SUPPLEMENTARY FILE

Trial Design

This was a single-center, randomized, placebo-controlled trial of 5mg immediate release Melatonin or matching Placebo for 5 nights with an additional 2 days of follow-up. The protocol of this trial is based on our pilot study. [1] The study was approved (MHHREC2019.043) by our institute Human Research Ethics Committee (HREC) and conducted in concordance with the Declaration of Helsinki. The trial is registered with the Australia New Zealand Clinical Trials Registry (ACTRN12619000034134). Melatonin and Placebo were supplied by Optima Ovest, Australia. The protocol was modified once after commencement to exclude patients in isolation due to COVID-19 diagnosis or suspicion. The trial and reporting conform to CONSORT 2010 guidelines.

Participants and Setting

The study was conducted in a 550-bed university hospital. The study population was internal medicine inpatients aged ≥65 years. This public hospital provides free healthcare to residents of north-west Melbourne, Australia, an area ethnically, culturally, and economically diverse.

Screening: The institution screens routinely for delirium at admission using the 4 'A's Test (4AT; a short delirium assessment tool designed for use without special training, which improves delirium detection in routine care) administered by trained nursing staff. [2] The 4AT tool scores range from 0 to 12, with pre-specified score of > 3 being positive. Positive 4AT cases were notified to the medical staff, and bedside nurses initiated the Delirium Observation Screening Scale (DOSS) assessment. [3]

The DOSS is a 13-item observation scale for non-intubated patients, positive DOSS is a score ≥ three. Research staff screened 4AT and DOSS data on the hospital electronic medical record and attended daily handover meetings to identify eligible patients. Potential participants were also identified by referral from treating teams following trial education. Research staff screened patients for eligibility following discussions with treating staff. Patients with delirium were considered to lack the capacity to consent, hence potential participants' surrogate decision makers were approached, given trial information and an invitation to assess further for potential participation.[2]

Eligible participants were next screened for delirium by trained, medically qualified research team members using the Confusion Assessment Method (CAM). [4] This is a validated tool with a sensitivity of 94% (95% confidence interval, CI, 91–97%), and specificity of 89% (95% CI, 85–94%). Where the diagnosis was not clear, consensus for a diagnosis was reached by referral to the medical notes, conversations with the treating health professional team and study principal investigator.

Participants could be included within 48 hours of admission if mixed or hyperactive delirium was diagnosed at admission. Participants could also be included if delirium was diagnosed during admission, within 48 hours of diagnosis. To be included, participants also needed to be 65 years or more years of age, CAM positive, and admitted under internal medicine. Patients were excluded if they had sensory impairment, dysphasia, or were unable to understand or communicate in English, had taken Melatonin or a Melatonin receptor agonist within the preceding 24 hours, were unlikely to complete the trial (defined as an expected prognosis of less than six months or planned admission to hospital of less than seven days), had a contraindication to Melatonin (severe hepatic failure, an active seizure disorder, or

concomitant cimetidine use), had an exclusively hypoactive motor state delirium (as melatonin metabolites are elevated in this condition suggesting an excess rather than deficiency [5]), an abnormal International Normalised Ratio with Warfarin therapy, or they were not able to participate in study measures for any other reason including isolation for infection control.

Randomization and Intervention

Participants were randomized to Melatonin or Placebo once nightly for five nights using a 1:1 ratio. An independent statistician generated randomized allocation codes with permuted blocks of six participants (three Placebo and three treatment) provided to pharmacy who allocated to intervention groups. To balance groups by cognitive impairment status these were in two lists for those with and without baseline cognitive impairment. The allocations were not known to research or treating staff. A separate code sheet was kept locked in the trial pharmacy to allow matching after trial completion or in the event of an adverse event.

The clinical trials pharmacy dispensed all study medication. Placebo and Melatonin were packed in identical HDPE containers. The study medication was prescribed by medically qualified research staff and given by nurses at ~ 8pm due to sedating effects. Supplementary Figure 1 illustrates the assessments and medication administration. Any remaining medication was returned to the hospital pharmacy.

