Understanding and interpreting funnel plots for the clinician

ABSTRACT

Funnel plots are an increasingly common graphical tool which are widely used in the literature. They were first introduced by Light and Pillemer in 1984. In scientific literature, funnel plots are used to identify the probability of bias in meta-analyses and compare institutional performance.

The ability to identify variation is better with graphical than tabular display. In addition, the way data are presented can directly influence the interpretation of results. This was demonstrated by Marshall et al (2004), who presented institutional mortality data in both a league table and control chart format. This study illustrated that when displayed as a league table, a greater number of units were identified for investigation than were actually required. The use of control charts or funnel plots may therefore show benefit in reducing the number of inappropriately labelled outliers.

This article explains how clinicians should read and interpret funnel plots, and discusses their considerations and limitations.

Funnel plots are simple scatterplots. They allow readers to identify outliers by analysing a diagram. In the case of a meta-analysis, a funnel plot can highlight studies that are subject to bias compared to those that are not. This enables readers to more accurately interpret conclusions drawn from the overall study.

While funnel plots are commonly used for this purpose in the scientific literature another use has grown considerably over recent years. Following the public inquiry into paediatric cardiac surgery at the Bristol Royal Infirmary, the NHS has placed increasing emphasis on transparency of clinical outcomes. As such, surgeon- and institution-specific data are now increasingly available across a range of specialties and are being used as a way to monitor performance. As a result, the funnel plot is now showing a second benefit as a way to display performance data.

For example the National Joint Registry was initially set up to provide information on joint replacements performed in the UK. It has since progressed to monitor the rates of revision surgery, the brands of implants used and individual surgeons. Funnel plots are commonly used in National Joint Registry analyses to identify outliers which need further investigation. The use of funnel plots can also inform surgeons and the public alike that the performance of a surgeon or institution is in accordance with accepted standards. For this reason, an understanding of the funnel plot is essential if the data that it contains are to be accurately interpreted.

Figure 1 gives an example of a funnel plot. They are constructed by plotting individual

Mr Charles Godavitarne, Registrar, Department of Trauma and Orthopaedics, Princess Royal Hospital, Brighton and Sussex University Hospitals NHS Trust, Haywards Heath, Sussex

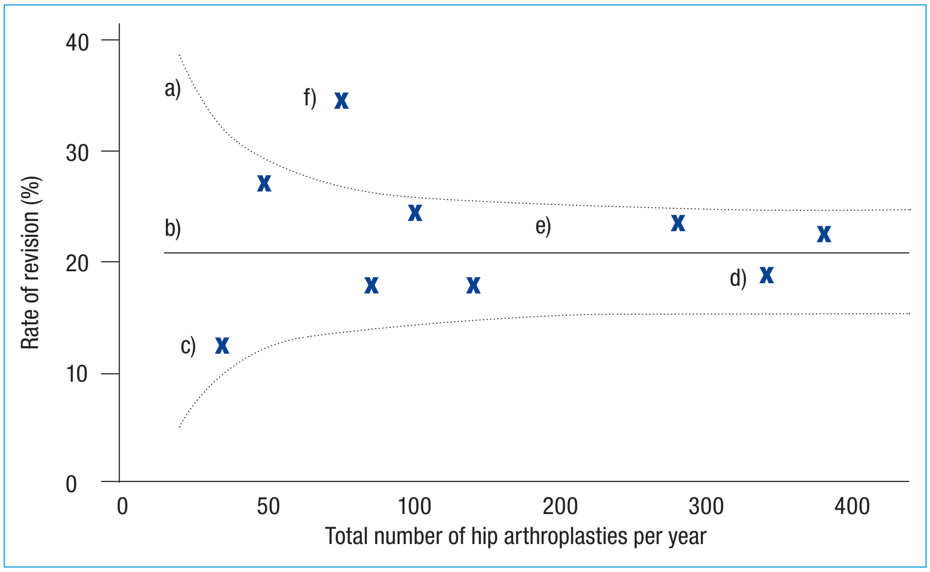
Mr Alastair Robertson, Registrar, Department of Trauma and Orthopaedics, Princess Royal Hospital, Brighton and Sussex University Hospitals NHS Trust, Haywards Heath, Sussex

Mr David M Ricketts, Consultant Surgeon, Department of Trauma and Orthopaedics, Princess Royal Hospital, Brighton and Sussex University Hospitals NHS Trust, Haywards Heath, Sussex

Mr Benedict A Rogers, Consultant Surgeon, Department of Orthopaedics, Brighton and Sussex University Hospital and Brighton University, Brighton BN2 5BE

Correspondence to: Mr BA Rogers (benedictrogers@hotmail.com)

