

# Hindsight bias critically impacts on clinicians' assessment of care quality in retrospective case note review

Authors: Edward Banham-Hall<sup>A</sup> and Sian Stevens<sup>B</sup>

## ABSTRACT

**Objective.** To determine whether hindsight bias impacts on retrospective case note review using a five point scoring system based on modern clinical governance toolkits. **Design.** Survey. **Setting.** Clinicians of varying grades invited to complete a short internet survey. **Participants.** Ninety three clinicians were invited to complete an anonymous survey in which they reviewed three case vignettes for the purposes of a fictional clinical governance meeting. For each vignette, participants were randomised to an outcome in which the patient made a full recovery or alternatively died shortly after discharge. **Main outcome measure.** Participants submit scores from 1 to 5 to indicate the quality of care provided to patients prior to their discharge. These scores were compared to determine whether judgements about the quality of antecedent care were biased by the description of a patient death. **Results.** In two out of three case vignettes clinicians exhibited marked hindsight bias. In a case of a patient with a swollen leg, identical antecedent care was scored as poor by participants when the patient died the next day, but good when the patient recovered ( $p < 0.00001$ ). In a case of headache, care was scored as poor when the patient died but adequate when the patient made a full recovery ( $p = 0.0003$ ). A third case of chest pain did not exhibit hindsight bias. Seniority of clinician had no impact on the tendency to exhibit hindsight bias when reviewing case notes. **Conclusion.** In some cases, clinicians are markedly more critical of identical healthcare when a patient dies compared to when a patient survives. Hindsight bias while reviewing care when a patient survives might prevent identification of learning arising from errors. Additionally, we predict hindsight bias combined with a legal duty of candour will cause families to be informed that patients died because of healthcare error when this is not a fact.

**KEYWORDS:** Structured judgement review, case note review, hindsight bias, morbidity mortality

## Introduction

The processes used to scrutinise patient records and disseminate learning following a patient death are of considerable importance to healthcare providers and patients. Recent well-documented flaws in investigations into patient deaths have attracted justifiable criticism.<sup>1</sup> Case note review provides a vital opportunity to recognise areas of weakness in quality of care, and identify interventions for improvement.<sup>2,3,4</sup>

Recent evidence from Care Quality Commission investigations shows that review of information after certain patient deaths lacks consistency, misses opportunities to learn, and fails to support families.<sup>5</sup> This has led to a welcome NHS-wide focus on learning from deaths with the goal of improving learning from incidents and the rigour of investigations themselves. It is hoped that this approach will improve the rigour with which case note reviews are conducted, and improve the identification and dissemination of learning within the NHS.

Hindsight bias (colloquially known as 'the retrospectroscope') is known to play a significant role in the evaluation of an antecedent event, and has been demonstrated in both medical and judicial settings.<sup>5</sup> It is human nature to be sympathetic to those who are suffering or have died. This can contribute to heuristic biases, as it is easy to inadvertently use knowledge gained after an event to critique it when information was not available for those involved at the time.<sup>6,7,8</sup> The recognition of a poor outcome can bias the ability to pass judgement on a less appraised perspective, heightening the perception of preventability.<sup>9,10,11</sup> This might lead to an unjustified evaluation based disproportionately on a poor outcome, and not because care was poor.<sup>12</sup>

We wanted to determine whether assessment of antecedent healthcare was susceptible to hindsight bias when clinicians are aware of a patient's outcome. We tested this by using a scoring system with similarities to a modern audit toolkit being developed in response to the learning from deaths report. We were also interested to test whether a physician's clinical experience (using the metric of professional seniority) mitigated any tendency to exhibit hindsight bias, if present.

## Methods

### Selection and description of participants

We recruited 93 doctors of varying seniority opportunistically while undertaking clinical work and during educational meetings, as well as through the internet. As a result of this approach, participants

**Authors:** <sup>A</sup>consultant in acute and general medicine, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK; <sup>B</sup>CT2 in emergency medicine, acute care common stem, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

were predominantly hospital doctors. By design, we recorded no details about participants other than professional seniority to facilitate anonymised participation.

### Patient and public involvement

There was no patient or public involvement in the design of this study.