Measurements

The trial Investigators and research staff were trained to assess outcomes prior to the trial opening with each of the instruments by teaching, observation, and review

with independent examination by the lead author, a geriatrician. The lead author also independently validated outcomes assessed by trial staff during the trial.

Upon receiving consent, participants baseline data including demographic and health-related characteristics such as age, sex, delirium severity with the Memorial Delirium Assessment Scale (MDAS), smoking status, alcohol consumption, psychiatric history including depressive symptoms with the Geriatric Depression Scale, 15 items (GDS15), medications and potential delirium precipitants (classified as infectious/inflammatory, pain, metabolic) were recorded. [6, 7] Electrolytes and liver function tests were recorded. Dementia status was assessed through the Informant Questionnaire on COgnitive Decline in the Elderly, short form (IQCODE), and chart review conducted for comorbidities quantified with the Charlson Comorbidity Index (CCI). [8, 9]

Activities of daily living (ADLs) at baseline were assessed by using the Katz index, a widely used 6-item self-report tool to assess functional capacity, (assessing independence in bathing, dressing, toileting, transferring, continence, and feeding). Scores range from zero (very dependent) to six (fully independent). The Katz scale has high validity even in patients with neurologic decline. [10, 11]

Delirium presence or absence was assessed at baseline and follow up visits using the CAM conducted by trained research staff. Delirium was considered to have ceased the day of the last visit at which the participant was positive on CAM, and recorded in days. The 4AT was also administered daily.

Delirium severity was measured at all timepoints by research staff using the MDAS, a validated scale, scored from 0 to 30, positively correlated with severity with good internal consistency and inter-rater reliability, that is derived from scores on the

Standardised Mini-Mental State Examination (sMMSE), and Digit Span Backwards (DSB) and Digit Span Forwards (DSF). [6, 12] MDAS has 10 subscales rated 0 (not present) to 3 (severe) for delirium features awareness, disorientation, short-term memory impairment, impaired digit span, attention, disorganized thinking, perceptual disturbance, delusions, psychomotor activity, and sleep-wake cycle disturbance. Unassessable MDAS items were derived pro rata from other items per the validation paper. [6] DOSS measurements by bedside nursing teams were recorded as another measure of severity, but did not contribute to this analysis.

Level of sedation, alertness, and agitation at baseline and during treatment was measured by the modified Richmond Agitation and Sedation Scale (mRASS). [13] The scale includes 10 points ranging from – 5 to +4. The score of 0 represents the alert and calm participant, spontaneously paying attention to the caregiver. Negative scores describe the level of sedation with – 1 representing drowsiness, characterized by not being fully alert, but does have sustained awakening as defined by eye-opening/eye contact to voice for more than 10 seconds. Levels of – 2 to – 5 describe light, moderate, and deep sedation, as well as being unrousable. Positive scores describe the level of agitation, ranging from + 1 to + 4, representing restlessness, agitation, pronounced agitation and combativeness. The motor subtype of delirium was assessed by the clinical judgement of the investigators and their consideration of mRASS and MDAS (item 9) data, due to the lack of a validated tool when planning the trial. Study participants who scored mRASS of > 2 or < -1 were deemed not assessable by sMMSE, DSB and DSF.

Cognitive performance was assessed at baseline and once daily for the duration of participation in the study with the sMMSE . [14] Number of falls and delirium management methods such as number of restraints and number and dose of rescue

medications, including benzodiazepines and antipsychotics, were recorded at follow up visits.

Following enrolment, participants were interviewed and assessed daily by trained investigators for 7 days (5 days of the study medication and 2 days after treatment) to collect primary and secondary outcome and assess safety (Supplementary Figure 1). This included assessment of falls and pressure areas. Adverse events were reported to an independent safety monitor for assessment for severity and relationship to the intervention.

Participants in the biomarker sub-study had blood drawn at entry, during the intervention, and where possible at 6-month follow-up. Bloods were sent directly to the laboratory for storage and processing.

