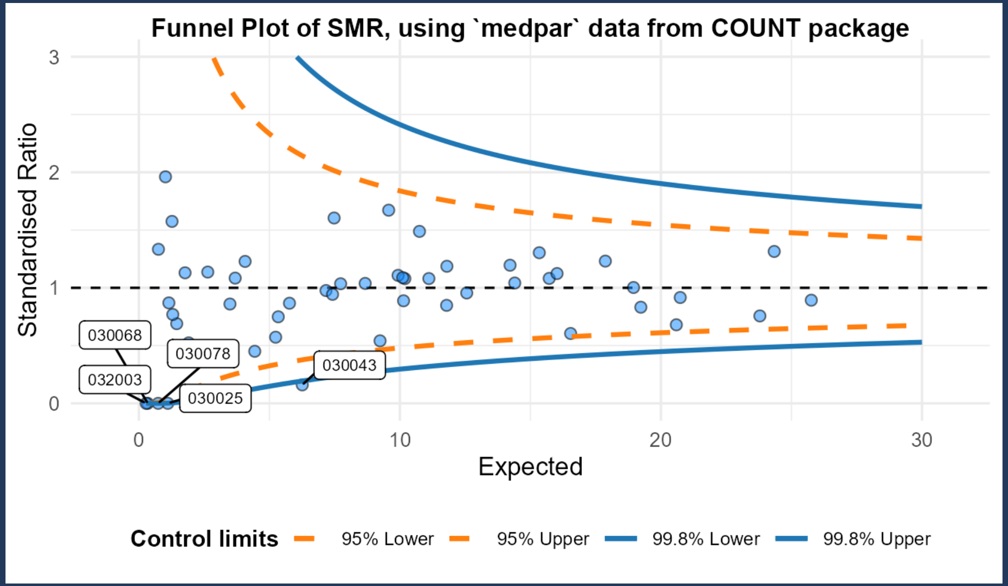
Understanding Standardised Mortality Ratios (SMRs)

SHMI and HSMR



Dr Chris Mainey

@chrismainey
chrismainey
chrismainey
0000-0002-3018-6171
www.mainard.co.uk

Presentation and code available: **https://github.com/chrismainey/understanding_standardised_mortality**

Measuring death

Why do we do it?

- 'Smoke alarm for the quality of care' - (Keogh, 2013)
- Tacit assumption that high mortality is bad

Does it work?

- **Yes:**
  - Increases power of comparison by reducing confounding.
  - Monitoring them has effects on hospital's vigilance and culture (Jarman, Bottle, Aylin, et al., 2005; Wright, Dugdale, Hammond, et al., 2006)
- **No:**
  - Poor proxy of avoidable death: (Girling, Hofer, Wu, et al., 2012)
  - Case-mix adjustment can exaggerate biases it tries to address: (Deeks, Dinnes, D'Amico, et al., 2003)

Desceptive argument:

"...of course we should be monitoring deaths": - but is this too simplistic a view? Is it a measure of 'quality?'

