

Hypnotic depth and postoperative death: a Bayesian perspective and an Independent Discussion of a clinical trial

Phillip E. Vlisides¹, John P. A. Ioannidis² and Michael S. Avidan^{3,*}

¹University of Michigan Medical School, Department of Anesthesiology, Ann Arbor, MI, USA, ²Stanford University, Meta-Research Innovation Center, Palo Alto, CA, USA and ³Washington University in Saint Louis School of Medicine, Department of Anesthesiology, St. Louis, MO, USA

*Corresponding author. E-mail: avidanm@wustl.edu

Several observational studies have found that measures of 'deep hypnosis' during general anaesthesia were associated with increased odds of intermediate-term postoperative mortality.^{1,2} A meta-analysis evaluated eight observational studies that included 40 317 patients; overall there was a statistically significant association between the depth of hypnosis, reflected by low bispectral index (BIS) values (<40–45), and mortality (pooled adjusted hazard ratio, 1.21; 95% confidence interval [CI], 1.07–1.38; $P=0.003$).¹ However, given that observational studies are vulnerable both to known and unknown confounders, it is difficult to disentangle whether or to what extent 'deep hypnosis': 1) directly causes harm, 2) mediates harm when patients are vulnerable, or 3) reflects patient morbidity, and therefore merely predicts 'inevitable' poor outcomes but has no causal relationship whatsoever. According to conventional wisdom, well-designed randomised trials would be required to solve this important mystery. In this issue of the *British Journal of Anaesthesia*, Sieber and colleagues³ report the secondary outcomes of the STRIDE (A STRategy to Reduce the Incidence of Postoperative Delirium in Elderly Patients) clinical trial, in which older adults undergoing hip fracture repair under spinal anaesthesia were randomised to 'lighter' or 'deeper' sedation.³ Primary outcome results (on the incidence of in-hospital delirium between days 1 and 5) have been

previously reported, showing no overall significant benefit with 'lighter' sedation, but a possible effect modification, based on severity of preoperative morbidity, was suggested.⁴ The secondary outcomes of interest in the current study were intermediate-term mortality and return to preoperative ambulation level. It has been suggested that an independent Discussion for scientific manuscripts can be useful in objectively assessing inferential reproducibility.⁵ In **Table 1**, we provide a structured independent Discussion for the study by Sieber and colleagues,³ and we comment on the similarities and differences between the original Discussion and the independent Discussion as an example of this approach. The author of the Independent Discussion (PEV) had access only to the manuscript's Introduction, Methods and Results, but not to the Abstract or the Discussion.

Regarding the specific hypotheses tested in the study by Sieber and colleagues,³ it is important to consider whether it is biologically plausible that 'deeper' hypnotic levels could increase the risk of intermediate-term postoperative death. There are several postulated mechanisms by which 'deeper' hypnosis could be injurious. First, anaesthetic agents might be directly neurotoxic, and increased administration could have adverse postoperative cognitive consequences. Second, 'deep' hypnosis could cause haemodynamic instability, which could increase the risk of postoperative organ injury, including the heart, brain, and kidneys. Third, increased anaesthetic agent concentrations could have immunosuppressant effects, which could increase susceptibility to infection or cancer growth. All

DOI of original article: doi: 10.1016/j.bja.2018.12.021.

Accepted: 17 January 2019

© 2019 Published by Elsevier Ltd on behalf of British Journal of Anaesthesia.
For Permissions, please email: permissions@elsevier.com

Table 1 Original and independent discussions, and interpretive commentary.

