# Funnel plots a graphical instrument for the evaluation of population performance and quality of trauma care: a blueprint of implementation

Driessen, M.L.S.; Zwet, E.W. van; Sturms, L.M.; Jongh, M.A.C. de; Leenen, L.P.H.

European Journal of Trauma and Emergency Surgery (2023) 49:513–522
https://doi.org/10.1007/s00068-022-02100-z

ORIGINAL ARTICLE

# Funnel plots a graphical instrument for the evaluation of population performance and quality of trauma care: a blueprint of implementation

**M. L. S. Driessen[1] · E. W. van Zwet[2] · L. M. Sturms[1] · M. A. C. de Jongh[3] · L. P. H. Leenen[4]**

Received: 4 May 2022 / Accepted: 26 August 2022 / Published online: 9 September 2022


## Abstract

**Background** Using patient outcomes to monitor medical centre performance has become an essential part of modern health care. However, classic league tables generally inflict stigmatization on centres rated as "poor performers", which has a negative effect on public trust and professional morale. In the present study, we aim to illustrate that funnel plots, including trends over time, can be used as a method to control the quality of data and to monitor and assure the quality of trauma care. Moreover, we aimed to present a set of regulations on how to interpret and act on underperformance or overperformance trends presented in funnel plots.

**Methods** A retrospective observational cohort study was performed using the Dutch National Trauma Registry (DNTR). Two separate datasets were created to assess the effects of healthy and multiple imputations to cope with missing values. Funnel plots displaying the performance of all trauma-receiving hospitals in 2020 were generated, and in-hospital mortality was used as the main indicator of centre performance. Indirect standardization was used to correct for differences in the types of cases. Comet plots were generated displaying the performance trends of two level-I trauma centres since 2017 and 2018.

**Results** Funnel plots based on data using healthy imputation for missing values can highlight centres lacking good data quality. A comet plot illustrates the performance trend over multiple years, which is more indicative of a centre's performance compared to a single measurement. Trends analysis offers the opportunity to closely monitor an individual centres' performance and direct evaluation of initiated improvement strategies.

**Conclusion** This study describes the use of funnel and comet plots as a method to monitor and assure high-quality data and to evaluate trauma centre performance over multiple years. Moreover, this is the first study to provide a regulatory blueprint on how to interpret and act on the under- or overperformance of trauma centres. Further evaluations are needed to assess its functionality.

**Level of evidence** Retrospective study, level III.

**Keywords** Funnel plot · Trauma system · Performance trend · Standardized mortality ratio · Trauma care

✉ M. L. S. Driessen
mls.driessen@lnaz.nl

E. W. van Zwet
e.w.van_zwet@lumc.nl

L. M. Sturms
lsturms@gmail.com

M. A. C. de Jongh
m.d.jongh@nazb.nl

L. P. H. Leenen
l.p.h.leenen@umcutrecht.nl

1 Dutch Network for Emergency Care (LNAZ), Newtonlaan 115, 3584 BH Utrecht, The Netherlands

2 Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, The Netherlands

3 Brabant Trauma Registry, Network Emergency Care Brabant, Tilburg, The Netherlands

4 Department of Surgery, University Medical Centre Utrecht, Utrecht, The Netherlands

## Introduction

Monitoring the performance of medical centres based on patient outcomes has become an essential part of modern health care [1]. Classic league tables are an established technique for displaying the comparative performance ranking of organizations. Previous research suggests that these rankings aimed to generate a stimulus to initiate improvements [2]; however, they generally inflict stigmatization on centres rated as "poor performers", which has a negative effect on public trust and professional morale [3, 4].

The use of funnel plots has been suggested as a standard method for institutional comparisons using cross-sectional data [5–9]. Funnel plots are a graphical tool used to present centre comparisons while avoiding ordering or ranking of centres [6]. Moreover, they clearly visualize the relationship between sample size and precision since the control limits and the distribution become narrower with higher volumes. The control limits indicate a range in which the values of the quality indicator would be expected to fall. Centres exceeding these control limits may be considered underperforming or overperforming, prompting an investigation into their practices. In addition, quality can be improved by learning from good performing centres (i.e., adopting best practice methods) and initiating improvement strategies.

The funnel plot methodology has been applied previously in a trauma setting to evaluate and compare mortality rates and hospital length of stay between centres [10, 11]. However, there are two main issues that need to be addressed when evaluating trauma centre performance. First, the quality of the data needs to be assured, as mortality rate prediction is less accurate in cases of incorrectly entered or missing data. Second, because trauma populations can vary widely between centres, it is important to ensure that a centre's performance is investigated rather than focusing on differences in case variability. Indirect standardization can be used to overcome problems resulting from comparing centres with different degrees of injury severity [12]. However, individual centres are, even after standardization, not directly comparable because each centre's' own population is used to calculate the expected outcomes. The indicator thus shows how well a centre performs within its own population in comparison to the performance of the reference standard. In current practice, the comparison between centres is overemphasized when evaluating healthcare-related outcomes. We believe that assessing a centre's performance trend for its designated trauma population over multiple years is extensively more interesting, assuming that the patient population remains relatively stable. Thus, the aim of the present study is to illustrate that funnel plots can be used as a method to control the quality of data and assure an optimal level of trauma care by evaluating centre performance trends over time. Moreover, a set of guidelines on how to interpret and act on results presented in funnel plots, including underperformance as well as overperformance trends, will be presented.

