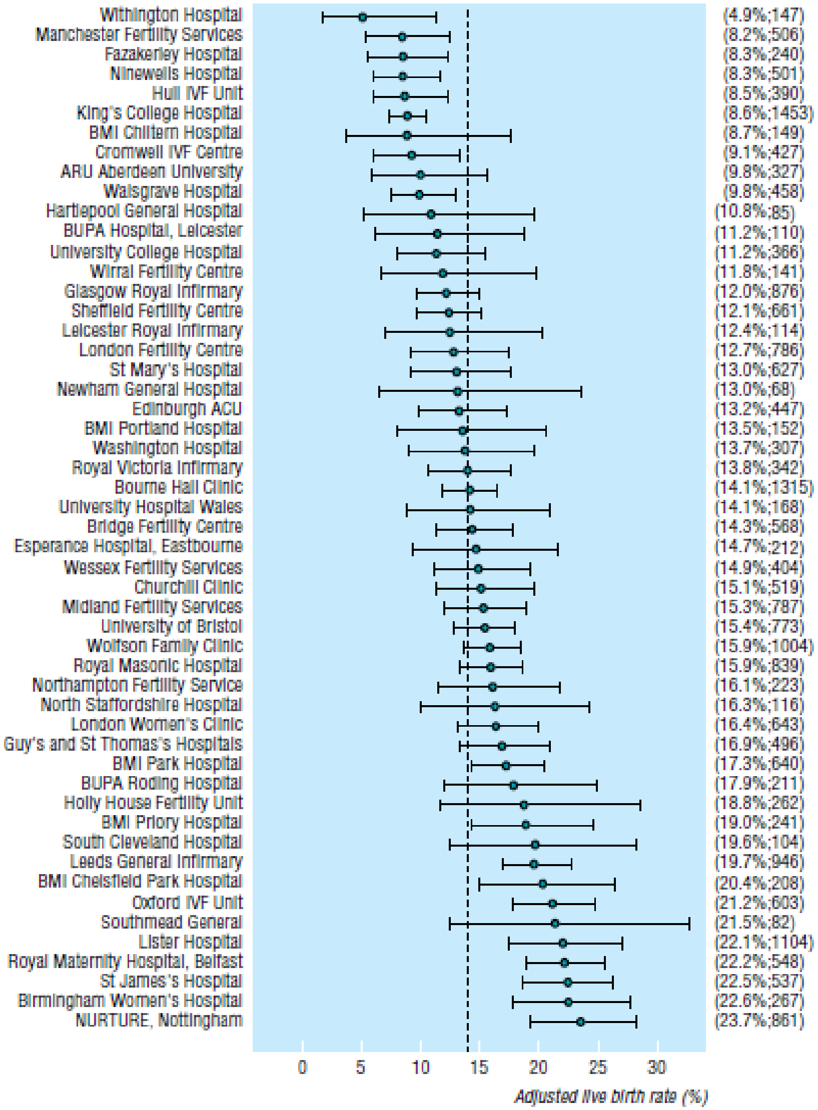
SPRT for detection of a doubling in mortality risk: age >64 and death in home/practice

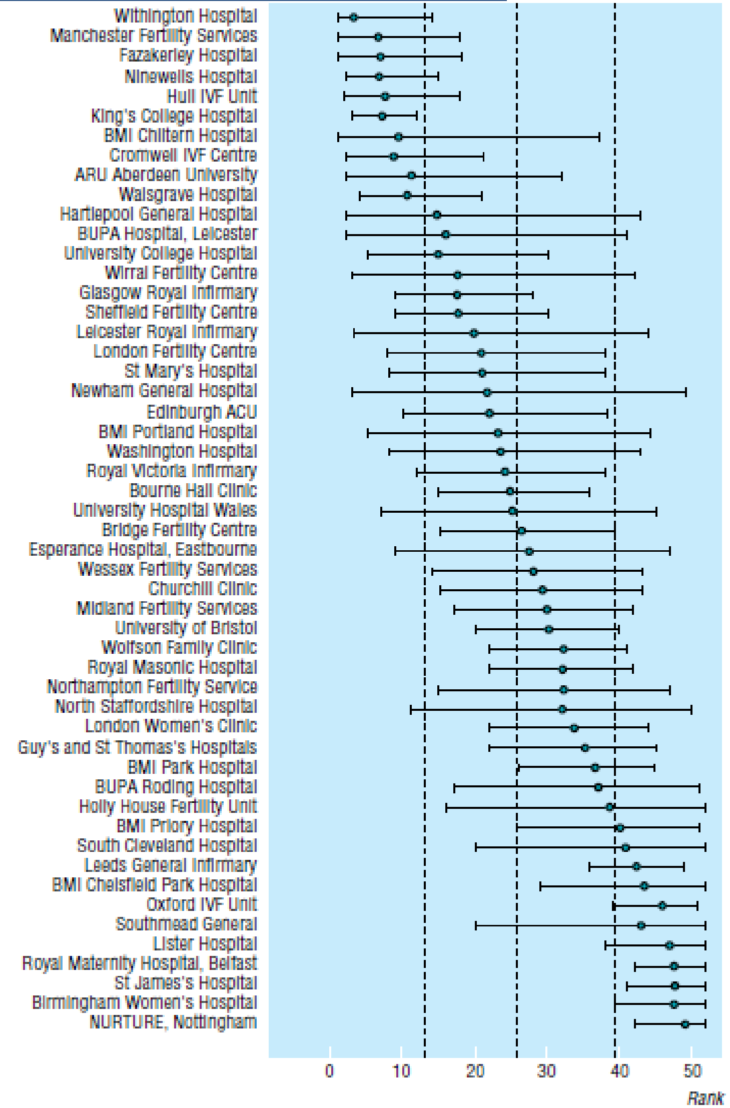


Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates

E Clare Marshall, David J Spiegelhalter

BMJ VOLUME 316 6 JUNE 1998

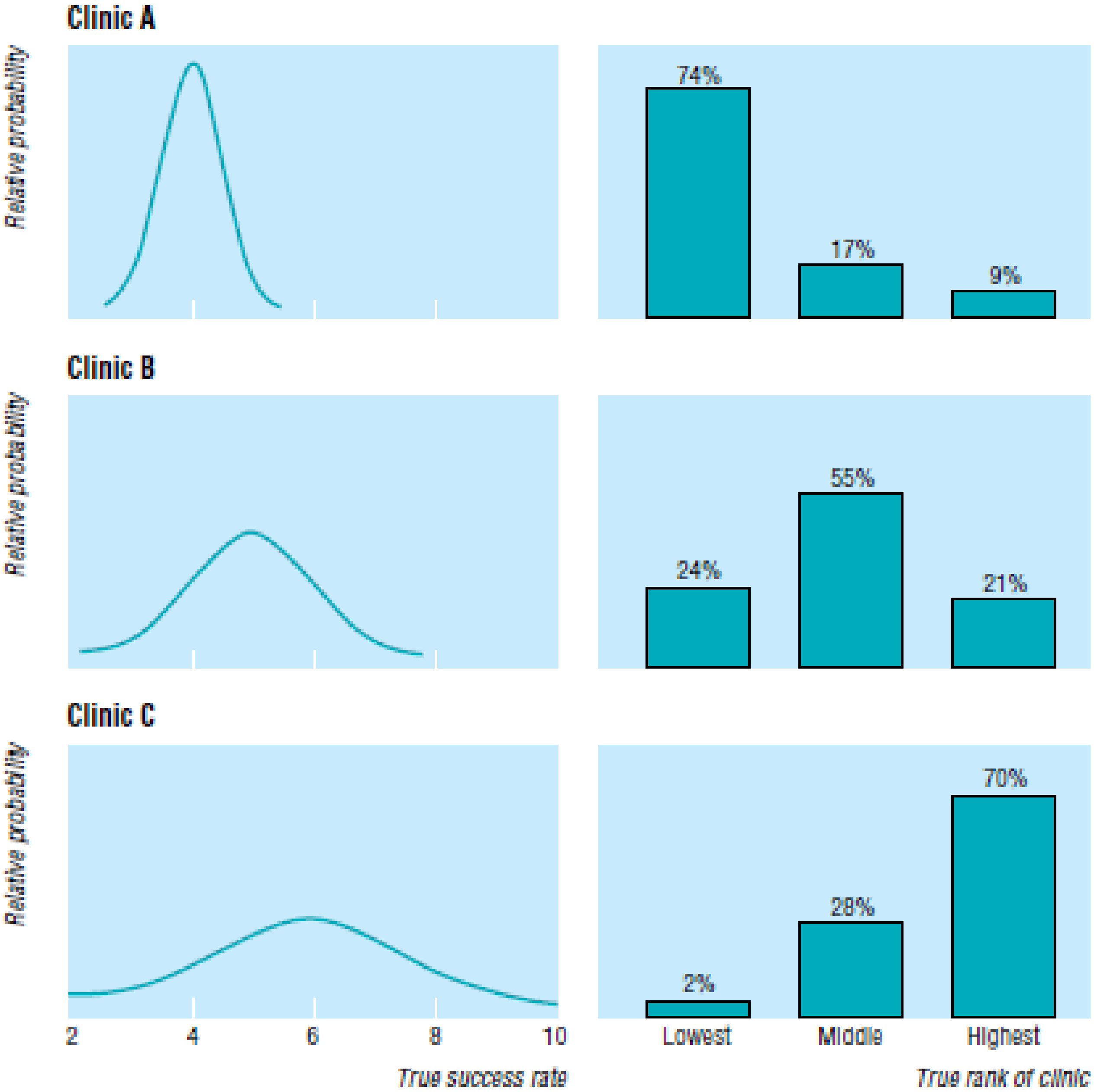




Estimates and 95% confidence intervals (in brackets, with no. of treatment cycles) for adjusted live birth rate in each clinic. Vertical dotted line represents national average of 14%.)

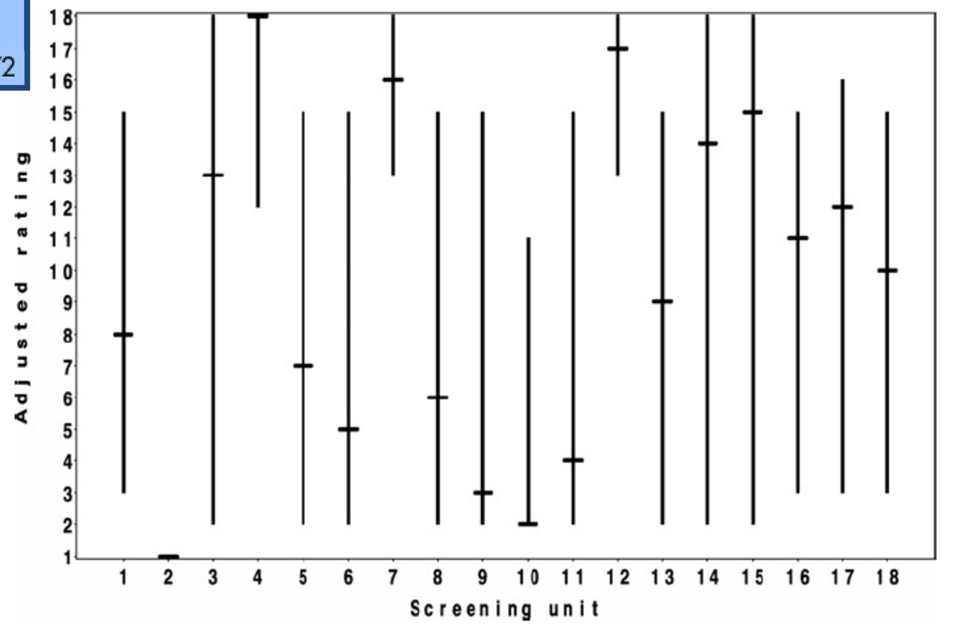
Median and 95% confidence intervals for rank of each clinic. Dashed vertical lines divide clinics into quarters according to rank.