<p>Title of manuscript: Depth of sedation as an interventional target to reduce postoperative delirium: mortality and functional outcomes of the STRIDE randomised clinical trial³</p> <p>Main finding</p> <p>Original Discussion: 'This manuscript reports the secondary outcomes of the STRIDE trial. When comparing heavier vs lighter sedation during spinal anaesthesia for hip fracture repair, there was no significant difference between study groups in delirium incidence at 30 days, functional outcomes at 12-month follow-up, and 1-year mortality. These results are reassuring in that level of intraoperative sedation comparable to those studied in STRIDE are probably not associated with worsened functional outcomes or increased mortality up to 1 year after surgery.'³</p> <p>Independent Discussion: In this secondary analysis of the STRIDE trial, depth of sedation during spinal anaesthesia was not associated with mortality, regardless of comorbidity burden. Likewise, sedative depth demonstrated no appreciable effect on return to pre-fracture ambulation level, persistent delirium, or additional functional outcomes after adjusting for relevant confounders. The analysis does not provide evidence to suggest that depth of intraoperative sedation influences mortality, functional outcomes, or persistent delirium after hip fracture surgery.</p> <p>Commentary: The interpretation of the main findings in the Original and Independent Discussions are concordant.</p> <p>Relationship of main finding to previous studies</p> <p>Original Discussion: 'Hip fracture repair is associated with high postoperative mortality. Large case series report a 5.1% 30-day and 25% 1-year mortality, with a mean survival of 3.75 years (95% CI, 3.13 to 4.54). The 14% 1-year mortality in the current study is comparable to other prospective hip fracture series, but lower than the smaller randomized trial conducted by members of the same research team as the STRIDE study. Exclusion criteria (severe dementia and medications which prevent the safe administration of spinal anaesthesia) were similar between these studies. However, anaesthetic management in STRIDE was performed exclusively by 4 senior anaesthesiologists and may have contributed to the different findings in the two studies. In addition, as discussed below, ongoing changes in management of hip fractures occurred at our institution over the time period of the two trials. The baseline risk factors of low BMI, incremental IADL decreases and incremental CCI increases for 1-year mortality observed in STRIDE are consistent with the literature. However, the previously reported higher mortality with male sex was not observed, possibly owing to our predominantly female population. Since the use of intraoperative electroencephalogram-based depth of anaesthesia monitoring has become commonplace, several observational case series report an association between deeper anaesthesia levels and mortality. However, it is unclear whether this link represents an epiphenomenon, where deeper anaesthesia is merely a marker of poor prognosis, particularly since these studies have not assessed for individual sensitivity to anaesthetics. We found that neither interventional group nor individual BIS values were risk factors for 1-year mortality. Furthermore, it is unlikely that anesthetic depth management alone would influence one year mortality given that patients in this age group and undergoing this type of surgery would likely have other events after the index surgery that may influence longer term survival. Our bias is that if any differences were to be found, they might assert themselves in the first 30 days postoperatively. We report 30 day complications in Table 1 and show no difference between intervention groups. For this reason we further investigated other predictors of 1-year mortality and reported them in the manuscript. Of note, we found, as others have, that severity of POD in hospital was predictive of mortality at 1-year, after multivariable adjustment. Heavier vs lighter sedation levels during hip fracture repair under spinal anaesthesia have been associated with higher mortality in patients with CCI score >4. STRIDE found no interaction between interventional group and CCI for 1-year mortality. The reasons behind study outcome differences are complicated, but could be explained by ongoing changes in practice as evidenced by the shorter time to surgery, shorter hospitalization following surgery, and fewer ICU days in the STRIDE trial compared to the previous, smaller trial. In addition, the incidence of pulmonary and cardiovascular complications was higher in subjects with CCI>4 comparing heavier vs lighter sedation in the previous trial; whereas no difference was found between treatment arms in STRIDE.'³</p> <p>Independent Discussion: One-year mortality is described as the main outcome of this secondary analysis, and depth of sedation was not associated with 1-yr mortality. Other than the prior small trial described by the investigators,⁶ there are no studies specifically examining the depth of sedation during spinal anaesthesia in relation to mortality in patients with hip fractures. However, observational, interventional, and meta-analytic studies have investigated whether anaesthetic technique (i.e. general vs regional anaesthesia) might be causally linked to outcomes such as mortality after hip fracture surgery. Two recent systematic reviews found no significant association between mortality and anaesthetic technique in adults presenting for hip fracture surgery, although considerable clinical and methodological heterogeneity was present among studies.^{7,8} Additionally, a 2016 Cochrane Review found no significant association between 30-day mortality and anaesthetic technique (risk ratio 0.78; 95% CI, 0.57–1.06; I²=24%).⁹ Quality of evidence, however, was rated very low. Subsequent to this review, Ahman and colleagues¹⁰ performed a registry-based retrospective cohort study that included more than 14 000 hip fracture patients, and type of anaesthesia was not independently predictive of mortality (hazard ratio 1.08; CI, 0.99–1.18, P=0.084); rather, comorbidity burden was most strongly predictive (ASA physical status 5: hazard ratio 12.57; 95% CI, 6.91–22.85, P<0.001; ASA physical status 4: hazard ratio 4.79; 95% CI, 4.30 to 5.33, P<0.001). Thus, it is not surprising that sedative depth was not found to be associated with mortality in this secondary analysis. Comorbidity burden may be a significant driver of postoperative mortality after hip fracture, as demonstrated here and in prior investigations.^{10,11}</p> <p>Commentary: In the Original Discussion, the investigators focus on reasons why the findings of the current study were different from a smaller previous study with a similar design done by the same team.⁶ The Independent Discussant notes that while there has been little research examining whether depth of sedation affects outcomes after hip fracture surgery, several clinical trials and observational studies have investigated whether the outcomes are different after general anaesthesia vs regional anaesthesia. These studies have generally found that regional anaesthesia was not associated with significant improvements in outcomes, including death. The Independent Discussant opines that the findings in the current study are consistent with this prior body of evidence. The Independent Discussion places the current trial in a broader and more systematic context than the investigators do, while the Original Discussion focuses more on juxtaposing the prior and current experience of the specific team, which conducted the studies. Changes in the practices of the team over time may not necessarily be generalisable to changes in practices in other teams.</p> <p>Additional (secondary) findings</p> <p>Original Discussion: 'Neither the occurrence of POD nor days of POD in hospital were predictors of 1-year mortality in STRIDE. Similarly, previous studies have found that, when accounting for confounders, incident delirium was not an independent risk</p>
--

