For *5th Congress of Polish Statistics*, Warsaw, Poland; 1–3 July 2025.

COMPARING INSTITUTIONAL PERFORMANCE

Nicholas T. Longford

SNTL Statistics Research and Consulting, London, UK

* Kolegium Analiz Ekonomicznych, SGH, Warszawa

Mohn Centre for Children's Health and Wellbeing, Imperial College London

sntlnick@sntl.co.uk, nlongf@sgh.waw.pl

Keywords:

Administrative database; audit; causal inference; decision theory; league table; report card and dashboard.

Introduction

Institutions — hospitals, schools, local authorities, police units

Data sources — routinely collected data for accounting

Performance — outcome measures

— indicators of processes and outcomes

Comparisons — against set standards, best – worst (your neighbours), against last year, league tables

Annual **audit reports**; report cards

— involving all stakeholders and the general public

Institutions' responses:

action plans, explanations, proposals for changes

Statistical methods

Afinity to small-area estimation

small areas (households/districts) — institutions (clients/hospitals)

Comparison of means and proportions

random coefficient models

causal inference — 'What if . . .' (*counterfactual*)

decision theory — . . . *consequences* of inferential errors

methods for constructing league tables

Emphasis on graphical presentation

dissemination of statistical principles (uncertainty/chance)

incorporation of perspectives, value judgements and remits

The setting

Institutions $j = 1, \ldots, m$; their clients $i = 1, \ldots, n_j$

outputs or outcomes y_{ij} defined on an ordinal scale (or binary)

sample means $\bar{y}_j$ — unbiased estimators of means/proportions μ_j

Q. Which institutions

— satisfy a standard $\mu_j > S$; fail to satisfy this standard

— are as good as the best; are as poor as the worst

— are outliers

The perspective:

— more liberal with praise

— higher statistical standard for pointing out deficiencies

League tables

The rank of institution j:

$$r\left(\mu_j \,|\, \boldsymbol{\mu}\right) = 1 + \sum_{h \neq j} I\left(\mu_j < \mu_h\right) \qquad\qquad (I = 0/1)$$

Estimate each summand by (Bayes/posterior) $\mathrm{P}\left(\mu_j < \mu_h \,|\, \hat{\boldsymbol{\mu}}\right)$

Standard error — Laird and Louis (1987)

Plausible ranks — $\sim$ confidence intervals for the (integer) ranks

Adaptations of winner relegation, and similar labels

Which institution should I go to when I need a service?

Uniform standards — there should be no postcode lottery.

Good service is a common good — a concern for all public

Application of causal inference

Potential outcomes framework:

Every client has a potential outcome for each institution

Define a synthetic set of clients — a *template*

How would each institution perform on this template?

fair comparisons

How would institution A fare if it had clients from institution B?

relevant comparison

League tables: Compare A with B; compare B with A (home & away)

