

Witness Name: David
Spiegelhalter
Statement No.: 2
Exhibits: N/A
Dated: 15/01/2025

THIRLWALL INQUIRY

WITNESS STATEMENT OF PROFESSOR SIR DAVID SPIEGELHALTER

I, David Spiegelhalter, will say as follows: -

- 1. The role of statistics.** Statistical analysis inevitably involves summarising events in terms of counts and rates, and this can appear a harsh way of dealing with tragic experiences that have meant so much to the families and staff involved. Statisticians do realise that underneath all the data lie individual lives, in all their humanity and complexity. But by aggregating those experiences into statistics, we can both get an idea of the magnitude of what has happened in the past, and also design systems that can hopefully detect future problems as soon as possible.
- 2. The need for formal monitoring systems.** As I say in my Witness Statement INQ0008966 "*Human intuition is inadequate for monitoring adverse events, as we are both poor at spotting trends and also tend to over-interpret specific events.*" Judging whether an apparently cluster of events is 'beyond coincidence' is not a simple issue, and requires careful statistical insights and professional input. Numerous inquiries have called for the establishment of statistical monitoring systems in clinical care, including Bristol, Shipman, Mid-Staffs and East Kent.
- 3. Fixed period vs continuous monitoring.** In my Witness Statement INQ0008966, I distinguish between two types of statistical monitoring systems, which I termed (A) *Retrospective audit* and (B) *Prospective, real-time monitoring*. In practice, the second type of system may only report monthly or even quarterly, but still use CUSUM or other methodology that continuously monitors events without regard to annual boundaries. So it might be clearer to label these as (A) *Fixed period* and (B) *Continuous monitoring*.
- 4. Limitations of statistics.**

- 4.1. In her witness statement, Dr Edile Murdoch (INQ0106962, paragraph 36) shows how a pilot continuous-time mortality monitoring system would have signalled for 12 General Practitioners between 1993 and 1998, one of which was serial-murderer Harold Shipman. The other 11 were possibly working in retirement communities, and so signing a large number of death certificates. These were caring doctors helping their elderly patients to die at home or in care homes rather than in hospital.
- 4.2. This demonstrates clearly that statistics can only point to something of interest and cannot on their own determine the underlying cause – as Dr Murdoch (para 30) says *“The signal is simply a prompt for a rapid assessment to understand what has caused the signal.”*

5. Some principles for monitoring systems. At the start of the MOSS project, I suggested some useful principles for monitoring systems:

- 5.1. Possibility of multiple indicators: including ‘never events’ and more common outcomes, which might be ‘basketed’ up.
- 5.2. Use cumulative data to identify trends and sudden shifts in performance
- 5.3. Use explicit thresholds for different levels of ‘unusualness’
- 5.4. Establish management actions when crossing thresholds, possibly using local and central scrutiny
- 5.5. Minimise burden on staff: ideally data should only be entered once
- 5.6. Rapid feedback, using attractive visualisations. This is particularly important to motivate encourage a feeling of local ownership of the system, shared by all staff.

6. The importance of variation

- 6.1. There is always substantial variation in the number of adverse events, both between units and within units over time. There are many possible reasons, including changes in case-mix and unavoidable ‘chance’ variation due to unforeseeable factors. So ‘good’ years may be followed by ‘bad’ years, and it is the role of statistical analysis to establish whether the changes are beyond what it might be reasonable to expect, and so justify investigation of any underlying cause. Human intuition alone can be poor at assessing what variation is ‘reasonable to expect’.
- 6.2. Very broadly, unusual observations are of two types, as illustrated in the Figure below.

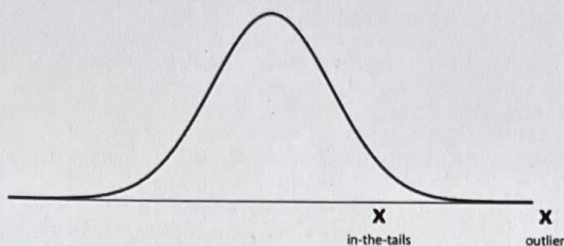


Figure: In the typical distribution of results, unusual observations may either be in the tails, or clear outliers.