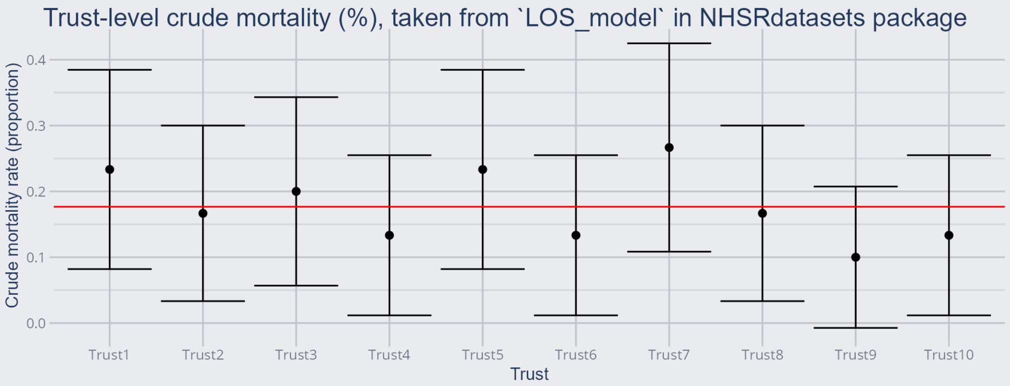
Crude Mortality

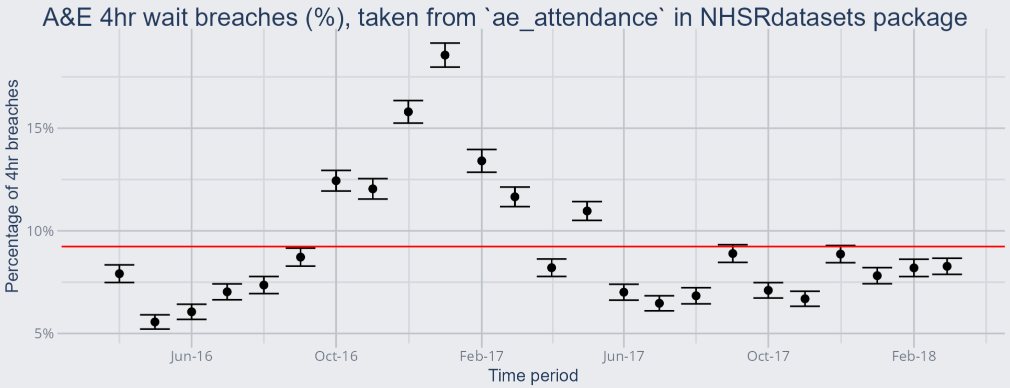
Crude mortality Rate

- Count the numbers of deaths? - Not a rate
- Give it some sort of scale: proportion
  - In hospital, patients, discharges, bed-days etc.
  - Often adjusted to a larger standard, e.g. per thousand

$$Crude\ Rate(p) = \frac{\Sigma Deaths}{n}$$

- **Strengths:**
  - Easy to calculate
  - Directly linked to real deaths
- **Weaknesses:**
  - Not really comparable across organisations
  - Case-mix confounds rate





Standardising mortality

Aim: reduce confounding and increase power of comparison

Direct standardisation

- Take our data and map them to a common population/structure.

- Example: Age-standardisation:

  - Calculate age-specific rates in groups (e.g. 10-year bands)
  - Identify a relevant standard population in corresponding groups. E.g. European Standard Population
  - Multiply age-specific rates by standard population bins
  - Sum and adjust to desired multiplier (e.g. per 100, per 100,000 etc.)

- Commonly used in public health cases, such as cancer incidence and mortality rates.

- **Strengths:**
  - Directly comparable between units/group/sites/countries
  - Does not require statistical model

- **Weaknesses:**
  - Harder to relate to local/observed numbers
  - Challenging to do for anything more than age and sex

Indirect standardisation

- Compare our data to expected averages
- E.g Calculate average rate, per-patient, across the dataset
    - Per trust, calculate 'expected rate': $average\ rate * n$
    - Present as grouped ratios of Observed / Expected
- Commonly uses a regression model
    - Predict risk of event, based on case-mix factors (predictors)
- **Strengths:**
    - Directly comparable between units/group/sites/countries
    - Usually require statistical model, e.g. regression
- **Weaknesses:**
    - Usually requires a statistical model, e.g. regression to calculate 'expected rate'
    - Susceptible to more forms of bias (Iezzoni, 1997; Deeks, Dinnes, D'Amico, et al., 2003)
    - Can be challenging to understand what changes in rates mean

Example SMR:

```r
# Load 'medpar' dataset from COUNT package.
data("medpar")

# build logistic regression for risk of death
mod1 <- glm(died ~ los + factor(type) + age80
            , data=medpar
            , family = "binomial")

# Predict risk of death back into data frame.
medpar$pred <- predict(mod1, type="response")

# SMRs
medpar %>%
  group_by(provnum) %>%
  summarise(Observered = sum(died),
            Expected = sum(pred),
            SMR = sum(died) / sum(pred))
```

```
## # A tibble: 54 x 4
##    provnum    Observered Expected   SMR
##    <labelled>      <int>    <dbl> <dbl>
##  1 030001             16    19.2  0.832
##  2 030002             19    20.7  0.916
##  3 030003              2     1.77 1.13
##  4 030006             23    25.8  0.893
##  5 030007              2     4.44 0.451
##  6 030008              5     9.24 0.541
##  7 030009              4     5.34 0.749
##  8 030010             22    17.9  1.23
##  9 030011             12    12.6  0.955
## 10 030012             12     7.48 1.60
## # ... with 44 more rows
```