Objectives, Outcomes and Delirium Assessment

The **primary outcome** measure was delirium severity assessed using the MDAS. This was calculated from change in MDAS scores from baseline. Delirium was defined as a positive episode detected by the CAM. The research team investigated the presence and severity of delirium for individual participants at the same time each day, where possible. We had *a priori* defined a clinically significant difference as three points improvement on the MDAS, by comparison to a placebo- controlled trial of an antipsychotic (quetiapine, a common treatment for agitation in delirium) for delirium demonstrating a change of similar extent. [15]

The **secondary outcomes** included the change in delirium severity post-treatment as indicated by MDAS scores over days 6-7 (non-treatment period) and delirium duration (total number of days with positive CAM). Other secondary outcomes during

treatment were: level of arousal as measured by daily observations with the mRASS (Proportion normal vs. abnormal), reduction in number and dose of rescue medications (specifically benzodiazepines and antipsychotics) and sleep quality as assessed in the MDAS (Item 10). Six-month outcomes included new dementia diagnosis, cognition (sMMSE) and cost-effectiveness. Neurofilament light levels were assessed in the cohort who consented to additional blood draws.

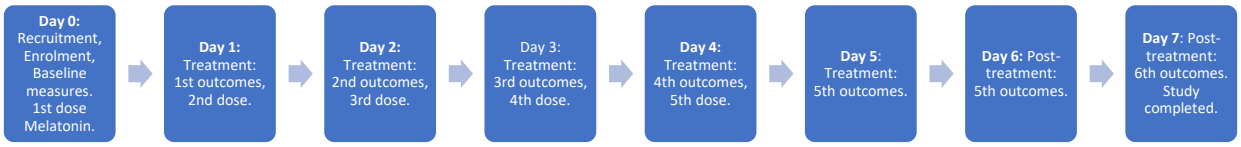
Power Calculation

The pilot study showed recruitment to be feasible, and the intervention could be administered reliably. The power calculation and 120 participants. [16] (allowing for drop-out of 30% as experienced in the pilot) would have 90% power to demonstrate a statistically significant outcome at the alpha = 0.05 level (two sided calculation). [16]

Statistical Analysis

Data was collected daily for 7 days. The primary and secondary outcome measures were compared using a Student's t-test for normally distributed data or Wilcoxon Rank-Sum test for non-normally distributed continuous variables. For discrete outcomes, results were tested using a Chi-square test. Where expected cell counts were less than 7 Fisher's exact test (two-sided) substituted for Chi squared test, with the addition method used where cell counts were zero. Missing outcome data were imputed by regression imputation where at least three data points were available. Secondary outcomes were considered exploratory and hypothesis generating, and not corrected for multiple comparisons.

Supplementary Figure 1: Schedule of procedures: Assessments and medication administration



Protocol violation:

There was a protocol violation due to an error in the dispensing instructions through the electronic medical record. Three participants (#52- #54) were ordered eight capsules per night. No adverse effect resulted and the participants and their next of kin were provided open disclosure. One participant's next-of-kin requested to be unblinded; the participant was receiving Placebo. The other two participants had completed participation and were discharged when the error was identified; on review (after trial enrolment was complete) they were both receiving Melatonin. Recruitment was stopped on identification of the error until the cause could be identified and rectified and restarted with the permission of the institution's HREC. Results were analyzed with and without these participant's results without substantial difference; the results presented include these participants.

Supplementary Table 1. Schedule of Procedures.

Procedure	Baseline (Day 0)	Treatment (Days 1-5)	Post treatment (Days 6-7)	Six month Follow-up
Consent	X	X (If capacity)	X (If capacity)	X (If capacity)
Demographics: Age, Sex, living circumstances	X			X
Medical history, medications and examination	X			X
Electrolytes and Liver function	X			
Delirium at recruitment (D#)	X			
Standardized Mini Mental State Exam	X	X	X	X
Richmond Agitation and Sedation Scale	X	X	X	
Digit Span Forwards and Digit Span Backwards	X	X	X	
Confusion Assessment Method	X	X	X	
4 A's Test	X	X	X	
Memorial Delirium Assessment Scale	X	X	X	
Motor subtype	X	X	X	
Delirium Observation and Screening Scale	X	X	X	
Informant Questionnaire on Cognitive Decline in the Elderly	X*			X
Charlson Comorbidity Index	X*			
Katz Index of ADLs	X			
Presence of restraints		X	X	
Number of falls		X	X	
Number of pressure areas	X**	X	X	
Rescue medication		X	X	
Adverse events		X	X	
Review International Normalized Ratio	X***	X***		

Clinician Diagnosis of Dementia	X

*the IQCODE and CCI was measured at entry if possible but could be measured at any point during the trial.