Continued

Table 1 Continued

factor for intermediate term mortality. However, several delirium-associated characteristics do appear to be associated with mortality. Specific to surgical patients, trend-level associations between delirium severity and 6-month mortality following hip fracture repair have been reported, but STRIDE is the first study to demonstrate statistically significant relationships between POD severity and 1-year mortality. Participants with lower pre-fracture ambulation level had greater odds of maintaining level of function by 1-year compared to participants with a higher pre-fracture ambulation level. This indicates a floor effect in terms of return to function with poorer pre-fracture ambulation status and is consistent with previous reports documenting decreased odds of change in ambulatory status as pre-fracture ambulatory function decreases.³

Independent Discussion: In terms of the primary functional outcome, return to pre-fracture ambulation level, adjusted odds of returning to baseline were not significantly different between the study groups. In contrast, preoperative CCI, preoperative ambulatory status, and postoperative delirium were independently predictive of lack of return to pre-fracture ambulation level. The overall incidence of persistent delirium was low (2%) and not significantly different between groups. Lastly, additional functional outcomes (e.g. Activities of Daily Living, Instrumental Activities of Daily Living (IADL), Grip Strength, Timed Chair Raise, and Timed 3-m Walk) were not significantly different between groups both 1 month and 1 yr after operation.

Commentary: In the Original Discussion, the investigators focus on predictors of 1-yr mortality and ability to ambulate. The Independent Discussant focuses on predictors of ambulation ability, and functional outcomes. In general, there is substantial concordance between the two Discussions, although there is only modest overlap regarding what is highlighted in the two Discussions.

Relationship of additional (secondary) findings to previous studies

Original Discussion: 'Other studies have reported an association between poor recovery of ambulation following hip fracture repair and CCI as well as delirium episodes, similar to the STRIDE results.'³

Independent Discussion: There are limited data on sedative depth, particularly during spinal anaesthesia, and functional outcomes after hip fracture surgery. In 2018, Patel and colleagues⁷ performed a systematic review that analysed a broad range of hip fracture outcomes related to anaesthetic technique. In the review, 11 studies (between 1985 and 2018) reported on various postoperative functional outcomes (e.g. time to ambulation, functional independence). No consistent benefits or harms were identified in association with anaesthetic technique, though the quality of evidence was limited by sample size, variability in (or unknown) assessment methodology, and non-randomised study designs. In a single-centre, randomised controlled trial, patients with hip fracture were randomised to general or spinal anaesthesia, and those in the spinal group demonstrated a slightly lower postoperative IADL score (2.35 [1.06]) compared with the general anaesthesia group (3.03 [1.55], $P=0.043$).¹² The absolute difference, however, was less than one point, and this was a secondary outcome measure. These investigators also examined cognitive function, using 10 neuropsychological tests, at 1 month after operation (which parallels the 'persistent delirium' designation in this study). No consistent differences in 30-day cognitive function were appreciated between groups. Overall, there is a dearth of high-quality evidence to support—or refute—associations between anaesthetic/sedative depth and long-term functional or cognitive outcomes after hip fracture surgery. However, evidence to date does not demonstrate associations between anaesthetic technique and functional or cognitive outcomes 1 month after operation.

Commentary: In the Original Discussion, the emphasis is mostly on the relationship between hip fracture surgery, comorbidities, and delirium metrics on the one hand, and recovery of ambulation on the other hand. In contrast, the Independent Discussant examines the evidence regarding anaesthetic technique and functional outcomes, and concludes that the quality of evidence in this regard is weak, after a more extensive and systematic assessment of the evidence.

Limitations

Original Discussion: '1) The sample size of STRIDE was based on the hypothesized intervention effect on the primary outcome of delirium incidence during postoperative Day 1 to Day 5, or to hospital discharge. Because mortality was the secondary outcome of greatest interest, the minimal size of intervention effect on mortality that could be detected after sample size determination was evaluated. However, there was no formal determination of sample size requirements in relation to any secondary outcomes. In addition, most of the modeling with adjustment was done for exploratory purposes, so there were also no power calculations for these. 2) A single site trial limits generalisability. 3) Exclusion of participants with MMSE <15 may have limited the sample to a more cognitively intact cohort than the usual hip fracture population. 4) Exclusion of anticoagulant users may have biased toward a population with fewer cardiovascular risk factors. 5) Pre-fracture walking ability was assessed by patient and family interview, and may be subject to bias. 6) Averaging BIS values over the intraoperative period can conceal potentially important short periods of low BIS as well as cumulative duration of low BIS, both of which have been associated with worse intermediate term survival.'³

Independent Discussion: This was a single-centre trial, which limits generalisability. Additionally, as a secondary analysis, there are inherent limitations. STRIDE was originally powered for postoperative delirium (primary outcome), reported previously.⁴ Although power calculations were performed for mortality (secondary outcome), the mortality incidence in each group (14%) was lower than the preliminary assumption (31.5%); in fact, the absolute risk reduction magnitude (15.5%) for which power was calculated was larger than the actual 1-yr mortality rate in either group. Accordingly, this secondary analysis was underpowered for detecting associations between sedative depth and mortality.

