

# The benefits and risks of risk-adjustment in paediatric cardiac surgery

Christina Pagel,<sup>1</sup> Sonya Crowe,<sup>1</sup> Katherine Brown,<sup>2</sup> Martin Utley<sup>1</sup>

Outcomes following paediatric cardiac surgery have long been the subject of clinical, regulatory, media and public scrutiny. There are several reasons for this. The work is among the most technically challenging, resource intensive and emotionally charged clinical activity undertaken. In the UK, past events, public inquiries and intentions to reduce the number of centres performing this surgery provide a rich source of back-stories and a level of public awareness that make paediatric cardiac surgery ripe for political comment and productive journalism.

In this context, collection and open reporting of outcome data at a national level is as fraught with difficulties as it is inescapable. Chief among these is a reasonable expectation from the profession that audit will be 'fair' to clinical teams. This translates to a view that, in the reporting of outcomes, account should be taken of the hugely diverse set of diagnoses and comorbid conditions that patients present with, the wide range of surgical procedures performed, differences in case mix between centres and the impact of the relatively small numbers of patients on what can reliably be inferred from data. These characteristics of the specialty make risk-adjustment in outcomes analysis deemed essential, but they also make it very difficult to achieve.

Efforts in a number of countries to collect standardised data on case mix and outcomes for paediatric cardiac surgery (including our own national audit in the UK) have led to a shift from the use of consensus-based risk stratification tools (eg, RACHS-1<sup>1</sup> and ARISTOTLE<sup>2</sup>) to risk estimates based on empirical data, for example, the STS-EACTS score.<sup>3</sup> This has subtly shifted how risk is conceived of. Earlier subjective methods took account of how intrinsically difficult and complex the operations were. Empirical methods

do not—they account for how successful clinical teams are at getting patients through their first month after surgery.

We are part of the research team that developed the empirical risk-adjustment method used by the National Institute of Cardiovascular Outcomes Research (NICOR) in its comparative analyses of UK centre outcomes in March and April 2013.<sup>4</sup> The Partial Risk Adjustment in Surgery (PRAiS) model<sup>5 6</sup> was developed for the purpose of in-house monitoring of short term outcomes by clinical teams.<sup>7</sup> In discussing here the use of PRAiS in comparative audit, we hope to illustrate several points important to the interpretation of risk-adjusted outcomes in this surgical specialty, and indeed other areas of clinical practice.

The original version of PRAiS was developed using 10 years of UK national audit data (2000–2010). While the UK is one of only three countries with mandatory data submission for national audit, and the data completeness and quality were very high for many data-fields, there are inevitable problems in using an audit database for research and then using that research for audit. For instance, comorbidity data were included in the model despite concerns about data quality. Given that comorbidity is clearly clinically relevant when considering the risk of a case, we hoped that its inclusion in a risk model would drive up standards of data completeness and quality. Indeed, a positive development following NICOR's April 2013 report is that, in the 3 years' worth of data (2009–2012) resubmitted to NICOR by UK centres, data completeness is markedly better than in the initial dataset used to develop PRAiS. In particular, the proportion of records with a recorded comorbidity (excluding Down syndrome) doubled from 15% to 30% of cases. For this reason, and because of seemingly improved raw survival rates since 2007–2010 and with the prospect of further comparative analysis by NICOR, we recalibrated the PRAiS model on this most recent 2009–2012 dataset at NICOR's request (details of the current calibration of PRAiS will always be available on: <http://www.ucl.ac.uk/operational-research/AnalysisTools/PRAiS>).

While we are proud of PRAiS and think it fit for the purpose of in-house monitoring, it is important to remember that a major pitfall in risk-adjustment is believing one has completely adjusted for risk. All clinical risk models can only partially account for risk and this must be borne in mind when interpreting risk-adjusted outcome data. Inevitably, some factors associated with risk of 30-day mortality are not accounted for in PRAiS (how many we do not know) and others are not accounted for fully. For example, given historically poor data quality, non-Down's comorbidities are treated equally within PRAiS, although intuitively clinicians are aware that some conditions are more adverse (eg, extreme vs non-extreme prematurity). Using PRAiS within a single centre, one can assume that the prevalence of factors not accounted for is relatively stable and that medium–long term changes (say a rise in the proportion of patients with extreme prematurity) would be recognised and understood by the local clinical team. If comparing PRAiS-adjusted outcomes *between* centres, one needs to recognise that case mix in terms of factors unaccounted for in the model may differ. Although partial risk-adjustment makes for fairer comparisons, it does not make comparisons fair.

In many audits, including the work of NICOR, risk-adjusted outcomes are put in the context of a funnel plot of prediction limits (often set to report 95% or 99.8% prediction intervals). It should always be remembered that observed outcomes may lie outside a prediction interval for one or more of a number of reasons including: data completeness and quality; aspects of case mix not accounted for in the risk model; changes in the underlying risks of surgery since the model was calibrated; a chance run of better or worse than predicted outcomes; or markedly better or worse than 'average' team performance.

