Supplementary Table 4 - Critical appraisal scores for included studies

Study ID	Study design	NOS total score	NOS Q1 representative of population	NOS Q1 score	NOS Q8 adequacy of follow-up	NOS Q8 score	NOS Q3 Ascertainment of treatment exposure	NOS Q3 score	NOS Q2 Selection of non-exposed group	NOS Q2 score	NOS Q5 comparability of cohorts	NOS Q5 part 1	NOS Q5 part 2	NOS Q7 follow-up duration	NOS Q7 score	NOS Q6 assessment of outcome	NOS Q6 score	NOS Q4 Demonstration outcome of interest not present at start	NOS Q4 score
Cohort																			
Achille 2020	Prospective cohort study	4	States that vast majority of eligible population entered the study, but no values or indication of what this means given. Recruited from single clinic.	0	116 participants entered study - 50 who completed questionnaires reported on. No information on those lost to follow-up.	0	Information in paper provides confidence that clinic data used.	1	Participants who had received either nothing at all or only puberty blockers.	1	Controlled for outcome at baseline, psychiatric medication and psychotherapy. Separate analysis by gender. Did not control for age. Did not control for prior receipt of pubertal suppression.	0.5	0.5	Follow-up not linked to treatment duration. Exposure to treatment at any time used as variable - not duration	0	Validated scales designed for self-completion.	1	N/A	
Becker-Hebly 2020	Retrospective cohort study	3	204 eligible and invited to take part. Large proportion not included. Single clinic.	0	Response rate for follow-up 37% (n=75).	0	Self-reported treatment path then controlled via clinicians' reports.	1	Those not treated with hormones from same clinic sample.	1	Did not control for gender / sex, other treatments or outcome at baseline. Used age-adjusted population norms to compare outcomes. Did not control for distribution in intervention group.	0	0	Treatment started after baseline but duration and start of treatment not reported or included in analysis	0	Validated scales - combination of self-report and clinician-report.	1	N/A	
Beking 2020	Prospective cohort study	5	Single clinic population but recruitment / response not described so no way of knowing what proportion of clinic population were included.	0	All treated participated at follow-up (4 of 41 controls dropped out before follow-up).	1	Clinic data on treatment reported.	1	Male and female controls presumed to not experience gender incongruence.	0.5	Controlled for testosterone levels and handedness (identified as confounders). Controlled for session effect in analysis (baseline). Matched for age and birth-registered sex.	1	0	Mean duration treatment / follow-up 9.8 months (sufficient for some)	0.5	Independent assessment using validated method.	1	N/A	
Burke 2016	Prospective cohort study	5	Single clinic population and recruitment / response not described. No way of knowing what proportion of clinic population were included.	0	All 21 girls with gender dysphoria followed-up (5 of 41 controls dropped out between sessions).	1	Clinic data on treatment reported.	1	Male and female controls presumed to not experience gender incongruence.	0.5	Controlled for testosterone levels and IQ (identified as confounder). Controlled for session effect in analysis (baseline). Matched for age and birth-registered sex.	1	0	Mean duration treatment / follow-up 10 months (sufficient for some)	0.5	Independent assessment using validated method.	1	N/A	
Cantu 2020	Retrospective cohort study	4	Single clinic population - likely to be a selected sample because participation depends on data availability. No information provided on numbers.	0	Participation depended on outcome data being available for analysis so all followed up. Two did not provide GAD-7 data.	1	Clinic data used to categorise treatment groups.	1	Drawn from same source as exposed participants.	1	Nothing reported about potential confounders being controlled for in the analysis.	0	0	Follow-up not linked to treatment - exposure to treatment occurred between baseline and follow-up	0	Validated measures for depression and anxiety used. Suicidality outcome not included in analysis on treatment differences.	1	N/A	