Commentary: In the Original Discussion, the investigators highlight several weaknesses mostly related to the study population, and these are not dealt with in the Independent Discussion. In contrast to this, the Independent Discussant focuses only on mortality, and suggests that the both the *a priori* estimated mortality incidence and the estimated mortality reduction with the intervention were unrealistically high.

Strengths

Original Discussion: '1) The study was a randomized, double blinded design with excellent intermediate term follow up to 1 year. 2) There was a high rate of enrollment. 3) Assessments of the delirium outcome were rigorously conducted by multidisciplinary consensus panel using all available data. 4) STRIDE was a large enough study to detect a clinically important 15% difference in mortality between treatment groups.'³

Independent Discussion: STRIDE was a double-blind, randomised clinical trial that was registered with pre-defined outcomes. To demonstrate the fidelity of the trial intervention, the investigators also report data on depth of sedation (both Observer's Assessment of Alertness/Sedation Scale and BIS). Rigorous delirium assessments were performed, which incorporated Diagnostic and Statistical Manual of Mental Disorders (DSM)-based criteria and a consensus panel for adjudication. The statistical plan for these secondary analyses was also finalised and approved before data review. Mortality and functional outcome regression models

Continued

Table 1 Continued

included multiple covariates for adjustment, in addition to the randomised nature of the intervention. Lastly, the investigators have made trial data available to interested researchers to facilitate transparency and reproducibility.

Commentary: There are multiple strengths that are highlighted in both Discussions. There is general agreement on these strengths.

Future directions

Original Discussion: 'In planning the STRIDE trial the mechanisms studied focused on intermediate term effects of delirium and its relationship to both mortality and function. The intervention did not affect overall delirium incidence in hospital, nor did it affect the intermediate term outcomes studied. Yet, the importance of postoperative delirium still was apparent when analyzing factors associated with mortality and function. In addition, the same theme emerges throughout both the short term and intermediate term study results, and that is the role of underlying co-morbidity in determining not only incident delirium in hospital, but intermediate term outcomes as well. Although the intervention did not produce an overall decrease in delirium incidence or intermediate term outcomes in the elderly hip fracture population, it is hoped that reporting the other covariates important in determining intermediate term outcomes will help to stimulate further work focusing on perioperative management based on comorbidity as a possible means of preventing postoperative delirium or reducing delirium severity, and thus improving surgical outcomes in the elderly.'³

Independent Discussion: Reducing mortality in this population may prove challenging given the urgent nature of surgery and comorbidity-driven risk factors.^{10,11} Surgical factors (e.g. bone cementation¹¹) could be further explored. Novel and early strategies involving physical and occupational therapy may help patients to recover function, particularly those with high pre-fracture ambulatory status (given the inherent difficulty with returning to high pre-fracture levels). Barriers to physical and occupational therapy participation could also be explored in this context. Regarding persistent delirium, now described as delayed neurocognitive recovery,¹³ additional neuroscientific investigation is warranted to understand the pathophysiology and risk factors associated with such syndromic states.

Commentary: Both Discussions address the importance going forward of improving our understanding of delirium, optimising perioperative care, and enhancing surgical outcomes in older adults. The Original Discussion continues to place emphasis on evaluating lines of research related to comorbidity, which was felt to be an important effect modifier in the prior reporting of the primary outcome of the STRIDE trial.⁴

Conclusion

Original Discussion: 'Nonetheless, the STRIDE results provide reassurance for the anaesthesiologist managing hip fracture patients. Particularly for 1-year mortality, given the ongoing controversy surrounding the association between anaesthetic depth and mortality, the STRIDE interventional groups had similar outcomes. Furthermore, the heavier levels of sedation studied in STRIDE were not extreme and are consistent with current practice in the United States.'³

Independent Discussion: In this secondary analysis of the STRIDE trial, depth of sedation during spinal anaesthesia was not associated with postoperative mortality, functional outcomes, or persistent delirium.

Commentary: The overall conclusion is concordant in the two Discussions.

Inferential reproducibility

Commentary: The major inferential difference between the Discussions is in relation to appropriateness of the sample size. In the Original Discussion the investigators opine that the study was large enough to detect a clinically meaningful reduction in mortality. In contrast, the Independent Discussion infers that the estimated mortality was too high and that the estimated decrease in mortality with the intervention was unrealistic; thus, with only 200 patients, the study was not sufficiently large to address the research question. There are also differences in emphasis in the Discussions regarding existing evidence and contextualization, and whether comorbidity should be a major issue for future research. In many other respects, there is inferential reproducibility between the Discussions.

of these mediating actions and early postoperative adverse events could in turn impact intermediate-term functionality and survival (Fig. 1).