## Materials and methods

### Study population

A retrospective observational cohort study was performed using the Dutch National Trauma Registry (DNTR) [13]. The DNTR documents all injured patients directly admitted to a centre through the emergency department (ED) within 48 h after trauma, regardless of their age, injury location and severity. Patients arriving at the ED without vital signs were excluded [13]. For this study, all patients recorded in the DNTR in 2020 were included. To illustrate trends in performance over multiple years, standardized mortality ratios of two level-I trauma centres between 2017 and 2019 were additionally calculated. Each dot in the comet plots shows the performance over 1 year, yet the time frame moves three months forward with each dot. In other words, from point to point, three quarter of the data is identical, and one quarter is new. By doing so, the points in the graph move slowly 'like a comet'. Without this feature (for example, if annual data were used without overlap), points would jump around more.

This study was exempted from ethics review board approval because the study used existing coded data from the DNTR, and patient anonymity was guaranteed. Neither patients nor members of the public were involved in the design, execution, reporting, or dissemination plans of our study. The DNTR dataset includes the Utstein template items for uniform reporting of data following major trauma and covers 100% of the trauma-receiving centres in the Netherlands [14].

The DNTR includes 86 trauma centres, 13 of which are designated level-I trauma centres [13, 15]. Injuries are coded according to the Abbreviated Injury Scale (AIS) 2005, update 2008 [16]. The Injury Severity Score (ISS) is calculated from the three most affected body regions as the sum of squares of the respective AIS severity levels [17]. Patients with ISS scores of 16 or above are classified as severely injured.

### Statistical analysis

In-hospital mortality was used as the main indicator of a centre's performance. To describe patient characteristics, centres were divided into two groups according to their level of expertise, following the ASCOT guidebook entitled Optimal Resources for Care of the seriously Injured [15]. Level-I trauma centres are fully equipped to deliver the highest level of emergency and surgical care for the most severely injured, with 24/7 coverage of all specialities, including thoracic and neurosurgery. Lower-level trauma centres (i.e., level-II and level-III) provide optimal care for moderately and mildly injured patients in a cost-effective manner.

Missing values were imputed using multiple imputation, assuming missing values at random [18]. We used Multivariable Imputation by Chained Equations (R-package mice) for multiple imputations of missing case-mix variables [19, 20]. To assess the value of complete data for evaluating funnel plots, we generated a second dataset where missing values were imputed with normal healthy values. For example, when the ISS is missing, the lowest possible score of 1 is recorded, and if the American Society of Anaesthesiologists (ASA) score for comorbidity is missing, a score of 1 (no comorbidity) is recorded. The number of missing values per variable is listed in Table 1 of the supplemental material.

The standardized mortality ratio (SMR) is the ratio of observed deaths to the expected number of deaths or the observed mortality rate to the expected mortality rate [12]. A ratio of 1 means that a particular centre performs exactly as expected based on its population characteristics. A value above one indicates more deaths recorded than the reference model predicts, while a value less than one represents fewer deaths recorded than expected.

To account for differences in patient characteristics between centres, the expected in-hospital mortality rate was calculated with the use of a recently published modified Dutch version of the Trauma Injury Severity Score (mTRISS-NL) [21]. This mortality prediction model uses polynomial transformation of classic TRISS variables and is able to accurately predict mortality rates for all acutely admitted trauma patients. The model includes the variables sex, ASA class, nonlinear transformations of age, systolic blood pressure (SBP), respiratory rate (RR), Injury Severity Score (ISS) and best motor response (BMR) [17, 22, 23]. Using this mortality prediction model, the expected probability of in-hospital mortality after trauma is determined for each patient, and these SMRs can be added to encompass all patients treated at a specific centre over a specified period of time.

## Control limits

The funnel plot is so named because of the funnel shape of the control limits or prediction intervals. The prediction interval is calculated around the SMR and is based on its precision. The precision of the SMR increases with sample size and injury severity. Therefore, wide prediction intervals occur with small patient numbers, and narrower prediction intervals occur with large patient numbers. The control limits for the funnel plots were set at 95% and 99.8% prediction intervals, corresponding to approximately 2 and 3 standard error widths, respectively. Centres that perform similarly to the reference population have a 5% chance of exceeding the limits, 2.5% at the upper limit and 2.5% at the lower limit. Estimates falling outside the control limits represent the centres showing a wider deviation from the estimate than the deviation expected because of chance alone.

**Table 1** Patient characteristics in the dataset set using healthy imputations, for patients treated at level-I or level-II and level-III trauma centers

| | All centers | Level-I | Level-II and level-III |
|---|---|---|---|
| Number of cases (%) | 71,613 (100) | 17,130 (23.9) | 54,483 (76.1) |
| Median age (IQR) | 66.1 (30.5–81.7) | 55.5 (24.6 -75.2) | 69.4 (34.7–83.0) |
| Male gender, *n* (%) | 35,229 (49.2%) | 9,997 (58.4%) | 25,232 (46.3%) |
| Blunt injuries, *n* (%) | 69,131 (96.5%) | 16,127 (94.1%) | 53,004 (97.3%) |
| Median ISS (IQR) | 6 (4–9) | 9 (4–12) | 5 (4–9) |
| Median RR | 14 (11–18) | 16 (14–20) | 14 (11–18) |
| Median SBP | 135 (109–157) | 139 (120–158) | 136 (109–158) |
| Median BMR | 6 (6–6) | 6 (6–6) | 6 (6–6) |
| ASA, *n* (%) | | | |
| I | 32,546 (45.4%) | 8624 (50.3%) | 23,922 (43.9%) |
| II | 24,054 (33.6%) | 5452 (31.8%) | 18,602 (34.1%) |
| III | 14,153 (19.8%) | 2897 (16.9%) | 11,256 (20.7%) |
| IV | 839 (1.2%) | 149 (0.9%) | 690 (1.3%) |
| V | 21 (0.0%) | 8 (0.0%) | 13 (0.0%) |
| ICU admission, *n* (%) | 3706 (5.2%) | 2438 (14.2%) | 1268 (2.3%) |
| Mortality, *n* (%) | 1963 (2.7%) | 859 (5.0%) | 1104 (2.0%) |