- 6.3. Observations that are extreme, but still part of the distribution, may be termed 'in-the-tails' or 'extreme'. Observations that are clearly different from the standard pattern may be termed 'outliers'.
- 6.4. In the Inquiry into deaths of children undergoing heart surgery at Bristol Royal Infirmary, Bristol was an outlier, with 'clear water' between it and other centres.
- 6.5. In the Inquiry into Harold Shipman's murders, Shipman's mortality rate was extreme but, as discussed above in Paragraph 3.1, he was not a clear outlier – some other GPs had even higher mortality rates.
7. **Risk-adjustment.** Ideally, monitoring systems for adverse events should make allowance for pre-existing conditions, so that more 'surprising' events would contribute more to a signal. With non-risk adjusted systems, everyone contributes equally. But proper risk-adjustment requires full collection of common data on risk factors across all centres, and is challenging to implement. Both adult and paediatric intensive care do have such systems, eg PICANET risk-adjusts for clinical features upon admission to paediatric intensive care, but this requires disciplined and standardised data collection across all units.
8. **What the MBRRACE system showed for 2015/2016 at the Countess of Chester Hospital.**
- 8.1. Among other outputs, the MBRRACE-UK programme produces retrospective analysis on neonatal deaths, classifying deaths by the place of birth rather than place of death. As well as 'crude' neonatal mortality rates, MBRRACE also records 'adjusted and stabilised' neonatal mortality rates, which adjusts for basic demographic risk factors of mother and child (for example ethnicity and gestational age), and stabilises estimates by 'shrinking' crude rates towards the group mean by

allowing for some unexplained variation between units. It is designed to identify extreme units in the tail of the distribution, rather than outliers.

- 8.2. For the Countess of Chester Hospital (CoCH), Lisa Annaly's Witness statement (INQ 20241220, 4.2.1.1) states that the MBRRACE-UK programme "*identified the trust as higher than average for neonatal mortality indicators for 2015 (more than 10% higher) and 2016 (up to 10% higher). This data was published in 2017 and 2018 respectively, which was after CQC's comprehensive inspection of Countess of Chester*".
- 8.3. Professor Marian Knight's witness statement (INQ0006757) is slightly different, showing that in both 2015 and 2016 CoCH was red (more than 10% higher) for both crude and 'adjusted and stabilised' neonatal mortality rates.
- 8.4. A substantial proportion of units are 'red' each year. For 2015, from visual inspection of the MBRRACE report (Figures 18 and 19) around 30% of units were classified as 'red' for both the crude rates and 'adjusted and stabilised' rates. CoCH's crude neonatal mortality rate of 2.96 per 1,000 placed it highest of the 27 units in its Tier (2,000 to 4,000 births per year) for which rates were reported. Blackpool, also with around 3,000 births, had a crude neonatal mortality rate of 2.73, translating to 9 deaths in CoCH and 8 deaths in Blackpool.
- 8.5. In 2015, for 'adjusted and stabilised' rates, CoCH's figure of 1.91 per 1,000 was highest of 43 units in its Tier, with Burton being the second highest with 1.67 per 1,000.
- 8.6. In 2016, Figures 19 and 20 of the MBRRACE report suggest around 30% of units were classified as 'red' for the crude neonatal mortality rates and 20% for 'adjusted and stabilised' rates. The crude rate for CoCH is not provided due to 'small number of deaths'. For 'adjusted and stabilised' rates, CoCH's figure of 1.49 per 1,000 was highest of 39 units in its Tier, with Sherwood Forest being the second highest with 1.26 per 1,000.
- 8.7. In conclusion, after the substantial delay typical of such retrospective audits, MBRRACE identified CoCH as having higher rates in both 2015 and 2016, along with many other units. CoCH's neonatal mortality rate was highest in its Tier (2,000 to 4,000 births per year), but I would not judge it a clear 'outlier'.