Although the notion of chance playing a role in determining programme-level outcomes is often unpalatable to clinical teams, it is central to the statistical interpretation of these data using tools such as prediction intervals and funnels. If one accepts the role of chance, presenting outcomes from *multiple* centres in the context of a prediction interval makes interpretation of outcomes that fall outside the interval even more complicated. Imagine throwing a coin four times in a row. The chance of you getting four heads in a row is about 6%. Now imagine 10 of your friends each throwing a coin four times in a row. There is almost a 50% chance that at least one of them would throw four

<sup>1</sup>Clinical Operational Research Unit, University College London, London, UK; <sup>2</sup>Cardiac Unit, Great Ormond Street Hospital for Children, London, UK

**Correspondence to** Dr Christina Pagel, Clinical Operational Research Unit, University College London, 4 Taviston Street, London WC1H 0BT, UK; [c.pagel@ucl.ac.uk](mailto:c.pagel@ucl.ac.uk)

heads, but importantly you would not know which friend(s) in advance. Just as it would make little sense to assume that any of your friends that got four heads is particularly good at achieving heads, in the absence of other information a single unit out of 10 being outside a 95% funnel is *not* evidence of especially good or poor performance. Indeed, if one accepts the role of chance in determining outcomes, there is about a 40% chance that at least one unit out of 10 will be outside its 95% prediction interval.

However fair or unfair, comparisons do have a role to play in quality assurance and quality improvement. In many areas of healthcare there are no absolute standards and so feedback in terms of outcomes has to rely on the relative standards of 'how are we doing compared with last year?' and 'how are we doing compared with the place down the road?' If comparisons are made in the spirit of improvement and with an understanding of the caveats, there is valuable learning to be had that could lead to improved outcomes. The approach taken by Queensland Health in Australia in their quality improvement programme<sup>8</sup> was to set quite a low threshold for a 'flag' so that most programmes went through the incremental process of review every now and then, with the result that any stigma soon faded. Such an approach is compatible with a 'safety first' governance ethos with an expectation that genuine problems would likely be identified and addressed in a timely way. Importantly, this was coupled with a sensible protocol for response that began with a review of the data concerned. Any further investigation proceeded as far as warranted through an ordered sequence of additional reviews (looking next at the patient case mix, then the health service structure and resources, followed by healthcare processes and finally professional performance).

Focusing only on comparative analysis and relative outcomes misses the wider context of changes in absolute outcomes. Completely missing from the media debate

in the spring of 2013 was the fact that 30-day survival following paediatric heart surgery in the UK is among the highest in the world (currently greater than 97%, NICOR<sup>4</sup>) and has improved steadily since the 1990s. What was miraculous a generation ago is now deemed a minimum standard. This highlights the importance of periodic recalibration of any risk model to incorporate the continuous evolution of practice and improving outcomes.

Finally, of 30-day survival suits timely in-house monitoring, which motivated the development of PRAiS. However, we must not forget that for many children and their families, the experience of congenital heart disease continues beyond 30 days. Many congenital heart conditions require several staged surgeries and all operated children require long term follow-up. The clinical intent in treating patients is to secure long term survival and improved quality of life for patients rather than 30-day survival. For this reason, national audit of 30-day outcomes should only ever be part of a larger programme of quality assurance and improvement, focused on delivering the best possible long term outcomes.

**Acknowledgements** The authors are representatives of the research team that developed and validated a risk adjustment model for paediatric cardiac surgery, 'PRAiS', using national audit data from National Institute of Cardiovascular Outcomes (NICOR) and funding from the National Institute of Health Research (NIHR). The Health Foundation contributed funding for the development of the software. Dr K Brown sits on the steering committee of the congenital heart audit at NICOR and contributed to this article in her capacity as a member of the research team.

**Contributors** CP, SC and MU wrote the first draft of the editorial. All authors contributed significantly to the final draft.

**Competing interests** MU, CP and SC developed the software that can be used to implement PRAiS and have received royalties from its purchase by UK hospitals. KB was a key part of the team that developed the PRAiS risk model and conducted the pilot study<sup>7</sup> and also sits on the NICOR congenital group steering committee.

**Funding** None of the authors received funding to support the writing of this editorial.

**Provenance and peer review** Not commissioned; externally peer reviewed.



## OPEN ACCESS

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

**To cite** Pagel C, Crowe S, Brown K, *et al.* *Heart* Published Online First: [please include Day Month Year] doi:10.1136/heartjnl-2013-304848

*Heart* 2013;0:1–2. doi:10.1136/heartjnl-2013-304848

## REFERENCES

- Jenkins KJ. Risk adjustment for congenital heart surgery: the RACHS-1 method. *Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu* 2004;7:180–4.
- Lacour-Gayet F, Clarke D, Jacobs J, *et al.* The Aristotle score for congenital heart surgery. *Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu* 2004;7:185–91.
- O'Brien SM, Clarke DR, Jacobs JP, *et al.* An empirically based tool for analyzing mortality associated with congenital heart surgery. *J Thorac Cardiovasc Surg* 2009;138:1139–53.
- Cunningham D, Franklin R, Bridgewater B, *et al.* Investigation of mortality from Paediatric Cardiac Surgery in England 2009–12: 8 April 2013 [Internet]. 2013 [cited 2013 Aug 17]. [http://www.ucl.ac.uk/nicor/paediatric\\_cardiac\\_surgery\\_report](http://www.ucl.ac.uk/nicor/paediatric_cardiac_surgery_report)
- Crowe S, Brown KL, Pagel C, *et al.* Development of a diagnosis- and procedure-based risk model for 30-day outcome after pediatric cardiac surgery. *J Thorac Cardiovasc Surg* 2013;145:1270–8.
- Pagel C, Brown KL, Crowe S, *et al.* A mortality risk model to adjust for case mix in UK paediatric cardiac surgery. *Heal Serv Deliv Res* 2013;1.
- Pagel C, Utley M, Crowe S, *et al.* Real time monitoring of risk-adjusted paediatric cardiac surgery outcomes using variable life-adjusted display: implementation in three UK centres. *Heart Br Card Soc* 2013;99:1445–50.
- Clinical Practice Improvement Centre, Queensland Government. *VLADs for Dummies*. Australia: Wiley, 2008; [Epub in print].