Question 3

1. Yes. Low risk of bias.

The randomisation sequence was generated by a computer with randomly varying block sizes. This helps to ensure equal numbers between the intervention and control groups. Participants are also stratified according to trial site and delirium motor subtype. Stratified randomisation helps to ensure an even distribution of patients from different sites and delirium motor subtypes in the intervention and control group. These are appropriate methods to produce a truly random sequence. Therefore, the risk of bias is low.

1. Yes. Low risk of bias.

Randomisation is performed at a central location, minimising the probability of interference with the allocation sequence by people from other sites. The use of randomly varying block sizes further ensures allocation concealment. For example, even if a clinician discovers that one of the ampules contains haloperidol or placebo, they would still be unable to guess the block size and deduce the next allocation.

Study group allocation is communicated through identical ampules with identical labelling. The study also specifies that the solutions are colourless and indistinguishable from each other. Therefore, no one involved in the study, including clinicians, can interpret the group allocation through the solutions.

Although no third-party system is mentioned for group allocation, performing group allocation at a central location and using indistinguishable ampules are reasonable approaches for allocation concealment. Sufficient efforts have been made, so the risk of bias is low and selection bias is unlikely.

1. Yes. Low risk of bias.

The authors stated in the ‘results’ section that the baseline characteristics were reasonably balanced between the groups. However, further analysis can be done using ‘Table 1’, which shows the distribution of baseline characteristics between the two study groups. It is important to make our own judgement based on the table than relying on the author’s statement. The table appears to cover all the important confounders, indicating that the groups were similar in terms of potential confounders at the start of the trial. All differences between the baseline characteristics were 5% or less.

The largest percentage difference is in the type of admission, with more surgical admissions in the intervention group. This may suggest that more patients in this group were in a more serious condition, potentially giving the intervention group an 'advantage.' However, the results in Figure 2 show that the intervention did not work and suggested that the placebo was better despite this potential advantage. Therefore, this baseline difference does not affect our decision about whether or not to use this treatment.

Overall, the study groups are similar in terms of baseline characteristics, indicating effective randomisation and a low risk of bias.

1. Unclear. Low risk of bias.

The study specifies it is a blinded trial in the 'methods' section, but verifying this claim independently is essential. Five groups need to be blinded: patients, healthcare providers, data collectors, outcome adjudicators, and data analysts. The 'randomization and blinding' section states that clinicians, patients, investigators, outcome assessors, statisticians, and members of the data and safety monitoring committee were unaware of the group assignments. Further analysis is provided below:

1. Patients

Randomisation was done centrally and group allocation was not disclosed to patients. Treatments were colourless and contained in identical ampules with identical labelling, making it difficult for patients to guess their group allocation. The use of varying block sizes further conceals this information. Thus, it is safe to conclude that patients were blinded, and the risk of bias is low.

1. Health care providers

Clinicians, as healthcare providers, are responsible for all other interventions and additional doses that are provided at their discretion. If clinicians were not blinded, they might treat groups differently, thereby introducing bias to the study. The study lacks details on how clinician blinding was ensured. Therefore, the risk of bias is unclear.

1. Data collectors

The 'outcome measures' section mentions that outcomes were obtained from medical records and hospital registries by trial investigators. However, there is no information on how these investigators were blinded. Despite this, most outcomes were objective and measured using predefined, widely used clinical scales. Thus, the risk of data collector interfering with the data collected is low, and consequently, the risk of bias is low.

1. Outcome adjudicators

Outcome adjudicators are responsible for assessing serious adverse reactions. If not blinded, their decisions could be biased by knowledge of group assignments. However, the study does not provide details on blinding procedures for outcome adjudicators. Therefore, the risk of bias is unclear.

1. Data analysts

The 'statistical analysis' section states that the first and tenth authors conducted the analysis and were unaware of group assignments. However, further validation is needed to confirm this. Having multiple analysts, with intervention statuses labelled differently, is a method to check consistency between the analyses. If any data analyst is unblinded, significant difference in the analyses would be detected. Moreover, the study stated that the statistical analysis plan was published before the last patient's enrolment, preventing the data analysts to make any adjustments to the plan. Reasonable efforts were made to ensure blinding, so the risk of bias is low.

Overall, the risk of bias is unclear for healthcare providers and outcome adjudicators. However, sufficient efforts were made to ensure blinding, even though not all processes were explicitly stated. Therefore, it is reasonable to consider the overall risk of bias as low.

1. Yes. Low risk of bias.

All missing data are shown in ‘Figure 1’. Missing data before randomisation do not result in selection bias, making them irrelevant. The reasons for missing outcome data are consistent across both study groups, primarily due to the absence of treatment and lack of consent.

According to the figure, 510 patients were assigned to the treatment group and 490 to the placebo group. In the treatment group, 9 patients did not receive haloperidol and were excluded from the study. Of the remaining patients, 10 did not provide consent or did not have consent provided and were excluded from the primary outcome analysis. Thus, the missing primary outcome data was 19/510 (3.7%) and the missing secondary outcome data was 9/510 (1.8%).

Similarly, in the placebo group, 4 patients did not receive the placebo and were excluded. Among the remaining patients, 14 did not provide consent or did not have consent provided and were excluded from the primary outcome analysis. The missing primary outcome data for this group was 18/490 (3.7%) and the missing secondary outcome data was 4/490 (0.8%).

The missing primary outcome data is identical in both groups, while the treatment group has slightly more missing secondary outcome data. However, given the study's size, the amount of missing data is minimal and relatively equal between the two groups. After excluding these missing outcome data, the remaining groups would have still been fairly equal. Although the study did not specify the baseline characteristics of patients who remained in the study, the small number of missing data suggests that any differences in baseline characteristics would be minor. Therefore, the risk of bias is low.

1. Yes. High risk of bias.

In the 'Statistical Analysis' section, the study states that primary and secondary outcomes were analysed using the intention-to-treat (ITT) population. The ITT population includes all patients who were randomised, received their assigned treatment, and provided consent for their data to be used. ITT is a highly regarded analysis method, but this study did not perform a true ITT analysis, which would include all patients randomised, regardless of protocol adherence. Instead, a modified ITT analysis was performed, as mentioned in the appendix, where participants who did not adhere to the intervention or did not provide consent were excluded, resulting in 4.5% missing data. While modified ITT is appropriate, it is less rigorous than true ITT.

The study also used per-protocol analysis, which includes patients in the ITT population but excludes those with major protocol violations. This method provides a more controlled view of treatment efficacy by excluding non-compliant patients but may introduce bias, as compliant patients may differ systematically from non-compliant ones. According to the appendix, the percentage of patients excluded due to major protocol violations is relatively large (21.8% for the intervention group and 20.4% for the placebo group). Although these percentages are fairly equal, the baseline characteristics of the remaining patients may differ significantly, potentially introducing bias.

Overall, the analysis methods used are appropriate, but the per-protocol analysis may introduce bias. This should be considered when making clinical decisions.

1. Yes, I would. The overall bias of this study is low. The study is well designed with appropriate randomisation, allocation concealment, blinding, handling of missing data and analysis methods. The funding organisation had no involvement in the study, eliminating the risk of industry bias. To ensure there is no publication bias, further research for systematic reviews of similar studies could be conducted and a funnel plot could be used to test for publication bias. I would also inform patients about the major protocol violations and the potential bias they might introduce. I will consider the patient's preferences before making clinical decisions.