— score these comparisons

A *league table* based on the scores

Making decisions

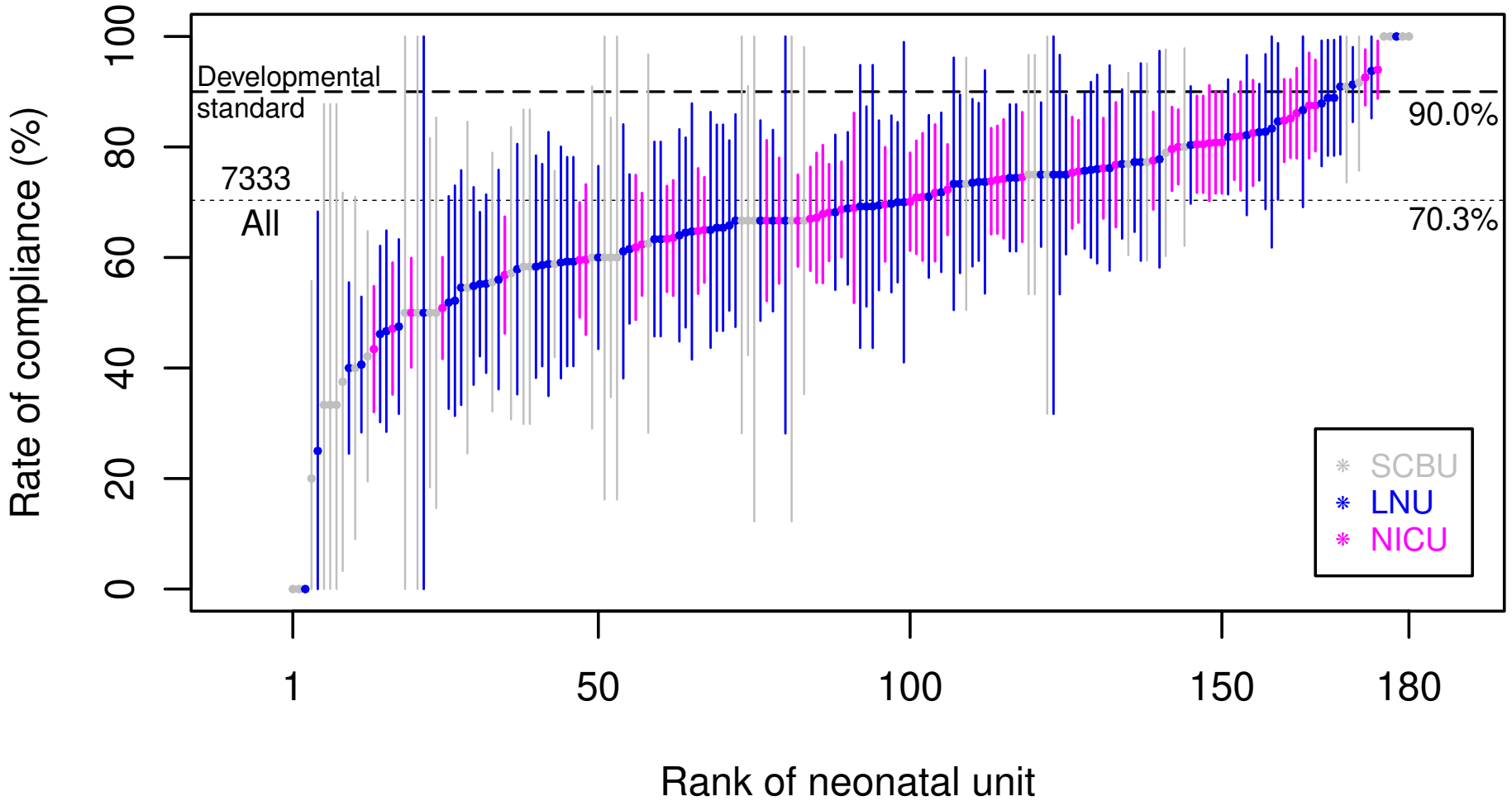
A decision has consequences — win/loss