SMR = 1: observed=predicted, >1: observed>predicted, <1: observed <predicted

Common (indirectly standardised) SMRs:

Summary Hospital-level Mortality Indicator (SHMI)

Dr Foster Hospital Standardised Mortality Ratio (HSMR)

Hospital Standardised Mortality Ratio (HSMR)

- Work in USA as early as 1970s demonstrated ability to calculate theses metrics.
- Prof. Sir Brian Jarman and others adapted these methods to English health care system, data coding standards and structures.
- Methods were heavily impacted and applied in aftermath of Bristol and Mid-Staffs enquiry.
- Controversy on some issues:
    - Commercial exploitation of method by Dr Foster Intelligence - accused of 'black box methods'
    - University of Birmingham and others published criticism
    - Imperial rebutted criticism, and won NIHR funding to build national monitoring system
    - Until recently, sent alerts to CQC and trusts for high mortality.

Key References:

- Original paper: (Jarman, Gault, Alves, et al., 1999)
- Birmingham's criticism: (Mohammed, Deeks, Girling, et al., 2009)
- Paul Taylor's long-form article on history and controversy: (Taylor, 2014)

Summary Hospital-level Mortality Indicator (SHMI)

With growing controversy, then NHS Medical Director, Prof. Sir Bruce Keogh, commissioned a review.

Recommended creating a new, NHS owned and transparently published indicator, however:

- Changed the remit to in-hospital or within 30-days of discharge
- Applied to all acute activity (except still birth)
- Cruder case-mix model deliberately to avoid controversial measure such as:
    - Palliative care coding
    - No adjustment for deprivation - political context
    - Co-morbidity score 'binned' rather than continuous to reduce change of gaming
    - Fewer, larger diagnosis groups

Key References:

- Review: National Quality Board (in national web archives)
- Sheffield paper: (Campbell, Jacques, Fotheringham, et al., 2012)
- NHSD SHMI: www.digital.nhs.uk/SHMI

Case-mix factors:

Factor	HSMR	SHMI
Inclusion	~20-30% inpatient activity	All inpatient activity excluding still births
Diagnosis Stratification	260 CCS groups	142 SHMI groups, groups of CCSs
Transfers considered as CIPS	Yes	No
Predictors / casemix variables:		
Age	5-year bands	5-year bands
Sex	Categorical	Categorical
Admission Method	Elective/Non-elective/ Transfer/Unknown	Elective / Non-elective /Unknown
CCS-sub group	Yes	No
Co-morbidity Score	Charlson score (continuous)	Charlson score (binned: 0, 1-5, >5)
Emergency admission in last 12-months	Yes	No
Admission Source	Yes	No
Deprivation	Yes	No
Specialist Palliative Care	Yes	No
Year of admission / index	Yes	Yes
Seasonality	No	Monthly
Birth Weight	No	Categorical for neonatal groups

