The ROYAL COLLEGE of OPHTHALMOLOGISTS

ARTICLE

Graphical comparison of surgeon outcomes for the audit of a national corneal transplant registry (OTAG study 32)

Cathy Hopkinson [1✉], Elinor Curnow[1], Daniel F. P. Larkin[2], Jeremy Prydal[3] and Stephen Tuft[2]

PURPOSE: To compare Kaplan–Meier survival curves and funnel plots for the audit of surgeon-specific corneal transplantation outcomes.

METHODS: We obtained data on all patients with Fuchs endothelial dystrophy (FED) receiving a first corneal transplant in one eye between January 2012 and December 2017. We produced 2-year Kaplan–Meier graft survival curves to compare a simulated individual surgeon's graft survival rate to national pooled data. We used funnel plots to compare all surgeon outcomes to the national graft survival rate with superimposed 95 and 99.8% confidence limits. We defined an outlier as a surgeon who performed ≥10 transplants and had graft survival below the 99.8% national lower limit. To assess the effect of the surgeon case mix, we also compared unadjusted and risk-adjusted graft survival rates.

RESULTS: There were 3616 first corneal transplants for FED patients with complete data, performed or overseen by 196 surgeons. The 2-year national graft survival rate was 88%. The median change from the unadjusted to the risk-adjusted graft survival rate for individual surgeons was 0% (IQR: 0%– −2%). Of the 108 surgeons who had performed ≥10 transplants, we identified two outliers based on the unadjusted graft survival funnel plot, compared to four outliers based on the risk-adjusted graft survival funnel plot.

CONCLUSION: Funnel plots provide a visually accessible method for comparing individual graft survival rates to the national rate. Risk-adjustment accounts for clinical factors, and this has advantages for audit and clinical governance.

INTRODUCTION

A corneal graft is one of the most frequent transplantation procedures performed worldwide. There were 4504 corneal transplants performed in the United Kingdom (UK) in the year 2019–2020, of which 35% were for Fuchs endothelial dystrophy (FED), 17% were re-grafts, 13% were for pseudophakic bullous keratopathy (PBK) and 13% were for keratoconus (KC) [1]. Endothelial keratoplasty (EK) rather than penetrating keratoplasty (PK) was performed in 97% of eyes with FED. The EK techniques are sub-divided as Descemet Stripping Endothelial Keratoplasty (DSEK) and Descemet Membrane Endothelial Keratoplasty (DMEK). Several factors influence the risk of graft failure, such as perioperative surgical complications, glaucoma, and allograft rejection, which is in turn affected by the number of prior transplants and corneal vascularisation [2–5]. Additional surgeon-specific risks for transplant failure are not well described but include surgeon experience, case workload, clinical support, and surgical competence [4].

The audit of transplant outcomes in the UK is an essential component of medical revalidation, and the NHS Standard Contract also mandates participation in clinical audits [6]. Regulatory authorities require eye banks to maintain records to ensure the traceability of all ocular tissues used for transplantation. Surgeons undertaking corneal transplantation must complete a standardised data collection form at the time of surgery; they are also strongly encouraged to return follow-up forms at 1, 2, and 5 years post-operatively. Transplant details include the primary corneal diagnosis, transplant type, previous ocular surgery, potential risk factors for transplant failure, transplanting centre and consultant. Follow-up data reports on critical events such as allograft rejection episodes, graft failure, visual acuity, and patient death. Corneal transplant outcomes are analysed for the three most frequent indications (FED, KC and PBK) and reported annually to each surgeon and transplant centre in the UK. A reference standard or benchmark is required to effectively audit outcomes, such as a national rate or an agreed clinical standard.

