## THIRLWALL INQUIRY

---

## WITNESS STATEMENT OF PROFESSOR SIR DAVID SPIEGELHALTER

---

I, Professor Sir David Spiegelhalter, will say as follows: -

1. I am Sir David Spiegelhalter FRS OBE, Emeritus Professor of Statistics at the University of Cambridge. I have mainly worked as a medical statistician, in particular on methods for monitoring and comparing performance in health services. I worked with surgeons at Great Ormond St Hospital on the analysis of a 'cluster of failures', which led to my being asked to lead the statistical team at the Inquiry into deaths of babies with congenital heart disease at the Bristol Royal Infirmary. I was then part of the statistical team for the Harold Shipman Inquiry where we found that, had data been collected and properly analysed, he could have been detected as 'unusual' after only around 40 deaths, possibly saving over 200 further victims.

2. I shall answer the questions I have been posed to the best of my knowledge, I but have not conducted extensive investigations and recommend further inquiry into the available data-sources.

3. *Question 1. What role can statistics play in monitoring and alerting to serious adverse incidents in a neonatal setting (whether a singular event or a cluster of events) and which may need further investigation? What are the potential benefits and limitations of a statistical monitoring system? Please draw on examples of existing statistical monitoring systems as you see fit, such as Picanet.*

   Human intuition is inadequate for monitoring adverse events, as we are both poor at spotting trends and also tend to over-interpret specific events. This has long been recognised within industry, where formal statistical process control (SPC) methods have been routinely used since before WW2. Over the last two decades, it has become apparent that these methods can be adapted for monitoring adverse clinical outcomes, which is not only more efficient for detecting 'clusters of failures', but permits proper accountability of units at both a local and national level. Numerous inquiries have called

for the establishment of statistical monitoring systems, including Bristol, Shipman, Mid-Staffs and East Kent.

There are two broad types of system, (A) retrospective audit and (B) real-time prospective monitoring, which can be illustrated using some existing systems for perinatal, neonatal and paediatric care.

**(A) Retrospective audit**: data is collected and aggregated over a period, centres compared, and results fed back to centres and made available for further analysis.
- Advantages: quality-control of data, centralised scrutiny.
- Disadvantages: historical, not designed to detect 'clusters of failures'.

There are a bewildering array of such audit systems for neonatal and perinatal outcomes, and I cannot provide a full review. Examples include:

- **MBBRACE** run by the National Perinatal Epidemiology Unit. This produces an annual perinatal mortality surveillance report, currently publicly available up to 2021. Their public interactive website shows raw and adjusted mortality outcome data for different organisations, allowing benchmarking against other trusts. Their Perinatal Mortality Review Tool (PMRT) is available to trusts and health boards.

- **National Maternity Dashboard** is run by NHS Digital and includes Clinical Quality Improvement Metrics (CQIM). This is based on monthly statistics from the National Maternity DataSet, which have a lag of around 3 months, although currently (January 2024) the dashboard shows data up to April 2023. The mortality data is based on MBBRACE.

- Other related audit and research tools include the National Neonatal Audit Programme (NNAP), National Maternity and Perinatal Audit (NMPA), the Office for Health Improvement and Disparities Child and Maternal Health dashboard, the National Child Mortality Database, while the National neonatal research database (NNRD) contains a wide range of indicators extracted from clinical information systems, and is available to researchers,

- The PICANET surveillance system for paediatric intensive care includes retrospective annual standardised mortality ratios which are publicly reported using 'funnel plots',

which allow visual assessment of outliers. An example from the latest report is shown below, from which they conclude "*There was no evidence that any PICU had an excess mortality rate compared to what would be expected based on the level of sickness at the time of admission across the three-year reporting period.*"