### Technical information

Participants were invited to complete a web-based survey in which they read three case vignettes describing patients presenting with a swollen leg, a headache, and chest pain respectively. Each vignette provided identical clinical details and information up to the point of a patient's discharge from hospital. However, in each scenario participants were randomised using a php-based algorithm either to an outcome in which the patient died suddenly shortly after discharge, or an outcome in which the patient made a full recovery. After reading each vignette, participants were invited to enter their opinion of the quality of care provided using a five point scale. Options provided were '1 – very poor care', '2 – poor care', '3 – adequate care', '4 – good care' and '5 – excellent care'. Participants were invited to record their seniority (for example, consultant, core trainee, foundation doctor etc).

To test whether more experienced physicians showed a difference in susceptibility to hindsight bias we pooled all responses to cases in which the patient died, and compared the phase-of-care scores provided by consultants to those provided by more junior staff.

### Statistics

Responses were recorded using Microsoft Excel and then analysed with a Mann-Whitney U test to calculate a p value using a Python script and the SciPy library. Power analysis was performed using R with the packages rcompanion and samplesize.

### Results

From the 93 participants recruited to complete the survey, a total of 262 responses were returned across all scenarios equating to

**Table 1. Self-reported grade of participant**

Grade of participant	Number of responses (total = 262)	Percentage of responses (%)
Consultant	65	24.8
Specialist trainee 3 +	63	24.1
Core trainee	126	48.1
Foundation year 1 + 2	2	0.8
Nurse practitioner	3	1.1
Medical student	3	1.1

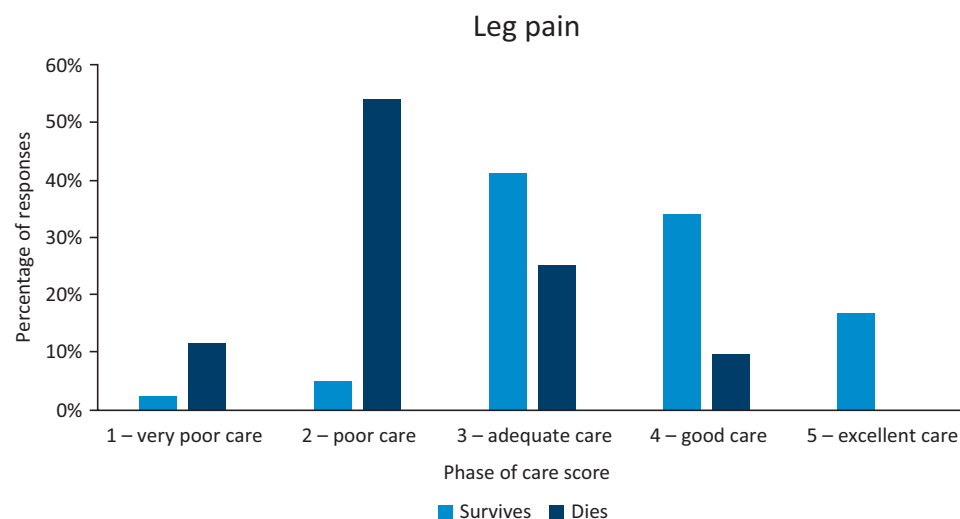
94% completion rate. A small number of participants withdrew as they progressed through the survey. For case one a total of 93 responses returned, case two received 85 responses, and case three received 84 responses. The participants varied in seniority, as shown in Table 1.

### Case one – leg pain

This vignette described a patient attending the emergency department (ED) with a painful left knee who was assessed by an emergency nurse practitioner. A clerical error means that the case notes erroneously document that the right leg was examined, despite the presenting complaint being a left sided problem. The patient was diagnosed with a soft tissue injury, given advice to rest, take pain relief and to attend their general practitioner if worsening. The patient was then discharged home, and participants were presented with one of two possible outcomes. Outcome one described resolution of symptoms within two weeks, whereas outcome two resulted in patient death the next day. A postmortem concluded 1a pulmonary embolism, 1b deep vein thrombosis in the left leg. See supplementary material S1 for the complete scenario with both alternative endings.