*ISS* Injury Severity Score, *SBP* systolic blood pressure, *RR* respiratory rate, *BMR* best motor response, *ASA* American Society of Anesthesiologists, *ICU* intensive care unit

## Results

A total of 71,613 acutely admitted patients were included in this study. Most ($n = 54,483$ (76.1%)) of these patients were admitted to a level II or III trauma centre (Table 1). A total of 4671 severely injured patients (ISS > 15) were included, and 3299 (70.6%) were treated at one of the 13 level-1 trauma centres, with a range of 76 to 452 severely injured patients per centre per year. The average mortality rate for all trauma-receiving hospitals was 2.7%, varying from 2.0% for level-II and level-III trauma centres and 5.0% for level-I trauma centres.

## Funnel plot

Two funnel plots showing the standardized mortality ratio for all trauma-receiving centres in the Netherlands are shown in Fig. 1. The funnel plot derived from the dataset with multiple imputations (Fig. 1a) shows a clear distribution of level-I, level-II and level-III trauma centres. Precision increases when a centre has a high patient volume or treats a large number of more severely injured patients. As level-I trauma centres treat a higher volume of severe patients, the precision of level-I trauma centres is generally higher than level-II and III centres, and is thus positioned at a narrower location between the prediction intervals of the funnel plot. In the upper prediction intervals (PI), there are three underperforming level-II or level-III centres and one at the

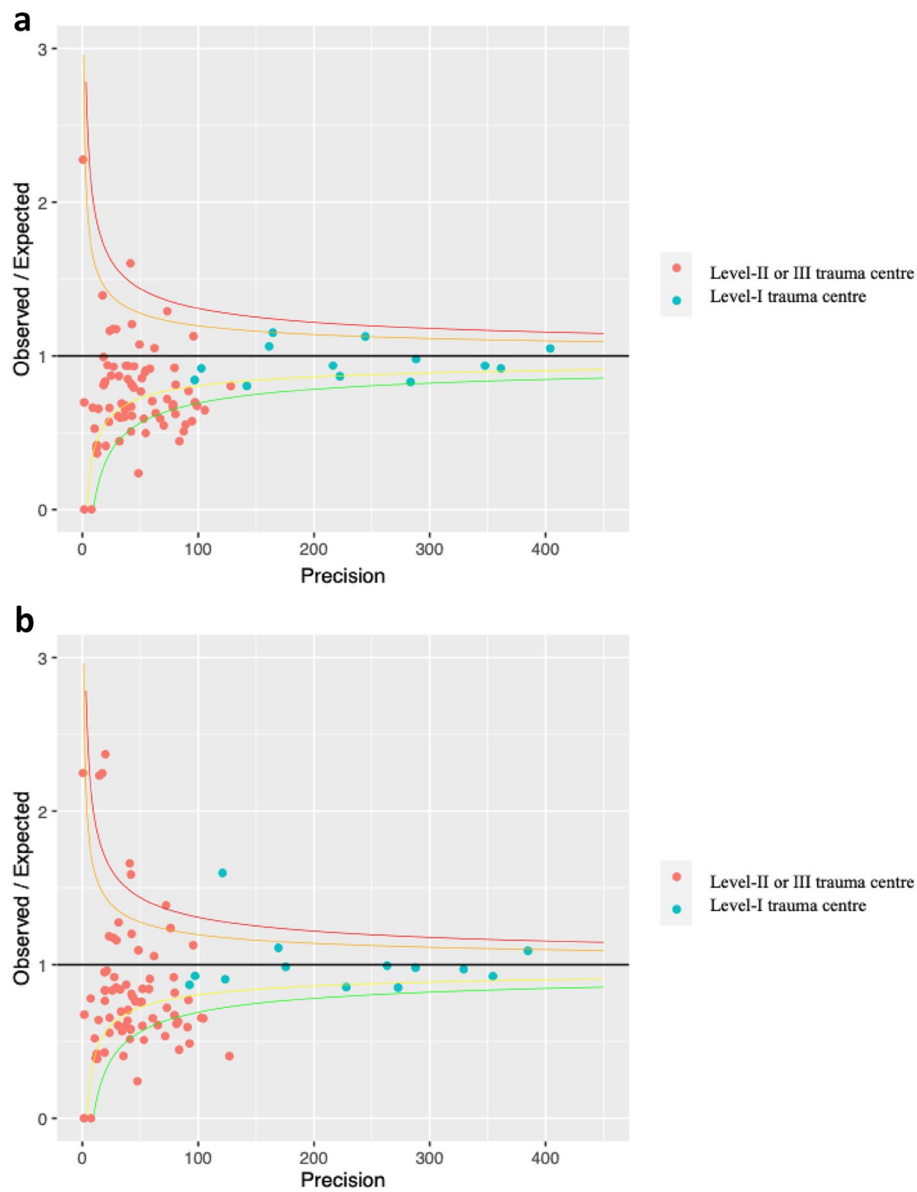


**Fig. 1** Funnel plot showing the standardized mortality ratio for all trauma receiving centres in the Netherlands, **a** for the multiple imputed dataset and **b** for the healthy imputed data set. The inner orange and yellow lines are the 95% and the outer red and green lines are the 99.8% confidence intervals. Note that several normal performing centres in **a** are underperforming in **b**