Simulation


Clinic A
Relative probability
74%
17%
9%
Clinic B
Relative probability
24%
55%
21%
Clinic C
Relative probability
2%
28%
70%
2
4
6
8
10
True success rate
Lowest
Middle
Highest
True rank of clinic

League tables of breast cancer screening units: worst-case and best-case scenario ratings helped in exposing real differences between performance ratings

Oscar Lemmers, Mireille Broeders, André Verbeek, Gerard den Heeten, Roland Holland and George F Borm
J Med Screen 2009;**16**:67–72

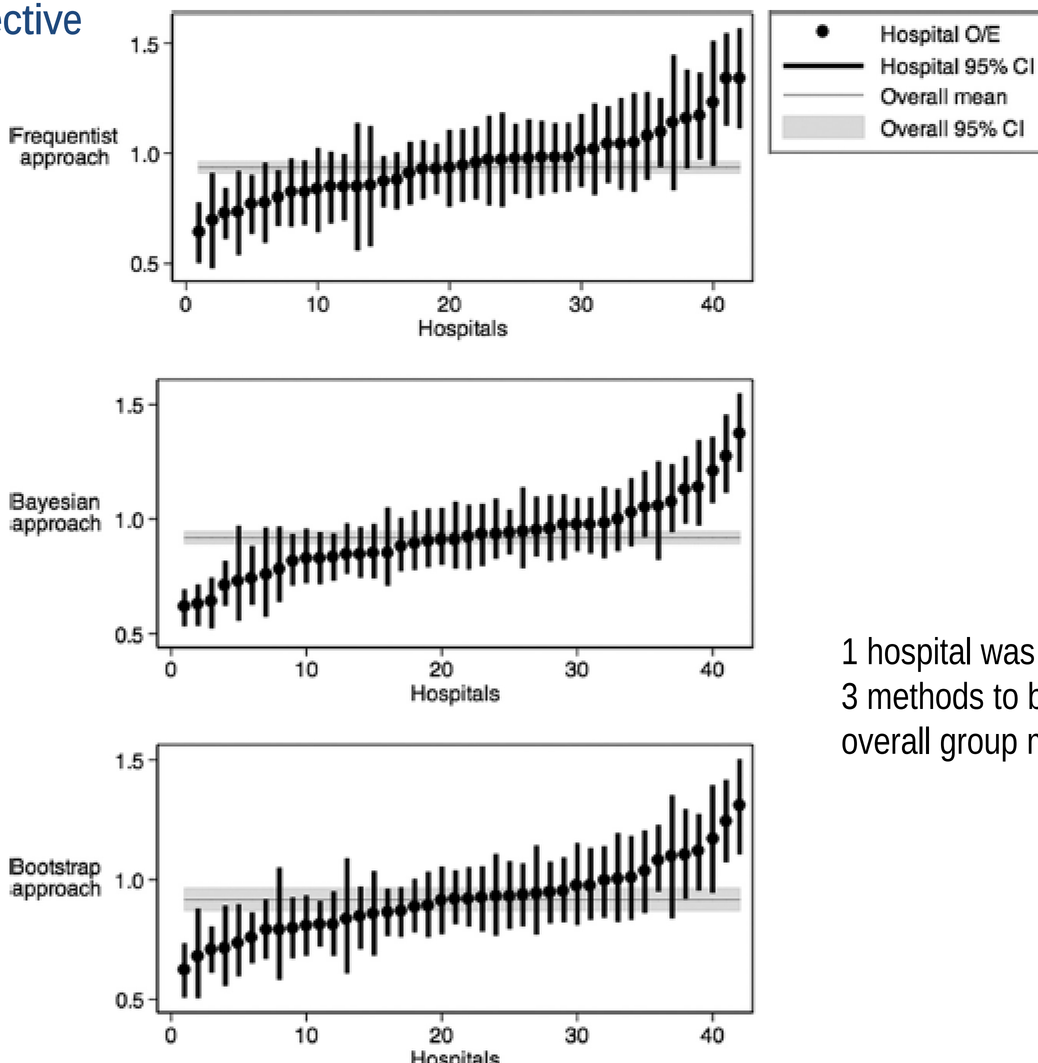


Screening unit	Unadjusted rating	Best-case scenario – worst-case scenario	Ratings adjusted for age*	Best-case scenario – worst-case scenario adjusted for age	Credible intervals (unadjusted)
2	1	1–1	1	1–1	1–1
10	2	2–9	2	2–11	2–5
11	3	2–15	**4**	2–15	2–11
9	4	2–15	**3**	2–15	2–10
8	5	2–15	**6**	2–15	2–12
6	6	2–15	**5**	2–15	3–13
13	7	2–17	**9**	2–15	3–15
5	8	3–17	**7**	2–15	3–15
1	9	3–15	**8**	3–15	4–15
18	10	3–17	10	3–15	3–16
16	11	3–17	11	3–15	5–15
17	12	3–17	12	3–16	4–16
14	13	2–18	**14**	2–18	2–18
15	14	2–18	**15**	2–18	3–18
3	15	3–18	**13**	2–18	3–18
7	16	8–18	16	13–18	13–18
12	17	13–18	17	13–18	14–18
4	18	8–18	18	12–18	12–18

Statistical Uncertainty of Mortality Rates and Rankings for Children's Hospitals

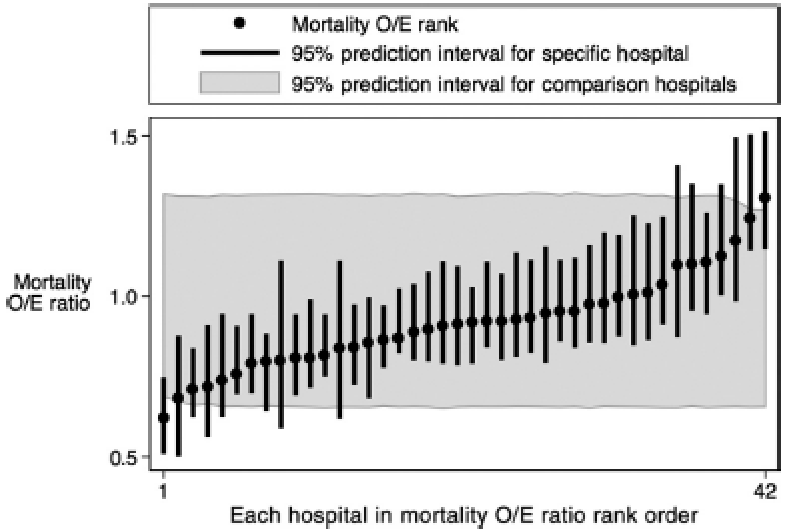
Feudtner *et al.*, *Pediatrics*
Vol. 128 No. 4 October 1, 2011

Central tendency perspective

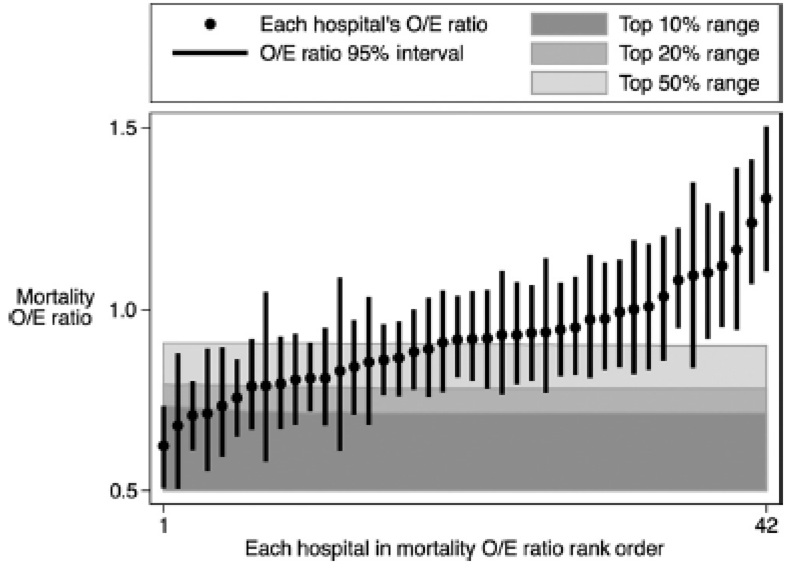


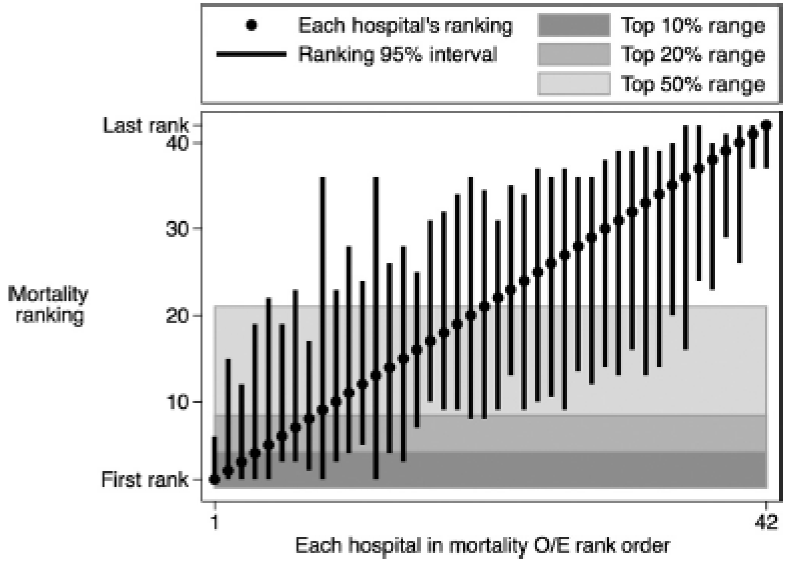
1 hospital was identified consistently by all 3 methods to be significantly below the overall group mean and 1 hospital above

Outlier perspective



Performance group perspective





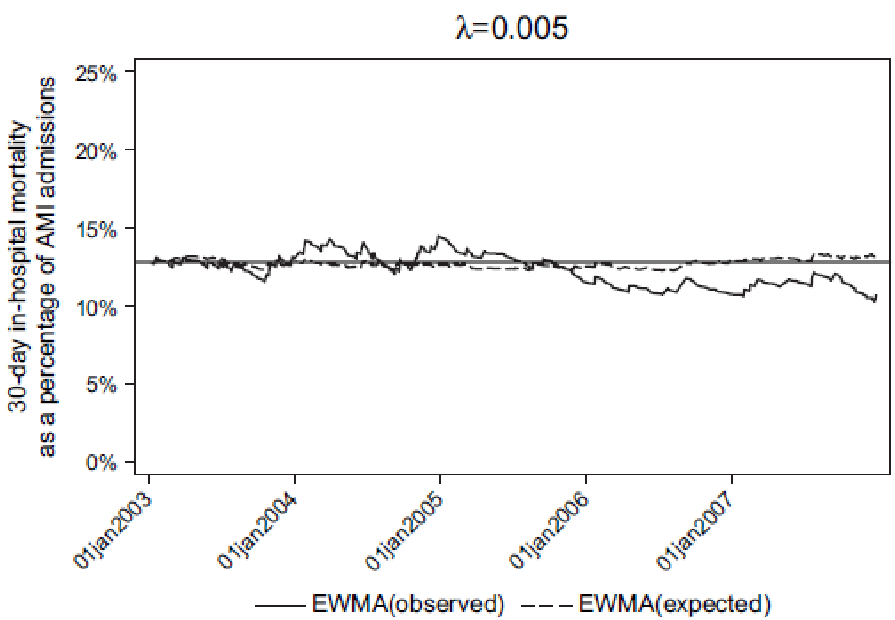
Exponentially weighted moving average charts to compare observed and expected values for monitoring risk-adjusted hospital indicators

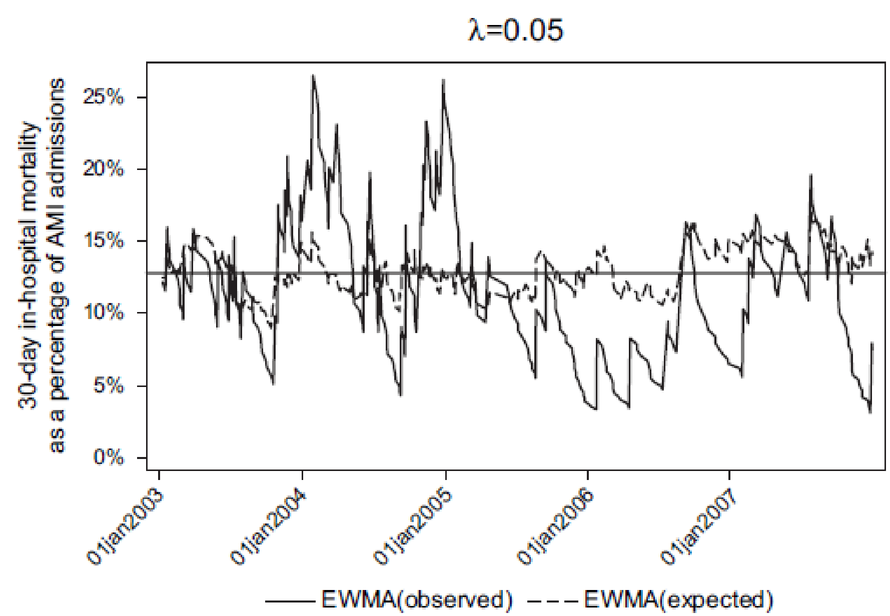
Cook DA, Coory M, Webster RA. *BMJ Qual Saf* (2011)