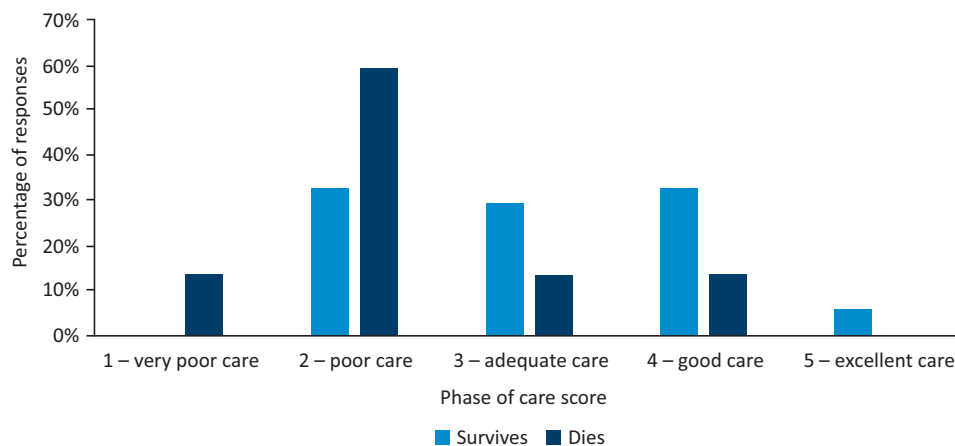
This first scenario had the total of 93 responses returned. Fig 1 shows the results of outcome one (full recovery) with n=41 and outcome two (patient death) n=52.

Respondents provided quality of care scores for patient full recovery spanning the entire scale 1 to 5, with the median at 4



**Fig 1. Participants assigned markedly different care quality scores when assessing the same vignette depending on the patient outcome in a case of leg pain.**

## Headache



**Fig 2. Participants assigned markedly different care quality scores when assessing the same vignette depending on the patient outcome in a case of headache.**

(good care). By comparison, participants assessing a mortality outcome still exhibited a wide range of 1 to 4 points, with a median score of 2 (poor care). There was a statistically highly significant difference in the rating of the same antecedent care provided depending on the patient outcome ( $p < 0.00001$ ).

### Case two – headache

This scenario presented a patient on anticoagulation with a sudden onset severe headache who was referred by their general practitioner to the ED. The patient was reviewed by a foundation trainee who recorded that they felt a subarachnoid haemorrhage (SAH) was unlikely. However, given the sudden onset of the headache SAH was nevertheless a diagnosis that ought to be excluded. The patient was then reviewed by a medical registrar who documented that the headache severity peaked at a number of hours and explained the patient the headache was most likely a tension headache, and did not fit the indication for computed tomography (CT) scanning or lumbar puncture. The patient was discharged with advice and analgesia. In outcome one, the patient made a full recovery. Outcome two described the patient becoming agitated overnight, then going into cardiac arrest. The postmortem showed a subarachnoid haemorrhage and a ruptured berry aneurysm. See supplementary material S2 for the full vignette and subsequent endings.

This scenario had a total of 85 responses, Fig 2 shows the results of outcome one (full recovery)  $n=34$  and outcome two (patient death)  $n=51$ . Similar to the first scenario the quality scores showed a statistically significant difference in the outcome of each scenario. These ranged from 2 to 5 for outcome one, with a median of 3 (adequate care). Outcome two scores ranged from 1 to 4 and a median of 2 (poor care) ( $p=0.0003$ ).

### Case three – chest pain

The final scenario presented a patient with a background of malignant hypertension with end organ damage who had sudden onset chest pain and collapsed. The patient was reviewed by a medical registrar who noted a raised troponin, deranged renal function and that the patient was non-compliant with regular

follow up. Discussion with a consultant resulted in a decision to rule out both pulmonary embolism and aortic dissection via CT scan, however the request for the scan only documented the differential of possible pulmonary embolism.