95% PI that warrant attention. Furthermore, there are eleven (15%) overperforming level-II or level-III centres positioned outside of the 99.8% PI, and twenty-one (28.8%) in between the 95% and 99.8% PI. Of the level-I trauma centres, two (15.4%) are on the upper 95% PI line, and three are positioned between the lower 95% and 99.8% PI intervals. Comparing Fig. 1a and b (i.e., comparing funnel plots derived from multiple and healthy imputed data) shows some interesting deviations. Figure 1b shows that six level-II or level-III centres and one level-I centre are positioned outside the upper 99.8% PI, indicating a worse performance than expected. By comparing the patient characteristics from the dataset using healthy imputations (Table 1) with those using multiple imputations (Table 2), we can see that the median systolic blood pressure and the ASA score for comorbidities vary between the datasets. Lower ASA scores (i.e., fewer preinjury comorbidities) as a result of using healthy imputations due to the number of missing values. The derived prediction (expected mortality) is lower, while the observed mortality is the same. This will increase the SMR, indicating worse performance. Only if there were no cases with missing data, the points would be exactly the same.

### Comet plot

A performance trend over multiple years gives additional insight into a centre's performance compared to a single measurement. Figure 2 shows the performance trend of a level-I trauma centre since 2018. The increasing dot size indicates its direction, with larger dots indicating more recent performance. Note that each dot in the SMR performance trend shows the performance over 1 year, yet the time frame moves three months forward with each dot. From this comet plot, we learn that this particular centre's precision is increasing (i.e., indicating more patients or more severely injured patients), while overall performance remains relatively stable. The comet shown in Fig. 3 illustrates an unfavourable performance trend of a level-I trauma centre. The evaluation at the time revealed a problem in the care of a specific trauma patient subgroup. After assessment and change in policy, the trend reverted.

## Discussion

This national retrospective observational study illustrates how funnel plots can be used to closely monitor the quality of data and trauma care. Moreover, by following and comparing standardized mortality trends in comet plots, we can identify both favourable and unfavourable effects that changes in, for example, a centres' organizational structure have on patient outcomes. Funnel and comet plots facilitate an independent evaluation of a centre's trauma performance, moving away from hierarchical intercentre comparison rankings.

### Regulations

We must strongly emphasize that crossing the upper or lower prediction interval does not directly indicate lower or higher "quality" of trauma care. Nevertheless, it should serve as a warning sign that prompts an investigation into the possible

**Table 2** The pooled numbers and medians of patient characteristics using the five multiple imputation data sets, for patients treated at level-I or level-II and level-III trauma centers

| | All centers | Level-I | Level-II and level-III |
|---|---|---|---|
| Number of cases (%) | 71,613 (100) | 17,130 (23.9) | 54,483 (76.1) |
| Median age (IQR) | 66.1 (30.5–81.7) | 55.5 (24.6–75.2) | 69.4 (34.7–83.0) |
| Male gender, *n* (%) | 36,527 (51.0%) | 10,015 (58.5%) | 25,713 (47.2%) |
| Blunt injuries | 69,008 (96.4%) | 16,126 (94.1%) | 52,980 (97.2%) |
| Median ISS (IQR) | 6 (4–9) | 9 (4–12) | 5 (4–9) |
| Median RR | 16 (14–20) | 16 (14–20) | 16 (14–20) |
| Median SBP | 140 (120–159) | 136 (120–156) | 140 (121–160) |
| Median BMR | 6 (6–6) | 6 (6–6) | 6 (6–6) |
| ASA, *n* (%) | | | |
| I | 27,325 (38.2%) | 7826 (45.7%) | 19,518 (35.8%) |
| II | 27,403 (38.3%) | 5997 (35.0%) | 21,405 (39.3%) |
| III | 15,924 (22.2%) | 3150 (18.4%) | 12,773 (23.4%) |
| IV | 928 (1.3%) | 169 (1.0%) | 769 (1.4%) |
| V | 25 (0.0%) | 9 (0.1%) | 16 (0.0%) |
| ICU admission, *n* (%) | 3709 (5.2%) | 2439 (14.2%) | 1270 (2.3%) |
| Mortality, *n* (%) | 1963 (2.7%) | 859 (5.0%) | 1104 (2.0%) |

*ISS* Injury Severity Score, *SBP* systolic blood pressure, *RR* respiratory rate, *BMR* best motor response, *ASA* American Society of Anesthesiologists, *ICU* intensive care unit