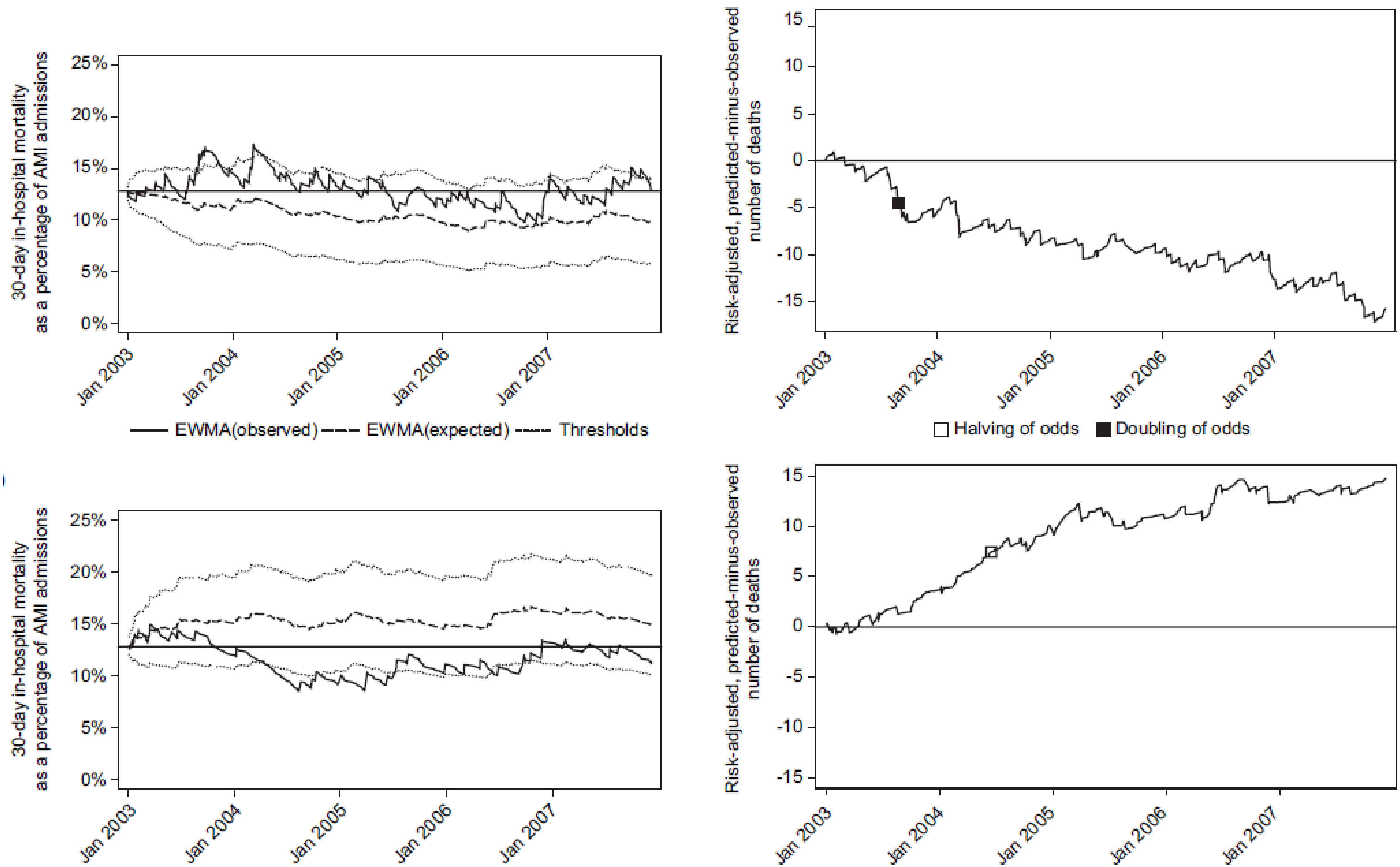
$$\text{EWMA(observed)}_j = \lambda y_j + (1-\lambda)\text{EWMA(observed)}_{j-1}$$

$$\text{EWMA(predicted)}_j = \lambda p_j + (1-\lambda)\text{EWMA(predicted)}_{j-1}$$

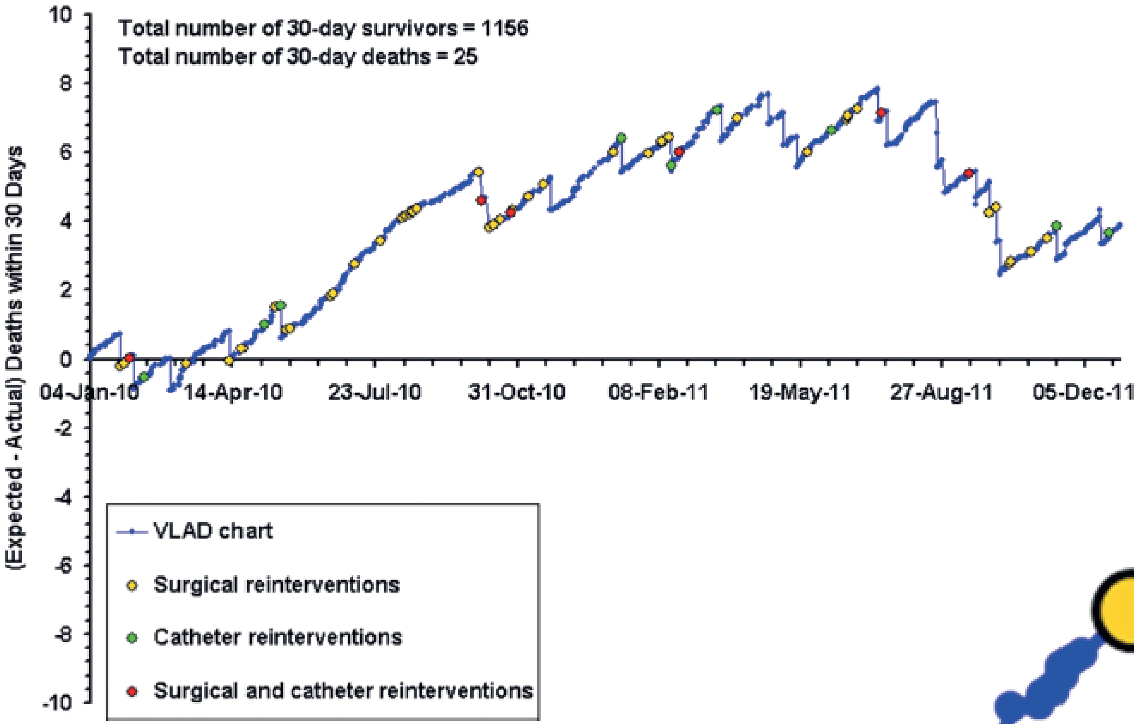
$$\text{CL}_j = \text{EWMA}_j^{pj} \pm L\lambda\sqrt{\sum_{k=1}^{j}\left[(1-\lambda)^{2(j-k)}p_k(1-p_k)\right]}$$

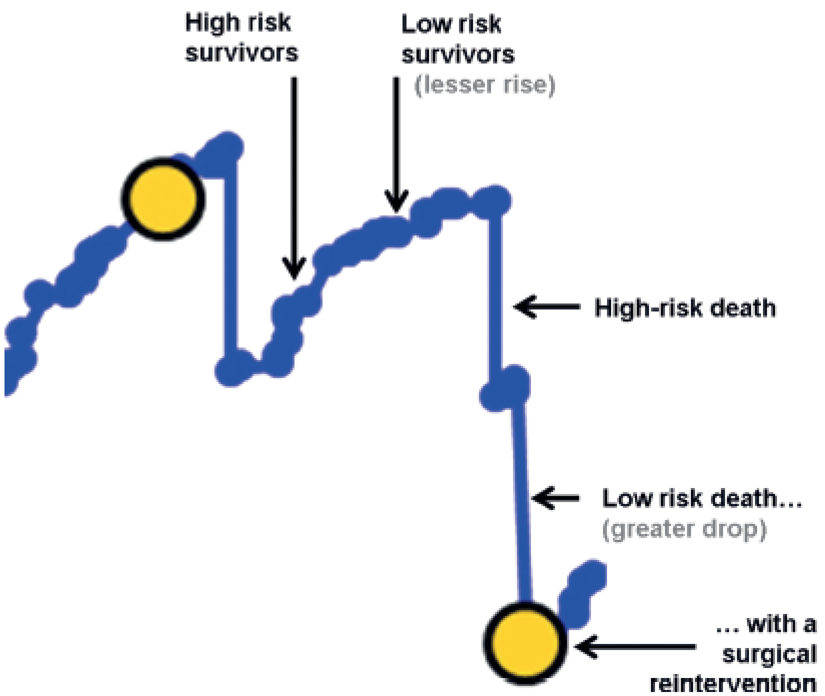





30-day in-hospital mortality as a percentage of AMI admissions
25%
20%
15%
10%
5%
0%
Jan 2003
Jan 2004
Jan 2005
Jan 2006
Jan 2007
EWMA(observed)
EWMA(expected)
Thresholds
Risk-adjusted, predicted-minus-observed number of deaths
15
10
5
0
-5
-10
-15
Halving of odds
Doubling of odds

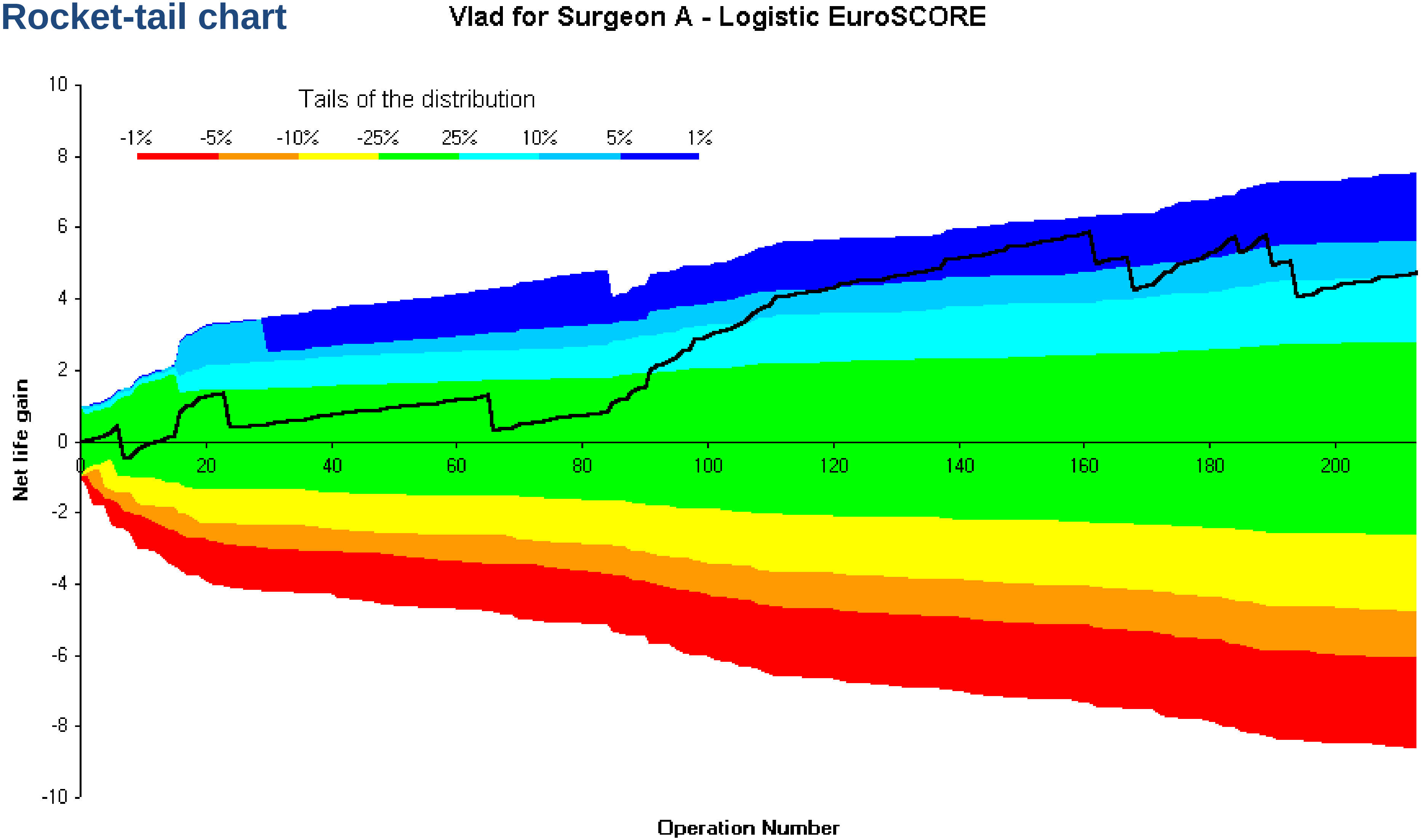
Real time monitoring of risk-adjusted paediatric cardiac surgery outcomes using variable life-adjusted display: implementation in three UK centres Pagel C, *et al. Heart* 2013



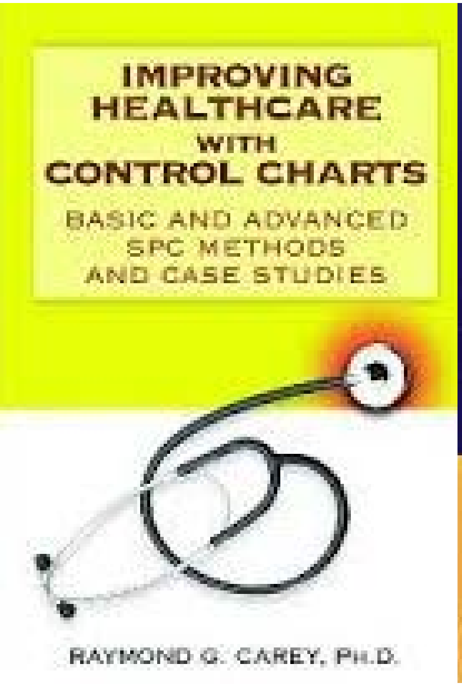


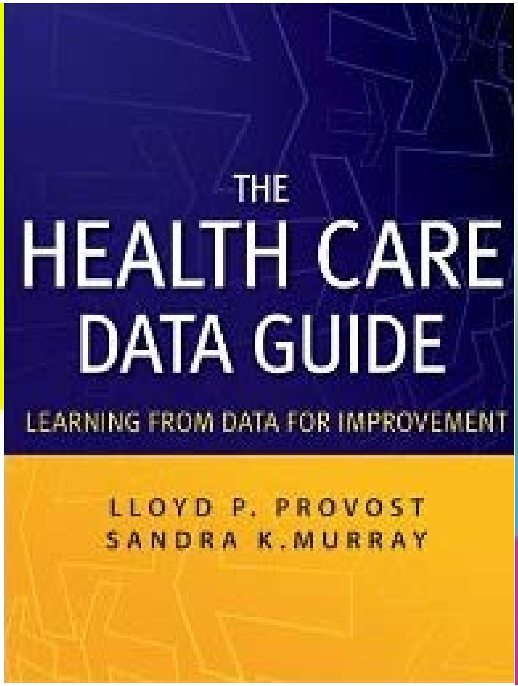
Real-time monitoring of coronary care mortality: A comparison and combination of two monitoring tools
Sherlaw-Johnson C, Morton A, Robinson MB, Hall A. *International Journal of Cardiology*, 100(2), 301-307, 20 April 2005