In addition to visual outcome, the rates of graft rejection and failure at a given time point post-transplant are used to quantify and compare the success of corneal transplantation. The standard analysis of these rates uses a time-to-event methodology, also known as survival analysis, which allows patients with incomplete follow-up data to be incorporated into survival rates. Kaplan–Meier estimates can make univariable rate comparisons, which is most effective when comparing a small number of independent groups, e.g., different graft types. When comparing numerous groups, e.g., multiple surgeons, Kaplan–Meier plots are not easy to interpret as multiple lines may overlap, and it is not possible to compare individual data to pooled national rates due to a lack of independence between the groups. In contrast, a funnel plot can display transplant activity and graft survival rates

[1]NHS Blood and Transplant, Fox Den Road, Stoke Gifford, Bristol BS34 8RR, UK. [2]Moorfields Eye Hospital, 162 City Road, London EC1V 2PD, UK. [3]Department of Ophthalmology, University Hospitals of Leicester NHS Trust, Infirmary Square, Leicester LE1 5WW, UK. ✉email: cathy.hopkinson@nhsbt.nhs.uk

for each surgeon and compare these to the national rate and its associated 95 and 99.8% confidence intervals. An unadjusted funnel plot reports the observed graft failure rate for each surgeon, but a disadvantage of this approach is that a surgeon with a low-risk case mix is likely to have a graft failure rate that is unfairly favourable when compared to the national average. Conversely, a surgeon with a high-risk case mix may be unfairly penalised. Using a risk-adjusted funnel plot can mitigate the effect of an unequal case mix. Based on each surgeon's case mix, the expected survival is compared with their observed survival, and the survival rate reported on the funnel plot adjusted accordingly. Visualisation of this data facilitates the interpretation of information for clinical audits and aids decision making [7, 8]. To illustrate the two methods, we compared Kaplan–Meier survival curves and funnel plots as tools for monitoring surgeon outcomes following the first corneal transplantation for FED.

MATERIALS AND METHODS
Data collection and cohort selection
Data were collected by NHS Blood and Transplant (NHSBT) for the UK Transplant Registry. To reduce potential variation in surgeon case mix we only included patients with FED who received a first corneal transplant (PK or EK) in either eye performed in the UK between January 2012 and December 2017. A transplant was recorded according to the consultant who was allocated the tissue, although surgery could have been performed by a trainee under their supervision. We did not specify that a surgeon had performed surgery throughout the 6-year audit period. All data were anonymised. In addition to the real data set, we also sampled this data to generate a simulated set of transplant data for a hypothetical surgeon. We used this simulated data to illustrate the two methods below. We performed a complete case analysis; hence, transplants with no follow-up data and missing risk factor data were excluded. The Ethics Committee of Moorfields Eye Hospital classified the study as an audit and exempt from review. The research methods and the analysis plan adhered to the tenets of the Declaration of Helsinki and UK law in relation to data collection and medical research.

Primary outcome
We analysed outcomes at 2 years post-transplant. The primary outcome measure was graft failure, caused by primary graft failure (no evidence of graft function post-transplant), irreversible rejection, infection, endothelial decompensation, a subsequent record in the database of a replacement graft in the same eye, or an unknown cause of failure. Individuals who died with a functioning graft were censored at the time of death, and patients with a functioning graft at the end of the observation period were censored at the last known follow-up.

Statistical methods
Kaplan–Meier plot. We used a Kaplan–Meier plot to show the change in graft survival over time from transplant until 2 years follow up. We did not use the log-rank test because the individual surgeon's data contribute to the pooled national data.

Funnel plot. A funnel plot shows individual survival rates plotted against the number of transplants performed (*n*) at a specified time point after transplantation. The funnel shape of the confidence limits reflects the increasing statistical precision as the number of transplants increases along the horizontal axis. In our example, we superimposed this with the national survival rate at 2 years post-transplant (estimated using the Kaplan–Meier method) and the 95 and 99.8% confidence limits around the national rate (calculated using the score method for the binomial distribution) [9]. We used the score method rather than the more commonly used normal approximation because the normal approximation is unsuitable for rates of 0 or 100% or a small *n*.