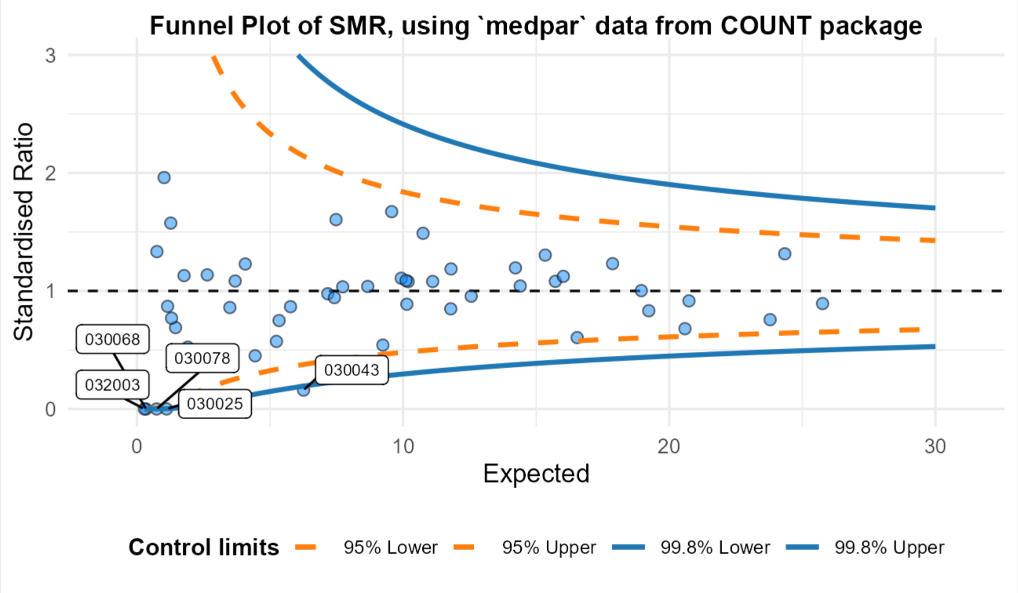
Criticisms (HSMR and/or SHMI)

- Link to quality of hospitals unclear,
  - For: (Cecil, Bottle, Esmail, et al., 2020; Cecil, Bottle, Esmail, et al., 2018)
  - Against: (Lilford and Pronovost, 2010; Black, 2010)
- They do not directly relate to avoidable death (Girling, Hofer, Wu, et al., 2012; Hogan, Zipfel, Neuburger, et al., 2015)
- Single number does not convey nuance
- Insensitive to who patients who survived
- Susceptible to 'gaming' - (Hawkes, 2013)
- Covid-19 pandemic, these models assume stability
- Case-mix adjustment fallacy (Mohammed, Deeks, Girling, et al., 2009)
- Constant risk fallacy (Mohammed, Deeks, Girling, et al., 2009; Nicholl, 2007)
- Potential for Simpsons paradox (Marang-van de Mheen and Shojania, 2014)

Cross-sectional

Comparison at single point in time

- Both HSMR and SHMI report on the final year of their modelling period.
- Snapshot of performance against expected
- League-tables are bad, as measure is relative (Goldstein and Spiegelhalter, 1996; Lilford, Mohammed, Spiegelhalter, et al., 2004)
- SPC principles applied in using funnel plot (Spiegelhalter, 2005a)
- Overdispersion