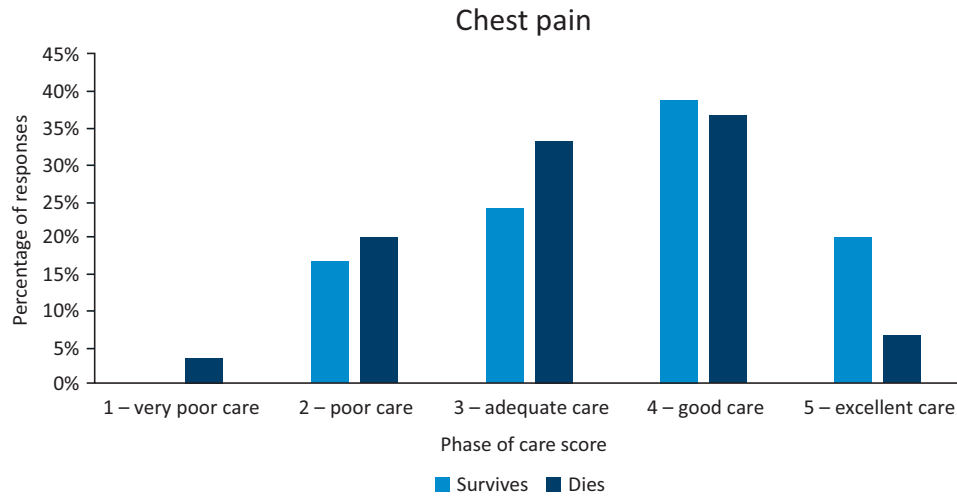
The patient was handed over at shift change and reviewed by a second medical registrar when the patient decided to self-discharge as they then felt well. Prior to self-discharge, the verbal report of the CT pulmonary angiogram scan stated that there was no pulmonary embolism or evidence of aortic dissection. The patient was documented to have capacity and was therefore allowed to leave. After this, a final written report noted ascending aorta dilatation needing vascular surgeon review.

In outcome one, the patient remained well and attended for an outpatient vascular surgical appointment. Outcome two described the patient being found deceased in bed the next day, with a postmortem showing the cause of death to be cardiac tamponade secondary to ascending aorta dissection. For the full scenario and endings, see supplementary material S3.

This scenario received a total of 84 responses. Fig 3 shows the results of outcome one (full recovery)  $n=54$ , outcome two (patient death)  $n=30$ . In this case, there was no statistically significant difference found. The quality scores for full recovery ranged from 2 to 5, a median of 4, with the poor outcome scores ranging from 1 to 5, and a median of 3 ( $p=0.1074$ ).

### No effect of seniority on susceptibility to hindsight bias

Having observed a marked tendency to exhibit hindsight bias when conducting case note review in two cases, we next tested whether (due to greater clinical experience) hospital consultants were less prone to this. When looking at all 133 responses with the end result of patient death, 32 were submitted from consultants and the remaining 101 from all other grades combined. The phase of care scores did not show any difference in their evaluation of the care received ( $p=0.34212$ ). This is shown in Fig 4. The spread for junior doctors spanned all five possible scores and the consultants spanned four scores. We also repeated this analysis for participants randomised to a normal patient outcome and again observed no difference (data not shown). Given the same sample size and proportion of consultants, a medium effect size ( $r=0.373$ )



**Fig 3. Participants assigned similar care quality scores when assessing a case vignette, irrespective of patient outcome, in a case of chest pain.**

was detectable at significance level 0.05 with a power of 0.987. Thus, in our sample the seniority of reviewer did not mitigate the tendency to exhibit hindsight bias.

### Discussion

Clinical review of case notes following a patient’s death is an established process firmly embedded within the clinical governance processes of NHS trusts. However, the rigour with which this takes place has recently attracted considerable scrutiny in the national media following the tragic death of Connor Sparrowhawk<sup>1</sup> and certain other high profile incidents.<sup>13,14</sup>

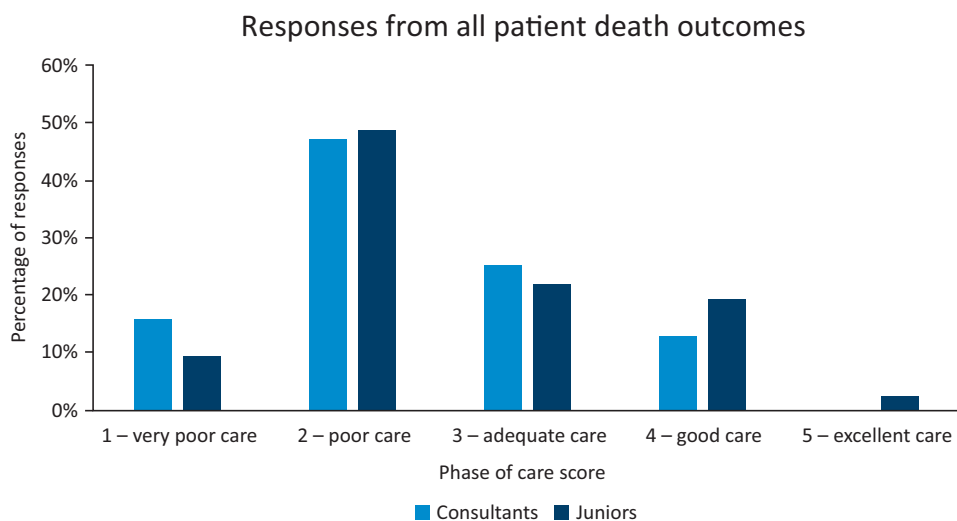
These tragedies have understandably focussed attention on NHS morbidity and mortality case note review, with a number of recent reports recommending reform to how these are conducted.<sup>9,15</sup> In response, the secretary of state for health endorsed robust reform of these processes as a national priority for the NHS.<sup>15</sup>