If 'deeper' hypnosis were harmful, guiding anaesthetic administration with a 'depth of hypnosis' monitor could potentially mitigate the damage. Several randomised trials, with variable follow-up periods, have been conducted using 'depth of hypnosis' monitors in the intervention group, allowing us to conduct an exploratory examination of this hypothesis. Most of these studies have used the BIS processed EEG monitor to guide anaesthetic titration. The primary outcomes have usually been intraoperative awareness, postoperative delirium, or postoperative cognitive dysfunction. The findings in relation to mortality (as a *post hoc* secondary outcome) are summarised as follows. 1) Leslie and colleagues¹⁴ followed up on patients in the B-Aware trial after a median of 4.1 yr: 548 patients (22.2%) had died since the index surgery. The overall mortality was 20.6% (252/1225) in the BIS-guided group and 23.9% (296/1238) in the control group (difference=-3.34%; 95% CI, -6.67 to 0.01 to, P=0.05).¹⁴ 2) In an analysis of cardiac surgery patients in the B-Unaware trial, Kertai and colleagues¹⁵ reported no significant difference in mortality rates (>3 yr) between patients who had been randomised to the BIS-guided group vs the routine care

group (mortality=19.7% [47/239] in the BIS group vs 15.8% [35/221] in the control group; difference=3.83%; 95% CI, -3.53 to 11.06, P=0.33). 3) In a separate analysis of non-cardiac surgery patients in the B-Unaware trial, Kertai and colleagues¹⁶ reported no significant difference in mortality rates (over approximately 4 yr) between patients who had been randomised to the BIS-guided group vs the routine care group (mortality=24.9% [180/723] in the BIS group vs 23.7% [178/750] in the control group; difference=1.2%, 95% CI, -3.3 to 5.6, P=0.63). 4) Avidan and colleagues¹⁷ reported 30-day mortality for patients in the BIS or anesthesia gas to reduce explicit recall (BAG-RECALL) trial. They found that 30-day mortality was 1.96% (57/2907 patients) in the BIS-guided group and 2.21% (64/2902 patients) in the control group (difference=-0.24%; 95% CI, -0.99 to 0.50, P=0.51). 5) In the cognitive dysfunction after anesthesia (CODA) trial, Chan and colleagues¹⁸ found 26/462 (5.6%) deaths at 3 months in the 'light' group, and 22/459 (4.8%) deaths in the 'deep' group (difference=0.8%; 95% CI, -2.26 to 3.95, P=0.57). 6) In the dexamethasone, light anesthesia and tight glucose control (DeLiT) study, mortality was the primary outcome.¹⁹ The trial was terminated early for futility; at 1 yr after surgery, Abdelmalak and colleagues¹⁹ found 22/194 (11.3%) deaths in the 'light' group and 23/187 (12.3%) in the 'deep' group

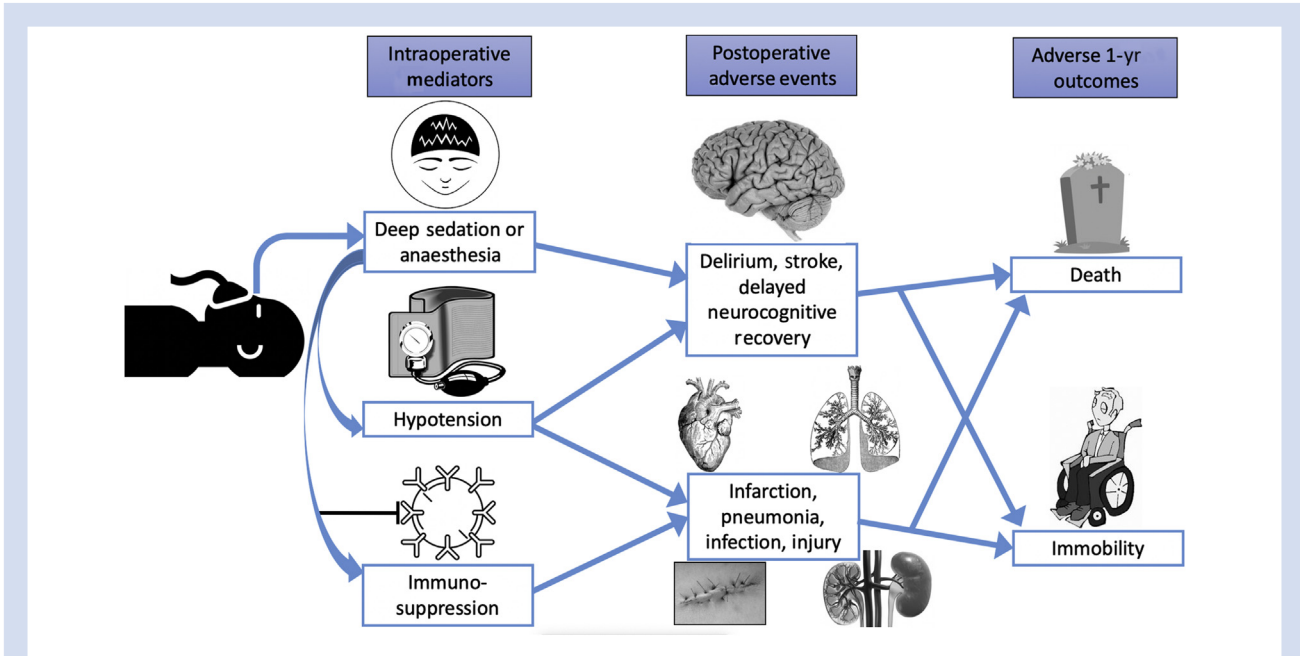


Fig 1. Deep sedation or anaesthesia and poor intermediate-term outcomes. This figure illustrates possible intraoperative mediators and postoperative adverse events associated with ‘deeper’ hypnosis during sedation or general anaesthesia, which could in turn increase the likelihood of intermediate-term immobility and death.