Metric 5: Mortality in PICU

**Figure 5: Risk-adjusted Standardised Mortality Ratio (SMR) by health organisation for under 16 year olds, 2019–2021**
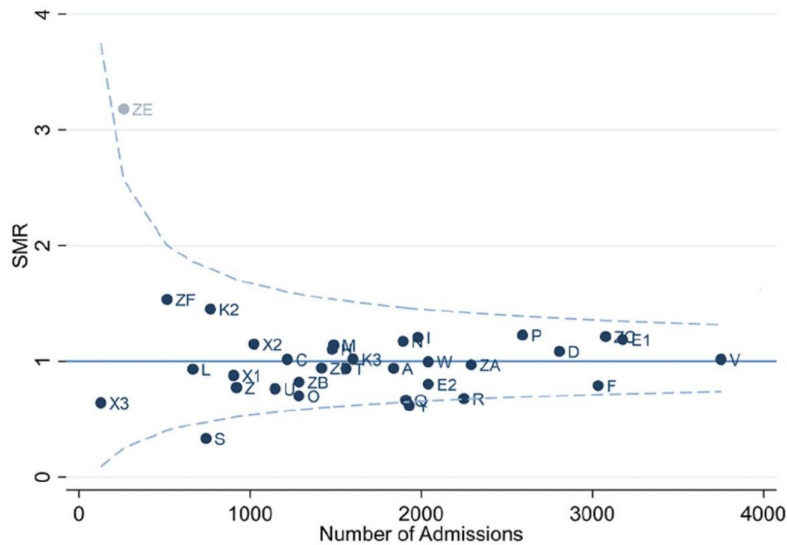


Organisation key

ZE (Harley Street) will not be flagged as an outlier as the unit closed in 2020 and an investigation previously took place
Unit X3 (Leicester CICU) was opened in 2021 and therefore has a low number of admissions

B) **Prospective, real-time monitoring.**  The intention is to have systems that can respond rapidly to changes in outcomes, based on continual monitoring of accumulated experience.  Ideally the system should have strong local support, reinforced by using the data routinely in team meetings.

- Advantages: ownership of local data, uses accumulated data, rapid response, with formal monitoring methods to trigger further investigation
- Disadvantages: requires buy-in from centres, rapid access to high-quality data, care in setting parameters and thresholds, and explanation and acceptance of more complex monitoring.

Statistical monitoring is generally based on comparing the cumulative observed number of adverse events (denoted $O$) since a specified starting point, with the

number (denoted *E*) that would be expected were the centre 'average'. The expected number ideally takes into account the risk factors of the mix of patients, and then the monitoring is said to be 'risk-adjusted'. A standard and easily-understood tool is to plot the cumulative difference between *O* and *E* – for historical reasons this is often called a VLAD plot (Variable Life-Adjusted Display). If positive, this is sometimes termed the 'excess' events, but this phrase is best avoided as it can suggest the excess was avoidable, whereas half of all centres will have 'excess' events by definition. This plot is shown below for deaths of patients of Harold Shipman from 1977 to his arrest in 1997.
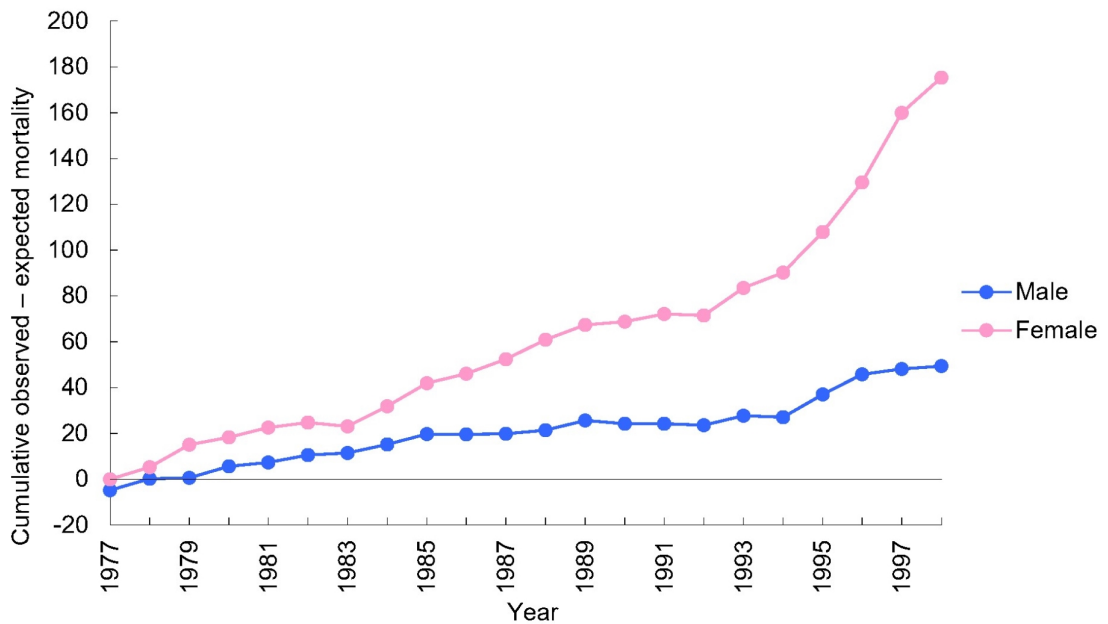