A funnel plot assumes that the mix of patients for each surgeon is the same as the national rate. We compared two different methods to address this assumption:

(1) We estimated unadjusted graft survival rates using the Kaplan–Meier method. We refer to this as the unadjusted funnel plot.
(2) We calculated risk-adjusted graft survival rates from a Cox regression model that included risk factors commonly associated with graft failure for FED patients (Table 1). The ratio of the observed to the expected failures was then multiplied by the national failure

Table 1. Clinical factors used to estimate the 2-year risk-adjusted graft survival rates for 3616 Fuchs endothelial dystrophy patients who received their first corneal transplant from 1 January 2012–31 December 2017.

Factor	Number	Failed	Hazard ratio	95% CI	*p* value
Graft type					
PK	224	26 (12%)	1.0	–	
DSEK	2823	287 (10%)	1.0	0.7–1.5	0.92
DMEK	569	99 (17%)	1.9	1.2–3.0	0.003
Pre-operative acuity					
Better than 6/60	2723	296 (11%)	1.0	–	
6/60 => <1/CF	444	47 (11%)	1.0	0.7–1.3	0.79
1/CF and worse	449	69 (15%)	1.4	1.1–1.8	0.02
Cataract procedure at the time of transplant					
No	2065	268 (13%)	1.0	–	
Yes	1551	144 (9%)	0.7	0.6–0.9	0.001
Procedure complications					
No	3585	404 (11%)	1.0	–	
Yes	31	8 (26%)	2.5	1.3–5.1	0.01
Risk of glaucoma					
No	3475	391 (11%)	1.0	–	
Yes	141	21 (15%)	1.3	0.8–2.0	0.25

The median patient age at the time of surgery was 73 years (IQR: 66–79). Note that glaucoma is a clinically relevant factor although it is not significant in this time period most likely due to small numbers.
PK penetrating keratoplasty, *DSEK* descemet stripping endothelial keratoplasty, *DMEK* descemet membrane endothelial keratoplasty, *CI* confidence interval, *CF* count fingers, *IQR* interquartile range.

rate. This value was then subtracted from 100% to give the risk-adjusted graft survival rate. We refer to this as the risk-adjusted funnel plot.
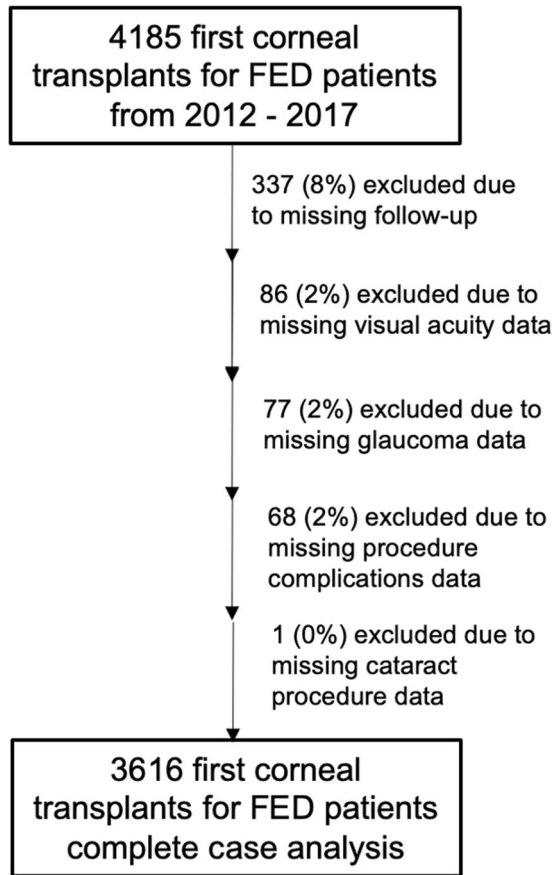

Fig. 1 Flow chart of data exclusion pathway. The total number of corneal transplants analysed in this study excluding transplants due to missing follow-up or risk factor data.