Figure 1. A funnel plot displaying the rate of revision hip surgery (%). The crosses represent data points (in this example institutions). The further along the x-axis an institution lies, the more cases it has performed. The further along the y-axis an institution lies, the higher the rate of revision at that centre. **a.** Control limit typically two or three standard deviations about the mean. **b.** The mean. **c.** An institution scattering widely from the mean representing increased variability, most likely the result of a smaller number of cases performed. **d.** An institution scattering close to the mean representing less variability as a result of a larger number of cases performed. **e.** Institutions contained within the control limits are likely to be accountable to common cause or natural variance and therefore performing within set standards. **f.** An institution lying outside of the control limits is highlighted as an outlier. This institution is displaying special cause variance and should be investigated to determine why it has a higher than expected revision rate.



data points followed by control limits, which are superimposed around the data. Plotted data can represent individual studies (in the case of meta-analyses) or an individual institution (in the case of performance comparison). The control limits represent the area two or three standard deviations either side of the mean. Data points lying within this area can be assumed to be the result of expected variation.

A small sample size typically displays a greater degree of variability resulting in wider control limits. As a sample size increases the degree of variability decreases resulting in narrower control limits, generating an inverted funnel shaped distribution. In the absence of bias, the scatter of data will be the result of sampling variation alone and the funnel plot should be symmetrical about the mean.

The scatter of data points lying either within the control limits or outside of them are the result of common cause or special cause variance respectively. Common cause or natural variance are the usual quantifiable variations within a system. They represent an inherent part of a process and no specific actions can be performed to prevent common cause variance. In other words common cause variance represents random variability or marginal differences in external factors. While common cause variance can account for differences in the data points, it cannot be used to overcome basic inadequacies in data quality (Mayer et al, 2009). Special cause variance refers to unusual or unplanned variance that has an assignable cause. It can be corrected by identifying weaknesses within a system or process. Data points in the funnel plot which lie within the control limits are said to be attributable to common cause variance whereas those lying outside the control limits are attributable to special cause variance.

Commonly the display of three standard deviations serves as the demarcation between common cause and special cause variance. An author may choose to include two standard deviations only to illustrate that all of the data points lie within those limits and variation is likely accountable to chance alone. However, data points lying outside of two standard deviations could still be accountable to chance. Conversely an author may decide to display control limits of two standard deviations with greatly spread data points to highlight the degree of variation among the data. With this in mind the choice of two or three standard deviations depends on the message the author is ultimately trying to convey.

Funnel plots can be presented with the inverted funnel in an up-down direction or in a horizontal direction, with control limits either being straight or curved. Typically when the funnel plot is used to identify bias in a meta-analysis, the measure of study size is found on the vertical axis and the treatment effect is found on the horizontal axis. This creates the up-down inverted funnel (*Figure 2*).

In contrast, when assessing performance outcomes, the health outcome is seen on the vertical axis with the sample size seen on the horizontal axis creating a horizontal funnel (*Figure 3*). While the data can be displayed as either an up-down or horizontal funnel, there is no statistical reason or significance for the difference in display. However, it is important that the reader acknowledges the difference in layout so that the funnel plot can be interpreted correctly.

Funnel plots can be used to identify the probability of bias within meta-analyses or in the comparison of performance, discussed in more detail below.

Funnel plots in meta-analyses

Meta-analysis is a statistical method of analysis combining results from a number

Figure 2. An up-down inverted funnel plot illustrating 14 studies estimating the risk ratio of postoperative surgical site infections after clean orthopaedic implant surgery in HIV-infected patients compared to HIV-negative patients. From Kigera et al (2012).

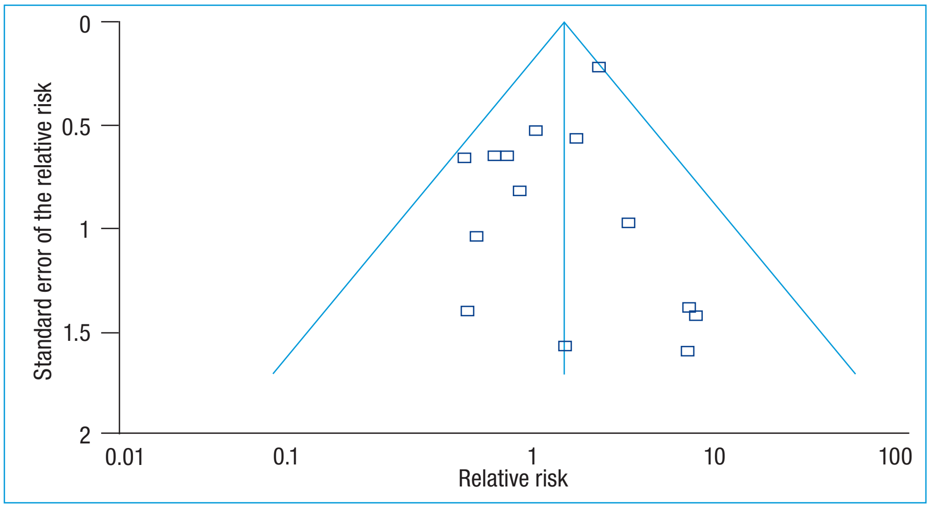


Figure 3. A horizontal funnel plot illustrating the risk-adjusted postoperative mortality against total caseload by trust from 0 to 30 days (90-day mortality outliers marked). From Byrne et al (2013).