(difference=−0.96%, 95% CI, −7.96 to 5.97 to, $P=0.77$). 7) In the surgery depth of anaesthesia cognitive outcome (SuDoCo) trial, Radtke and colleagues²⁰ reported 5.4% 90 day mortality (31) in the BIS-guided group and 5.3% in the control group (difference=0.05%; 95% CI, −2.73 to 2.83, $P=0.97$). 8) In a 114 patient study by Brown and colleagues,⁶ which provided motivation for the STRIDE trial, mortality did not differ significantly between groups at 1 yr after surgery. There were 11/57 (19.3%) deaths in the ‘light’ sedation group, and 17/57 (29.8%) in the ‘deep’ sedation group (difference=−10.53%,

95% CI, −26.87 to 6.53; $P=0.19$).⁶ 9) The STRIDE trial reported 1-yr mortality as a pre-specified secondary outcome. There were 14/100 (14%) deaths in each group (difference=0%, 95% CI, −10.5 to 10.5).³ 10) Finally, in the electroencephalography guided anaesthesia to alleviate geriatric syndromes (ENGAGES) trial, Wildes and colleagues²¹ reported that 30-day postoperative mortality was four of 614 (0.7%) in the EEG-guided group and 19/618 (3.1%) in the usual care group (difference=−2.42%, 95% CI, −4.25 to −0.81). **Figure 2** shows a meta-analytic summary of these 10 trials.

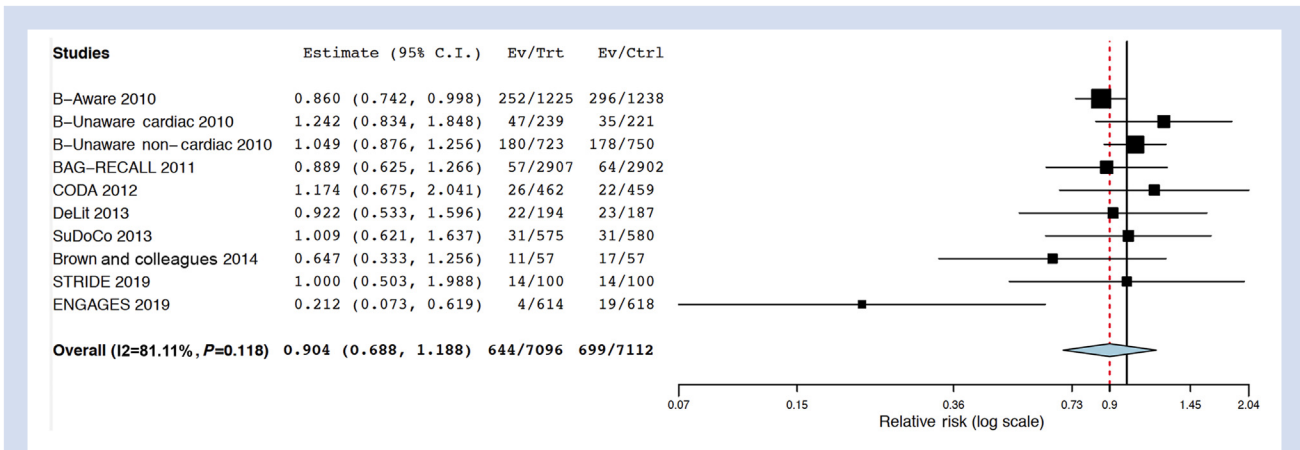


Fig 2. Meta-analysis summarising 10 trials in which the intervention group had received EEG or bispectral index (BIS) guidance, with or without the explicit goal of ‘light’ anaesthesia or sedation. This analysis was conducted using OpenMetaAnalyst.²² It is a binary, random effects, Hartung-Knapp-Sidik-Jonkman model.^{23,24} The $I^2=81\%$, $\tau^2=0.131$, $Q(df=9)=14.135$, and heterogeneity P -value=0.118. As shown in the figure, the estimated overall risk ratio for death with the intervention (BIS-guided [reduction in] sedation/anaesthesia)=0.904 (95% confidence interval, 0.688–1.188, $P=0.471$).