**Initial number and location of pressure areas was recorded at entry in order to derive the number of new areas at each subsequent visit.

***If participant was on warfarin, the INR was performed every 1–2 days as per standard practice of the treating team for acutely unwell inpatients on warfarin.

Supplementary Figure 2: CONSORT Checklist

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	3-4
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	5
	2b	Specific objectives or hypotheses	5-6
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	7
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	7
Participants	4a	Eligibility criteria for participants	7-9
	4b	Settings and locations where the data were collected	7-9
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	9
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	10-12
	6b	Any changes to trial outcomes after the trial commenced, with reasons	7
Sample size	7a	How sample size was determined	13
	7b	When applicable, explanation of any interim analyses and stopping guidelines	14
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	9
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	9
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	9
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	9
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	9

CONSORT 2010 checklist *Page 1*

[1] Clayton-Chubb DI, Lange PW. Moderate dose melatonin for the abatement and treatment of delirium in elderly general medical inpatients: study protocol of a placebo controlled, randomised, double blind trial. *BMC Geriatr*. 2016;**16**: 54.

[2] Bellelli G, Morandi A, Davis DH*, et al.* Validation of the 4AT, a new instrument for rapid delirium screening: a study in 234 hospitalised older people. *Age Ageing*. 2014;**43**: 496-502.

[3] Schuurmans MJ, Shortridge-Baggett LM, Duursma SA. The Delirium Observation Screening Scale: a screening instrument for delirium. *Res Theory Nurs Pract*. 2003;**17**: 31-50.

[4] Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med*. 1990;**113**: 941-948.
[5] Balan S, Leibovitz A, Zila SO*, et al.* The relation between the clinical subtypes of delirium and the urinary level of 6-SMT. *J Neuropsychiatry Clin Neurosci*. 2003;**15**: 363-366.
[6] Breitbart W, Rosenfeld B, Roth A, Smith MJ, Cohen K, Passik S. The Memorial Delirium Assessment Scale. *J Pain Symptom Manage*. 1997;**13**: 128-137.
[7] Lesher EL, Berryhill JS. Validation of the Geriatric Depression Scale--Short Form among inpatients. *J Clin Psychol*. 1994;**50**: 256-260.
[8] Jorm AF. A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): development and cross-validation. *Psychol Med*. 1994;**24**: 145-153.
[9] Maes M, Vandoolaeghe E, Degroote J, Altamura C, Roels C, Hermans P. Linear CT-scan measurements in alcohol-dependent patients with and without delirium tremens. *Alcohol*. 2000;**20**: 117-123.
[10] Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of Illness in the Aged. The Index of Adl: A Standardized Measure of Biological and Psychosocial Function. *JAMA*. 1963;**185**: 914-919.
[11] Ferretti-Rebustini RE, Balbinotti MA, Jacob-Filho W*, et al.* Validity of the Katz Index to assess activities of daily living by informants in neuropathological studies. *Rev Esc Enferm USP*. 2015;**49**: 946-952.
[12] Molloy DW, Alemayehu E, Roberts R. Reliability of a Standardized Mini-Mental State Examination compared with the traditional Mini-Mental State Examination. *Am J Psychiatry*. 1991;**148**: 102-105.
[13] Sessler CN, Gosnell MS, Grap MJ*, et al.* The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. *Am J Respir Crit Care Med*. 2002;**166**: 1338-1344.
[14] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;**12**: 189-198.
[15] Tahir TA, Eeles E, Karapareddy V*, et al.* A randomized controlled trial of quetiapine versus placebo in the treatment of delirium. *J Psychosom Res*. 2010;**69**: 485-490.
[16] Lange PW, Clayton-Chubb DI, Watson R, Maier AB. Results from a double blinded, randomized, placebo-controlled, feasibility trial of melatonin for the treatment of delirium in older medical inpatients. *Intern Med J*. 2020.