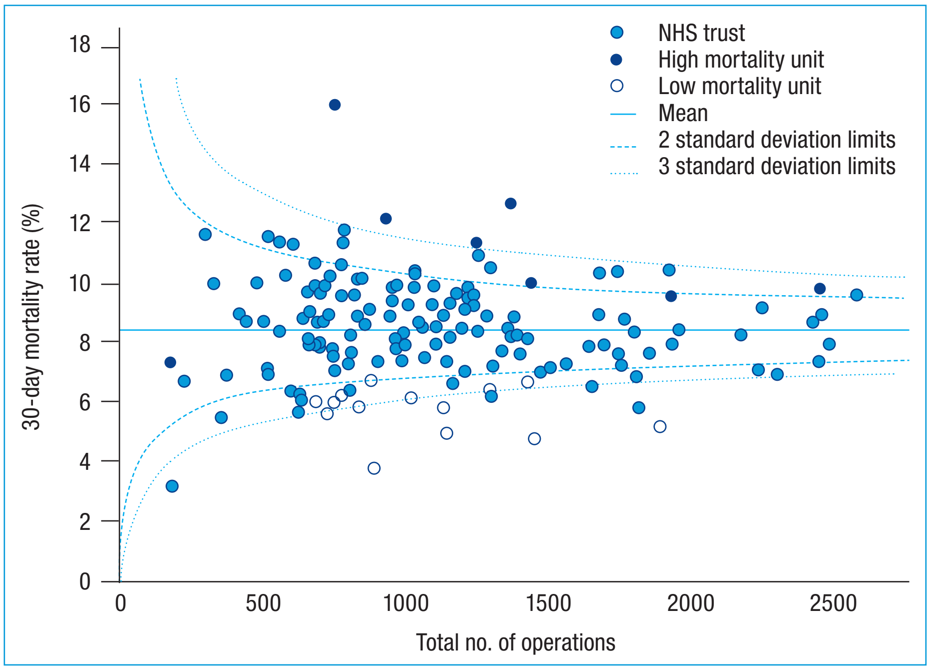


Figure 4. Funnel plot illustrating asymmetry. The log of the risk ratios are plotted against the standard error of the risk ratio of each study to identify asymmetry in the distribution of trials. From Ritchie and Romanuk (2012).

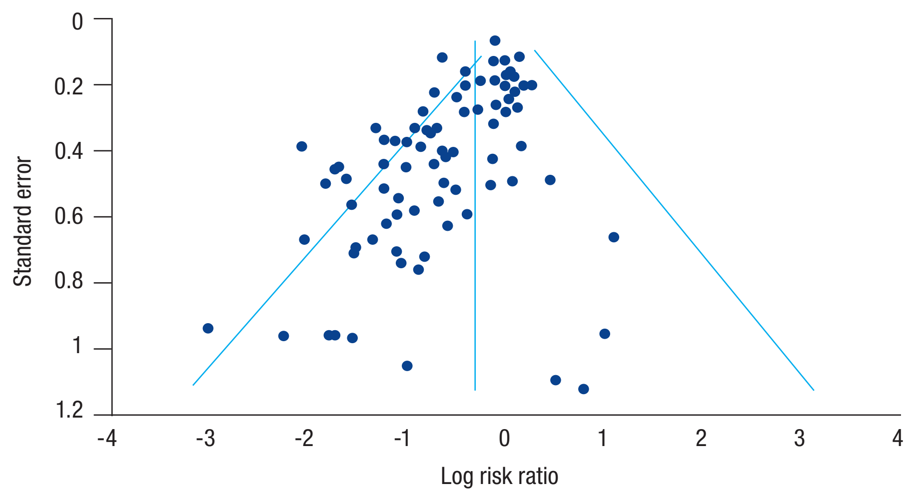


Table 1. Different types of reporting bias

Type of bias	Description
Publication bias	More likely to be published than statistically insignificant studies
Time lag bias	More likely to be published into the scientific literature at a quicker rate
Multiple publication bias	Published in numerous locations which could lead to overestimation of a treatment effect
Location bias	Published in journals with better ease of access, indexing standards, impact factor
Citation bias	References included are most likely to those studies supporting the treatment effect
Language bias	Published in journals that are written in English so can be more widely read

From Hopewell et al (2009)

of different studies on a specific topic. The aim is to draw a conclusion about the overall effect of a process or intervention. While this method of analysis can produce estimates of treatment effect, it can be subject to bias. This can be identified by using a funnel plot.