Some of the randomised trials were also included in the observational meta-analysis,¹ but with a focus on overall low BIS values rather than on randomisation groups. The individual trials constituting this meta-analysis reported mortality over widely differing time intervals, and there is considerable heterogeneity in the results. Much of the heterogeneity is driven by the recent ENGAGES trial which, in a *post hoc* exploratory analysis, found a very large decrease in 30-day mortality with EEG guidance of anaesthesia. Excluding the ENGAGES trial from the meta-analysis reduces the heterogeneity I^2 estimate from 81.1% to 30.5%. The overall point estimate of risk ratio for mortality with EEG-guided anaesthesia, excluding the ENGAGES trial from the meta-analysis, would be 0.964 (95% CI, 0.839–1.109, $P=0.610$). Taken individually and jointly, the assumption is made that any effect of the intervention in the trials on mortality is persistent and relatively constant over time (proportionality assumption). Also, the difference in anaesthetic management between the intervention and control groups varied markedly across the trials in the meta-analysis. The total sample size across all 10 trials is still only ~14 000 participants (approximately one-third of the sample size of the meta-analysed observational evidence¹), and the 95% CI of the summary estimate from the 10 trials includes at one end an increased risk ratio for death in the control group that is higher than the point estimate of increased risk of death (with deep anaesthesia) suggested by the observational evidence ($1/0.688=1.45$ vs 1.21). Therefore, some substantial benefit (and some substantial harm) in terms of mortality with BIS or EEG-guided anaesthesia cannot be excluded based on these data.

Despite these important caveats, the result of this meta-analysis weighs somewhat against 'hypnotic depth' having a large effect on postoperative mortality. The summary point estimate for relative risk is relatively close to the null (0.904). Thus, from a Bayesian perspective, any study evaluating the hypothesis that 'light' sedation or anaesthesia substantially decreases postoperative death would have a low pre-study probability of being true, even if an individual study were to find a large and 'significant' decrease in mortality with this approach,²⁵ such as the study by Wildes and colleagues.²¹ With a completed enrolment of 6500 patients, the Balanced Anaesthesia Trial²⁶ will make an important contribution to resolving a controversy that arose almost 15 yr ago when, in an observational study, Monk and colleagues²⁷ first reported an association between hypnotic depth during surgery and 1-yr postoperative mortality. However, the difficulty in attempting to answer this question 'definitively' is emphasised by the small estimated effect size of the intervention: approximately 10% relative risk reduction in death with 'light' sedation or hypnosis is suggested by the summary point estimate of the 10 trials. The practical implications are illustrated in the following conceptualisation. In designing a trial to detect an absolute decrease in 1-yr mortality from 10% to 9% (10% relative reduction) with >80% power and a statistical significance level of <0.005,²⁸ the trial would require ~23 000 patients per group. Even with a less stringent and more conventional statistical significance level of <0.05, an adequately powered (>80%) trial would require about 13 500 patients per study group. Yet a 1% absolute reduction in death should be considered clinically meaningful, as this would mean that for every 100 patients treated with 'lighter' anaesthesia, one life would be saved.

Overall, it is probably reasonable to conclude that biologically plausible 'negative' findings, such as those of Sieber and colleagues,³ are likely to reflect the 'truth': depth of hypnosis does not meaningfully affect intermediate-term postoperative outcomes. However, on the other hand, we must also realise that it is very difficult to detect and demonstrate a small effect of an intervention on a clinically important outcome such as death, and conventionally designed clinical trials of modest sample size might be inadequate for this purpose. Large pragmatic trials with far larger sample sizes, done in multi-centre settings, and using point of care randomisation to optimise the yield of recruitment, may have a role for securing more definitive answers to such questions.

Authors' contributions

Writing paper: MSA.

Contributed to writing paper: MSA, JPAI, PEV.

Writing Independent Discussion: PEV.

Writing commentary on the Discussions: MSA, JPAI.

Agree with the paper's results and conclusions: MSA, JPAI, PEV.

Approval of final version of this paper: all authors.

Have read, and confirm that they meet ICMJE criteria for authorship: all authors.

Declaration of interest

The authors declare that they have no conflicts of interest. MSA is an editor of the *British Journal of Anaesthesia*.

References

- Zorrilla-Vaca A, Healy RJ, Wu CL, Grant MC. Relation between bispectral index measurements of anesthetic depth and postoperative mortality: a meta-analysis of observational studies. *Can J Anaesth* 2017; **64**: 597–607
- Willingham M, Ben Abdallah A, Gradwohl S, et al. Association between intraoperative electroencephalographic suppression and postoperative mortality. *Br J Anaesth* 2014; **113**: 1001–8
- Sieber FE, Neufeld KJ, Gottschalk A, et al. Depth of sedation as an interventional target to reduce postoperative delirium: mortality and functional outcomes of the STRIDE randomized clinical trial. *Br J Anaesth* 2019; **122**: 480–9
- Sieber FE, Neufeld KJ, Gottschalk A, et al. Effect of depth of sedation in older patients undergoing hip fracture repair on postoperative delirium: the STRIDE randomized clinical trial. *JAMA Surg* 2018; **153**: 987–95
- Avidan MS, Ioannidis JPA, Mashour GA. Independent discussion sections for improving inferential reproducibility in published research. *Br J Anaesth* 2019; **122**: 413–20
- Brown 4th CH, Azman AS, Gottschalk A, Mears SC, Sieber FE. Sedation depth during spinal anesthesia and survival in elderly patients undergoing hip fracture repair. *Anesth Analg* 2014; **118**: 977–80
- Patel V, Champaneria R, Dretzke J, Yeung J. Effect of regional versus general anaesthesia on postoperative delirium in elderly patients undergoing surgery for hip fracture: a systematic review. *BMJ Open* 2018; **8**: e020757