de Nie 2022	Retrospective cohort study	5	Single clinic population - only those who underwent bilateral orchiectomy combined with vaginoplasty were eligible - random sample taken.	0	Of the 263 sampled, 49 were then excluded due to no tissue being stored or no testicular parenchyma on slides. Not likely to influence results.	1	Authors state data was collected from medical records.	1	Adults presenting to the same clinic who underwent bilateral orchiectomy combined with vaginoplasty.	0.5	Study participants are birth-registered males only, groups are split by puberty stage (Tanner stage 2-3; Tanner stage 4-5; Adult). Study does not control for use of medications such as contraceptive pill.	1	0	Information on follow-up not explicit - indications follow-up sufficient for some participants (12 months plus)	0.5	Detailed information on sample testing given.	1	N/A	
Grimstad 2021b	Retrospective cohort study	3.5	Single-clinic population, very small number excluded due to incomplete height data (6/195 eligible patients).	0.5	Only those with complete height data were included, those had not reached adult height by the end of the study were excluded.	0.5	Data obtained from medical records.	1	Drawn from same source as exposed participants.	1	Only participants who were assigned female sex at birth were included. Unadjusted analyses were carried out.	0.5	0	No information provided on time between start of treatment and measurement of final height	0	Height measured in triplicate at clinic visits. Participants defined as growing if demonstrated growth velocity of 0.5cm per year. Parental height used - unclear how outcome data were calculated.	0	N/A	
Jensen 2019	Retrospective cohort study	6	Single-clinic study, one participant excluded due to non-binary gender identification and two participants excluded from analysis.	0.5	No loss to follow-up.	1	Data extracted from medical records.	1	Drawn from same source as exposed participants.	1	Separate analyses were carried out for sex. No other covariates were adjusted for.	0.5	0	Median duration and range of duration of hormone treatment and follow-up indicates sufficiency	1	Extracted from medical records.	1	N/A	
Lopez de Lara 2020	Prospective cohort study	4.5	No information given on the number eligible or the response rate.	0	Table 3 indicates all participants were successfully followed up.	1	Treatment data from clinic records used.	1	Controls presumed to not experience gender incongruence matched on age, ethnicity and socio-economic status. Outcomes are psychosocial.	0	The non-exposed group was matched on age, however men and women were analysed together. No other covariates were controlled for.	0.5	0	Follow up was carried a year after initiation of treatment	1	Validates scales were used.	1	N/A	
Tordoff 2022	Prospective cohort study	3.5	Single-clinic study. 30% of eligible patients did not take part.	0	Follow-up rates less than 90% at each follow-up timepoint.	0	Data on hormone use was collected via self-report.	0	Drawn from same source as exposed population.	1	Gender but not sex was controlled for as confounder. Ethnicity was also controlled for. The analysis controlled for receipt of mental health therapy.	0.5	0.5	3, 6, 9, 12 month - follow-up not linked to treatment initiation but some participants with sufficient follow-up	0.5	Collected via validated scales.	1	N/A	
Valentine 2021	Retrospective cohort study	3.5	Single-clinic study. Nearly 20% excluded due to only being seen twice.	0	Only patients who had been seen twice were included, although only 28 had lipid panels and 18 pre- and post-testosterone.	0	Information obtained from primary care database.	1	Females presumed to not experience gender incongruence drawn from a primary care database.	0.5	Only birth-registered females were included. Unadjusted analyses were used.	0.5	0	Average total follow-up 10.8 months (range 2.6 - 25.7), therefore sufficient for some but not all.	0.5	Data collected at clinical visits.	1	N/A	
van de Grift 2020	Retrospective cohort study	5	Single-clinic study. Patients lost to follow-up (n=68) were excluded.	0	Only participants with complete follow-up data were included.	1	Registry and patient record data collection used.	1	Drawn from same source as exposed group.	1	Separate analyses were carried out for males and females.	0.5	0	Initiation of different therapies in treatment protocol indicate follow-up sufficient for most but not all	0.5	Data collected as part of routine clinical practice.	1	N/A	
Pre-post																			