The funnel plot is based on the estimation that the underlying treatment effect will increase as the sample size increases. Studies with small sample sizes therefore scatter widely at the base of the funnel (further from the mean), while studies with larger sample sizes spread more narrowly at its apex (closer to the mean) (*Figure 3*). In the absence of bias and other factors, a symmetrical funnel shape will be seen. It follows that funnel plots can be used to assess the quality of collated data within a meta-analysis. The resultant symmetry or asymmetry (*Figure 4*) can indicate the presence of bias. Publication bias for example is a known cause of funnel plot asymmetry. It is well discussed in the literature and often attributed as the sole cause of asymmetry. This is not the case, however, and other factors including heterogeneity and chance must also be considered.

Reporting or selection bias

Reporting or selection bias is the selective revealing or suppression of information that is typically influenced by the nature and direction of results. This means that studies with findings that support or favour an outcome are more likely to be published than those that do not. There are many different types of reporting bias as outlined in *Table 1*.

The significance of reporting bias is simply that the data collected by a meta-analysis may be skewed, leading to funnel plot asymmetry. This asymmetry will be exaggerated if the smaller studies (more prone to selective suppression as a result of insignificant findings) are concealed.

Heterogeneity

Heterogeneity refers to the variation in study outcomes between studies which is not attributable to chance. This can be either clinical (types of patients, implementation or intervention) or statistical (control over bias). Heterogeneity is an important consideration as studies used in a meta-analysis may not be estimating the same effect for the same intervention in the same way. This can result in asymmetry of a funnel plot, particularly if a treatment effect is larger in smaller studies. This could occur if, for example, interventions were implemented less thoroughly in a larger study resulting in more positive results in a smaller study.

Numerous methods exist to help investigate the presence of heterogeneity in meta-analyses. Statistical testing can be used to identify the probability of between-study variation, represented as a '*P* value'. Two common methods include the DerSimonian–Laird test and the Breslow–Day test. The main issue is that this form of testing can be underpowered as it relies on two main factors: the number of studies included and the respective weight allocated to each study. Therefore true differences may be hard to identify particularly if there is only a small number of studies (Song et al, 2001).

Other methods commonly used to investigate heterogeneity are subgroup analysis and meta-regression. Subgroup analysis separates the studies into subgroups according to study-level variables allowing separate analysis of each subgroup. Its usefulness may be limited in meta-analyses containing only a small number of studies. It is also possible that differences between subgroups could simply be the result of chance. Meta-regression is a more sophisticated approach that uses the estimate of study results as the dependent variable and study-level variables as the independent variable. If a linear relationship is seen between the two, regression analysis will yield a greater statistical power than subgroup analysis (Song et al, 2001).

If there is little difference between studies a fixed effect model can be used which assumes that the studies have been conducted under similar conditions with similar subjects. Random effect models allow study outcomes to vary in a normal distribution and are often used to incorporate heterogeneity in meta-analysis (Sterne et al, 2011).

Chance

Chance must be considered when interpreting funnel plots. As meta-analyses can contain a limited number of studies, they can be prone to uncover false positive findings affecting the symmetry of the plot.

Assessing funnel plot asymmetry

Assessing for asymmetry is done visually. The reader may note the absence of data points in one quadrant of the plot and assess for asymmetry by counting studies in each quadrant for comparison. The differences in data points between quadrants can be termed as imbalance, but the imbalance must be interpreted relative to the total number of studies in the analysis (Simmonds, 2015).

Imbalance does not take into account the distance of the data points from the mean. Therefore assessing asymmetry distance can be useful. This is done by comparing the sum of the distance of data points on one side of the mean to the sum of the distance of data points on the other side of the mean and dividing this by the total distance. A score of zero implies perfect symmetry while a score of one implies maximum asymmetry (Simmonds, 2015).

Using visual assessment alone to identify the presence of asymmetry in a funnel plot can be misleading particularly when a meta-analysis has a relatively low number of studies. This can lead to incorrect conclusions being drawn as readers may struggle to appropriately interpret the funnel plot (Terrin et al, 2000) or become confused by differing axes and outcome measures (Tang and Liu, 2000). In this case more formal statistical tests such as the rank correlation test or Egger's regression test (see below) can be used to identify the presence of asymmetry which can then be illustrated on the funnel plot.

If asymmetry is found, one method to identify the possibility of bias is for the author to establish contour lines (*Figure 5*) that correspond to statistical significance (Peters et al, 2008). If data points are missing from areas of 'non significance', it is possible that reporting bias may be present but other factors should also be considered. Conversely if data points are missing from 'high significance' areas then reporting bias is unlikely.