We defined an outlier as a graft survival rate outside the 99.8% confidence limits of the national rate. Due to the sensitivity of being identified as an underperforming outlier, the graft survival rates for outlying individuals, and the number of procedures, were slightly altered without affecting the study conclusions.

RESULTS

Between 2012 and 2017, there were 3616 first corneal transplants for FED with follow-up and complete risk factor data at 2 years, performed or overseen by 196 consultants. We excluded from the analysis 337 (8%) transplants without follow-up data and 232 (6%) transplants with missing risk factor data (Fig. 1). The return of transplant record forms was 100%, with 89% for 1-year follow-up and 86% for 2-year follow-up. The number of procedures allocated to an individual surgeon ranged from 1 to 126. The data set included 2823 (78%) DSEK, 569 (16%) DMEK and 224 (6%) PK procedures. The mean number of DSEK grafts performed was 18 (95% CI: 15–21), 8 for DMEKs (95% CI: 6–10) and 3 for PKs (95% CI: 2–4).

Kaplan–Meier plots (Fig. 2) show the estimated graft survival and 95% confidence limits versus time since transplant. The national graft survival rate was 88% at 2 years, compared with 70% for an individual surgeon (transplants, $n = 69$). A visual comparison of the confidence limits for these two curves suggests that graft survival for the surgeon is below the national survival rate. The unadjusted funnel plot compares all surgeons to the national rate 2 years post-transplant (Fig. 3). Of the 108 surgeons who performed ≥10 grafts, two (2%) had a graft survival rate below the 99.8% lower confidence limit. One of these two outlying surgeons was the individual described in Fig. 2, confirming the difference in graft survival suggested by the Kaplan–Meier plot.

Table 1 shows the Cox regression model estimates used to calculate the risk-adjusted graft survival rates at 2 years for each surgeon. After risk-adjustment, the graft survival rate of 70 (36%) surgeons stayed the same, 74 (38%) decreased and 52 (27%) increased. The median change (percentage points) from the unadjusted to the risk-adjusted graft survival rate was 0% (IQR: 0%– −2%). Those with the greatest change in graft survival performed fewer transplants. After risk-adjustment, four surgeons were identified as an outlier, suggesting that, for these surgeons,
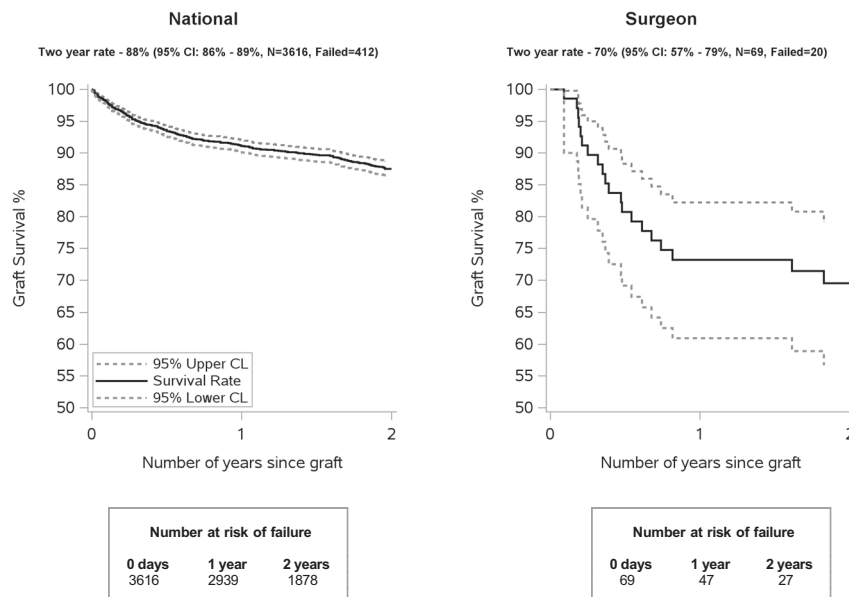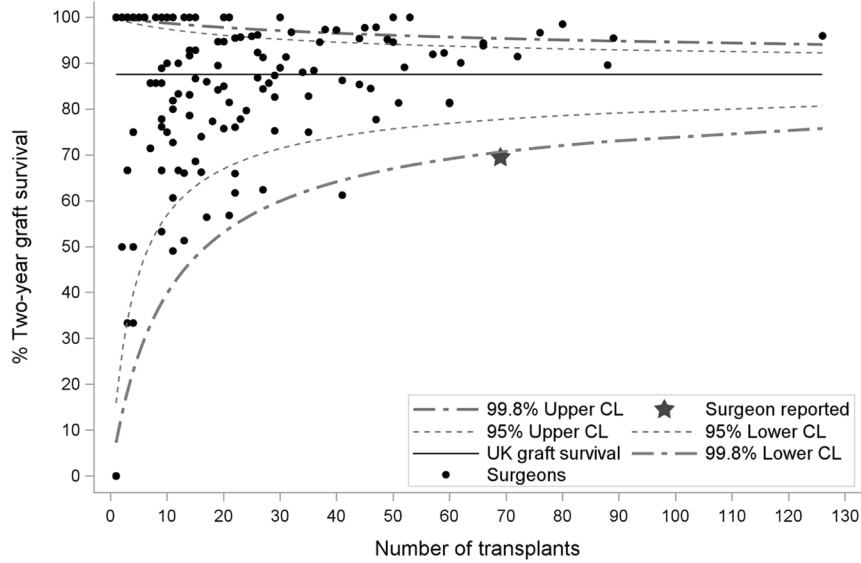