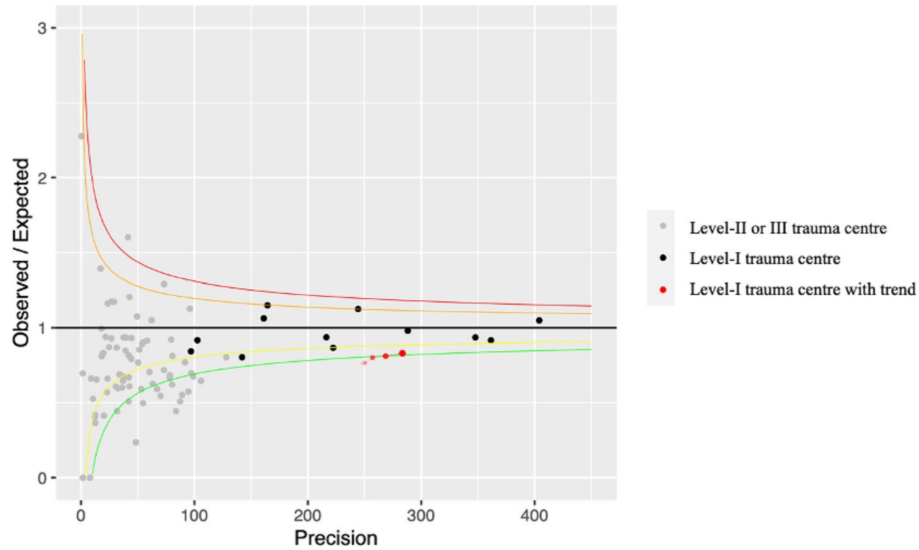


**Fig. 2** Funnel plot showing the standardized mortality ratio for all trauma receiving centres in the Netherlands. The inner orange and yellow lines are the 95% and the outer red and green lines are the 99.8% confidence intervals. Note that for one level-1 trauma center the trend since 2018 is shown

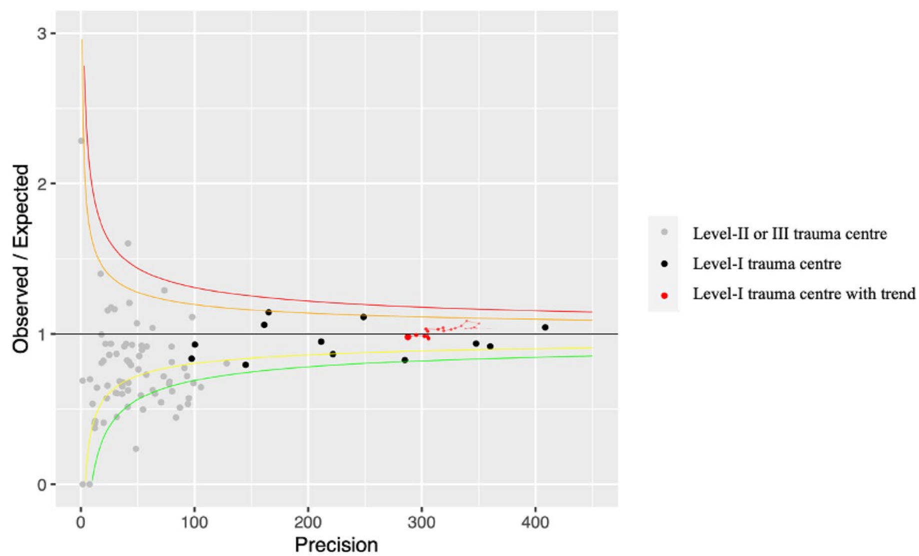


**Fig. 3** Funnel plot showing the standardized mortality ratio for all trauma receiving centres in the Netherlands, highlighting the trend for one level-1 trauma since 2017. The inner orange and yellow lines are the 95% and the outer red and green lines are the 99.8% confidence intervals

causes that could inflict this deviation. To successfully implement and regulate the use of funnel plots, an independent party should be appointed. In the example of the Dutch trauma system, this leading party is the Dutch Network for Emergency Care (Landelijk Netwerk Acute Zorg, in Dutch (LNAZ)). The LNAZ is an organization charged with overseeing and coordinating acute care within the Netherlands. Moreover, it is the overarching network organization of the eleven trauma regions in the Netherlands. Each trauma region has at least one level-I trauma centre. The leading trauma surgeons from the eleven trauma networks (Dutch Trauma Council of the LNAZ) authored a regulatory flowchart on how to manage trauma centres whose performance deviates from expectations (Fig. 4). This flowchart has been adopted by the board of the LNAZ, and the funnel and comet plots are generated based on DNTR data and distributed in a yearly report. If necessary, according to the flowchart, a centre has to investigate and clarify its SMR in case of underperformance.

## Importance of complete data

The evaluation of trauma care is highly dependent on the quality of the data. The accuracy of a funnel plot is as reliable as the data supporting it. For this reason, funnel plots were generated based on both multiple and healthy imputations. The dataset using healthy imputations is intended to expose centres that might have an issue in their registration. The imputation process using healthy scores can lead to an increased underestimation of expected mortality in the case of missing values. This becomes clear after comparison with funnel plots based on multiple imputations for missing values. In the dataset used for this study, the evaluation process showed that the ASA score (the variable for comorbidities)