8. O'Donnell CM, McLoughlin L, Patterson CC, et al. Perioperative outcomes in the context of mode of anaesthesia for patients undergoing hip fracture surgery: systematic review and meta-analysis. *Br J Anaesth* 2018; **120**: 37–50
9. Guay J, Parker MJ, Gajendragadkar PR, Kopp S. Anaesthesia for hip fracture surgery in adults. *Cochrane Database Syst Rev* 2016: CD000521
10. Ahman R, Siverhall PF, Snygg J, et al. Determinants of mortality after hip fracture surgery in Sweden: a registry-based retrospective cohort study. *Sci Rep* 2018; **8**: 15695
11. Kilci O, Un C, Sacan O, et al. Postoperative mortality after hip fracture surgery: a 3 years follow up. *PLoS One* 2016; **11**: e0162097
12. Tzimas P, Samara E, Petrou A, Korompilias A, Chalkias A, Papadopoulos G. The influence of anesthetic techniques on postoperative cognitive function in elderly patients undergoing hip fracture surgery: general vs spinal anesthesia. *Injury* 2018; **49**: 2221–6
13. Evered L, Silbert B, Knopman DS, et al. Recommendations for the nomenclature of cognitive change associated with anaesthesia and surgery-2018. *Br J Anaesth* 2018; **121**: 1005–12
14. Leslie K, Myles PS, Forbes A, Chan MT. The effect of bispectral index monitoring on long-term survival in the B-aware trial. *Anesth Analg* 2010; **110**: 816–22
15. Kertai MD, Pal N, Palanca BJ, et al. Association of perioperative risk factors and cumulative duration of low bispectral index with intermediate-term mortality after cardiac surgery in the B-Unaware Trial. *Anesthesiology* 2010; **112**: 1116–27
16. Kertai MD, Palanca BJ, Pal N, et al. Bispectral index monitoring, duration of bispectral index below 45, patient risk factors, and intermediate-term mortality after noncardiac surgery in the B-Unaware Trial. *Anesthesiology* 2011; **114**: 545–56
17. Avidan MS, Jacobsohn E, Glick D, et al. Prevention of intraoperative awareness in a high-risk surgical population. *N Engl J Med* 2011; **365**: 591–600
18. Chan MT, Cheng BC, Lee TM, Gin T, Group CT. BIS-guided anesthesia decreases postoperative delirium and cognitive decline. *J Neurosurg Anesthesiol* 2013; **25**: 33–42
19. Abdelmalak BB, Bonilla A, Mascha EJ, et al. Dexamethasone, light anaesthesia, and tight glucose control (DeLiT) randomized controlled trial. *Br J Anaesth* 2013; **111**: 209–21
20. Radtke FM, Franck M, Lendner J, Kruger S, Wernecke KD, Spies CD. Monitoring depth of anaesthesia in a randomized trial decreases the rate of postoperative delirium but not postoperative cognitive dysfunction. *Br J Anaesth* 2013; **110**: i98–105
21. Wildes TS, Mickle AM, Ben Abdallah A, et al. Effect of electroencephalography-guided anesthetic administration on postoperative delirium among older adults undergoing major surgery: the ENGAGES randomized clinical trial. *JAMA* 2019; **321**: 473–83
22. Wallace B, Dahabreh I, Trikalinos T, Lau J, Trow P, Schmid C. Closing the gap between methodologists and end-users: R as a computational back-end. *J Stat Softw* 2012; **49**: 1–15
23. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol* 2014; **14**: 25
24. Serghiou S, Goodman SN. Random-effects meta-analysis: summarizing evidence with caveats. *JAMA* 2018. <https://doi.org/10.1001/jama.2018.19684>. Advance Access published on December 19, 2018
25. Nuzzo R. Scientific method: statistical errors. *Nature* 2014; **506**: 150–2
26. Short TG, Leslie K, Chan MT, Campbell D, Frampton C, Myles P. Rationale and design of the balanced anesthesia study: a prospective randomized clinical trial of two levels of anesthetic depth on patient outcome after major surgery. *Anesth Analg* 2015; **121**: 357–65
27. Monk TG, Saini V, Weldon BC, Sigl JC. Anesthetic management and one-year mortality after noncardiac surgery. *Anesth Analg* 2005; **100**: 4–10
28. Ioannidis JPA. The proposal to lower P value thresholds to .005. *JAMA* 2018; **319**: 1429–30

Handling editor: H. C. Hemmings Jr