Fig. 2 Two-year graft survival for 3616 Fuchs endothelial dystrophy patients who received their first corneal transplant from 1 January 2012 to 31 December 2017. The left Kaplan–Meier plot shows the national pooled data, and the right compares a simulated individual surgeon with relatively low graft survival.

A. Unadjusted funnel plot
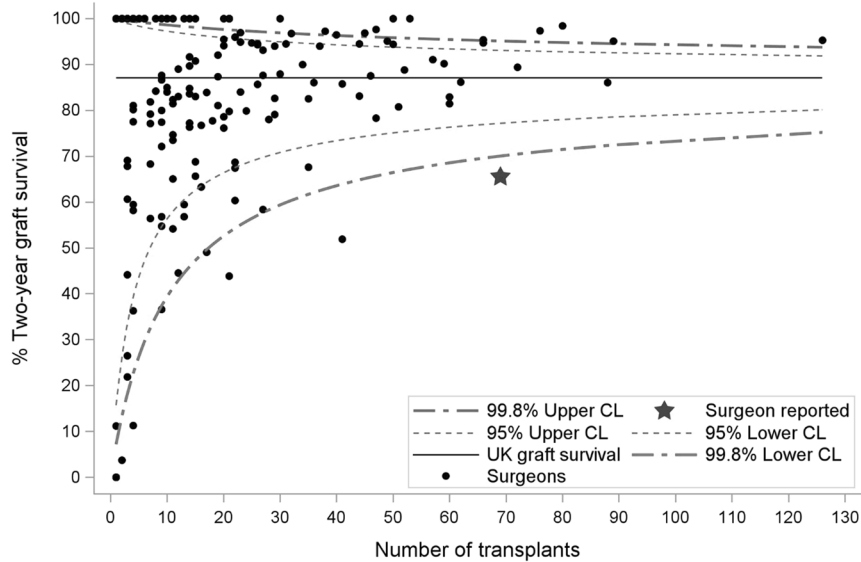

B. Risk-adjusted funnel plot

Fig. 3   Two-year graft survival for 3616 Fuchs endothelial dystrophy patients who received their first corneal transplant from 1 January 2012 to 31 December 2017. Funnel plot **A** shows all surgeons' unadjusted graft survival rates. Funnel plot **B** shows the risk-adjusted graft survival rates accounting for variation in surgeon case mix. We derived the national rate from Fig. 2. The simulated individual surgeon from Fig. 2 (star), had a risk-adjusted graft survival rate of 66%, slightly lower than their unadjusted rate of 70%.