minimise the expected (posterior) loss

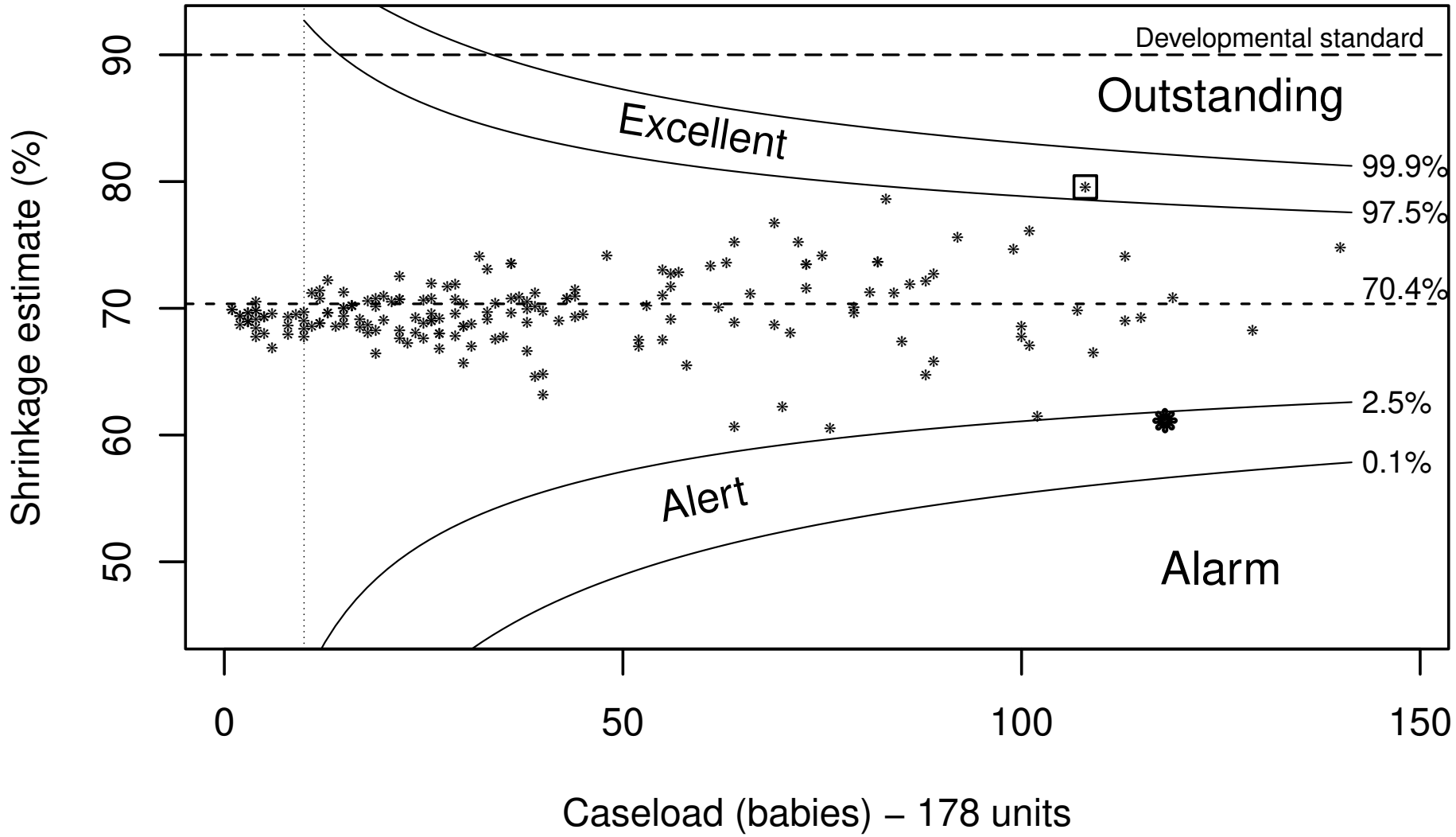
The loss matrix (example)

		Verdict		
		Smaller	∼Equal	Larger
Truth:	Smaller	0	2	5
	∼Equal	1	0	2
	Larger	3	1	0