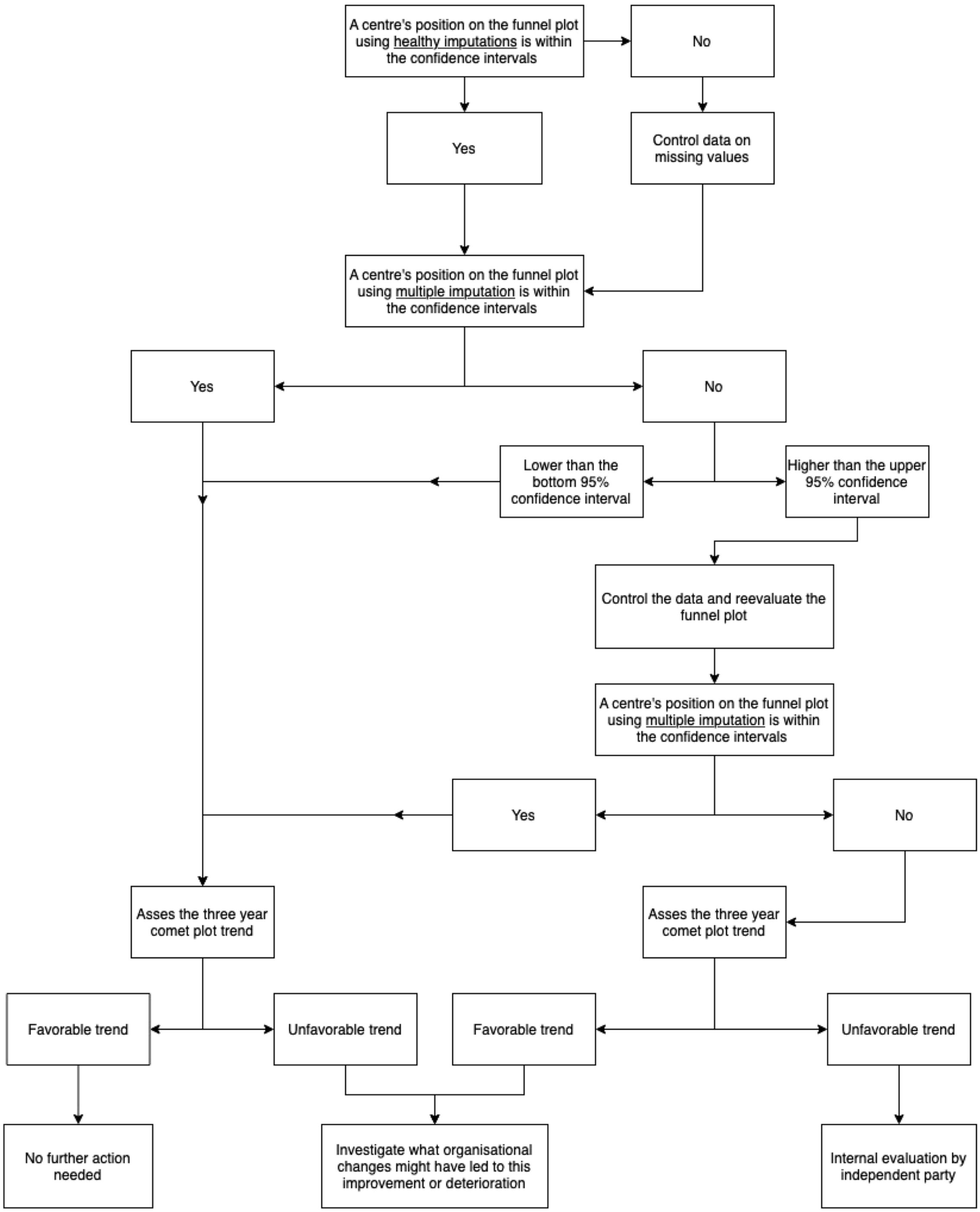


**Fig. 4** Dutch regulatory flowchart

was missing for several centres. Healthy imputations resulted in SMRs above the upper prediction intervals in the funnel plot. However, if there were no cases with missing data, the points for multiple and healthy imputations would be exactly the same.

After the initial evaluation of funnel plots using healthy imputations designed to filter out poor performance due to missing values, funnel plots using multiple imputations were generated and distributed to the individual trauma regions. Similar to the process in the first step, the centres that significantly deviate from what is expected are notified of the possibility of lacking data, giving them the opportunity to review and improve the quality of their supplied data. The accuracy of in-hospital mortality rate predictions can be assessed by evaluating if the data on the deceased are entered correctly, for example, if the AIS- or ASA-scores are accurately registered. After this quality control step, new funnel plots were generated. Moreover, in addition to evaluating the performance of a particular year, the general trend over multiple years is assessed.

There are two situations in which a centre is asked to self-conduct a local investigation to assess whether any intentional or unintentional organizational changes have been made that could have either led to an improvement in or deterioration of trauma care. This will be initiated in cases, where after carefully reviewing the data, a centre is positioned below the lower 95% prediction interval of the funnel plot and the general performance trend shows an unfavourable path (trend 03 in Fig. 5), or the centre is positioned above the 95% prediction interval and the trend shows a favourable path (trend 04 in Fig. 5).

In the unfortunate scenario where a hospital finds itself positioned above the 95% prediction interval and the general performance trend shows an unfavourable path (trend 06 in Fig. 5), an independent party (such as the LNAZ) is asked to investigate the locally delivered trauma care. During such an investigation, the injury characteristics of the deceased, as well as the surgical interventions executed and admission descriptions, need to be evaluated. Any organizational changes made in the previous years that might have impacted patient outcomes need to be reported and reassessed based on their initial targets. In this way, any unfavourable effects

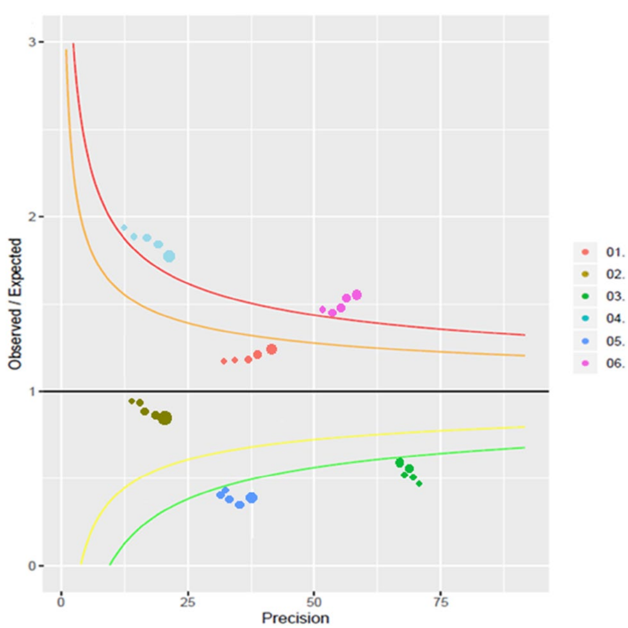


**Fig. 5** Six possible hospitals' standardized mortality ratio performance trends, based on fictive data. The inner orange and yellow lines are the 95% and the outer red and green lines are the 99.8% prediction intervals

can be detected at an early stage, and transparent reporting will serve as a learning opportunity for other centres. Moreover, the funnel plot works both ways, overperforming centres with descending favourable trends (trend 02 and 05 in Fig. 5) on the comet plots are asked to assess what organizational or quality improvement changes might have led to the improved outcomes for their patients.

