NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE (Adapted cohort study scale)

Adapted from Wells G, Shea B, O'Connell D, et al. (2021) The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available at https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Selection

1) Representativeness of the paediatric gender dysphoric group

Score 1 if:
a) truly representative of the average child or adolescent with gender dysphoria / incongruence, e.g., nationally representative community sample, population-based medical database, national gender service, multiple gender services covering different localities

Score 0.5 if:
b) somewhat representative of the average child or adolescent with gender dysphoria / incongruence, e.g., single gender service (in a country where there are several), locality representative community sample

Score 0 if:
c) a selected group of users e.g., convenience sample, self-selection, sub-sample due to data availability
or
d) no description of the derivation of the cohort

Q1. TIP: If paper does not include information on sample size / representativeness / response rate in relation to eligible clinic population, score 0 as we must assume it is a selected group of users.

Q1. NOTE: When scoring take into consideration that puberty blockers are not available to pre-pubertal children and cross-sex hormones not until mid-adolescence, therefore an adolescent sample would be representative of those seeking medical treatment for gender dysphoria.

2) Selection of the non-exposed group

Score 1 if:
a) drawn from the same community as the exposed group

Score 0.5 if:
b) drawn from a different source that is comparable e.g., population norms for all adolescents or matched cisgender controls where outcome is expected to be the same for both groups such as BMI or adult height

Score 0 if:
c) drawn from a different source that is not comparable e.g., cisgender population for studies measuring psychosocial outcomes
or
d) no description of the derivation of the non-exposed cohort

Score N/A if:
e) single group study (e.g., pre-post treatment design)

Q2. TIP: Some studies include two plus comparator groups – score 1 if all appropriate, score 0.5 if mix of comparable and non-comparable, and score 0 if none are comparable.

3) Ascertainment of exposure (medical treatment for gender dysphoria)
Score 1 if:
a) secure record, e.g., clinic or medical records
or
b) structured interview

Score 0 if:
c) written self-report
or
d) no description

Q3. TIP: Use of medical records can be inferred from methods reporting overall, e.g., if sample eligibility was based on medical records and this includes treatment, or if detailed treatment information provided from medical records.

4) Demonstration that outcome of interest was not present at start of study
Score 1 if:
a) yes

Score 0 if:
b) no

Score N/A if:
c) study is measuring an outcome such as quality of life or severity of anxiety for which there would be an intrinsic value for each participant at any time during the study

Comparability

5) Comparability of cohorts based on the design (e.g., matched controls, inclusion criteria) or analysis (e.g., propensity score matching, regression analysis)

PART A
Score 1 if:
a) study controls for [age OR puberty stage] AND [natal sex OR gender]

Score 0.5 if:
b) one (but not both) of the above confounders is controlled for
or
c) other important sociodemographic confounders are controlled for, e.g., family support for studies measuring mental health / psychosocial outcomes

Score 0 if:
d) study does not control for important sociodemographic confounders

PART B
Score 1 if:
a) study controls for co-interventions expected to affect the outcome(s), e.g., psychosocial support, psychiatric medication, use of other medication likely to alter outcome (e.g., contraceptive pill),

other aid or intervention designed to address gender dysphoria or to modify body (e.g., social transition, binders, voice therapy)
or
b) there are no co-interventions that are expected to affect the outcome(s)

Score 0.5 if:
c) at least one (but not all) of the important co-intervention confounders are controlled for

Score 0 if:
d) study does not control for important co-intervention confounders

Q5. TIP (from manual): Statements of no differences between groups or that differences were not statistically significant are not sufficient for establishing comparability.
This question needs to be answered for single group studies (i.e., how do these studies control for potential demographic and treatment confounders that occur between baseline and follow-up).

Outcome

6) Assessment of outcome
Score 1 if:
a) validated scale or standardised assessment tool/method
or
b) record linkage, e.g., medical/clinic/administrative records

Score 0.5 if:
c) combination of validated / standardised and non-validated / unstandardised assessment methods

Score 0 if:
d) self-report
or
e) no description

7) Was follow-up long enough for outcomes to occur?
Score 1 if:
a) follow-up is sufficient for all reported outcomes

Guidance on follow-up:
For puberty suppressants follow-up should be at least 3 months to assess desired / expected effect, gender dysphoria / incongruence or psychosocial outcomes.

For cross-sex hormones follow-up should be at least 6 months to assess desired / expected effect, gender dysphoria / incongruence, or psychosocial outcomes.

For both treatments, follow-up should be at least 3 months to assess safety, side-effects or cardiometabolic risk; and at least 12 months for cognitive development, bone health or fertility.

Score 0.5 if:
b) follow-up is sufficient for some outcomes but not others, e.g., studies that examine multiple outcomes requiring different follow-up
or

c) follow-up is sufficient for some participants but not others, e.g., where follow-up or treatment duration varies between participants

Score 0 if:
d) follow-up is not sufficient
or
e) no clear description, e.g., follow-up duration unclear

Score N/A if:
f) cross-sectional design (i.e., no follow-up)

8) Adequacy of follow up of cohorts
Score 1 if:
a) complete follow up or all subjects accounted for in analysis of outcomes
or
b) subjects lost to follow up or outcome analyses unlikely to introduce bias - small number lost (≤10%) or description provided of those lost justifies that there is no potential bias due to loss to follow-up

Score 0.5 if:
c) there are multiple sufficient follow-up timepoints (based on Q7) and follow-up is adequate for some but not others, e.g., all retained at 12 months but considerable attrition at 24 months

Score 0 if:
c) follow up / analysis rate < 90% and no description of those lost
or
d) no statement

TOTAL SCORING
Cohort studies – total score = 8
Pre-post single group studies – total score = 7 (Q2 not relevant)
Cross-sectional studies with comparators – total score = 7 (Q7 not relevant)

Any study for which Q4 is relevant, total score would be as above plus 1

Low: ≤ 50%
Moderate: >50 to 75%
High – >75%