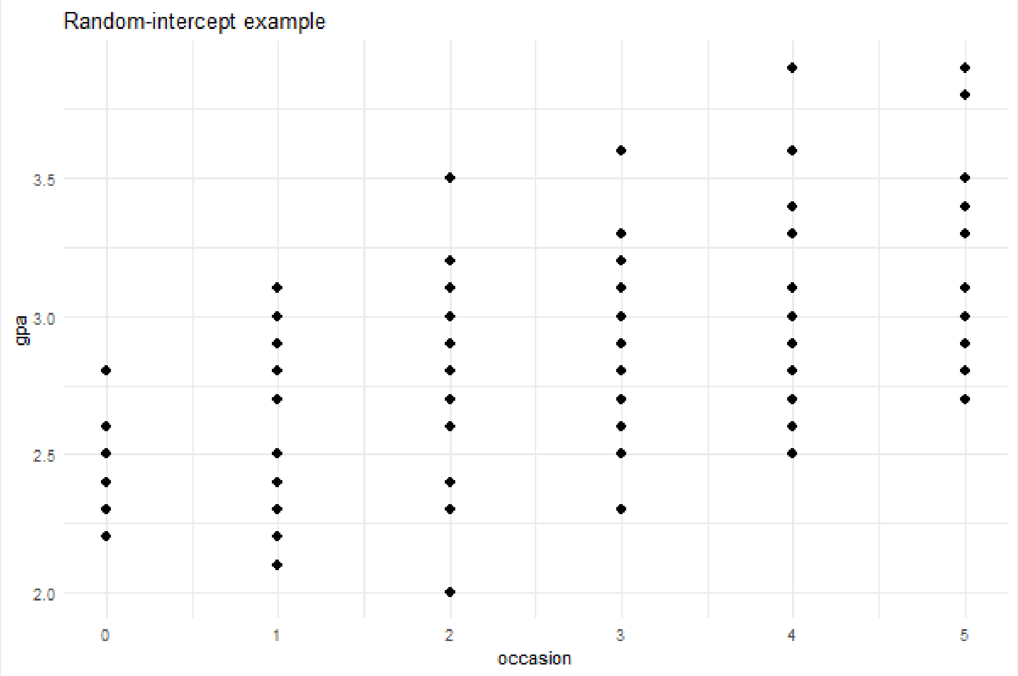
Overdispersion

Overdispersion, where conditional variance is greater than conditional mean, occurs when:

1. Aggregation / Discretization
2. Mis-specified predictors/model
3. Presence of outliers
4. Variation between response probabilities (heterogeneity)

Repeated measures (correlation)

- Regression assumes all points independent
- Sampling from same organisations repeatedly
- Clustered: - local means



Dealing with overdisperion

- Ignore it: use-case dependent. False alarm rate too high here, as error is underestimated
- Improve the model with more information: Some room for this.
- Build a model with clustered assumption (Mainey, 2020):
  - Quasi-likelihood methods (with multiplicative scale factor) (Wedderburn, 1974)
  - Compound distribution model: beta binomial (Skellam, 1948)
  - Random-intercept model: model 'within' and 'between' variance
- Apply tools based on meta-analysis methods: (Spiegelhalter, Sherlaw-Johnson, Bardsley, et al., 2012)
  - Designed to summarise studies fo different size
  - Akin to hospitals of different sizes
  - Additivity assumption - more like random-intercept than scale factor

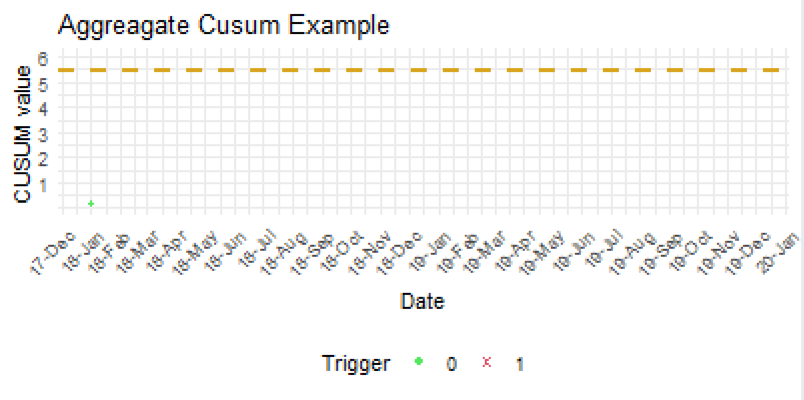
Longitudinal

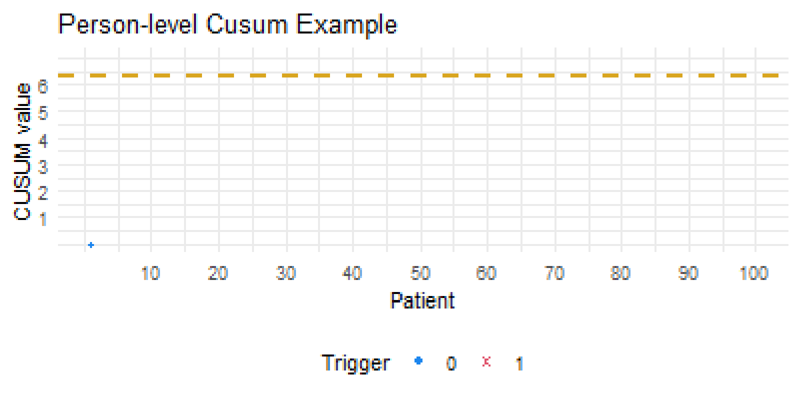
How?