Generally, a hospital's performance will be positioned between the 95% prediction intervals. However, this does not mean there is nothing that can be achieved or questioned. The particular case presented in Fig. 3 of this study is an excellent example of this. This hospital was situated within the 95% prediction intervals, just slightly above the midline. However, when assessing the trend of the funnel plot, an unfavourable deviation was found. Their internal evaluation revealed that an increased number of deaths among patients admitted after sustaining a hip fracture caused the centre's performance to deviate from its trend. After carefully evaluating the situation, an improved postoperative care path was established. As illustrated, this initiative quickly reversed its unfavourable trend in the following years.

## Future perspectives

The funnel plot methodology is currently being tested in the Netherlands. Therefore, knowledge of how to interpret and act upon a specific position and a started trend in the funnel plot is essential to successfully implement and regulate the use of funnel plots. The examples and the comprehensive flowchart presented in this paper will serve as a supporting guideline for surgeons, data managers and others involved. The next step after implementation will be creating a transparent platform (i.e., in the context of a digital environment) that facilitates hospitals to ask questions and browse through historical funnel plot content. Moreover, organizing an annual meeting to exchange both (un)successfully implemented organizational or quality improvements, would be of the essence to optimize the use of the presented methodology.

Funnel plots provide centres with insight into their performance on a specific outcome variable within their own patient population. Moreover, funnel plots clearly visualize the relation between sample size and precision, i.e., the control limits and the distribution of centre outcomes decrease with increased patient volume. The presentation of volume on the x-axis also provides the opportunity to observe an association between volume and outcome. Currently, major trauma patient volumes in the Netherlands are too low to demonstrate an allegedly beneficial effect of volume on mortality. Future analyses are needed to assess whether there is an volume effect on mortality or functional outcomes [24].

Even though approximately 95% of the Dutch trauma population survive their injury, mortality remains the main outcome measure [13]. Other outcomes, such as the Glasgow Outcome Scale [25] or the patient-reported outcome measure (PROM), would be interesting options to use to measure trauma care quality. However, several obstacles remain. First, when selecting a PROM, the purpose of healthcare quality evaluation must be taken into account. It is possible that a PROM serves as an important measure on an individual level, but it is not suitable for comparing health-related quality of care. To evaluate healthcare quality, PROMs should be selected for situations where an association with healthcare quality is plausible or established [26]. Second, adequate adjustment for case-mix correction is needed, as multiple influential factors, such as age, sex, educational level, type of injury, injury severity, frailty, comorbidity and duration of hospital stay, are relevant in predicting health status after injury [27, 28]. Ideally, a case-mix model is developed, enabling comparison between observed outcome and preinjury health status [26]. Previous efforts to develop such prediction models in trauma care resulted in models with an explained variance of almost 50 percent [27]. Third, PROMs can be more challenging to obtain than clinical outcomes. Because PROMs can only be observed and registered by the patients themselves, it is more difficult and time-consuming to collect complete data on fixed time points. Moreover, predictors such as education levels, preinjury health status and frailty are generally not routinely assessed in trauma registries.

### Strengths and limitations

This study has several strengths. First, we used data from the national trauma registry, which includes detailed data on all acutely admitted trauma patients in all Dutch centres registered by trained data managers [13]. By imputing missing values using “normal” healthy scores, we aimed to increase the quality of our data and avoid the presentation of overly optimistic results by punishing those who failed to deliver complete data. Third, the method of indirect standardization enables centres to reflect on the quality of trauma care given to the population they were designated to treat. This offers the opportunity to directly evaluate initiated improvement strategies. Fourth, to the best of our knowledge, this is the first study that describes regulatory proceedings related to a centre’s performance illustrated in a funnel plot using trends over multiple years.

This study also has limitations. Although the model used for case-mix correction has good accuracy, it does slightly overestimate mortality for severely injured individuals, possibly showing a more positive centre performance [21]. Second, because this study presents the initial blueprint of the regulations yet to be fully implemented in the Netherlands, experiences or flaws in the system have not yet been reported. Further evaluations are needed to assess its functionality. Third, lower-level trauma centres with a low number of cases may not observe a trauma-related death. The SMR makes its position within the funnel plot more susceptible to volatility. In these particular cases, the difference between observed and predicted mortality would be the preferred method to be used. However, further studies are needed to show whether the performance of any centre will be significantly divergent and whether any serious repercussions are needed. Fourth, a more detailed alternative to illustrate individual hospital performance could be achieved with the use of risk-adjusted cumulative sum charts. However, we deliberately chose to use funnel plots because they illustrate the position of an individual centre among other centres. Although centres cannot be compared due to differences in case types, it is of interest to observe a centre’s position within the acute care landscape.