Statistical tests for funnel plot asymmetry

More formal statistical tests can be performed to test for funnel plot asymmetry. The rank correlation test developed by Begg and Mazumdar (1994) and the Egger's regression test (Egger et al, 2001) are used to identify publication bias. As a general rule tests for asymmetry should not be used when there are fewer than 10 studies in the meta-analysis. Owing to a low test power formal statistical tests are unable to distinguish between chance and real asymmetry. This means that even if asymmetry is ruled out, bias cannot be excluded (Sterne et al, 2011).

Funnel plots as a comparison of institutional performance

League tables and caterpillar plots (*Figure 6*) are typically used in the literature to rank institutional and individual performance. Caterpillar plots are side-by-side bar plots displaying 95% confidence intervals. The specific purpose of this plot is to identify differences in performance that may be accountable to chance. While both types of data display are used for institutional comparison they do have limitations including not displaying sample size or accounting for over-adjustment of case-mix. Anell et al (2016) demonstrated that when league tables were used to compare institutional performance, units were over-identified for further investigation. As a result funnel plots are becoming an increasingly popular graphical tool in the comparison of institutions and identification of outliers.

Funnel plots for performance comparison are constructed by placing the performance indicator value on the vertical axis and the accuracy of the measured indicator or number performed by a surgeon or institution on the horizontal axis. As in the case for funnel plots used in meta-analyses (*Figure 2*), lines for target outcomes and control limits are superimposed onto the plot. The control limits (typically three standard deviations from the mean) account for the bulk of variation, which is assumed to be the result of common cause variation, as explained above.

In contrast to league tables, funnel plots can identify institutions that are performing outside of the set control limits, indicating special cause variation. This allows readers to identify specific institutions and determine possible reasons for their significant variance. Both funnel plots and cumulative funnel plots have been used to present in-hospital outcome data. The funnel plot is used to display summaries of overall performance while the cumulative funnel plot is used to display a case series of an individual clinician's performance (Kunadian et al, 2008).

Figure 5. Funnel plots **(a)** with and **(b)** without contour lines. **b** has shaded regions corresponding to different *P* values with varying levels of significance as identified in the legend. From Crossley et al (2014).

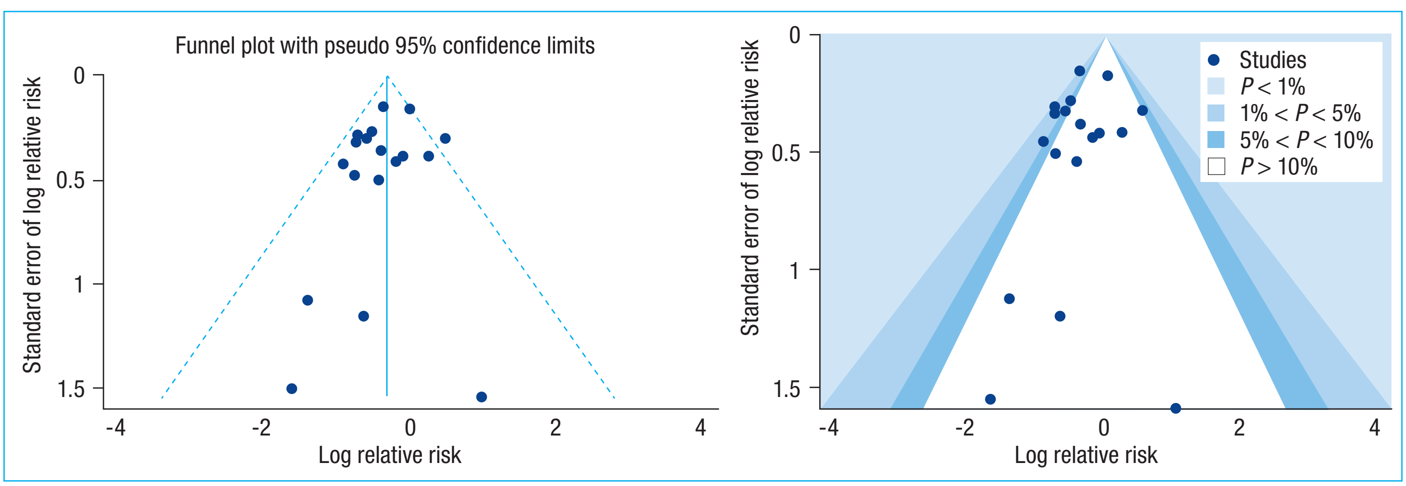
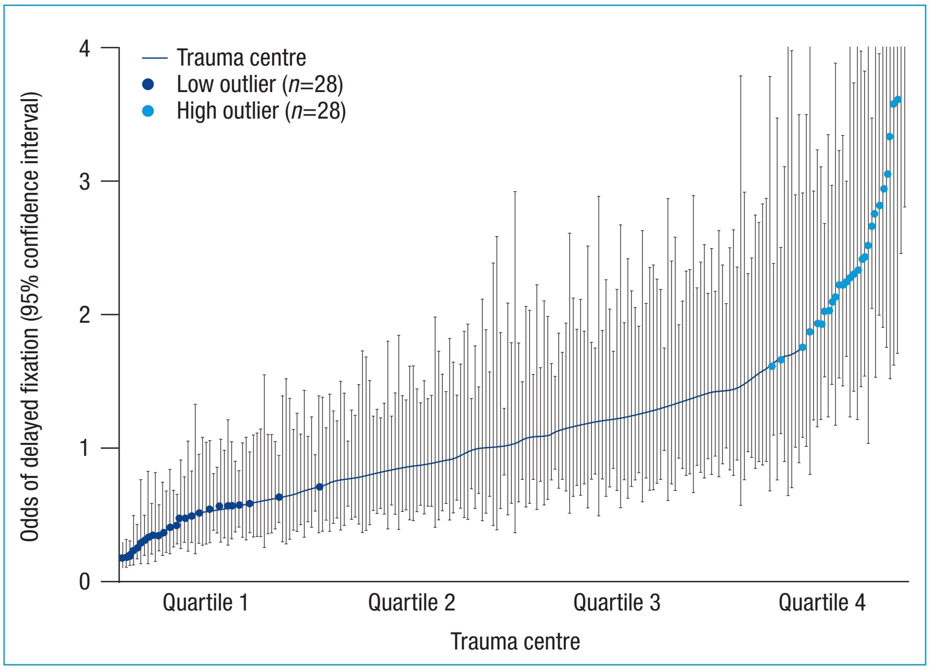