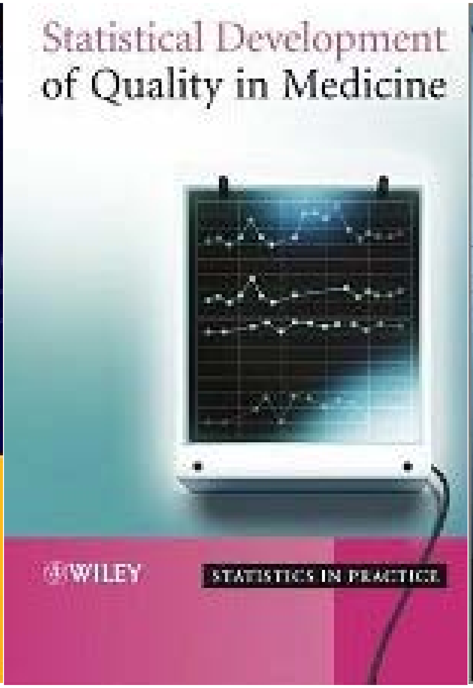
Rocket-tail chart

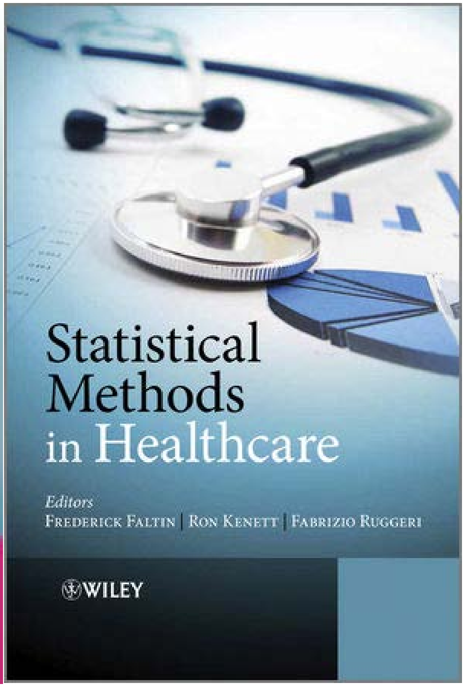


Recommended literature









Part II

Our research

Applied Stochastic Models in Business and Industry

Erratum

(wileyonlinelibrary.com) DOI: 10.1002/asmb.1982

Received 15 November 2010, Revised 22 April 2011, Accepted 25 April 2011 Published online 30 June 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/asmb.905

Business indicators of healthcare quality: Outlier detection in small samples

Gaj Vidmar[a,b*†], Rok Blagus[b], Luboš Střelec[c] and Milan Stehlík[d]

Special Issue Article

Quality and Reliability Engineering International

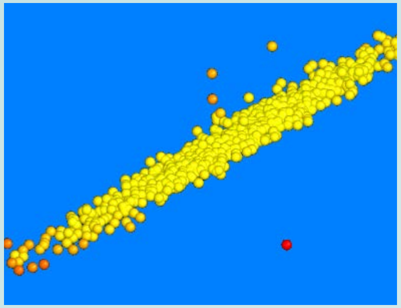
(wileyonlinelibrary.com) DOI: 10.1002/qre.1581

Published online in Wiley Online Library

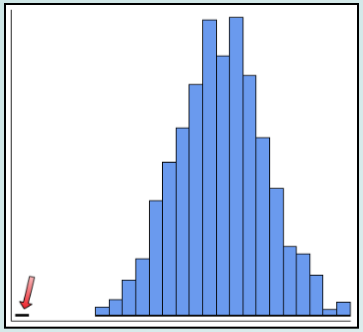
Outlier Detection for Healthcare Quality Monitoring – A Comparison of Four Approaches to Over-Dispersed Proportions

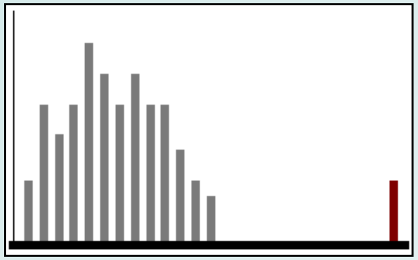
Gaj Vidmar[a,b*†] and Rok Blagus[b]

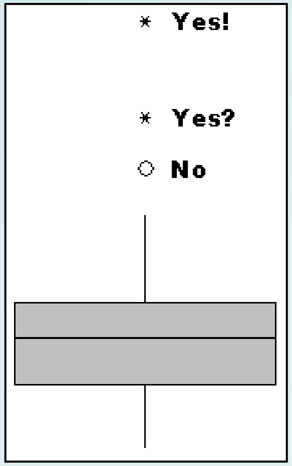
Background



vs.

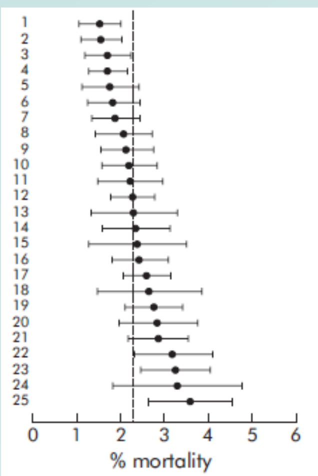




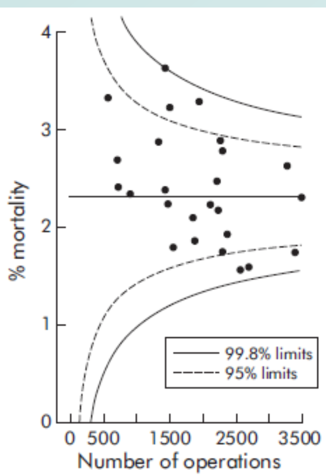


Background

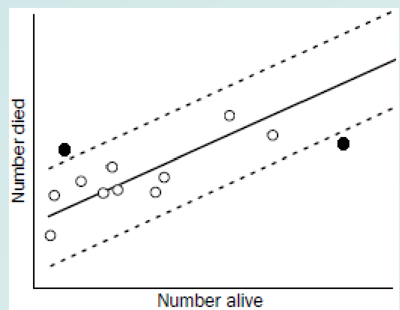
- Outlier detection in health care quality monitoring – milestones:
  - **UK inquiries** (Bristol cardiac surgery, Harold Shipman)
  - **Mohammed**: control charts instead of league tables
  - **Spiegelhalter**: funnel plots, regression modelling
  - **Laney**: p‘ chart for cross-sectional data

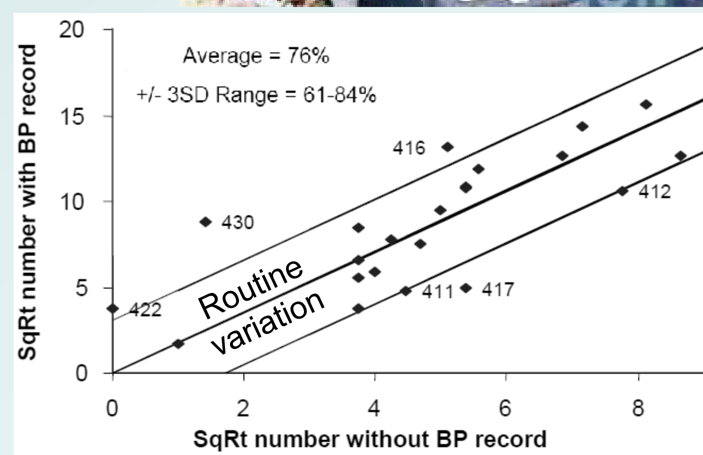


LEAGUE TABLE



FUNNEL PLOT





DOUBLE-SQUARE-ROOT CHART

Background

- QA activities of the Ministry of Health in Slovenia
  - since 2006
  - 100+ **business indicators** of
    - economy
    - efficiency
    - funding allocation
  - annually monitored in 26 hospitals
- Statistical analysis entrusted to the IBMI MF UL
  - essence is outlier identification
  - exploratory approach

Background

- All the indicators are same-quantity ratios (with highly correlated numerator and denominator) $\Rightarrow \in [0, 1]$ yet they are not appropriately treated
  - either as proportions (like e.g. success of operation) because of huge denominators (1000's of m^2, 10^6 €) $\Rightarrow$ **overdispersion**
    $\Leftrightarrow$ too narrow CI for mean π
    $\Leftrightarrow$ in funnel plots, virtually all points would be proclaimed outliers
  - or as fixed-denominator ratios (~ mortality rates)
- **Double Square Root Chart**:
  √numerator vs. √denominator–numerator

General-purpose outlier detection tests

- For small samples **assuming normality**
  - Grubbs (1969)

$$G = \frac{\max |Y_i - \bar{Y}|}{s}$$

$$G > \frac{(N-1)}{\sqrt{N}} \sqrt{\frac{t^2_{(\alpha/(2N),N-2)}}{N-2+t^2_{(\alpha/(2N),N-2)}}}$$

  - Dean & Dixon (1951)

$Q = r_{10} = \frac{x_2 - x_1}{x_n - x_1}$	$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$
$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$	$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$

  - Nalimov (1980)

$$q = \left| \frac{x_1 - \bar{x}}{s} \right| \sqrt{\frac{n}{n-1}}$$

- **Tukey‘s Boxplot rule**: >1.5 IQR from Q_1 or Q_3

Spiegelhalter‘s approach – Theory

Multiplicative (GLIM)

- overdispersion factor
- winsorizing

$$\overline{p} = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{k} n_i}$$

$$z_i = \frac{p_i - \overline{p}}{\frac{\overline{p}(1-\overline{p})}{n_i}}$$

10% winsorising

$$\phi^w = \frac{1}{N} \sum_{i=1}^{k} (z_i^w)^2$$

$$\begin{matrix} UCL = \overline{p} + \\ LCL = \overline{p} - \end{matrix} z_{\alpha/2} * \sqrt{\phi^w \frac{\overline{p}(1-\overline{p})}{n_i}}$$

Additive (random effects)

- method of moments estimator
  [DerSimonian & Laird, 1986 – for meta analysis]

$$s_i^2 = \frac{p_i(1-p_i)}{n_i}$$

$$w_i = \frac{1}{s_i^2}$$

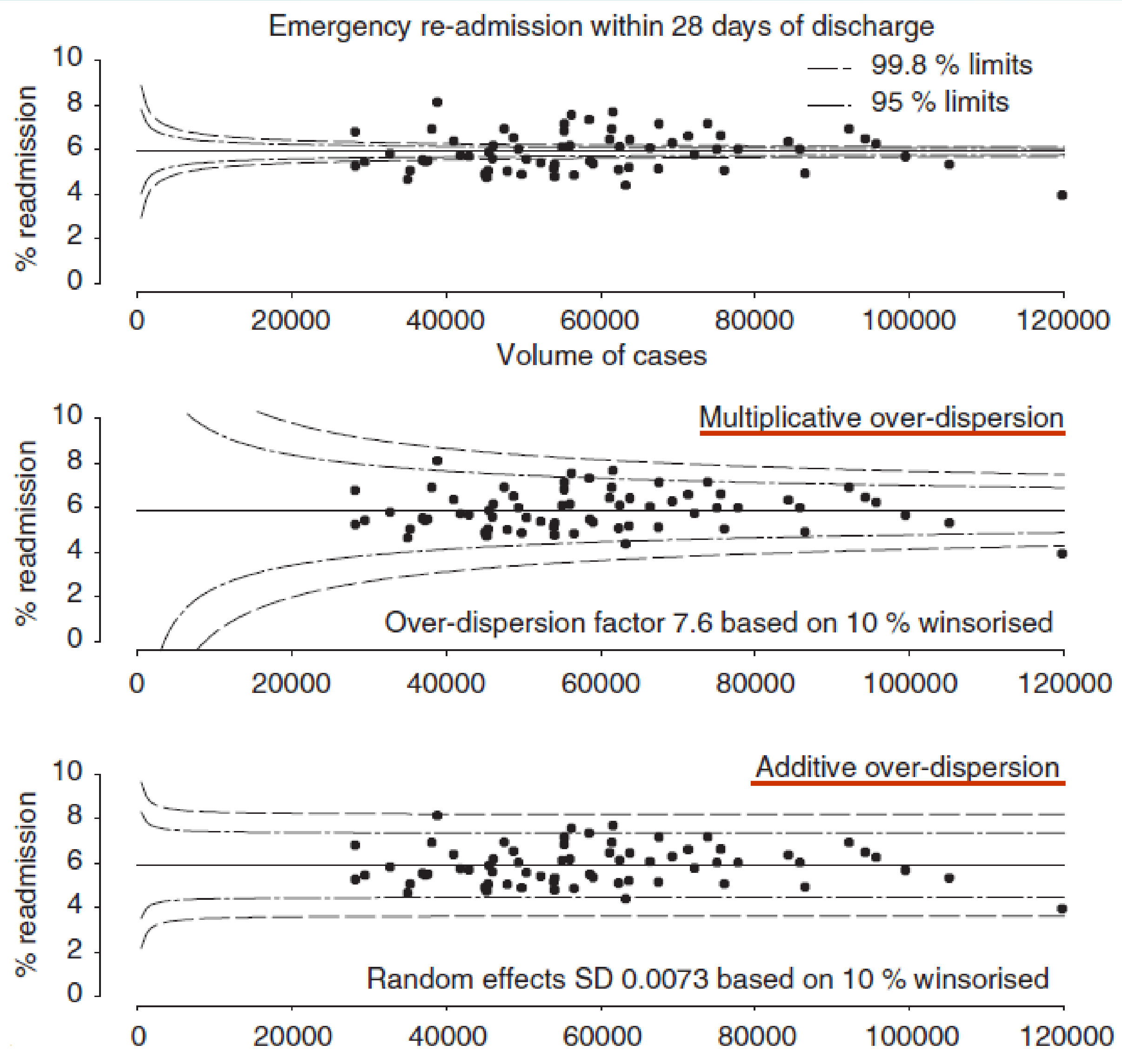
$$\text{if } \phi^w < \frac{N-1}{N} \text{ set } \hat{\tau} = 0$$

$$\text{otherwise } \quad \hat{\tau} = \frac{N \cdot \phi^w - (N-1)}{\sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}}$$

$$\begin{matrix} UCL = \overline{p} + \\ LCL = \overline{p} - \end{matrix} z_p * \sqrt{\frac{\overline{p}(1-\overline{p})}{n_i} + \hat{\tau}}$$

z_p set to 2.58 ⇔ $1 - \alpha = 99\%$

Spiegelhalter‘s approach – Examples



Laney‘s approach – p‘ chart

- Developed from [modified] *z*-chart for proportions & Wheeler‘s work
  - p-charts [& u-] too many false alarms – binomial [Poisson] assumption not true
  - parameter changes over time and/or between subgroups (in cross-sectional data)
  - like ANOVA: common cause **variation** = **within** (intra-subgroup) + **between**
  - becomes p-chart if data are binomial, XmR if subgroup sizes are equal

