

Non-RCT Literature Quality Assessment

Based on Newcastle-Ottawa Scale (NOS)

15 Studies Assessment Report

Table 1: Complete NOS Assessment Results for 15 Studies

No.	Study	Study design	Selection	Comparability	Outcome	Summary	Quality grade
1	Beitinger 2012	Group	★★★	★★	★★★	8/9	High
2	Vgontzas 2000	Group	★★★★	★★	★★	8/9	High
3	Abelleira 2024	Group	★★★	★★	★★★	8/9	High
4	Terzi 2016	Group	★★★	★★	★★	7/9	Medium
5	Sforza 2015	Group	★★★★	★★	★★★	9/9	High
6	Birmingham 2025	Group	★★★	★★	★★★	8/9	High
7	Martin 2017	Group	★★★	★★	★★	7/9	Medium
8	Tseng 2017	Group	★★★	★★	★★★	8/9	High
9	Sforza 2011	Group	★★★★	★★	★★★	9/9	High
10	Buxton 2012	Crossover	★★★★	★★	★★★	9/9	High
11	Leproult 2015	Crossover	★★★	★★	★★★	8/9	High
12	Reynolds 2012	Crossover	★★★	★★	★★	7/9	Medium
13	Schüssler 2006	Crossover	★★★	★	★★	6/9	Medium
14	Cintra 2011	Crossover	★★★	★★	★★★	8/9	High
15	van Holst 2016	Crossover	★★★	★★	★★	7/9	Medium

Table 2: Selection Domain (Maximum 4 stars)

Study	Representativeness	Non-exposed Group Selection	Exposure Ascertainment	Outcome Not Present	Score
Beitinger 2012	★	★	★	-	3/4
Vgontzas 2000	★	★	★	★	4/4
Abelleira 2024	★	★	★	-	3/4
Terzi 2016	★	★	★	-	3/4
Sforza 2015	★	★	★	★	4/4
Birmingham 2025	★	★	★	-	3/4
Martin 2017	★	★	★	-	3/4
Tseng 2017	★	★	★	-	3/4
Sforza 2011	★	★	★	★	4/4
Buxton 2012	★	★	★	★	4/4
Leproult 2015	★	★	★	-	3/4
Reynolds 2012	★	★	★	-	3/4
Schüssler 2006	★	★	★	-	3/4
Cintra 2011	★	★	★	-	3/4
van Holst 2016	★	★	★	-	3/4

Table 3: Comparability Domain (Maximum 2 stars)

Study	Main Confounder Control	Other Confounder Control	Score
Beitinger 2012	★ (BMI matched)	★ (Age, sex)	2/2
Vgontzas 2000	★ (BMI matched)	★ (Age, sex)	2/2
Abelleira 2024	★ (BMI, age, sex)	★ (Comorbidities)	2/2
Terzi 2016	★ (Age, BMI)	★ (Smoking)	2/2
Sforza 2015	★ (Age matched)	★ (Sex, BMI)	2/2
Birmingham 2025	★ (Age, BMI)	★ (Multiple covariates)	2/2
Martin 2017	★ (Age, BMI, sex)	★ (Activity level)	2/2
Tseng 2017	★ (Age, sex)	★ (1:2 matching)	2/2
Sforza 2011	★ (Age matched)	★ (Sex, community-based)	2/2
Buxton 2012	★ (Within-subject)	★ (Controlled conditions)	2/2
Leproult 2015	★ (Within-subject)	★ (Sleep duration)	2/2
Reynolds 2012	★ (Within-subject)	★ (Diet controlled)	2/2
Schüssler 2006	★ (Within-subject)	- (Small n)	1/2
Cintra 2011	★ (Age, sex matched)	★ (BMI, CPAP trial)	2/2
van Holst 2016	★ (Age, BMI matched)	★ (Disease controls)	2/2

Table 4: Outcome Domain (Maximum 3 stars)

Study	Assessment Method	Follow-up Duration	Adequacy of Follow-up	Score
Beitinger 2012	★ (Validated)	★ (Follow-up)	★ (>80%)	3/3
Vgontzas 2000	★ (Lab measures)	- (Cross-sectional)	★ (Complete)	2/3
Abelleira 2024	★ (Validated)	★ (Follow-up)	★ (>90%)	3/3
Terzi 2016	★ (Validated)	- (Cross-sectional)	★ (Complete)	2/3
Sforza 2015	★ (Validated)	★ (Follow-up)	★ (>95%)	3/3
Bermingham 2025	★ (Validated)	★ (Follow-up)	★ (>85%)	3/3
Martin 2017	★ (Validated)	- (Cross-sectional)	★ (Complete)	2/3
Tseng 2017	★ (Validated)	★ (Follow-up)	★ (>90%)	3/3
Sforza 2011	★ (Validated)	★ (Follow-up)	★ (>95%)	3/3
Buxton 2012	★ (Lab measures)	★ (Controlled)	★ (Complete)	3/3
Leproult 2015	★ (Lab measures)	★ (Controlled)	★ (Complete)	3/3
Reynolds 2012	★ (Lab measures)	★ (Controlled)	★ (Complete)	2/3
Schüssler 2006	★ (Validated)	- (Short follow-up)	★ (Complete)	2/3
Cintra 2011	★ (Lab measures)	★ (Controlled)	★ (Complete)	3/3
van Holst 2016	★ (Validated)	- (Cross-sectional)	★ (Complete)	2/3

Table 5: Summary Statistics

Category	Value
Total Studies	15
Group Design	9 (60%)
Crossover Design	6 (40%)
High Quality (7-9 stars)	12 (80%)
Medium Quality (4-6 stars)	3 (20%)
Low Quality (0-3 stars)	0 (0%)
Mean NOS Score	7.7/9
Studies with Full Selection Score (4/4)	4 (26.7%)
Studies with Full Comparability Score (2/2)	14 (93.3%)
Studies with Full Outcome Score (3/3)	9 (60%)

Table 6: NOS Quality Assessment Criteria

Domain	Criterion	Rating	Maximum Score
Selection	Representativeness of exposed cohort	Truly representative (★) vs. Somewhat repre	Max 1 star
Selection	Selection of non-exposed cohort	Drawn from same community (★) vs. Differ	Max 1 star
Selection	Ascertainment of exposure	Secure record (★) vs. Structured interview (Max 1 star
Selection	Outcome not present at start	Yes (★) vs. No (-)	Max 1 star
Comparability	Control for main factor	Yes (★) vs. No (-)	Max 1 star
Comparability	Control for additional factors	Yes (★) vs. No (-)	Max 1 star
Outcome	Assessment of outcome	Independent blind assessment (★) vs. Reco	Max 1 star
Outcome	Follow-up duration	Long enough for outcomes (★) vs. Not long	Max 1 star
Outcome	Adequacy of follow-up	>80% follow-up (★) vs. <80% or no descrip	Max 1 star