- Can't simply plot in XmR chart, as risk-adjustment forms denominator (and overdispersion)
- Can use the observed and predicted in risk-adjusted control chart
- Common is 'risk-adjusted CUSUM'
  - Continuous log-likelihood ratio test

$$C_t = max(C_{t-1} + w_t, 0)$$

- C CUSUM value at time-point t (e.g. a monthly at a trust)
- w is a weighting, in this case the log-likelihood ratio (observation v.s. England) to calculate the CUSUM weight/value (C) at time point (t)





Differences in CUSUM methods

Until recently, mortality outlier programme and Imperial college sent monthly alerts to Trusts.

CQC - (aggregated)

- Data are transformed to z-scores
- Overdispersion adjustment based on additive model (Spiegelhalter, Sherlaw-Johnson, Bardsley, et al., 2012)
- Can convert average run-length to FDR, and set threshold (Grigg and Spiegelhalter, 2008; Care Quality Commission, 2014)
- Global trigger (5.48) - set to marginal 0.01% FDR.
- Applicable to other indicators and groupings, subject to same transformation

DFI/Imperial - (person-level)

- Binomial assumption and threshold set through simulation of average run-length to false positives. (Bottle and Aylin, 2011)
- Unique to each trust / group / reporting period.
- Intractable to calculate each month, and authors fitted a set of descriptive equations give a decent approximation for conditions where mortality rate 30% or lower.
- Formula can then be solved through optimisation methods and give threshold value for each group.

Problem with CUSUM charts

- A common criticism of CUSUMs is that the are opaque and hard to interpret
- What does the CUSUM value mean?

Variable Life-Adjusted Display (VLAD)

- Originally used to visualise surgical outcome more intuitively (Lovegrove, Sherlaw-Johnson, Valencia, et al., 1999)
- Can actually add limits to plot using cusums (Sherlaw-Johnson, 2005)

$$Vn = \sum_{i=1}^{n} y_n - \sum_{i=1}^{n} X_n$$

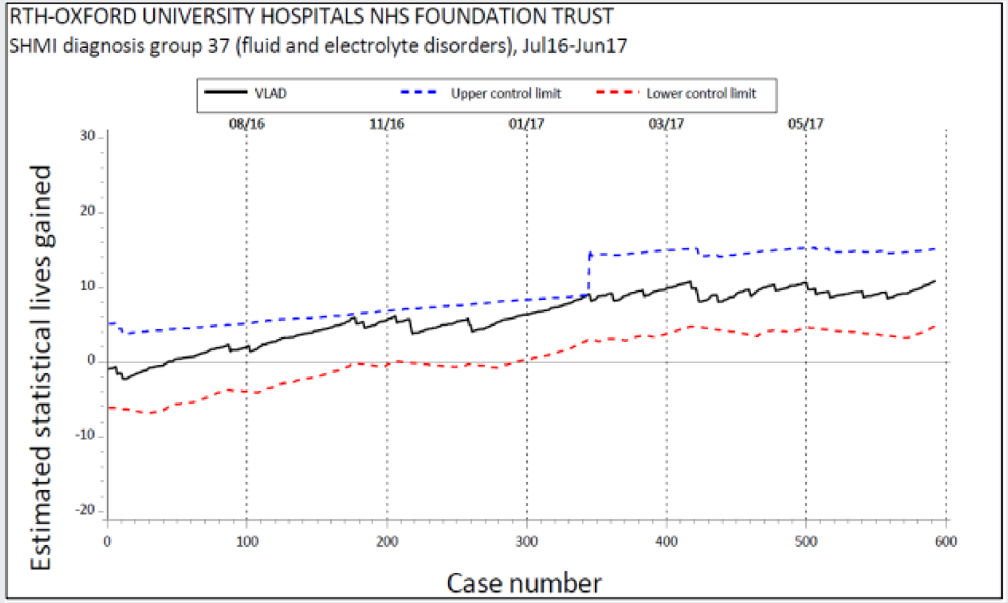


Chart sourced from https://www.ouh.nhs.uk/about/trust-board/2018/january/documents/MRG2017.149a-shmi-update.pdf

Summary

- Mortality monitoring is common, but it's use as a global measure (rather than specific conditions) is unclear.
- Not directly linked to avoidable deaths
- Crude mortality can be sensibly used in some cases, but confounded by case-mix
- Case-mix adjusted mortality is usually done by indirect standardisation, with regression model
- SMRs as grouped sums of observed / 'expected' (or 'predicted')
- HSMR was first national measure in UK - narrow scope and extensive case-mix adjustment
- SHMI is NHS-owned indicator - wider scope and less case-mix adjustment
- Criticisms remain of both - and SMRs broadly
- Cross-sectional comparisons usually by funnel plot, longitudinal with cusums and vlad.