Figure 6. A caterpillar plot illustrating the odds ratios and 95% confidence intervals for 216 trauma centres when exploring the variability in delayed fixation of femoral shaft fractures. From Byrne et al (2017).



Consideration of funnel plots used for institutional comparison

Data used in funnel plots should be adequately adjusted to represent the patient population. In others words, where possible the data used should be as homogenous as possible. While funnel plots will allow for factors that result in random variation to fall within the predetermined control limits, obviously heterogeneous data will not, and may instead be highlighted as an outlier. For this reason funnel plots should be interpreted with caution and readers should consider this when outlying units are identified.

Funnel plots for institutional comparison represent a snapshot analysis for a given period, meaning any variation seen could be the result of chance alone (Mayer et al, 2009). However, modifications in the construction of funnel plots can allow for the identification of performance over time, allowing readers to determine changes in performance as a continuous process.

Over-dispersion is noted when the majority of data points on the funnel plot are situated outside the control limits. This is clearly not conducive to identifying outliers and could occur as a result of unmeasured variables that have not been taken into account. This could be for a number of reasons, including inability of the variables to be measured, the variables being deemed insignificant to the displayed data or that they simply were not considered. While each individual unmeasured variable may not be enough to cause significant variation, their summation could lead to significant variability. When looking for institutions with divergent performance the data can be adjusted to bring the data points within the control limits. It should be remembered, however, that the aim is to identify divergent performance from an overall standard or target. There are a number of methods by which the data can be controlled. These include not using a given indicator (as not sensitive to compare institutions), improving risk stratification (by measuring factors causing excess variability), analysis by clustering (creating more homogenous groups to compare like with like), estimating an over-dispersion factor or assuming a random effects model (Spielgelhalter, 2005). An example of the effects of correcting over-dispersion can be seen in *Figure 7*.

The volume–outcome relationship was first coined by Luft in 1987. It describes the notion that institutions with higher volumes of cases will have better outcomes. This might be the result of either the 'practice makes perfect' hypothesis or the 'selective-referral pattern' hypothesis whereby patients are referred to a centre as it is perceived to have better results (Luft et al, 1987). Funnel plots are ideally suited to illustrate this relationship by placing the observed event rate on the vertical axis and the volume of cases on the horizontal axis. An example of the benefit of data being displayed in this way is in the identification of mortality rates for a given institution. While it is known that hospitals with smaller case loads are likely to exhibit statistically significant higher mortality rates, standardized mortality ratios can instead be plotted. This enables volume–outcome relationships to be identified, while also allowing the reader to identify divergent data points. This makes funnel plots an ideal graphical tool in the identification of institutions that may need further investigation for service improvement.