Table 2. Comparison between Kaplan–Meier and funnel plots for presentation of outcome data.

	Kaplan–Meier plot	Funnel plot
Time-point	Any point since transplant	Single fixed time point
Possible comparisons	A small number of independent groups	Many and can be compared to national pooled data
Allows for incomplete follow-up	Yes	Yes
Easy identification of outliers	No	Yes
Easy to interpret	Yes	Yes
Displays confidence limits	Yes	Yes, for the national rate but not for individual graft survival rates

the transplant survival rate was worse than expected even after accounting for the surgeon case mix (Fig. 3). As before, the outlying surgeons included the individual described in Fig. 2.

DISCUSSION

Reporting outcomes is a good medical practice that provides evidence that a clinical service meets benchmark standards [10]. Since 1999 the NHSBT UK Transplant Registry has collected data on over 100,000 corneal transplants. Centre-specific reports have been generated since 2010. These give surgeons an overview of their centre's activity and outcomes, including graft survival, rejection rates and visual acuity. Initially, Kaplan–Meier survival curves were used to compare a single centre and all other centres, but this gave no information on the distribution of individual surgeon rates. In 2019, performance reports for individual surgeons were introduced that included funnel plots of 2-year graft survival outcomes. In this paper, we have demonstrated that funnel plots facilitate a comparison of the survival rates for individual surgeons versus the national rate, and they allow easier identification of outliers than Kaplan–Meier plots (Table 2). We have also shown that risk-adjustment can account for surgeon case mix by incorporating patient and transplant factors that may influence graft survival. While our unadjusted funnel plot does not distinguish the type of graft (PK or EK), the risk-adjusted model does (Table 1), although surgeon preference may determine the choice of procedure.

The task of improving health care by reporting outcomes is not straightforward. Identifying outliers based solely on statistical probability without regard for clinical circumstances may not be a reliable indicator of poor performance. The Health Quality Improvement Partnership has published quality measures relevant to the individual consultant, team, and surgical unit in the UK [11]. They have developed a protocol to manage instances where results are significantly lower than the national rate, with a definition of outliers based on a two-sided statistical approach with a 95% threshold (equivalent to a p value of 0.05) for 'alert' and a 99.8% (equal to a p value of 0.002) for 'alarm' [12]. NHSBT, supported by The Royal College of Ophthalmologists (UK), have adopted a similar approach to manage corneal transplant outliers [13]. Importantly, corneal surgeons have an opportunity to review their data annually and submit additional forms, particularly if they have been identified as an outlier. Upon reviewing the data, an outlying surgeon may also identify patients with risk factors for graft failure, a learning effect of adopting a new procedure, or an effect due to surgical supervision of trainees. Following data validation, additional clinical or statistical advice may be recommended, and this may also be an opportunity for surgeons with excellent outcomes to share techniques.

As a first iteration, we have used unadjusted funnel plots to monitor outcomes for surgeons who perform ≥10 first grafts for each indication (KC, FED and PBK) in a 6-year cycle, with results reported separately. We have only reported here the results of first grafts for FED, which excludes second eye grafts or re-grafts, and thus represents only a proportion of the total number of grafts performed by the surgeon in that period. This may explain the relatively low number of procedures for some surgeons. Notably, the threshold of ≥10 first grafts was determined from a visual inspection of the funnel plots rather than clinical evidence. Applying a threshold may avoid identifying surgeons with inconclusive results, but it also excludes 88 (45%) surgeons from this audit, some of whom may be low volume, poorly performing surgeons [10]. An alternative strategy to increase the number of surgeons above the audit threshold would be to consider all grafts performed by a surgeon and include the initial pathology as part of the risk-adjustment. Our aim is to apply risk-adjusted models to each surgeon's graft survival rate so that the funnel plots will