Worth knowing history and limitations of indicators before using them

References (1)

Black, N. (2010). “Assessing the Quality of Hospitals”. In: BMJ 340, p. c2066. DOI: 10.1136/bmj.c2066.

Bottle, A. and P. Aylin (2011). “Predicting the false alarm rate in multi-institution mortality monitoring”. In: The Journal of the Operational Research Society 62.9, pp. 1711–1718. ISSN: 01605682, 14769360. DOI: 10/cbr4rq.

Campbell, M. J., R. M. Jacques, J. Fotheringham, et al. (2012). “Developing a summary hospital mortality index: retrospective analysis in English hospitals over five years”. In: BMJ 344, p. e1001. DOI: 10/gb3r9t.

Care Quality Commission (2014). NHS acute hospitals: Statistical Methodology. Care Quality Commission (CQC).

Cecil, E., A. Bottle, A. Esmail, et al. (2020). “What Is the Relationship between Mortality Alerts and Other Indicators of Quality of Care? A National Cross-Sectional Study”. In: Journal of Health Services Research & Policy 25.1, pp. 13–21. ISSN: 1355-8196. DOI: 10.1177/1355819619847689.

Cecil, E., A. Bottle, A. Esmail, et al. (2018). “Investigating the Association of Alerts from a National Mortality Surveillance System with Subsequent Hospital Mortality in England: An Interrupted Time Series Analysis”. In: BMJ Quality & Safety 27.12, p. 965. DOI: 10.1136/bmjqs-2017-007495.

Deeks, J. J., J. Dinnes, R. D'Amico, et al. (2003). “Evaluating Non-Randomised Intervention Studies.” In: Health technology assessment (Winchester, England) 7.27, pp. iii-173. ISSN: 1366-5278.

Girling, A. J., T. P. Hofer, J. Wu, et al. (2012). “Case-Mix Adjusted Hospital Mortality Is a Poor Proxy for Preventable Mortality: A Modelling Study”. In: BMJ Quality & Safety 21.12, pp. 1052–1056. DOI: 10/f4fr3b.

Goldstein, H. and D. J. Spiegelhalter (1996). “League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance”. In: Journal of the Royal Statistical Society. Series A (Statistics in Society) 159.3, pp. 385–443. ISSN: 09641998, 1467985X. DOI: 10/chf9kj.

Grigg, O. and D. Spiegelhalter (2008). “The null steady-state distribution of the CUSUM statistic”. In: Technometrics : a journal of statistics for the physical, chemical, and engineering sciences. DOI: 10/bgvkdx.

Hawkes, N. (2013). “How the Message from Mortality Figures Was Missed at Mid Staffs”. In: BMJ : British Medical Journal 346, p. f562. DOI: 10.1136/bmj.f562.

Hogan, H., R. Zipfel, J. Neuburger, et al. (2015). “Avoidability of Hospital Deaths and Association with Hospital-Wide Mortality Ratios: Retrospective Case Record Review and Regression Analysis”. In: BMJ 351. DOI: 10/gb3swm.

References (2)

Iezzoni, L. I. (1997). “The Risks of Risk Adjustment”. In: JAMA 278.19, pp. 1600–7. ISSN: 0098-7484 (Print) 0098-7484 (Linking). DOI: 10/c2trv4.

Jarman, B., A. Bottle, P. Aylin, et al. (2005). “Monitoring Changes in Hospital Standardised Mortality Ratios”. In: BMJ (Clinical research ed.) 330.7487, pp. 329–329. ISSN: 1756-1833. DOI: 10.1136/bmj.330.7487.329.