A number of tools have been proposed to improve the quality of case note review by clinicians. Among these is the Structured

Judgement Review (SJR),<sup>16</sup> which has been supported by the Royal College of Physicians as a means to provide more detailed scrutiny of case notes to identify potential instances of harm and learning which may prevent similar occurrences.<sup>17,18,19</sup> One element of this tool is the application of a five point scale to different phases of care, with supporting statements to indicate why a particular phase of care may be poor, adequate or good.

We wanted to determine whether physicians exhibited hindsight bias when reviewing case vignettes using a Likert scale similar to that used in the SJR. To do this we asked clinicians to review three sets of case notes, but randomised them in each case to an outcome in which the patient dies shortly after discharge or makes a full recovery. Irrespective of the final outcome, participants were invited to provide a score relating to exactly the same antecedent care provided until the patient’s discharge.

In two out of three cases we observed marked hindsight bias, in which clinicians exhibited significantly more critical assessments of the same antecedent care when the patient had died. In the first case, this corresponded to a median response rating identical care



**Fig 4. Participant seniority did not have a significant effect on the tendency to exhibit hindsight bias in this study.**

as poor when the patient had died, but as good when the patient made a full recovery. In the second case the median responses corresponded to poor care when the patient died, and adequate when the patient made a full recovery. It is notable that in all cases participants' scores were highly variable and always spanned at least four points out of a five point scale, irrespective of the vignette outcome.

In the final case, no hindsight bias was observed. The reason for this is not clear and was not tested. One possible explanation is that as participants encountered the scenarios in the same order (leg pain, headache, then chest pain) they increasingly self-censored hindsight bias as they progressed through the vignettes. Another explanation might be that aortic dissection is simply regarded by clinicians as an easier diagnosis to miss. However, it is our opinion that another explanation is most likely. In the third case, the vignette described a patient who had not attended prior appointments and then self-discharged. We suggest that in all three cases, participants identified a human agent to censure for a negative outcome – only in the third case, it was the patient and not a healthcare provider. It is possible that even though the wrong test for exclusion of aortic dissection was arranged, criticism was withheld owing to participants having judged the patient pejoratively and refraining from identifying further causal factors.

These observations demonstrate that in certain (but not all) cases, clinicians are prone to judging identical antecedent healthcare more harshly when a patient dies compared to when the patient survives. In our opinion, this could have both positive and negative consequences for doctors, patients and the NHS, which merit further discussion.

### Positive consequences

An important goal of morbidity and mortality processes is to identify opportunities for learning and improvement.<sup>20</sup> This intent would clearly be best served if clinicians are biased towards identifying things that could have been done better during a patient's assessment. Thus, heuristic biases that lead doctors to be more critical of antecedent care when a patient dies should enhance the opportunity to identify process and systems changes.<sup>4,20</sup> This might be expected to reduce the likelihood of future errors, and is a foreseeable positive consequence of the clinician hindsight bias observed in this study.

Given that the NHS only routinely reviews patient case notes when a patient dies, the tendency to exhibit hindsight bias when reviewing case notes may serve to improve future patient care by being highly sensitive to identifying areas for improvement and learning within a healthcare service.

### Negative consequences

It is possible that the tendency of doctors to exhibit hindsight bias could have negative consequences, too.

Firstly, participants' exhibited more lenient assessment of care when a patient was described as surviving. This raises the possibility that case note review might fail to identify healthcare errors when a patient's care might have been substandard, but the patient did not come to harm. Thus, the case note reviewer might lose an opportunity to learn from a near-miss event.