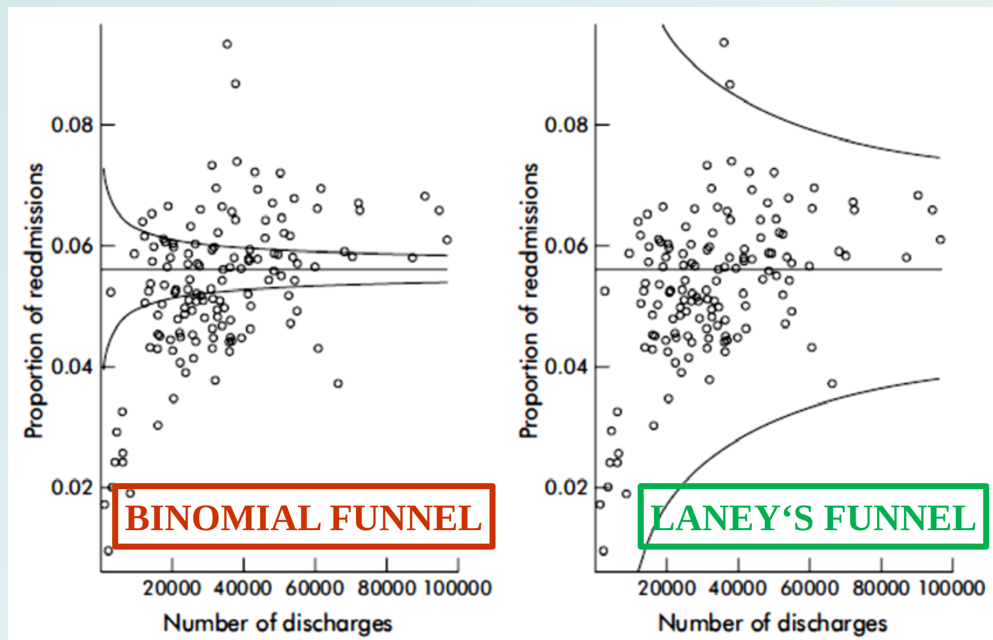
$$\sigma_{p_i} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}$$

$$z_i = \frac{p_i - \bar{p}}{\sigma_{p_i}}$$

$$p'_i = \bar{p} + \sigma_{p_i} z_i$$

$$\sigma_{p'_i} = \sigma_{p_i}\sigma_z$$

$$UCL(p'_i) = \bar{p} + \boxed{3}\sigma_{p_i}\sigma_z$$



- Many critical typos in Mohammed & Laney paper [QSHC 2006,15:383-4] ☹

Our approach

- **Double-square-root chart (DSQRTC)**
  - introduced by Shewhart, Fisher, Tukey, Mosteller in 1940's with CLs based on binomial probability paper
  - **slope (b) reflects binomial parameter (p)**: $p = \frac{b^2}{1+b^2}$
  - NEW: **control limits = prediction limits from regression through origin**
    - slightly curved (hourglass-like shaped – not parallel to the regression line)
    - in Study #1 we used $1 - \alpha = 95\%$
  - estimates LCL < 0 less often than Laney's or Spiegelhalter's ☺
  - **translates into asymmetric funnel** (alike exact binomial CI) ☺
  - estimates p as slightly less than $\bar{p}$ with right-skewed data ☺
- NEW in Study #2: **confidence level adjustment**: $1 - \alpha = \frac{N-0.5}{N}$
  - E (no. of outliers flagged under the null, i.e., binomial model) = 0.5
  - examples:

N	**$1 - \alpha$**
10	0.95
30	0.983
50	0.99
100	0.995
140	0.996

Median rule

- Also known as Carling's resistant rule [CSDA 2000,33:249-58]
  - **like boxplot rule with median instead of** q_1

$$c^{\mathrm{U}} = q_2 + k_2(q_3 - q_1)$$

  - compute quartiles using ideal (machine) fourth

$$q_1 = (1 - g)x_{(j)} + gx_{(j+1)}$$
$$j + g = n/4 + \frac{5}{12}$$

  - default k_2=2.3 with **adjustment for sample size**

$$k_2 = \frac{17.63n - 23.64}{7.74n - 3.71}$$

  - backed by profound theory and extensive simulations
  - successfully applied in neuroscience [J Neurosci Methods 2011,194:374-9]

Methods – Study #1

- Distributions of Numerator, Denominator & Ratio based on real data
  - right-skewed bounded, based on best fit (EasyFit by MathWave)
  - **Pert, Burr, 3P Loglogistic**
- **Outlier** generation
  - Mode (Pert) | Scale (Burr) | μ (3PLGL) ↑ by 50%, 100%, 150%, 500%
  - other parameters fixed (shape for Pert & Burr; ζ & σ for 3PLGL)
- 2 types of **ratios**
  - Small [0, 0.2]
  - Large [0.5, 1]
- 3 **situations**
  - no outliers
  - 1 outlier on the right
  - 2 outliers on the right
- **DSQRTC compared** with
  - Grubbs, Dean & Dixon and Nalimov test, all applied conditionally if Kolmogorov-Smirnov-Lilliefors & Shapiro-Wilk tests for normality passed
  - Tukey‘s boxplot rule

Methods – Study #2

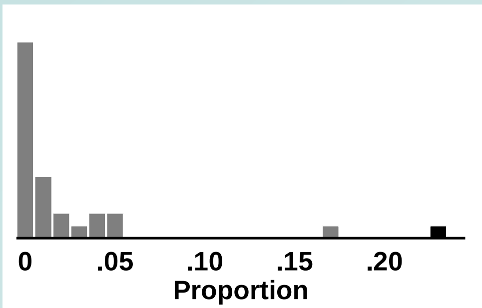
- **Real data**
  - **hospital readmissions from UK** presented by both ‚rivals' (N=140)
    - Spiegelhalter [QSHC 2002,11:390-1; Stat Med 2005,24:1185-202; QSHC 2005,14:347-51]
    - Mohammed & Laney [QSHC 2006,15:383-4]
  - **financial indicators** from Slovenian hospitals
- **Simulated data** (in R)
  - proportions
    - **small** [0, ~0.2] (right-skewed)
    - **large** [~0.5, 1] (more symmetrically distributed)
    - **3-parameter log-logistic distribution** with fixed scale (σ) and shape (ξ) parameter

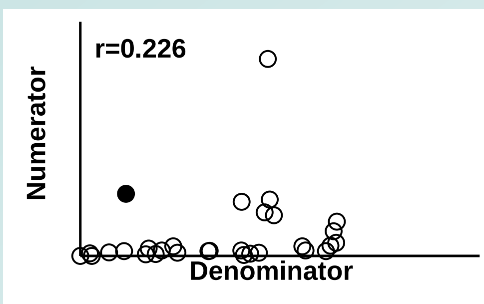
$$f_{3\text{Ploglog}}(x) = \frac{\left(1+\frac{\xi(x-\mu)}{\sigma}\right)^{-(1/\xi+1)}}{\sigma\left[1+\left(1+\frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}\right]^{2}}$$

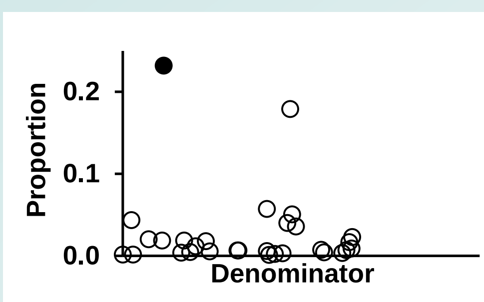
  - 1000 samples under each experimental condition {sample size, outlier}
    - ***N*=10, 30, 50, 100**
    - **no outlier | one outlier with μ×2 (at 100%) | one outlier with μ×6 (at 500%)** (outlier always the largest value – generated using rejection method)
  - **measures: 4 flags|counts & 3 diagnostic measures**
    - outlier found, overall result correct, no. of false outliers, total no. of outliers identified
    - predictive accuracy (PA), sensitivity (Sens), specificity (Spec)

Samples from simulations [N=30, outlier at 100%]

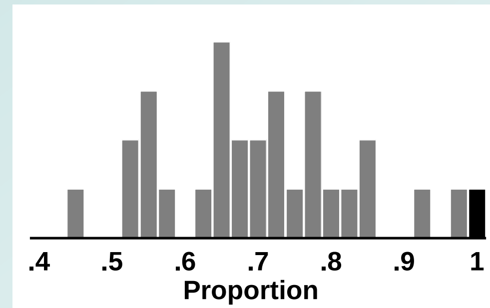
Small ratios

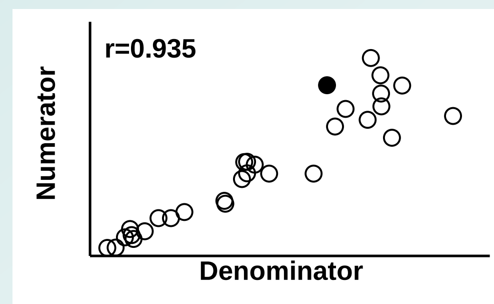


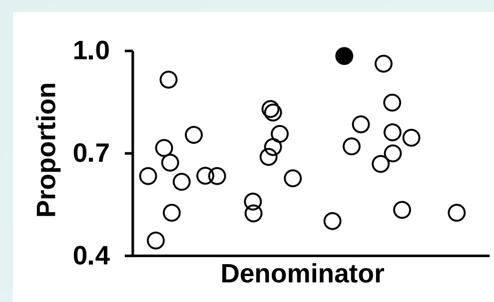




Large ratios







Results – Study #1

Under H_0

	N	5					10					20					25					30				
Ratio Type	Test	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK
Small	Grubbs		0,09	0	2	91%		0,11	0	3	89%		0,10	0	2	90%		0,08	0	2	92%		0,09	0	3	91%
	Dixon	86%	0,02	0	1	98%	69%	0,04	0	1	96%	55%	0,04	0	2	96%	51%	0,04	0	2	96%	49%	0,04	0	2	96%
	Nalimov		0,33	0	3	73%		0,70	0	5	52%		1,27	1	6	28%		1,50	1	6	23%		1,69	2	7	16%
	Boxplot		0,47	0	2	61%		0,60	0	4	55%		0,97	1	6	47%		1,22	1	6	43%		1,42	1	6	40%
	DSQRTC		0,00	0	0	100%		0,11	0	1	89%		0,71	1	3	35%		0,99	1	3	21%		1,26	1	4	12%
Large	Grubbs		0,06	0	2	94%		0,06	0	2	94%		0,07	0	2	93%		0,07	0	3	93%		0,07	0	3	94%
	Dixon	87%	0,04	0	2	96%	74%	0,07	0	2	93%	63%	0,04	0	3	96%	63%	0,04	0	3	96%	63%	0,03	0	2	97%
	Nalimov		0,23	0	3	81%		0,60	0	7	62%		1,89	1	12	29%		2,38	2	15	21%		3,07	3	21	14%
	Boxplot		0,35	0	2	73%		0,29	0	10	80%		0,32	0	8	81%		0,31	0	9	82%		0,34	0	9	80%
	DSQRTC		0,00	0	0	100%		0,10	0	1	90%		0,71	1	3	37%		0,96	1	4	24%		1,30	1	4	12%

1 out. to the right

	N	5					10					20					25					30				
Ratio Type	Test	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK
Small	Grubbs		0,17	0	2	15%		0,22	0	3	20%		0,29	0	2	27%		0,27	0	2	26%		0,28	0	2	27%
	Dixon	47%	0,05	0	2	4%	21%	0,07	0	2	6%	10%	0,10	0	2	10%	8%	0,09	0	2	9%	7%	0,09	0	1	9%
	Nalimov		0,49	0	3	29%		1,06	1	6	37%		1,93	2	7	32%		2,29	2	7	23%		2,47	2	6	21%
	Boxplot		0,76	1	2	60%		1,09	1	4	60%		1,62	1	7	49%		1,88	2	9	45%		2,09	2	8	42%
	DSQRTC		0,00	0	0	0%		0,48	0	1	47%		1,00	1	3	70%		1,18	1	3	64%		1,35	1	4	57%
Large	Grubbs		0,12	0	2	10%		0,24	0	3	20%		0,32	0	3	26%		0,33	0	3	27%		0,30	0	3	24%
	Dixon	78%	0,06	0	2	5%	62%	0,16	0	4	13%	52%	0,18	0	3	15%	52%	0,20	0	3	16%	52%	0,18	0	3	15%
	Nalimov		0,40	0	3	24%		1,02	1	7	29%		2,39	2	14	20%		3,01	3	14	16%		3,62	3	22	13%
	Boxplot		0,49	0	2	26%		0,57	0	10	25%		0,62	0	20	25%		0,66	0	25	24%		0,63	0	30	22%
	DSQRTC		0,00	0	0	0%		0,48	0	1	45%		0,89	1	3	57%		1,07	1	3	53%		1,28	1	4	45%