account for patient mix across the three main indications. We also hope to explore the performance of low volume surgeons in more detail to determine whether they should be audited separately. Finally, cases with missing risk factor data were excluded from this study, but in future applications of funnel plots we will evaluate multiple imputation methods to handle missing data to improve bias and precision estimates of graft survival [14].

The role of surgical volume or surgeon experience on graft survival has been reported previously, although it is limited to grouped surgeon data and first PK grafts [4, 15–17]. In KC, FED and PBK, there was no difference in transplant survival at 5 years between high volume (>30 transplants per year) compared to low volume surgeons (<10 transplants) [15]. Similarly, there was no difference in graft survival between centres performing more or less than 10 PK procedures per year [17]. Low-activity surgeons are unlikely to be a homogenous group because low activity can occur for several reasons, including early-career corneal surgeons, maternity leave, or retirement during the 6-year audit cycle. Surgeons with low activity tend to have greater uncertainty surrounding their graft survival estimate, demonstrated by a wide confidence interval. Although it is not a formal statistical test, overlapping confidence intervals for an outlying surgeon's rate and the national rate would indicate uncertainty about the true difference between the two. However, a limitation of the funnel plot is that it does not show the 95% confidence interval for an individual surgeon's graft survival rate.

An inherent weakness of any registry study is that the accuracy of the data depends on surgeon engagement for the completeness and quality of the information collected, with a risk that poorly performing surgeons may not submit returns. Our registry has the advantage that all transplants undertaken in the UK are registered, although disparities remain between surgeons in returning follow up forms. With both the Kaplan–Meier method and funnel plot, reporting data at 1 or 2 years also means an inherent lag in the performance monitoring process that may allow a poor practice to remain undetected for a period. Whilst we have chosen to use funnel plots to monitor corneal transplant outcomes, other methods for monitoring outcomes exist, e.g., for risk of death following liver transplantation, where outlying survival rates are primarily identified using cumulative sum control chart (CUSUM) methods which monitor survival in real-time [18]. This methodology requires reporting data for graft failure or death at the first follow-up or 3 months.

In conclusion, funnel plots offer more information and context for statistical analysis than raw data or Kaplan–Meier plots. Risk-adjustment may help to account for surgeon case mix, but it is essential that the basis for the statistical adjustment is transparent and fully explained for the clinician. We have demonstrated that visualisation of data using funnel plots can facilitate the audit of corneal transplant outcomes, although there is uncertainty about the graft survival estimates of low-activity surgeons. Finally, funnel plots have previously been used to audit cataract surgery [19], and it may be feasible to use funnel plots for outcome analysis for other ophthalmic procedures in which post-surgical events, such as disease recurrence or procedure failure, can be defined.

SUMMARY

What was known before

- Reporting graft outcomes is a good medical practice that provides evidence that a clinical service meets benchmark standards.
- Previous reports have used Kaplan–Meier survival analysis.
- Survival analysis is difficult to interpret when there are multiple participating surgeons.

What this study adds

- We have compared Kaplan–Meier and funnel plots to demonstrate graft outcomes.
- A risk-adjusted funnel plots has advantages for the identification of surgical outliers.
- Identifying low volume poorly performing surgeons is still problematic.