Allen 2019	Retrospective pre-post single group study	4	Required pre- and post-treatment data to be available. States that around half of eligible youth did not have pre-test data so were not included. Single clinic.	0	Only those with data available at both time-points were included.	1	Clinic data on treatment reported.	1	N/A		Controlled for treatment duration and outcome at baseline (ANCOVA). Separate analysis by birth-registered sex.	0.5	0	At least 3 months and treatment duration mean 349 days. Potentially not all participants sufficient.	0.5	Questionnaires administered by mental health professional as part of clinical care assessment. Validated scales used.	1	N/A	
Chiniara 2018	Retrospective pre-post single group study	3.5	Single-clinic study, 15/218 excluded due to missing data.	0.5	Low follow-up rates reported.	0	Retrospective review of medical records.	1	N/A		Separate analyses were conducted by sex.	0.5	0	Repeat hormonal levels measured after 4.7 (SD 3.7) months of hormone therapy - sufficient for some not all	0.5	Validated scales and clinical record data used.	1	N/A	
de Vries 2014	Prospective pre-post single group study	3.5	National clinic. 111 prescribed GnRHa. 70 participants approached one-year post-surgery, 55 took part. Large proportion of eligible population now missing.	0	Not all 70 provided data (some questionnaires added partway through). Response rate: CBCL, YSR: 54; BDI, TPI, STAI, CGAS, and UGS: 41; BIS: 57.	0	Information presented on start of treatment - medical records data.	1	N/A		Separate analyses were conducted by sex, and age was adjusted for.	1	0	Final follow-up takes place one year after surgery	1	All validated scales except 'self-constructed' objective measure of wellbeing.	0.5	N/A	
Delemarre-van de Waal 2006	Prospective pre-post single group study	1.5	Single clinic, inadequate information on response rates given.	0	Inadequate information on follow-up given.	0	Follow-up protocol integrated into clinical practice.	1	N/A		No adjustment made for age, sex, co-interventions or socio-demographic confounders.	0	0	Insufficient information provided on duration of follow-up	0	Clinical measurements presented, but no information given on how this information was obtained.	0.5	N/A	
Grimstad 2021a	Retrospective pre-post single group study	4.5	Single-site clinic. Inclusion based on medical charts explicitly documenting occurrence or absence of breakthrough bleeding, for which 5 were excluded.	0.5	Only those with complete outcome data were included.	1	Data collected from medical records.	1	N/A		Unadjusted analyses used.	0	0	Breakthrough bleeding measured up to 12 months following start of treatment	1	Breakthrough bleeding recorded on medical records.	1	N/A	
Hannema 2017	Prospective pre-post single group study	4	Single-clinic population. Participants were invited to participate but no information is given on response rates.	0	Complete data available at earlier timepoints but attrition at later timepoints.	0.5	Data obtained from medical records.	1	N/A		The study only included birth-registered males. Unadjusted analyses were used.	0.5	0	Participants had been treated for at least 12 months	1	Detailed information given on physical, laboratory and radiological investigations.	1	N/A	
Hisle-Gorman 2021	Retrospective pre-post single group study	4	All eligible participants were included from the Military Health System.	0.5	No information given on missing data.	0	Obtained from pharmacy records.	1	N/A		Analyses adjusted for age and sex. Some important covariates such as parental rank adjusted for.	1	0	Median follow-up post-treatment was 1.5 years (IQR 0.7 to 2.7) - potentially not sufficient for all participants	0.5	Outcome data collected from Military Healthcare Data Repository.	1	N/A	
Jarin 2017	Retrospective pre-post single group study	4.5	Four large local clinics. Data extracted from pre-existing databases.	1	Low follow-up rates at each timepoint.	0	Data extracted from pre-existing databases.	1	N/A		Separate analyses were conducted for gender groups. No other covariates were adjusted for.	0.5	0	Outcomes were assessed beyond 6 months	1	Assessed as part of clinical practice.	1	N/A	
Kaltiala 2020	Retrospective pre-post single group study	4.5	Recruited from one of two gender services in Finland. Of 57 diagnosed, 5 were excluded.	1	All followed up.	1	Information in paper provides confidence that clinic data used.	1	N/A		Logistic regression examined whether age and sex were predictive of outcomes, but these variables were not taken into account when examining change pre-post.	0	0	Time between start of treatment and follow-up given as approximately one year	1	Collected by the clinical team via structured and free format assessments, through evaluation of existing files.	0.5	N/A	