Jarman, B., S. Gault, B. Alves, et al. (1999). “Explaining differences in English hospital death rates using routinely collected data”. In: BMJ 318.7197, pp. 1515–1520. ISSN: 0959-8138 1468-5833. DOI: 10/fkkfm9.

Keogh, B. (2013). Keogh Review on Hospital Deaths Published - NHSUK.

Lilford, R., M. A. Mohammed, D. Spiegelhalter, et al. (2004). “Use and Misuse of Process and Outcome Data in Managing Performance of Acute Medical Care: Avoiding Institutional Stigma”. In: Lancet 363.9415, pp. 1147–54. ISSN: 0140-6736. DOI: 10/c97xd2.

Lilford, R. and P. Pronovost (2010). “Using Hospital Mortality Rates to Judge Hospital Performance: A Bad Idea That Just Won't Go Away”. In: BMJ 340, p. c2016. ISSN: 1756-1833 (Electronic) 0959-535X (Linking). DOI: 10/frq62g.

Lovegrove, J., C. Sherlaw-Johnson, O. Valencia, et al. (1999). “Monitoring the Performance of Cardiac Surgeons”. In: The Journal of the Operational Research Society 50.7. Publisher: Palgrave Macmillan Journals, pp. 684–689. ISSN: 01605682, 14769360. DOI: 10/dtsvbg.

Mainey, C. (2020). “Statistical methods for NHS incident reporting data”. London. URL: https://discovery.ucl.ac.uk/id/eprint/10094736/.

Marang-van de Mheen, P. J. and K. G. Shojania (2014). “Simpson's Paradox: How Performance Measurement Can Fail Even with Perfect Risk Adjustment”. In: BMJ Quality & Safety 23.9, p. 701. DOI: 10.1136/bmjqs-2014-003358.

Mohammed, M. A., J. J. Deeks, A. Girling, et al. (2009). “Evidence of Methodological Bias in Hospital Standardised Mortality Ratios: Retrospective Database Study of English Hospitals”. In: BMJ 338, p. b780. ISSN: 1756-1833 (Electronic) 0959-535X (Linking). DOI: 10/bv4sh8.

Nicholl, J. (2007). “Case-Mix Adjustment in Non-Randomised Observational Evaluations: The Constant Risk Fallacy”. In: Journal of Epidemiology and Community Health (1979-) 61.11, pp. 1010–1013. ISSN: 0143005X, 14702738. DOI: 10/d7f9jh.

Sherlaw-Johnson, C. (2005). “A Method for Detecting Runs of Good and Bad Clinical Outcomes on Variable Life-Adjusted Display (VLAD) Charts”. In: Health Care Management Science 8.1, pp. 61–65. ISSN: 1572-9389. DOI: 10/dpvrft.

References (3)

Skellam, J. G. (1948). "A Probability Distribution Derived from the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials". In: Journal of the Royal Statistical Society. Series B (Methodological) 10.2, pp. 257–261. ISSN: 00359246.

Spiegelhalter, D. J. (2005a). "Funnel plots for comparing institutional performance". In: Statistics in Medicine 24.8, pp. 1185–202. ISSN: 0277-6715 (Print) 0277-6715 (Linking). DOI: 10/fq7z8t.

Spiegelhalter, D., C. Sherlaw-Johnson, M. Bardsley, et al. (2012). "Statistical methods for healthcare regulation: Rating, screening and surveillance". In: Journal of the Royal Statistical Society: Series A (Statistics in Society) 175.1, pp. 1–47. ISSN: 09641998. DOI: 10/dqpqpk.

Taylor, P. (2014). "Standardized Mortality Ratios". In: International Journal of Epidemiology 42.6, pp. 1882–1890. ISSN: 0300-5771. DOI: 10/f5pxfw.

Wedderburn, R. W. M. (1974). "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method". In: Biometrika 61.3, pp. 439–447.

Wright, J., B. Dugdale, I. Hammond, et al. (2006). "Learning from Death: A Hospital Mortality Reduction Programme". In: Journal of the Royal Society of Medicine 99.6, pp. 303–308. ISSN: 0141-0768. DOI: 10.1258/jrsm.99.6.303.