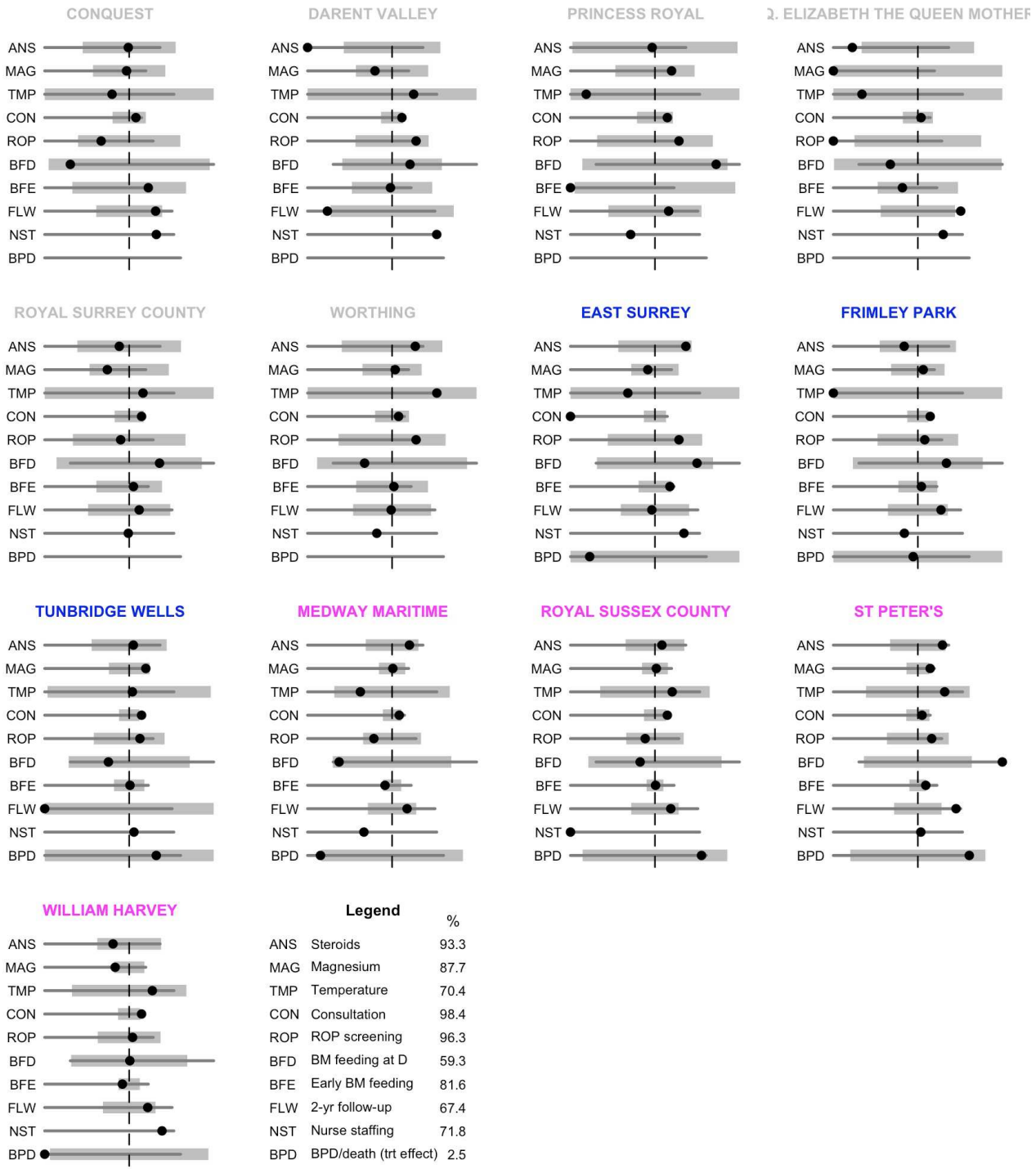
— combined with *sensitivity analysis*

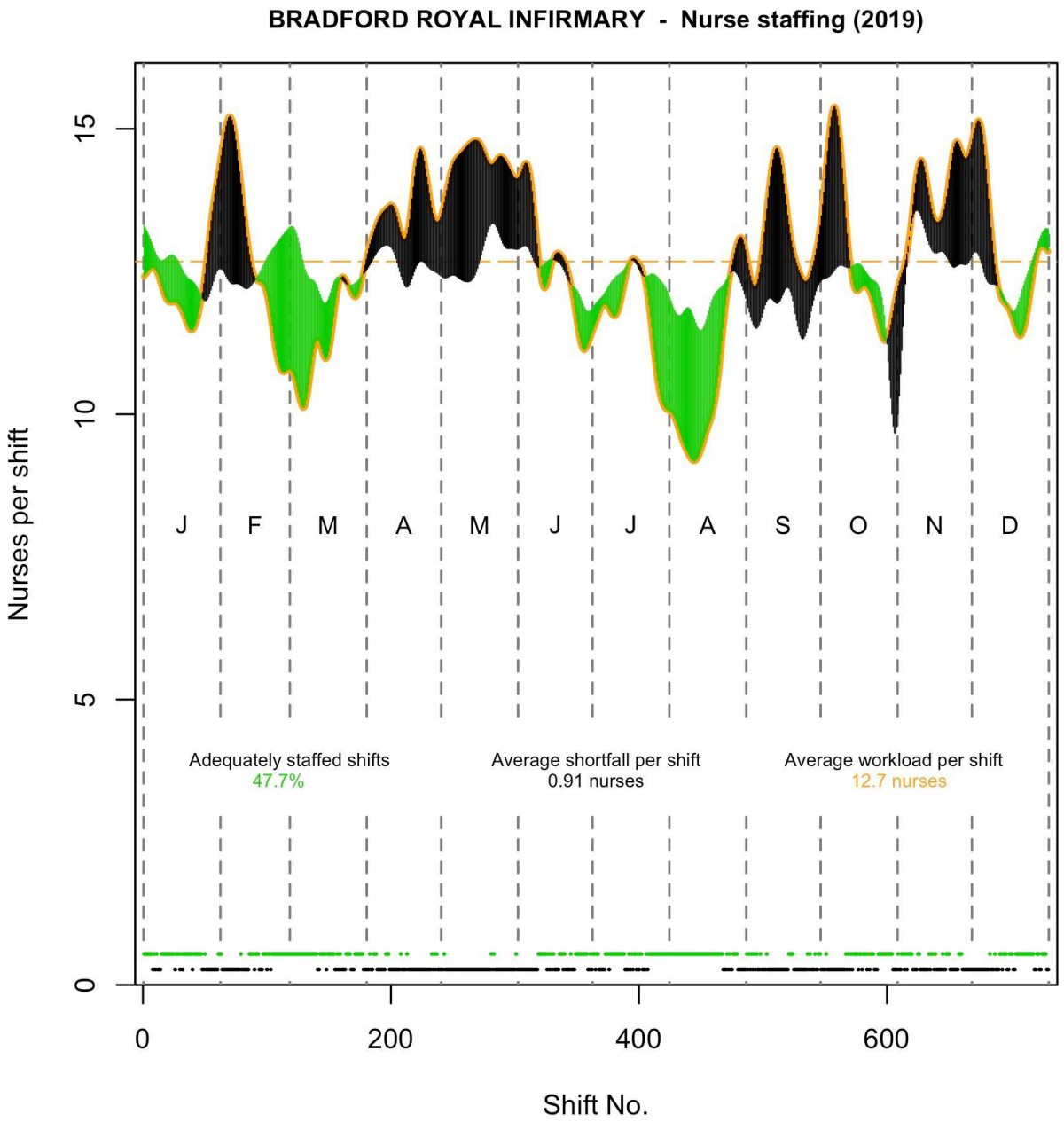


National Neonatal Audit Programme (NNAP) 2020.
Timely measurement of temperature upon admission
— (hospital) unit-level analysis. Caterpillar plot.



NNAP 2020. Timely measurement of temperature upon admission — (hospital) unit-level analysis. Funnel plot.


CONQUEST
DARENT VALLEY
PRINCESS ROYAL
ELIZABETH THE QUEEN MOTHER
ROYAL SURREY COUNTY
WORTHING
EAST SURREY
FRIMLEY PARK
TUNBRIDGE WELLS
MEDWAY MARITIME
ROYAL SUSSEX COUNTY
ST PETER'S
WILLIAM HARVEY
ANS
MAG
TMP
CON
ROP
BFD
BFE
FLW
NST
BPD
Legend
%
ANS Steroids 93.3
MAG Magnesium 87.7
TMP Temperature 70.4
CON Consultation 98.4
ROP ROP screening 96.3
BFD BM feeding at D 59.3
BFE Early BM feeding 81.6
FLW 2-yr follow-up 67.4
NST Nurse staffing 71.8
BPD BPD/death (trt effect) 2.5



NNAP 2020. Adequacy of nurse staffing.

References

Aitkin, M., and Longford, N.T. (1986). Statistical modelling of school effectiveness studies *Journal of the Royal Statistical Society* Series A **149**, 1–43.

Imbens, G.W., and Rubin, D.B. (2015). *Causal Inference for Statistics, Social and Biomedical Sciences. An Introduction.* Cambridge University Press, New York.