REFERENCES

1. NHS Blood and Transplant. Annual Activity Report. Section 10—Corneal Activity. 2022. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/24048/section-10-cornea-activity.pdf.
2. Aboshiha J, Jones MNA, Hopkinson CL, Larkin DFP. Differential survival of penetrating and lamellar transplants in management of failed corneal grafts. JAMA Ophthalmol. 2018;136:859–65.
3. Sibley D, Hopkinson CL, Tuft SJ, Kaye SB, Larkin DFP. Differential effects of primary disease and corneal vascularisation on corneal transplant rejection and survival. Br J Ophthalmol. 2020;104:729–34.
4. Greenrod EB, Jones MN, Kaye S, Larkin DF. Center and surgeon effect on outcomes of endothelial keratoplasty versus penetrating keratoplasty in the United Kingdom. Am J Ophthalmol. 2014;158:957–66.
5. Armitage WJ, Jones MN, Zambrano I, Carley F, Tole DM. The suitability of corneas stored by organ culture for penetrating keratoplasty and influence of donor and recipient factors on 5-year graft survival. Invest Ophthalmol Vis Sci. 2014;55:784–91.
6. NHS Standard Contract 2020/21. General Conditions (Shorter Form). 2022. https://www.england.nhs.uk/wp-content/uploads/2020/03/6-SF-GCs-100320.pdf.
7. Moore J. Data visualization in support of executive decision making. Interdiscip J Inf, Knowl, Manag. 2017;12:125–38.
8. Collett D. Modelling survival data in medical research. 3rd ed. Boca Raton: CRC Press, Taylor & Francis Group; 2015.
9. Vollset SE. Confidence intervals for a binomial proportion. Stat Med. 1993;12:809–24.
10. Keogh B, Spiegelhalter D, Bailey A, Roxburgh J, Magee P, Hilton C. The legacy of Bristol: public disclosure of individual surgeons' results. BMJ. 2004;329:450–4.
11. HQIP. Identification and management of outliers (English data). 2022. https://www.hqip.org.uk/wp-content/uploads/2021/11/Appendix-10-HQIP-Outlier-guidance-v4.pdf.
12. HQIP. Detection and management of outliers for National Clinical Audits. 2022. https://www.hqip.org.uk/resource/detection-and-management-of-outliers-for-national-clinical-audits/#.Xvru5JNKg1g.
13. Prydal J, Larkin F, Hopkinson C. NHSBT Divergent Outcomes Policy—Response to Signals Arising from Audit of Corneal Transplantation Outcomes. 2022. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/24643/pol308.pdf.
14. Carpenter JR, Smuk M. Missing data: a statistical framework for practice. Biom J. 2021;63:915–47.
15. Williams KA, Lowe M, Bartlett C, Kelly TL, Coster DJ. Risk factors for human corneal graft failure within the Australian corneal graft registry. Transplantation. 2008;86:1720–4.
16. Kasbekar SA, Jones MN, Ahmad S, Larkin DF, Kaye SB. Corneal transplant surgery for keratoconus and the effect of surgeon experience on deep anterior lamellar keratoplasty outcomes. Am J Ophthalmol. 2014;158:1239–46.
17. Larkin DF, Mumford LL, Jones MN. Centre-specific variation in corneal transplant outcomes in the United Kingdom. Transplantation. 2011;91:354–9.
18. Neuberger J, Madden S, Collett D. Review of methods for measuring and comparing center performance after organ transplantation. Liver Transpl. 2010;16:1119–28.
19. Sparrow JM, Taylor H, Qureshi K, Smith R, Johnston RL. The cataract national data set electronic multi-centre audit of 55,567 operations: case-mix adjusted surgeon's outcomes for posterior capsule rupture. Eye. 2011;25:1010–5.

AUTHOR CONTRIBUTIONS

CH and ST conceived the study and ST wrote the first draft of the manuscript. ST and CH performed the literature search. CH and EC performed the statistical analysis. All authors were responsible for the interpretation of the data. DFPL and JP developed the clinical guidelines. All authors reviewed and approved the final manuscript and had responsibility for the decision to submit for publication.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Cathy Hopkinson.

Reprints and permission information is available at http://www.nature.com/reprints