Secondly, case notes are generally only reviewed in clinical governance audits when a patient has died or something has gone wrong. This causes ascertainment bias, as patients are already

known to have died or come to harm.<sup>21</sup> When combined with the hindsight bias observed in this study, it is possible that clinicians might conclude care was poor and patients came to harm in certain cases purely as a consequence of bias.

As many healthcare organisations now have a statutory duty of candour to inform patients and their relatives if they have come to harm as a result of poor care, this could cause potentially incorrect, unnecessary and significant emotional harm to families who are likely to be distressed already.<sup>22</sup>

Thus, we predict that families will be told that patients died because of healthcare errors when this impression arose from a combination of hindsight bias and ascertainment bias, but is not necessarily a fact.

Thirdly, in the two cases in which hindsight bias was observed, we noted that the median phase of care scores differed significantly depending on the outcome described. In the case of a painful leg, the scores represented 'good' and 'poor' care. In the case of headache the median scores represented 'adequate' and 'poor' care. These differences in assessment straddle an important medicolegal and regulatory inflection point with respect to consequences and restrictions imposed by courts and regulators, for example the General Medical Council.

It is important to emphasise that our survey explicitly directed participants to assume the role of a physician conducting case note review for an internal clinical governance meeting and was not designed to allow generalisation of these observations more broadly. It is imaginable, however, that these observations might also apply in the regulatory or medicolegal spheres of influence. For broader application of these findings, we suggest three considerations. First, more work would be useful to determine whether the same heuristic biases exist when physicians conduct case note review for medicolegal or regulatory reasons, rather than solely for clinical governance audit. Second, it would be important to establish whether controls to censor hindsight bias in standard medicolegal and regulatory processes have been omitted. Our observations show that the existence of such controls would be extremely important. Finally, judgements and sanctions would need to depend heavily upon biased retrospective review of patient case notes by clinical peers, rather than allegations relating to demonstrable fraud, seriously impaired probity or other misdemeanour.

We suggest that if all these tests were met, then certain professional restrictions or legal convictions could be based upon evidence that is not sound.

### Strengths and limitations

This study describes an area of clinical governance that is of great importance to healthcare systems and patients alike. Strengths of this paper include blinding of participants to alternative outcomes, a relatively good number of recruitees that permitted clear statistical analysis, and high survey completion rates.

It is also important to note that our study has limitations. Firstly, our approach was not identical to completing a comprehensive SJR as it used only a single phase of care score for the patients' admission assessment, and did not involve providing explicit judgements as supporting text. Thus it is possible, although not certain, that a formal and rigorous SJR might better censor hindsight bias. This would be worthy of further study. Secondly, as the cases were always presented in the same order (leg pain, then headache, then chest pain) and only the outcome varied for each

case it is possible that participants began to self-censor hindsight bias as they progressed through the vignettes, contributing to the final case showing no hindsight bias. Thirdly, we randomised participants to a positive outcome of recovery or a negative outcome of patient death. In our design, there was no completely blinded outcome in which no details of subsequent events were provided. Thus, the extent to which hindsight bias caused lenient assessments when a patient survived, as opposed to critical assessments when they died, was not directly tested. Finally, although a clear hindsight bias was observed in two cases, the reason why it was not apparent in the third case was not tested, and could be accounted for by varying explanations. ■

### Supplementary material

Additional supplementary material may be found in the online version of this article at [www.clinmed.rcpjournals.org](http://www.clinmed.rcpjournals.org):

- S1 – Case one – leg pain.
- S2 – Case two – headache.
- S3 – Case three – chest pain.