## Conclusion

This study describes the use of funnel and comet plots as a method to assure high-quality data and to monitor and evaluate trauma centre performance over multiple years. Moreover, this is the first study to provide a regulatory blueprint on how to interpret and act on the under- or over-performance of trauma centres.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00068-022-02100-z.

**Acknowledgements** The authors would like to thank the leading trauma surgeons that participate in the Dutch Trauma Council for their efforts in realizing these guidelines.

**Author contributions** van Zwet, Driessen, de Jongh, Leenen and Sturms conceived and designed the study. van Zwet analyzed the data. Interpretation of the data and writing the first draft of the paper: Driessen. All authors contributed to writing the paper and approved the final version. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

**Transparency statement** The lead author (Mitchell Driessen) affirms that the manuscript is an honest, accurate and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned have been explained.

## Declarations

**Conflict of interest** All authors declare to have no conflict of interest, including financial, consultant, institutional, and other relationships that might lead to bias or a conflict of interest.

## References


1. Porter ME. What is value in health care? N Engl J Med US. 2010;363:2477–81.
2. Brook RH. Health care reform is on the way: do we want to compete on quality? Ann Intern Med. 1994;120:84–6.
3. Adab P, Rouse AM, Mohammed MA, Marshall T. Performance league tables: the NHS deserves better. BMJ. 2002;324:95–8.
4. Smith P. On the unintended consequences of publishing performance data in the public sector. Int J Public Adm. 1995;18:277–310.
5. Spiegelhalter D. Funnel plots for institutional comparison. Qual Saf Health Care. 2002;11:390–1.
6. Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med Engl. 2005;24:1185–202.
7. Mohammed MA, Deeks JJ. In the context of performance monitoring, the caterpillar plot should be mothballed in favor of the funnel plot. Ann Thorac Surg Neth. 2008;86:348.
8. Mayer EK, Bottle A, Rao C, Darzi AW, Athanasiou T. Funnel plots and their emerging application in surgery. Ann Surg US. 2009;249:376–83.
9. van Dishoeck AM, Looman CWN, van der Wilden-van Lier ECM, Mackenbach JP, Steyerberg EW. Displaying random variation in comparing hospital performance. BMJ Qual Saf Engl. 2011;20:651–7.
10. Wang W, Dillon B, Bouamra O. An analysis of hospital trauma care performance evaluation. J Trauma US. 2007;62:1215–22.
11. Kirkham JJ, Bouamra O. The use of statistical process control for monitoring institutional performance in trauma care. J Trauma US. 2008;65:1494–501.
12. Naing NN. Easy way to learn standardization : direct and indirect methods. Malays J Med Sci. 2000;7:10–5.
13. Driessen MLS, Sturms LM, Bloemers FW, ten Duis HJ, Edwards MJR, den Hartog D, et al. The Dutch nationwide trauma registry: the value of capturing all acute trauma admissions. Injury. 2020;51(11):2553–9.
14. Ringdal KG, Coats TJ, Lefering R, Di Bartolomeo S, Steen PA, Røise O, et al. The Utstein template for uniform reporting of data following major trauma: a joint revision by SCANTEM, TARN, DGU-TR and RITG. Scand J Trauma Resusc Emerg Med. 2008;16:7.
15. Rotondo M, Cribari C SS. Resources for optimal care of the injured patient. Bull Am Coll Surg. 2014.
16. Gennarelli TA, Wodzin E. AIS 2005: a contemporary injury scale. Injury. 2006;37(12):1083–91.
17. Baker SP, O'Neill B, Haddon W, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma. 1974;14:187–96.
18. Shrive FM, Stuart H, Quan H, Ghali WA. Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. BMC Med Res Methodol. 2006;6:57.
19. R Core Team (2020) R: a language and enviroment for statistic computing. R Foundation for Statistical Computing, Vienne Austria. 2020.
20. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw. 2011. https://doi.org/10.18637/jss.v045.i03.
21. Driessen MLS, van Klaveren D, de Jongh MAC, Leenen LPH, Sturms LM. Modification of the TRISS: simple and practical mortality prediction after trauma in an all-inclusive registry. Eur J trauma Emerg Surg. 2022. https://doi.org/10.1007/s00068-022-01913-2.
22. Dripps RD. New classification of physical status. Anesthesiology. 1963;24:111.
23. Teasdale G, Jennett B. Assessment and prognosis of coma after head injury. Acta Neurochir (Wien). 1976. https://doi.org/10.1007/BF01405862.
24. Sewalt CA, Venema E, van Zwet E, van Ditshuizen JC, Schuit SCE, Polinder S, et al. The relationship between hospital volume and in-hospital mortality of severely injured patients in dutch level-1 trauma centers. J Clin Med. 2021;10(8):1700.
25. Jennett B, Bond M. Assessment of outcome after severe brain damage. A practical scale. Lancet. 1975;305(7905):480–4.
26. van der Willik EM, van Zwet EW, Hoekstra T, van Ittersum FJ, Hemmelder MH, Zoccali C, et al. Funnel plots of patient-reported outcomes to evaluate health-care quality: basic principles, pitfalls and considerations. Nephrology (Carlton). 2021;26:95–104.
27. de Munter L, Polinder S, van de Ree CLP, Kruithof N, Lansink KWW, Steyerberg EW, et al. Predicting health status in the first year after trauma. Br J Surg Engl. 2019;106:701–10.
28. Maxwell CA, Mion LC, Mukherjee K, Dietrich MS, Minnick A, May A, et al. Preinjury physical frailty and cognitive impairment among geriatric trauma patients determine postinjury functional recovery and survival. J Trauma Acute Care Surg. 2016;80:195–203.