*Cumulative 'observed – expected' deaths in Harold Shipman's practice.*

Decisions on triggering further actions cannot be left to local discretion - experience has shown that units are capable of considerable self-deception about emerging problems. Statistical methods need to be carefully developed and evaluated – for example, simple hypothesis testing is inappropriate due to well-known problems with multiple-testing, and so 'sequential analysis' techniques such as risk-adjusted Sequential Probability Ratio Tests (SPRT) or Cusums are required, both of which are obtained from a simple formula based on *O* and *E*, and trigger an alert when the statistic goes above or below a pre-set level. This does require some sophistication, and in particular the parameters in the system and the thresholds for 'action' need to be carefully tuned. The plot below shows the use of a SPRT for monitoring Shipman, showing that an alarm could have been raised in 1985, after only around 40 deaths, and with a very low error rate.

Unfortunately no monitoring procedure was in place.
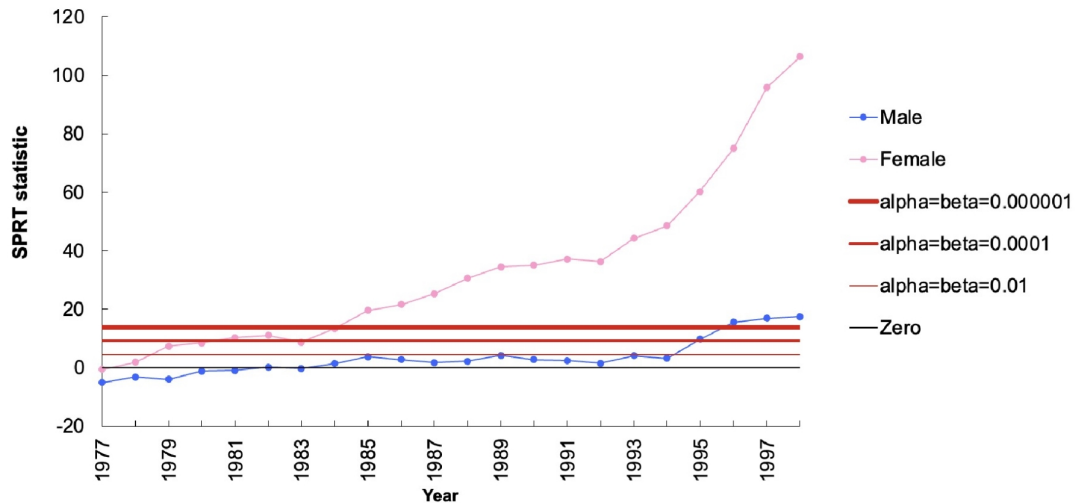


*Retrospective application of a Sequential Probability Ratio Test procedure to Harold Shipman's mortality data.*

Relevant examples of real-time monitoring systems include:

- The Perinatal Quality Surveillance Model (PQSM), run by NHS England, is planned to enable timely identification and escalation of issues in trusts, including outlier status for perinatal and/or neonatal mortality. It is in development, with the aim of producing local dashboards on a monthly basis, and Maternity Safety Champions can present to their boards.