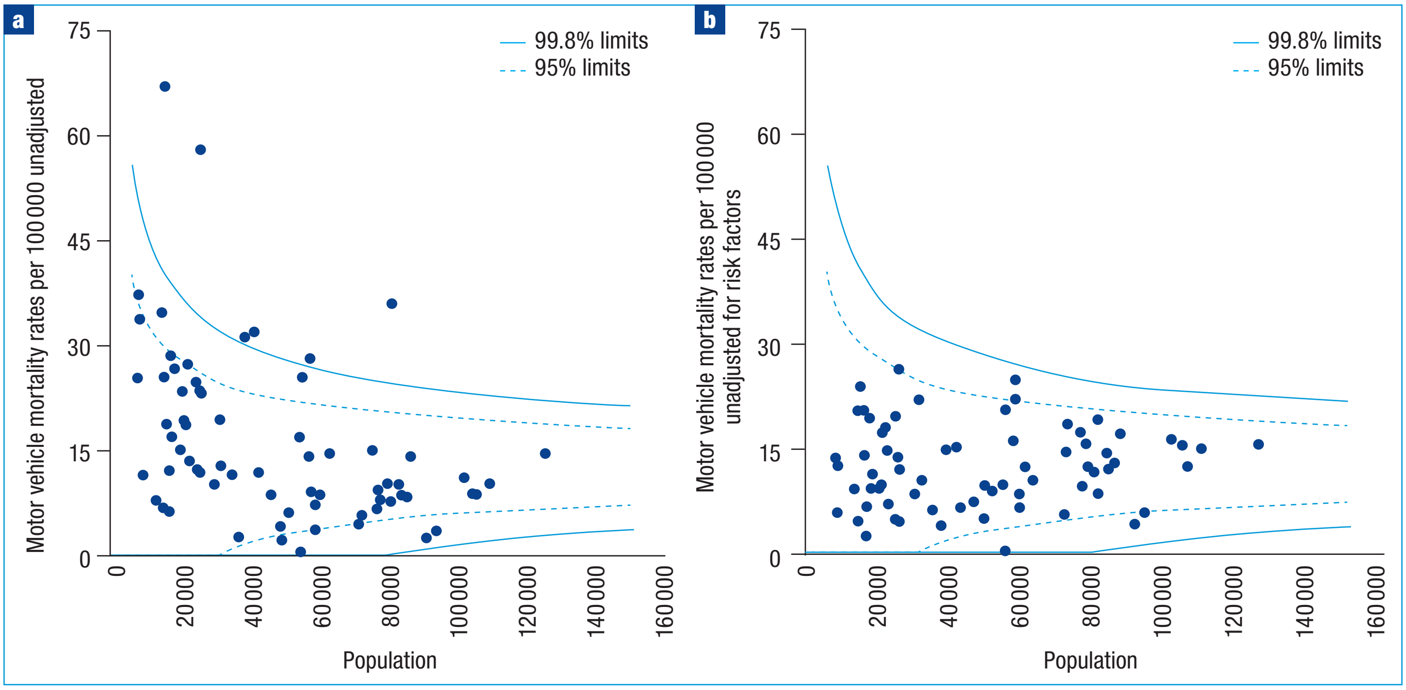
Conclusions

Funnel plots are becoming more prevalent in the scientific literature. They represent a graphical method by which data points can be categorized into normal or special cause variation. Through statistical analysis, they can be easily constructed and provide an easy way for readers to identify divergent data points. In the case of meta-analyses, outlying studies can be easily identified. By considering outlying studies as well as the presence of asymmetry the conclusions drawn from studies can be more accurately interpreted. This is imperative considering the importance of meta-analyses in evidence-based medicine. An increasing use of the funnel plot is in the display of performance outcome data either at surgeon or institution level. This is becoming ever more prevalent across a range of hospital specialties. Funnel plots used for this purpose can illustrate units performing outside of accepted control limits. This will allow them to be identified and further investigated to see if improvement is required. However, funnel plots can also be used to reassure readers of units that are performing within accepted standards. Owing to its increasing use in the literature the reader must understand the basic principles of the funnel plot for accurate interpretation of the data. **BJHM**

Conflict of interest: none.

Anell A, Hagberg O, Liedberg F, Ryden S. A randomized comparison between league tables and funnel plots to inform health care decision-making. Int J Qual Health Care. 2016 Dec 1;28(6):816–823. https://doi.org/10.1093/intqhc/mzw125
Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics. 1994 Dec;50(4):1088–1101.
Byrne BE, Mamidanna R, Vincent CA, Faiz O. Population-based cohort study comparing 30- and 90-day institutional mortality rates after colorectal

Figure 7. Funnel plots displaying unadjusted and adjusted motor vehicle mortality rates in Alberta based on 70 subregions (represented by black dots). **a.** Of 70 subregions, 16 lie on or outside of the 95% confidence limit with six identified as anomalies lying outside of the 99.8% limit. **b.** The authors then adjusted for demographic factors (age and sex) as well as behavioural risk factors (seat belt use and drink driving). Following factor adjustment only six subregions now lie on or outside of the 95% confidence limits with 0 subregions identified as anomalies. From Dover and Schopflocher (2011).