Laird, N.M., and Louis, T.A. (1987). Empirical Bayes ranking methods. *Journal of Educational Statistics* **14**, 29–46.

Longford, N.T. (2020). Performance assessment as an application of causal inference. *Journal of the Royal Statistical Society* Series A **183**, 1363–1385.

Longford, N.T. (2021). *Statistics for Making Decisions.* Chapman and Hall/CRC, New York.

Longford, N.T. (2025a). Statistical balancing as an unconstrained optimisation problem. *Australian and New Zealand Journal of Statistics* **67**; to appear.

Longford, N.T. (2025b). Small-area estimation — a confusion of paradigms. Submitted.

Rosenbaum, P.R. (2020). *Design of Observational Studies.* 2nd ed. Springer-Verlag, New York.

Royal College of Paediatrics and Child Health (2024). *National Neonatal Audit Programme — Summary report on 2023 data.* RCPCH, London, UK.

Rubin, D.B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics* **3**, 808–840.

Silber, J.H., Rosenbaum, P.R., Ross, R.N., Ludwig, J.M., Wang, W., Niknam, B.A., Mukherjee, N., Saynisch, P.A., Even-Shoshan, O., Kelz, R.R. and Fleisher, L.A. (2014). Template matching for auditing hospital cost and quality. *Health Services Research* **49**, 1446–1474.

Spiegelhalter, D.J. (2005). Funnel plots for comparing institutional performance. *Statistics in Medicine* 24, 1185–1202.

THANK YOU **DZIĘKUJĘ BARDZO**