- The Care Quality Commission (CQC) had a prospective mortality monitoring tool that I helped set up in 2007, which used sophisticated statistical methods to monitor thousands of mortality indicators and automatically alert a human when a possible outlier was detected (the design issues were covered in an academic paper in 2012, which has become a standard reference for health-care surveillance). I understand that neonatal metrics were included, but that the system was paused at the start of the Covid pandemic and has not been resumed, but this requires further investigation.   The Dr Foster unit at Imperial College ran a similar system, but I understand that this has also now ceased operation.

- The real-time PICANET surveillance system for paediatric intensive care is based on the risk-adjusted sequential probability ratio test (SPRT) methodology we established during the Shipman Inquiry.  Each centre monitors its own trajectory, but there is also central scrutiny, which automatically alerts centres to outlying results.  Thresholds are

set for triggering further scrutiny of either 'good' or 'poor' performance, and their outlier policy is explicitly laid out.

4. *Question 2. Can a statistical monitoring system work in real time? How is this done? What do you think the benefits of a statistical monitoring system within maternity and neonatal care are or would be?*

Real-time systems can be achieved, as PICANET has shown, but they need buy-in from centres, and ready access to data, which should ideally only be entered once.

There are obvious benefits for such a system within maternity and neonatal care - one only has to look at successive 'scandals' and inquiries which repeatedly identify units being reluctant to acknowledge problems.

5. *Question 3. Has a statistical monitoring system within neo-natal care been recommended before now, and if so by whom?*

I understand that neonatal metrics were part of the CQC system, but do not have further details. The Perinatal Quality Surveillance Model (PQSM) is essentially such a model, and the need for such a system has been reinforced by Recommendation 1 of Bill Kirkup's investigation into East Kent "*The prompt establishment of a task force with appropriate membership to drive the introduction of valid maternity and neonatal outcome measures capable of differentiating signals among noise to display significant trends and outliers, for mandatory national use.*". This has been accepted by government.

6. *Question 4. What work is currently being done by the Reading the Signals Data Co-ordination Group in order to create a real-time statistical monitoring system within maternity and neonatal care? Please give us a brief summary of its work to date and suggest who we might contact about the detail of that work in due course.*

A statistical monitoring system is in development. Details of the work so far can be obtained from the Working Group.

I feel it would be very valuable for the system being developed by the Working Group to be applied to data from the Countess of Chester unit. This would require examining a series of historical data on adverse events back to, for example, five years before Letby joined the unit and for the period following her departure. It would be of some interest to see whether, and when, the system issues any form of 'alert'.

5. *Question 5. Chapter 5 of Sir Robert Francis' Volume 1 Mid-Staffs Inquiry report is dedicated to Mortality statistics. At paragraph 5.180 Sir Robert Francis recorded that he had received voluminous and detailed evidence about the Hospital Standardised Mortality rate and "whatever may be the vulnerabilities of HSMR to coding and data quality deficiencies, the Trust's results were significantly and consistently high over many years." Furthermore, when concluding the chapter Sir Robert noted: "There is now a consensus that significantly high HSMR/SHMI (Summary Hospital-level Mortality Indicator) results should trigger a serious consideration of whether poor care is an explanation for them, whatever other steps are also taken (para 5.192)."*

*a. Do you agree that national, statistical mortality data is useful in flagging up when something is wrong and may need further attention?*

Sir Robert's remarks concern hospital-level aggregate mortality using a single index, an example of a retrospective audit tool. This can be a useful, although blunt, instrument, and surveillance at a more refined level is required in order to identify issues with maternity care.

*b. Is mortality data still collected and utilised for this purpose as far as you are aware?*

The *Summary Hospital-level Mortality Indicator (SHMI)* programme is still operating.

*c. Has it ever been collected in neo-natal units?*

This is essentially what MBRRACE does.

**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 of its truth.

**Signed**:

**Personal Data**

**Dated**: 8th January 2024.