surgery. Br J Surg. 2013 Dec;100(13):1810–1817. https://doi.org/10.1002/bjs.9318

Byrne JP, Nathens AB, Gomez D, Pincus D, Jenkinson RJ. Timing of femoral shaft fracture fixation following major trauma: A retrospective cohort study of United States trauma centers. PLoS Med. 2017 Jul 5;14(7):e1002336. https://doi.org/10.1371/journal.pmed.1002336

Crossley S, Reid J, McLatchie R, Hayton J, Clark C, MacDougall M, Andrews PJ. A systematic review of therapeutic hypothermia for adult patients following traumatic brain injury. Crit Care. 2014 Apr 17;18(2):R75. https://doi.org/10.1186/cc13835

Dover DC, Schopflocher DP. Using funnel plots in public health surveillance. Popul Health Metr. 2011 Nov 10;9(1):58. https://doi.org/10.1186/1478-7954-9-58

Egger M, Davey Smith G, Altman DG. Systematic Reviews in health care: Meta Analysis in context. 2nd edn. London: BMJ Publishing Group.

Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev. 2009 Jan 21;1(1):MR000006. https://doi.org/10.1002/14651858.MR000006.pub3

Kigera JW, Straetemans M, Vuhaka SK, Nagel IM, Naddumba EK, Boer K. Is there an increased risk of post-operative surgical site infection after orthopaedic surgery in HIV patients? A systematic review and meta-analysis. PLoS One. 2012;7(8):e42254. https://doi.org/10.1371/journal.pone.0042254

Kunadian B, Dunning J, Roberts AP et al. Cumulative funnel plots for the early detection of interoperator variation: retrospective database analysis of observed versus predicted results of percutaneous coronary intervention. BMJ. 2008 Apr 26;336(7650):931–934. https://doi.org/10.1136/bmj.39512.529120.BE

Light RJ, Pillemer DB. 1984. Summing up: The Science of Reviewing Research. Cambridge, Massachusetts: Harvard University Press.

Luft HS, Hunt SS, Maerki SC. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? Health Serv Res. 1987 Jun;22(2):157–182.

Marshall T, Mohammed MA, Rouse A. A randomized controlled trial of league tables and control charts as aids to health service decision-making. Int J Qual Health Care. 2004 Aug;16(4):309–315. https://doi.org/10.1093/intqhc/mzh054

Mayer EK, Bottle A, Rao C, Darzi AW, Athanasiou T. Funnel plots and their emerging application in surgery. Ann Surg. 2009 Mar;249(3):376–383. https://doi.org/10.1097/SLA.0b013e31819a47b1

Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol. 2008 Oct;61(10):991–996. https://doi.org/10.1016/j.jclinepi.2007.11.010

Ritchie ML, Romanuk TN. A meta-analysis of probiotic efficacy for gastrointestinal diseases. PLoS One. 2012;7(4):e34938. https://doi.org/10.1371/journal.pone.0034938

Simmonds M. Quantifying the risk of error when interpreting funnel plots. Syst Rev. 2015 Mar 11;4:24. https://doi.org/10.1186/s13643-015-0004-8

Song F, Sheldon TA, Sutton AJ, Abrams KR, Jones DR. Methods for exploring heterogeneity in meta-analysis. Eval Health Prof. 2001 Jun;24(2):126–151. https://doi.org/10.1177/016327870102400203

Spiegelhalter DJ. Handling over-dispersion of performance indicators. Qual Saf Health Care. 2005 Oct;14(5):347–351. https://doi.org/10.1136/qshc.2005.013755

Sterne JA, Sutton AJ, Ioannidis JP et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011 Jul 22;343:d4002. https://doi.org/10.1136/bmj.d4002

Tang JL, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. J Clin Epidemiol. 2000 May;53(5):477–484. https://doi.org/10.1016/S0895-4356(99)00204-8

Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol. 2005 Sep;58(9):894–901. https://doi.org/10.1016/j.jclinepi.2005.01.006

KEY POINTS

- Funnel plots are used primarily as a visual tool to assess for the presence of bias.
- Studies with small sample sizes tend to scatter widely at the base, while larger studies scatter more narrowly towards the mean.
- Asymmetrical funnel plots can be the result of reporting bias, as well as factors such as heterogeneous data and chance.
- When comparing institutional performance, funnel plots may more accurately identify significant outliers that need further investigation than traditional league tables or caterpillar plots.
- Modifying the funnel plot can allow comparison of institutions over time *vs* a snapshot analysis.
- Funnel plots are ideally suited to identify the volume–outcome relationship.