2 out. to the right

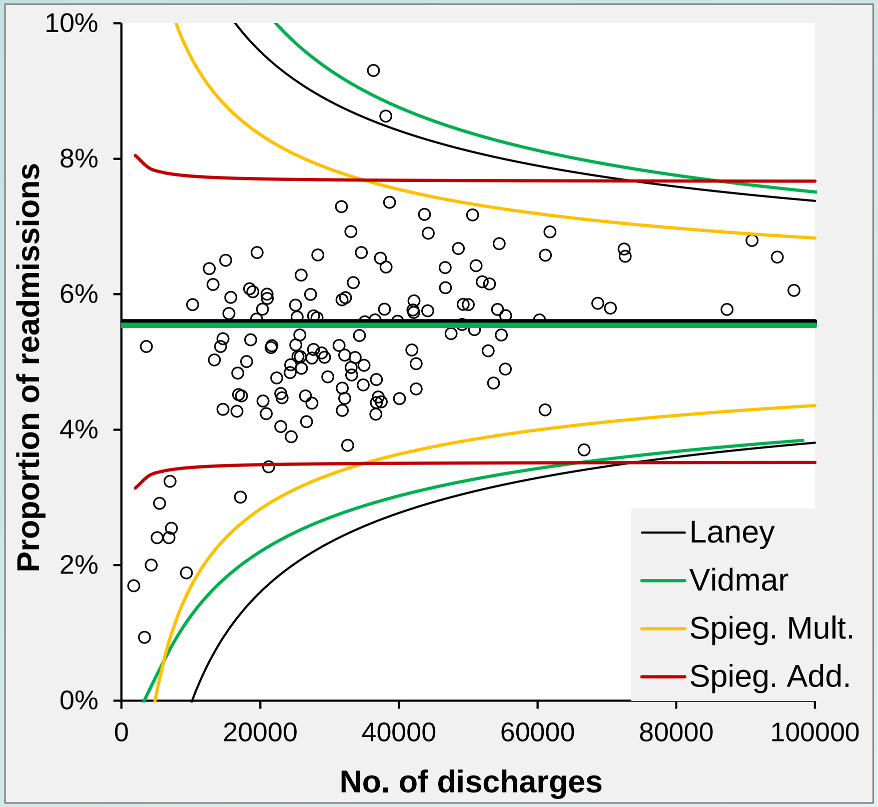
	N	5					10					20					25					30				
Ratio Type	Test	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK	Valid	M	Me	Max	OK
Small	Grubbs		0,10	0	2	0%		0,15	0	3	1%		0,24	0	2	1%		0,20	0	2	0%		0,24	0	2	0%
	Dixon	67%	0,03	0	2	0%	18%	0,03	0	2	0%	6%	0,05	0	2	0%	4%	0,06	0	2	0%	3%	0,05	0	1	0%
	Nalimov		0,42	0	3	8%		1,00	1	6	13%		2,07	2	7	21%		2,44	2	7	16%		2,63	2	7	18%
	Boxplot		0,43	0	2	0%		1,26	1	4	46%		2,21	2	6	49%		2,52	2	9	45%		2,75	2	8	43%
	DSQRTC		0,00	0	0	0%		0,32	0	1	0%		1,14	1	3	18%		1,40	1	4	29%		1,62	2	4	36%
Large	Grubbs		0,02	0	2	0%		0,02	0	4	0%		0,04	0	4	0%		0,04	0	2	0%		0,04	0	3	0%
	Dixon	72%	0,15	0	2	0%	58%	0,27	0	4	1%	45%	0,05	0	3	0%	43%	0,04	0	3	0%	45%	0,04	0	3	0%
	Nalimov		0,11	0	3	0%		0,39	0	6	4%		2,21	2	15	14%		3,04	3	15	13%		3,59	3	16	11%
	Boxplot		0,11	0	2	0%		0,61	0	10	15%		0,84	0	7	21%		0,89	0	9	20%		0,83	0	8	18%
	DSQRTC		0,00	0	0	0%		0,37	0	1	0%		1,11	1	3	25%		1,37	1	3	34%		1,59	2	4	37%

Results – Study #1

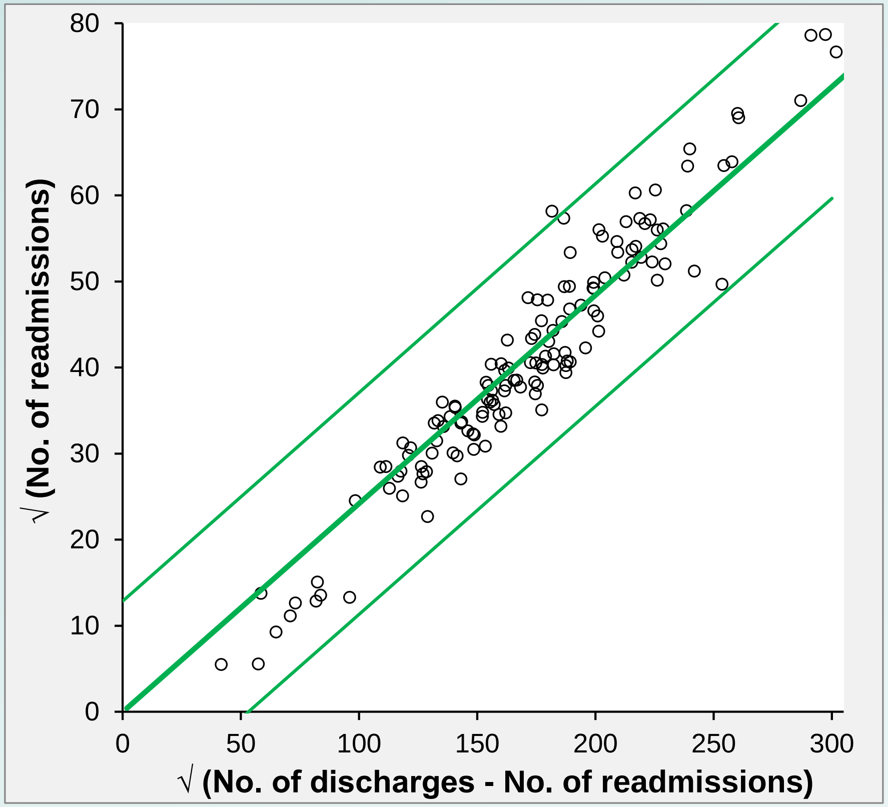
- ↑ outliers ⇒ **performance gets clearly worse**
- each approach excells/brakes down in certain conditions
- small-sample tests not applicable as sample size increases (sic!)
- simple boxplot is very unstable
- DSQRTC the best (though far from perfect) as N gets larger

Results S#2 – UK hospitals – Readmissions data

FUNNEL PLOTS



DOUBLE-SQUARE-ROOT CHART



Results S#2 – SI hospitals – Financial indicators

- With any real data, **the truth is unknown**
- **7 indicators (2 groups)**
  - 4 <large> of space utilisation (**S**): active area as part of
    1 total area, **2** inpatient clinic area, **3** outpatient clinic area, **4** dialysis unit area
  - 3 <small> of costs structure (**C**): part of total income for
    1 IT costs, **2** education costs, **3** energy costs
  - *N*=25 (except for S4: *N*=12)
- Performance of the tested methods **similar to the simulations**

No. of outliers	**Vidmar**	**Laney**	**Spiegelh. Mul.**	**Spiegelh. Add.**	**Median rule**
S1	0	0	1 (1↓)	3 (1↑,2↓)	1 (1↑)
S2	0	0	1 (1↓)	4 (3↑,1↓)	0
S3	0	0	1 (1↓)	8 (6↑,2↓)	0
S4 (*N*=12)	0	0	0	0	0
C1	1 (1↑)	1 (1↑)	2 (2↑)	4 (3↑,1↓)	3 (3↑)
C2	0	0	2 (2↑)	3 (2↑,1↓)	0
C3	0	0	3 (2↑,1↓)	7 (7↑,)	0

Results S#2 – Small ratios [0, ~0.2] – Flags|counts

Ratios	Outliers	Sample size	Vidmar					Laney					Spiegelhalter Multiplicative					Spiegelhalter Additive					Median rule					Measure
			Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	
Small [0,0.2]	One at 100%	10	25%	No	No			0%	No	No			64%	Yes	Yes			96%	Yes	Yes			68%	Yes	Yes			Outlier found
		30	68%	Yes	Yes			74%	Yes	Yes			93%	Yes	Yes			100%	Yes	Yes			99%	Yes	Yes			
		50	74%	Yes	Yes			85%	Yes	Yes			98%	Yes	Yes			100%	Yes	Yes			100%	Yes	Yes			
		100	82%	Yes	Yes			89%	Yes	Yes			98%	Yes	Yes			100%	Yes	Yes			100%	Yes	Yes			
	One at 500%	10	36%	No	No			1%	No	No			72%	Yes	Yes			98%	Yes	Yes			78%	Yes	Yes			
		30	75%	Yes	Yes			79%	Yes	Yes			95%	Yes	Yes			100%	Yes	Yes			99%	Yes	Yes			
		50	79%	Yes	Yes			87%	Yes	Yes			97%	Yes	Yes			100%	Yes	Yes			100%	Yes	Yes			
		100	85%	Yes	Yes			93%	Yes	Yes			99%	Yes	Yes			100%	Yes	Yes			100%	Yes	Yes			
	None	10	80%	Yes	Yes			100%	Yes	Yes			32%	No	No			7%	No	No			44%	No	No			Overall result correct
		30	25%	No	No			19%	No	No			2%	No	No			0%	No	No			7%	No	No			
		50	11%	No	No			4%	No	No			0%	No	No			0%	No	No			1%	No	No			
		100	1%	No	No			0%	No	No			0%	No	No			0%	No	No			0%	No	No			
	One at	10	25%	No	No			0%	No	No			63%	Yes	Yes			40%	No	No			42%	No	No			
		30	54%	Yes	Yes			65%	Yes	Yes			10%	No	No			0%	No	No			10%	No	No			
		50	33%	No	No			45%	No	No			1%	No	No			0%	No	No			2%	No	No			
		100	11%	No	No			16%	No	No			0%	No	No			0%	No	No			0%	No	No			
	One at 500%	10	36%	No	No			1%	No	No			71%	Yes	Yes			38%	No	No			52%	Yes	Yes			
		30	59%	Yes	Yes			70%	Yes	Yes			16%	No	No			0%	No	No			10%	No	No			
		50	37%	No	No			52%	Yes	Yes			2%	No	No			0%	No	No			1%	No	No			
		100	17%	No	No			21%	No	No			0%	No	No			0%	No	No			0%	No	No			
	None	10	0,2	0	0	1	0	0,0	0	0	1	0	0,7	1	1	2	0	1,7	1	1	10	0	0,7	1	1	2	0	Number of false outliers
		30	0,8	1	1	2	0	0,8	1	1	2	0	2,0	2	2	3	0	6,1	5	4	25	1	2,4	2	2	6	0	
		50	1,2	1	1	3	0	1,3	1	1	3	0	3,3	3	3	5	1	9,9	8	6	36	3	4,1	4	4	10	0	
		100	2,1	2	2	4	0	2,3	2	2	5	1	6,7	7	7	10	1	20,1	16	13	65	7	8,5	8	8	16	0	
	One at 100%	10	0,0	0	0	1	0	0,0	0	0	0	0	0,1	0	0	1	0	1,3	1	0	9	0	0,3	0	0	1	0	
		30	0,3	0	0	2	0	0,3	0	0	2	0	1,3	1	1	4	0	7,1	4	3	26	0	1,9	2	2	6	0	
		50	0,8	1	1	3	0	0,6	1	1	3	0	2,4	2	2	5	0	12,4	9	6	47	3	3,6	4	4	9	0	
		100	1,4	1	1	5	0	1,4	1	1	4	0	5,4	6	6	10	1	25,7	21	12	82	7	7,9	8	7	15	1	
	One at 500%	10	0,0	0	0	1	0	0,0	0	0	0	0	0,1	0	0	1	0	1,7	1	0	9	0	0,3	0	0	1	0	
		30	0,3	0	0	2	0	0,2	0	0	2	0	1,1	1	1	3	0	7,9	5	3	26	1	1,9	2	2	6	0	
		50	0,7	1	1	3	0	0,5	0	0	3	0	2,3	2	2	5	0	13,5	11	6	42	1	3,7	4	4	10	0	
		100	1,4	1	1	4	0	1,2	1	1	4	0	5,4	5	6	9	1	28,1	25	12	86	6	8,0	8	8	15	1	
	None	10	0,2	0	0	1	0	0,0	0	0	1	0	0,7	1	1	2	0	1,7	1	1	10	0	0,7	1	1	2	0	Total number of outliers flagged
		30	0,8	1	1	2	0	0,8	1	1	2	0	2,0	2	2	3	0	6,1	5	4	25	1	2,4	2	2	6	0	
		50	1,2	1	1	3	0	1,3	1	1	3	0	3,3	3	3	5	1	9,9	8	6	36	3	4,1	4	4	10	0	
		100	2,1	2	2	4	0	2,3	2	2	5	1	6,7	7	7	10	1	20,1	16	13	65	7	8,5	8	8	16	0	
	One at 100%	10	0,3	0	0	1	0	0,0	0	0	1	0	0,7	1	1	2	0	2,3	2	1	10	0	0,9	1	1	2	0	
		30	1,0	1	1	2	0	1,0	1	1	2	0	2,2	2	2	5	0	8,1	5	4	27	1	2,9	3	3	7	0	
		50	1,5	1	1	4	0	1,5	1	1	3	0	3,4	3	3	5	1	13,4	10	7	48	4	4,6	5	5	10	0	
		100	2,3	2	2	5	0	2,2	2	2	5	1	6,4	7	7	10	2	26,6	22	13	83	8	8,9	9	8	16	2	
	One at 500%	10	0,4	0	0	1	0	0,0	0	0	1	0	0,8	1	1	2	0	2,7	2	1	10	0	1,1	1	1	2	0	
		30	1,1	1	1	2	0	1,0	1	1	2	0	2,1	2	2	4	0	8,9	6	4	27	2	2,9	3	3	7	0	
		50	1,5	1	1	4	0	1,4	1	1	3	0	3,3	3	3	5	1	14,5	12	7	43	2	4,7	5	5	11	1	
		100	2,2	2	2	5	0	2,2	2	2	5	1	6,4	6	7	10	2	29,1	26	13	87	7	9,0	9	9	16	2	