### References

- 1 Dyer C. Psychiatrist who failed in care of teenager with epilepsy who died in bath is suspended for 12 months. *BMJ* 2018;360:k908.
- 2 Hutchinson A, Coster JE, Cooper KL *et al*. Comparison of case note review methods for evaluating quality and safety in health care. *Health Technol Assess* 2010;14:1–170.
- 3 Hutchinson A, Coster JE, Cooper KL *et al*. A structured judgement method to enhance mortality case note review: development and evaluation. *BMJ Qual Saf* 2013;22:1032–40.
- 4 David G. To make good use of medical error. *Bull Acad Natl Med* 2003;187:129–36.
- 5 Caplan RA, Posner KL, Cheney FW. Effect of outcome on physician judgments of appropriateness of care. *JAMA* 1991;265:1957–60.
- 6 Hugh TB, Tracy GD. Hindsight bias in medicolegal expert reports. *Med J Aust* 2002;176:277–8.
- 7 Hugh TB, Dekker SW. Hindsight bias and outcome bias in the social construction of medical negligence: a review. *J Law Med* 2009;16:846–57.
- 8 Fischhoff B. Hindsight ≠ foresight: the effect of outcome knowledge on judgment under uncertainty. *BMJ Quality & Safety* 2003;12:304–11.
- 9 Care Quality Commission. *Learning, candour and accountability*. CQC, 2018. [www.cqc.org.uk/publications/themed-work/learning-candour-and-accountability](http://www.cqc.org.uk/publications/themed-work/learning-candour-and-accountability) [Accessed 30 April 2018].
- 10 Birch SAJ, Brosseau-Liard PE, Haddock T *et al*. A ‘curse of knowledge’ in the absence of knowledge? People misattribute fluency when judging how common knowledge is among their peers. *Cognition* 2017;166:447–58.
- 11 Arkes H. The consequences of the hindsight bias in Medical decision making. *Psychological science* 2013;22:356–60.
- 12 Manaseki-Holland S, Lilford RJ, Bishop JRB *et al*. Reviewing deaths in British and US hospitals: a study of two scales for assessing preventability. *BMJ Qual Saf* 2017;26:408–416.
- 13 Iacobucci G. Troubled mental health trust is still putting patients at risk, warns CQC. *BMJ* 2016;353:i2483.
- 14 Torjesen I. Southern Health is fined £2m over deaths of two patients. *BMJ* 2018;360:k1411.
- 15 National Quality Board. *National guidance on learning from deaths*. National Quality Board, 2018. [www.england.nhs.uk/wp-content/uploads/2017/03/nqb-national-guidance-learning-from-deaths.pdf](http://www.england.nhs.uk/wp-content/uploads/2017/03/nqb-national-guidance-learning-from-deaths.pdf) [Accessed 30 April 2018].
- 16 Royal College of Physicians. *Using the structured judgement review method. A guide for reviewers*. RCP, 2016. [www.rcplondon.ac.uk/sites/default/files/media/Documents/NMCCR%20clinical%20governance%20guide\\_1.pdf](http://www.rcplondon.ac.uk/sites/default/files/media/Documents/NMCCR%20clinical%20governance%20guide_1.pdf) [Accessed 30 April 2018].
- 17 Roberts AP, Morrow G, Walkley M *et al*. From research to practice: results of 7300 mortality retrospective case record reviews in four acute hospitals in the North-East of England. *BMJ Open Qual* 2017;6:e000123.
- 18 National Confidential Enquiry into Patient Outcome and Death. *Caring to the End? A review of the care of patients who died in hospital within four days of admission*. NCEPOD, 2009. [www.ncepod.org.uk/2009report2/Downloads/DAH\\_report.pdf](http://www.ncepod.org.uk/2009report2/Downloads/DAH_report.pdf) [Accessed 30 April 2018].
- 19 Academy of Medical Royal Colleges. *A framework of operating principles for managing invited reviews within healthcare*. Academy of Medical Royal Colleges, 2016. [www.aomrc.org.uk/wp-content/uploads/2016/05/Invited\\_reviews\\_210116.pdf](http://www.aomrc.org.uk/wp-content/uploads/2016/05/Invited_reviews_210116.pdf) [Accessed 30 April 2018].
- 20 Rafter N, Hickey A, Condell S *et al*. Adverse events in healthcare: learning from mistakes. *QJM* 2015;108:273–7.
- 21 Rydon-Grange M. ‘What’s Psychology got to do with it?’ Applying psychological theory to understanding failures in modern health-care settings. *Journal of Medical Ethics* 2015;41:880–4.
- 22 General Medical Council. *Openness and honesty when things go wrong: The professional duty of candour*. GMC, 2015. [www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/candour---openness-and-honesty-when-things-go-wrong](http://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/candour---openness-and-honesty-when-things-go-wrong) [Accessed 30 April 2018].

**Address for correspondence: Dr Edward Banham-Hall, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge CB2 0QQ.**  
**Email: [edward.banhamhall@addenbrookes.nhs.uk](mailto:edward.banhamhall@addenbrookes.nhs.uk)**