9. What CQC's outlier system showed for 2015/2016 at the Countess of Chester Hospital.

- 9.1. This is the system that I helped design in 2007, and uses Hospital Episode Statistics to continuously monitor a wide range of outcomes. It is a centralised system

designed to detect clear outliers, and not observations that are simply extreme – otherwise it would be overwhelmed by signals. It was available quarterly.

9.2. Lisa Annaly's Witness statement (INQ 20241220, 4.2.2.1) reports "*The measures for neonatal deaths and neonatal non-elective readmissions did not flag as outliers for the Trust in 2015 or 2016*".

9.3. In particular, at the CQC inspection in February 2016 there was "*no information as to increased neonatal mortality at the hospital*" - using data up to December 2015 (para 3.6, Witness Statement of Ann Ford, INQ0107911)

9.4. This reinforces the findings from MBRRACE that while there was an increased number of neonatal deaths in 2015 and 2016, this did not constitute an outlier or 'spike' sufficient to trigger an alert in an outlier detection system.

10. What monitoring systems would have showed for 2015/2016 at Countess of Chester – MOSS

10.1. The prototype MOSS system only looks at adverse events for full-term births, and so does not include outcomes for premature babies. This is because it is looking for sub-optimal maternal care, not sub-optimal neonatal care.

10.2. Noting this caveat, Dr Edile Murdoch's Witness statement (INQ0106962, para 75), shows that MOSS would have generated a 'local alert' around Aug/Sept 2015. She reports that "*If MOSS were in place at the time, the signals would have prompted a standardised critical safety assessment to understand why the signal occurred.*"

11. What CoCh mortality data on its own might have shown

11.1. The monitoring systems discussed above all make use of comparisons with other units. In contrast, some statistical process control (SPC) systems use only the data from the unit alone, and specify criteria for identifying an alert should certain changes be observed in the series of observations. It should be in principle be possible to look at the mortality data for CoCH alone, and see whether any observed changes might have triggered an alert if such a system had been in place.

11.2. The MBRRACE real-time monitoring tool has now added SPC methods, but I do not know whether it would have triggered an alert, had this been in place in 2015/2016.

11.3. I have just been notified of Document INQ0108781, which contains a table of neonatal deaths in the CoCH unit from 2010 to June 2016, apparently collected in a

systematic way. Between 2010 and 2014, the number of deaths recorded each year was 1,3,3,2,3, which shows surprising consistency. Document INQ0004593 then reports 8 deaths in 2015, and 5 up to June 2016.

- 11.4. One basic approach is to see how surprising 8 deaths would be, if we assumed that circumstances had not changed between 2010 and 2015 – that is, it was just a ‘chance’ event. Conventionally, an ‘alert’ signal is generated if the probability of such an extreme result is less than 0.025, and an ‘alarm’ if the probability is less than 0.001.
- 11.5. Assuming the data between 2010 and 2014 are observations from a Poisson distribution, we can estimate the ‘underlying’ rate as 2.4, with standard error 0.7. From this we can assess the probability of getting 8 or more deaths in 2015 as 0.008, assuming there has been no change in the underlying rate [technical note, this is from a Poisson-gamma distribution]. This would generally be sufficient to trigger an ‘alert’ signal, although not extreme enough to be considered an outlier. To put this in perspective, out of around 150 neonatal units in the UK, we would expect on average one such a signal every year, by chance alone.
- 11.6. If we assume an underlying rate of 3 neonatal deaths per year, then the probability of getting 8 or more deaths in 2015 is 0.02. This would generally be considered as constituting an ‘alert’ signal. Again, to put this in perspective, we would expect around 3 such signals each year in the UK, just by chance alone.
- 11.7. As with all monitoring systems, signals can only trigger a more careful examination of the outcomes – they cannot reach any conclusions about any background causes.

Statement of Truth

I believe that the facts stated in this witness statement are true. I understand that proceedings may be brought against anyone who makes, or causes to be made, a false statement in a document verified by a statement of truth without an honest belief in its truth.

Signed:

Personal Data

Dated: 15 January 2025