Results S#2 – Small ratios – Diagnost. measures

Ratios	Outliers	Sample size	Vidmar					Laney					Spiegelhalter Multiplicative					Spiegelhalter Additive					Median rule					Measure
			Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	
Small [0,0.2]	None	10	0,98	1,00	1,00	1,00	0,90	1,00	1,00	1,00	1,00	0,90	0,93	0,90	0,90	1,00	0,80	0,83	0,90	0,90	1,00	0,00	0,93	0,90	0,90	1,00	0,80	Predictive accuracy
		30	0,97	0,97	0,97	1,00	0,93	0,97	0,97	0,97	1,00	0,93	0,93	0,93	0,93	1,00	0,90	0,80	0,83	0,87	0,97	0,17	0,92	0,93	0,93	1,00	0,80	
		50	0,98	0,98	0,98	1,00	0,94	0,97	0,98	0,98	1,00	0,94	0,93	0,94	0,92	0,98	0,90	0,80	0,84	0,86	0,94	0,28	0,92	0,92	0,92	1,00	0,80	
		100	0,98	0,98	0,98	1,00	0,96	0,98	0,98	0,98	0,99	0,95	0,93	0,93	0,93	0,99	0,90	0,80	0,84	0,87	0,93	0,35	0,92	0,92	0,92	1,00	0,84	
	One at 100%	10	0,92	0,90	0,90	1,00	0,80	0,90	0,90	0,90	1,00	0,90	0,96	1,00	1,00	1,00	0,80	0,86	0,90	1,00	1,00	0,10	0,94	0,90	0,90	1,00	0,90	
		30	0,98	1,00	1,00	1,00	0,90	0,98	1,00	1,00	1,00	0,90	0,96	0,97	0,97	1,00	0,87	0,76	0,87	0,90	1,00	0,13	0,94	0,93	0,93	1,00	0,80	
		50	0,98	0,98	0,98	1,00	0,92	0,98	0,98	1,00	1,00	0,92	0,95	0,96	0,96	1,00	0,88	0,75	0,82	0,88	0,94	0,06	0,93	0,92	0,92	1,00	0,82	
		100	0,98	0,99	0,99	1,00	0,94	0,99	0,99	0,99	1,00	0,95	0,95	0,94	0,94	0,99	0,89	0,74	0,79	0,88	0,93	0,18	0,92	0,92	0,93	0,99	0,85	
	One at 500%	10	0,93	0,90	0,90	1,00	0,80	0,90	0,90	0,90	1,00	0,90	0,96	1,00	1,00	1,00	0,80	0,82	0,90	1,00	1,00	0,10	0,95	1,00	1,00	1,00	0,90	
		30	0,98	1,00	1,00	1,00	0,90	0,99	1,00	1,00	1,00	0,90	0,96	0,97	0,97	1,00	0,87	0,74	0,83	0,90	0,97	0,13	0,93	0,93	0,93	1,00	0,80	
		50	0,98	0,98	0,98	1,00	0,92	0,99	1,00	1,00	1,00	0,92	0,95	0,96	0,96	1,00	0,88	0,73	0,78	0,88	0,98	0,16	0,93	0,92	0,92	1,00	0,80	
		100	0,98	0,99	0,99	1,00	0,95	0,99	0,99	0,99	1,00	0,95	0,95	0,95	0,94	0,99	0,90	0,72	0,75	0,88	0,94	0,14	0,92	0,92	0,92	0,99	0,85	
	One at 100%	10	0,25	0,00	0,00	1,00	0,00	0,00	0,00	0,00	1,00	0,00	0,64	1,00	1,00	1,00	0,00	0,96	1,00	1,00	1,00	0,00	0,68	1,00	1,00	1,00	0,00	Sensitivity
		30	0,68	1,00	1,00	1,00	0,00	0,74	1,00	1,00	1,00	0,00	0,93	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	0,00	0,99	1,00	1,00	1,00	0,00	
		50	0,74	1,00	1,00	1,00	0,00	0,85	1,00	1,00	1,00	0,00	0,98	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	0,00	
		100	0,82	1,00	1,00	1,00	0,00	0,89	1,00	1,00	1,00	0,00	0,98	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	
	One at 500%	10	0,36	0,00	0,00	1,00	0,00	0,01	0,00	0,00	1,00	0,00	0,72	1,00	1,00	1,00	0,00	0,98	1,00	1,00	1,00	0,00	0,78	1,00	1,00	1,00	0,00	
		30	0,75	1,00	1,00	1,00	0,00	0,79	1,00	1,00	1,00	0,00	0,95	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	0,00	
		50	0,79	1,00	1,00	1,00	0,00	0,87	1,00	1,00	1,00	0,00	0,97	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	
		100	0,85	1,00	1,00	1,00	0,00	0,93	1,00	1,00	1,00	0,00	0,99	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	
	None	10	0,98	1,00	1,00	1,00	0,90	1,00	1,00	1,00	1,00	0,90	0,93	0,90	0,90	1,00	0,80	0,83	0,90	0,90	1,00	0,00	0,93	0,90	0,90	1,00	0,80	Specificity
		30	0,97	0,97	0,97	1,00	0,93	0,97	0,97	0,97	1,00	0,93	0,93	0,93	0,93	1,00	0,90	0,80	0,83	0,87	0,97	0,17	0,92	0,93	0,93	1,00	0,80	
		50	0,98	0,98	0,98	1,00	0,94	0,97	0,98	0,98	1,00	0,94	0,93	0,94	0,92	0,98	0,90	0,80	0,84	0,86	0,94	0,28	0,92	0,92	0,92	1,00	0,80	
		100	0,98	0,98	0,98	1,00	0,96	0,98	0,98	0,98	0,99	0,95	0,93	0,93	0,93	0,99	0,90	0,80	0,84	0,87	0,93	0,35	0,92	0,92	0,92	1,00	0,84	
	One at 100%	10	1,00	1,00	1,00	1,00	0,89	1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	0,89	0,85	0,89	1,00	1,00	0,00	0,97	1,00	1,00	1,00	0,89	
		30	0,99	1,00	1,00	1,00	0,93	0,99	1,00	1,00	1,00	0,93	0,96	0,97	0,97	1,00	0,86	0,75	0,86	0,90	1,00	0,10	0,93	0,93	0,93	1,00	0,79	
		50	0,98	0,98	0,98	1,00	0,94	0,99	0,98	0,98	1,00	0,94	0,95	0,96	0,96	1,00	0,90	0,75	0,82	0,88	0,94	0,04	0,93	0,92	0,92	1,00	0,82	
		100	0,99	0,99	0,99	1,00	0,95	0,99	0,99	0,99	1,00	0,96	0,94	0,94	0,94	0,99	0,90	0,74	0,79	0,88	0,93	0,17	0,92	0,92	0,93	0,99	0,85	
	One at 500%	10	1,00	1,00	1,00	1,00	0,89	1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	0,89	0,81	0,89	1,00	1,00	0,00	0,97	1,00	1,00	1,00	0,89	
		30	0,99	1,00	1,00	1,00	0,93	0,99	1,00	1,00	1,00	0,93	0,96	0,97	0,97	1,00	0,90	0,73	0,83	0,90	0,97	0,10	0,93	0,93	0,93	1,00	0,79	
		50	0,99	0,98	0,98	1,00	0,94	0,99	1,00	1,00	1,00	0,94	0,95	0,96	0,96	1,00	0,90	0,72	0,78	0,88	0,98	0,14	0,92	0,92	0,92	1,00	0,80	
		100	0,99	0,99	0,99	1,00	0,96	0,99	0,99	0,99	1,00	0,96	0,95	0,95	0,94	0,99	0,91	0,72	0,75	0,88	0,94	0,13	0,92	0,92	0,92	0,99	0,85	

Results S#2 – Large ratios [~0.5, 1] – Flags|counts

Ratios	Outliers	Sample size	Vidmar					Laney					Spiegelhalter Multiplicative					Spiegelhalter Additive					Median rule					Measure
			Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	Mean	Median	Mode	Max	Min	
Large [0.5,1]	One at 100%	10	55%	Yes	Yes			0%	No	No			36%	No	No			96%	Yes	Yes			10%	No	No			Outlier found
		30	61%	Yes	Yes			3%	No	No			50%	No	No			100%	Yes	Yes			5%	No	No			
		50	57%	Yes	Yes			5%	No	No			50%	Yes	Yes			100%	Yes	Yes			2%	No	No			
		100	58%	Yes	Yes			3%	No	No			53%	Yes	Yes			100%	Yes	Yes			0%	No	No			
	One at 500%	10	54%	Yes	Yes			0%	No	No			38%	No	No			97%	Yes	Yes			9%	No	No			
		30	61%	Yes	Yes			5%	No	No			47%	No	No			100%	Yes	Yes			5%	No	No			
		50	59%	Yes	Yes			6%	No	No			53%	Yes	Yes			100%	Yes	Yes			3%	No	No			
		100	56%	Yes	Yes			3%	No	No			55%	Yes	Yes			100%	Yes	Yes			0%	No	No			
	None	10	61%	Yes	Yes			100%	Yes	Yes			55%	Yes	Yes			41%	No	No			92%	Yes	Yes			Overall result correct
		30	18%	No	No			94%	Yes	Yes			18%	No	No			7%	No	No			94%	Yes	Yes			
		50	6%	No	No			88%	Yes	Yes			6%	No	No			1%	No	No			97%	Yes	Yes			
		100	1%	No	No			80%	Yes	Yes			1%	No	No			0%	No	No			99%	Yes	Yes			
	One at 100%	10	55%	Yes	Yes			0%	No	No			25%	No	No			23%	No	No			7%	No	No			
		30	37%	No	No			3%	No	No			14%	No	No			0%	No	No			2%	No	No			
		50	17%	No	No			4%	No	No			5%	No	No			0%	No	No			0%	No	No			
		100	3%	No	No			2%	No	No			0%	No	No			0%	No	No			0%	No	No			
	One at	10	54%	Yes	Yes			0%	No	No			27%	No	No			19%	No	No			7%	No	No			
		30	39%	No	No			5%	No	No			13%	No	No			0%	No	No			1%	No	No			
		50	16%	No	No			5%	No	No			4%	No	No			0%	No	No			1%	No	No			
		100	3%	No	No			3%	No	No			1%	No	No			0%	No	No			0%	No	No			
	None	10	0,4	0	0	1	0	0,0	0	0	0	0	0,6	0	0	2	0	1,4	1	0	10	0	0,1	0	0	3	0	Number of false outliers
		30	1,0	1	1	3	0	0,1	0	0	1	0	1,6	1	1	6	0	11,2	6	6	30	0	0,1	0	0	5	0	
		50	1,6	2	2	4	0	0,1	0	0	2	0	2,7	3	3	8	0	27,9	41	44	50	0	0,1	0	0	7	0	
		100	2,6	3	2	5	0	0,2	0	0	3	0	5,4	5	4	12	0	81,6	88	88	97	3	0,0	0	0	3	0	
	One at 100%	10	0,1	0	0	1	0	0,0	0	0	0	0	0,3	0	0	2	0	3,3	2	1	9	0	0,0	0	0	2	0	
		30	0,6	1	1	3	0	0,0	0	0	1	0	1,1	1	1	5	0	18,8	24	26	29	0	0,1	0	0	4	0	
		50	1,1	1	1	3	0	0,1	0	0	1	0	2,2	2	2	7	0	38,1	43	44	48	3	0,0	0	0	4	0	
		100	2,2	2	2	5	0	0,2	0	0	2	0	4,9	5	5	12	0	85,4	87	87	94	17	0,0	0	0	4	0	
	One at 500%	10	0,1	0	0	1	0	0,0	0	0	0	0	0,2	0	0	2	0	3,3	2	1	9	0	0,0	0	0	1	0	
		30	0,6	1	0	3	0	0,0	0	0	1	0	1,2	1	1	5	0	19,0	25	26	29	0	0,1	0	0	3	0	
		50	1,2	1	1	4	0	0,0	0	0	2	0	2,3	2	2	8	0	38,3	43	43	48	2	0,1	0	0	5	0	
		100	2,2	2	2	6	0	0,2	0	0	2	0	4,9	5	5	12	0	85,5	87	87	96	14	0,0	0	0	7	0	
	None	10	0,4	0	0	1	0	0,0	0	0	0	0	0,6	0	0	2	0	1,4	1	0	10	0	0,1	0	0	3	0	Total number of outliers flagged
		30	1,0	1	1	3	0	0,1	0	0	1	0	1,6	1	1	6	0	11,2	6	6	30	0	0,1	0	0	5	0	
		50	1,6	2	2	4	0	0,1	0	0	2	0	2,7	3	3	8	0	27,9	41	44	50	0	0,1	0	0	7	0	
		100	2,6	3	2	5	0	0,2	0	0	3	0	5,4	5	4	12	0	81,6	88	88	97	3	0,0	0	0	3	0	
	One at 100%	10	0,6	1	1	1	0	0,0	0	0	0	0	0,6	0	0	2	0	4,2	3	1	10	0	0,1	0	0	3	0	
		30	1,2	1	1	3	0	0,0	0	0	1	0	1,6	1	1	6	0	19,8	25	27	30	1	0,1	0	0	5	0	
		50	1,7	2	2	4	0	0,1	0	0	2	0	2,7	3	2	8	0	39,1	44	45	49	4	0,1	0	0	5	0	
		100	2,8	3	3	6	0	0,2	0	0	2	0	5,5	5	5	13	0	86,4	88	88	95	18	0,0	0	0	5	0	
	One at 500%	10	0,6	1	1	1	0	0,0	0	0	0	0	0,6	0	0	2	0	4,3	3	2	10	0	0,1	0	0	2	0	
		30	1,2	1	1	3	0	0,1	0	0	1	0	1,6	1	1	6	0	20,0	26	27	30	1	0,1	0	0	4	0	
		50	1,8	2	2	4	0	0,1	0	0	2	0	2,8	3	2	9	0	39,3	44	44	49	3	0,1	0	0	6	0	
		100	2,8	3	3	6	0	0,2	0	0	3	0	5,5	5	5	13	0	86,5	88	88	97	15	0,0	0	0	8	0	

Results S#2 – Large ratios – Diagnost. measures

Ratios	Outliers	Sample size	Vidmar Mean	Vidmar Median	Vidmar Mode	Vidmar Max	Vidmar Min	Laney Mean	Laney Median	Laney Mode	Laney Max	Laney Min	Spiegelhalter Multiplicative Mean	Spiegelhalter Multiplicative Median	Spiegelhalter Multiplicative Mode	Spiegelhalter Multiplicative Max	Spiegelhalter Multiplicative Min	Spiegelhalter Additive Mean	Spiegelhalter Additive Median	Spiegelhalter Additive Mode	Spiegelhalter Additive Max	Spiegelhalter Additive Min	Median rule Mean	Median rule Median	Median rule Mode	Median rule Max	Median rule Min	Measure
Large [0.5,1]	None	10	0,96	1,00	1,00	1,00	0,90	1,00	1,00	1,00	1,00	1,00	0,94	1,00	1,00	1,00	0,80	0,86	0,90	1,00	1,00	0,00	0,99	1,00	1,00	1,00	0,70	Predictive accuracy
		30	0,97	0,97	0,97	1,00	0,90	1,00	1,00	1,00	1,00	0,97	0,95	0,97	0,97	1,00	0,80	0,63	0,80	0,80	1,00	0,00	1,00	1,00	1,00	1,00	0,83	
		50	0,97	0,96	0,96	1,00	0,92	1,00	1,00	1,00	1,00	0,96	0,95	0,94	0,94	1,00	0,84	0,44	0,18	0,12	1,00	0,00	1,00	1,00	1,00	1,00	0,86	
		100	0,97	0,97	0,98	1,00	0,95	1,00	1,00	1,00	1,00	0,97	0,95	0,95	0,96	1,00	0,88	0,18	0,12	0,12	0,97	0,03	1,00	1,00	1,00	1,00	0,97	
	One at 100%	10	0,95	1,00	1,00	1,00	0,80	0,90	0,90	0,90	0,90	0,90	0,91	0,90	0,90	1,00	0,70	0,67	0,80	0,90	1,00	0,10	0,91	0,90	0,90	1,00	0,80	
		30	0,97	0,97	1,00	1,00	0,87	0,97	0,97	0,97	1,00	0,93	0,95	0,97	0,97	1,00	0,80	0,37	0,20	0,13	1,00	0,03	0,97	0,97	0,97	1,00	0,87	
		50	0,97	0,98	0,98	1,00	0,92	0,98	0,98	0,98	1,00	0,96	0,95	0,94	0,96	1,00	0,84	0,24	0,14	0,12	0,94	0,04	0,98	0,98	0,98	1,00	0,92	
		100	0,97	0,97	0,97	1,00	0,94	0,99	0,99	0,99	1,00	0,97	0,95	0,95	0,94	1,00	0,88	0,15	0,13	0,13	0,83	0,06	0,99	0,99	0,99	0,99	0,96	
	One at 500%	10	0,94	1,00	1,00	1,00	0,80	0,90	0,90	0,90	0,90	0,90	0,91	0,90	0,90	1,00	0,70	0,66	0,80	0,90	1,00	0,10	0,91	0,90	0,90	1,00	0,80	
		30	0,97	0,97	1,00	1,00	0,87	0,97	0,97	0,97	1,00	0,93	0,94	0,93	0,97	1,00	0,80	0,37	0,17	0,13	1,00	0,03	0,97	0,97	0,97	1,00	0,90	
		50	0,97	0,98	0,98	1,00	0,90	0,98	0,98	0,98	1,00	0,94	0,95	0,94	0,94	1,00	0,84	0,23	0,14	0,14	0,96	0,04	0,98	0,98	0,98	1,00	0,90	
		100	0,97	0,97	0,98	1,00	0,93	0,99	0,99	0,99	1,00	0,97	0,95	0,95	0,95	1,00	0,88	0,15	0,13	0,13	0,86	0,04	0,99	0,99	0,99	0,99	0,93	
	One at 100%	10	0,55	1,00	1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,36	0,00	0,00	1,00	0,00	0,96	1,00	1,00	1,00	0,00	0,10	0,00	0,00	1,00	0,00	Sensitivity
		30	0,61	1,00	1,00	1,00	0,00	0,03	0,00	0,00	1,00	0,00	0,50	0,00	0,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	0,05	0,00	0,00	1,00	0,00	
		50	0,57	1,00	1,00	1,00	0,00	0,05	0,00	0,00	1,00	0,00	0,50	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	1,00	0,02	0,00	0,00	1,00	0,00	
		100	0,58	1,00	1,00	1,00	0,00	0,03	0,00	0,00	1,00	0,00	0,53	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	0,00	0,00	0,00	0,00	1,00	0,00	
	One at 500%	10	0,54	1,00	1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,38	0,00	0,00	1,00	0,00	0,97	1,00	1,00	1,00	0,00	0,09	0,00	0,00	1,00	0,00	
		30	0,61	1,00	1,00	1,00	0,00	0,05	0,00	0,00	1,00	0,00	0,47	0,00	0,00	1,00	0,00	1,00	1,00	1,00	1,00	0,00	0,05	0,00	0,00	1,00	0,00	
		50	0,59	1,00	1,00	1,00	0,00	0,06	0,00	0,00	1,00	0,00	0,53	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	0,00	0,03	0,00	0,00	1,00	0,00	
		100	0,56	1,00	1,00	1,00	0,00	0,03	0,00	0,00	1,00	0,00	0,55	1,00	1,00	1,00	0,00	1,00	1,00	1,00	1,00	0,00	0,00	0,00	0,00	1,00	0,00	
	None	10	0,96	1,00	1,00	1,00	0,90	1,00	1,00	1,00	1,00	1,00	0,94	1,00	1,00	1,00	0,80	0,86	0,90	1,00	1,00	0,00	0,99	1,00	1,00	1,00	0,70	Specificity
		30	0,97	0,97	0,97	1,00	0,90	1,00	1,00	1,00	1,00	0,97	0,95	0,97	0,97	1,00	0,80	0,63	0,80	0,80	1,00	0,00	1,00	1,00	1,00	1,00	0,83	
		50	0,97	0,96	0,96	1,00	0,92	1,00	1,00	1,00	1,00	0,96	0,95	0,94	0,94	1,00	0,84	0,44	0,18	0,12	1,00	0,00	1,00	1,00	1,00	1,00	0,86	
		100	0,97	0,97	0,98	1,00	0,95	1,00	1,00	1,00	1,00	0,97	0,95	0,95	0,96	1,00	0,88	0,18	0,12	0,12	0,97	0,03	1,00	1,00	1,00	1,00	0,97	
	One at 100%	10	0,99	1,00	1,00	1,00	0,89	1,00	1,00	1,00	1,00	1,00	0,97	1,00	1,00	1,00	0,78	0,64	0,78	0,89	1,00	0,00	1,00	1,00	1,00	1,00	0,78	
		30	0,98	0,97	0,97	1,00	0,90	1,00	1,00	1,00	1,00	0,97	0,96	0,97	0,97	1,00	0,83	0,35	0,17	0,10	1,00	0,00	1,00	1,00	1,00	1,00	0,86	
		50	0,98	0,98	0,98	1,00	0,94	1,00	1,00	1,00	1,00	0,98	0,95	0,96	0,96	1,00	0,86	0,22	0,12	0,10	0,94	0,02	1,00	1,00	1,00	1,00	0,92	
		100	0,98	0,98	0,98	1,00	0,95	1,00	1,00	1,00	1,00	0,98	0,95	0,95	0,95	1,00	0,88	0,14	0,12	0,12	0,83	0,05	1,00	1,00	1,00	1,00	0,96	
	One at 500%	10	0,99	1,00	1,00	1,00	0,89	1,00	1,00	1,00	1,00	1,00	0,97	1,00	1,00	1,00	0,78	0,63	0,78	0,89	1,00	0,00	1,00	1,00	1,00	1,00	0,89	
		30	0,98	0,97	1,00	1,00	0,90	1,00	1,00	1,00	1,00	0,97	0,96	0,97	0,97	1,00	0,83	0,34	0,14	0,10	1,00	0,00	1,00	1,00	1,00	1,00	0,90	
		50	0,98	0,98	0,98	1,00	0,92	1,00	1,00	1,00	1,00	0,96	0,95	0,96	0,96	1,00	0,84	0,22	0,12	0,12	0,96	0,02	1,00	1,00	1,00	1,00	0,90	
		100	0,98	0,98	0,98	1,00	0,94	1,00	1,00	1,00	1,00	0,98	0,95	0,95	0,95	1,00	0,88	0,14	0,12	0,12	0,86	0,03	1,00	1,00	1,00	1,00	0,93	

Discussion

- **Laney**
  - fewest false alarms, but could not detect the outlier if N=10 with small p
  - failed to detect the outlier regardless of sample size with large p
- **Spiegelhalter**
  - very high false-alarm rates (additive unusable), except multiplicative if N=10
  - additive fails if p_i=1 or p_i=0 (division by zero) **!?**
    and plugging in 0.99 or 0.9999 (0.01 or 0.0001) makes big difference **!?**
    (while winsorising p's is not a solution in that case:
    those points may be outliers plus it would still fail if >10% 0's or 1's) **!?**
  - additive made it to the UK legislation (NHS) **!?**
- **Median rule**
  - better than Laney with small p if N=10
  - higher false-alarm rate if N=30+
- **Vidmar**
  - with small p joint overall best with Laney but not insensitive if N=10 ☺
  - best overall for large p (followed by Spiegelhalter multiplicative) ☺
  - **clearly the best overall** ☺

Conclusions

- We followed the reasoning of Mohammed and Laney that "*it may be reasonable to undertake desktop comparisons of these different statistical approaches by using existing or simulated data sets*"
- We abide by their guideline that "*the basic aim of SPC is to support continual improvement and not the construction of optimum statistical models*"
- Being (apparently) effective and simple, our approach appears to be useful for the purpose of continual improvement, which is the only answer that we can give to dismissing it as *ad hoc* and lacking a profound theoretical basis
- Our approach follows the spirit of Shewhart by being simple; it also follows the Bayesian spirit by adjusting the α-level to sample size
- A possible critique of our studies is that we could have used other simulation setups (different distributions, outliers simulated in different ways, including more outliers than just one, etc.), but there are infinite possibilities for simulations

Further work

- Theoretical relations between our, Laney's and Spiegelhalter's funnel-plot behaviour (at least under the binomial model)
- Our approach is not "tuned" in the sense of Spiegelhalter's, where winsorising is applied to eliminate (or at least reduce) the influence of potential outliers on the estimates
  - winsorising before fitting the regression line
  - robust regression methods
  - most of such procedures would compromise simplicity
  - two other approaches can also be tuned: Laney's approach by replacing the traditional Shewhart's multiplier 3 with $z_{1-\alpha}$ (with α possibly adjusted for sample size), and the median rule by changing the factor a
- Other alternatives
  - robust control charts based on diagnostic statistics
  - control charts for overdispersed data based on the COM-Poisson distribution
  - bootstrap tolerance intervals

Limitations

- We did not address longitudinal monitoring
- Our approach fits into the general frameworks devised by Spiegelhalter and associates
  - Hierarchical modelling framework for identifying unusual performance in health care providers (Ohlssen, Sharples & Spiegelhalter, *JRSSA*, 2007) – **3 stages**
    - Exploratory ⬅
    - Modelling
    - Confirmatory
  - Comprehensive system of statistical strategies for healthcare rating, screening and surveillance (Spiegelhalter *et al.*, *JRSSA*, 2012) – **3 regulatory functions**
    - Rating organizations
    - Screening to target inspections ⬅
    - Continuous surveillance for arising problems
  - our approach could be extended along the time dimension by means of mixed (i.e., multi-level) models

Wider issues in SPC

- What is an outlier?!
- "Technical fixes" for overdispersion vs. why it exists?!
- It is of major benefit for wide adoption of a statistical method if – in addition to theoretical soundness and good performance in realistic simulations – it is simple and thus understandable to a wide audience
  - implemented in Excel without VBA macros
  - R code used for simulations