Khatchadourian 2014	Retrospective pre-post single group study	3	Single-clinic study, included all patients.	0.5	No information given on missing data.	0	Data obtained from clinical records.	1	N/A		Descriptive summaries were presented separately for males and females.	0.5	0	No information given on time between start of treatment and assessment of outcomes	0	Clinical outcomes assessed as part of routine medical care.	1	N/A	
Klaver 2018	Retrospective pre-post study	3.5	Single-clinic study, participants who had not undergone wholebody DXA were excluded (n=5). 66 participants excluded on different treatment protocol - reason unclear.	0	No information given on follow-up rates.	0	Data collected from medical records.	1	N/A		Analyses were carried out separately for sex. No other covariates were adjusted for.	0.5	0	Final follow-up at 22 years old, and average age at start of CSH was under 17, with an SD of approximately 1	1	Collected from medical records.	1	N/A	
Klaver 2020	Retrospective pre-post study	3.5	Single-clinic study. Study excluded those without whole-body dual-energy radiograph absorptiometry and with no consultation in early adulthood. No numbers reported.	0	No information given on follow-up timepoint.	0	Data collected from medical records.	1	N/A		Separate analyses were carried out for males and females. No other covariates were adjusted for.	0.5	0	At 22 years olds - sufficient follow-up indicated by age at starting hormones	1	Collected from medical records.	1	N/A	
Klink 2015	Retrospective pre-post study	4.5	Single-clinic study. Study only included participants for whom data available at each timepoint. Number of patients excluded not reported.	0	High follow-up rates at final timepoint.	1	Detailed information on timing of treatment given.	1	N/A		Separate analyses were carried out for males and females. No other covariates were adjusted for.	0.5	0	Mean age at start was <17 SD 1.4-2.3, with final follow-up at 22 years old	1	Collected from medical records.	1	N/A	

Kuper 2020	Prospective pre-post single group study	3	Single-clinic study that excluded 22/209 patients due to missing follow-up.	0.5	Despite those with follow-up data being excluded, less than 90% of participants included in analysis of each outcome.	0	Clinician data were entered into a research database.	1	N/A		Hypothesis testing carried out (not separately by age or sex). Regression controlling for confounders planned, but no correlations between change scores and demographic/treatment variables.	0	0	Mean treatment duration 10.9 months, range 1-18 - not all participants sufficient follow-up	0.5	Validated scales used.	1	N/A		
Laurenzano 2021	Retrospective pre-post single group study	4	Single-center study. Included all eligible.	0.5	No information given on follow-up rates.	0	De-identified clinical data entered institutional review board approved endocrine database.	1	N/A		Only birth-registered females were included in the study. Unadjusted analyses were used.	0.5	0	Minimum follow-up was 6 months, with the median being 1.9 years.	1	Collected as part of clinical practice.	1	N/A		
Madsen 2021	Retrospective pre-post single group study	4.5	Large database from single-clinic. 431 were excluded due to missing data, 894 were excluded due to lack of available laboratory results.	0	All included in analysis.	1	Clinic data on treatment reported.	1	N/A		The study only included birth-registered females. Prevalences were summarised descriptively.	0.5	0	Every 3 to 6 months in first year, then yearly / 2-yearly after - follow-up varied although sufficient.	1	Outcomes collected via routine clinical measurements.	1	Data not presented on participants with erythrocytosis at baseline.	0	
Millington 2019	Retrospective pre-post single group study	3	Single clinic study - excluded 5/90 patients for indications other than gender transition (n=2) and no outcome measurements (n=3).	0.5	Thirty-six subjects (42%) had at least one potassium measurement, 48 subjects (56%) had a measurement within the first 6 months.	0	Data collected from medical records.	1	N/A		Unadjusted analyses were used.	0	0	42% within first 3 months of therapy, 56% within first 6 months. Up to 7 years follow-up for some.	0.5	Routine clinical testing.	1	Hyperkalemia was present in one baseline measurement.	0	
Millington 2022	Prospective pre-post study	4.5	Small number of clinics In USA. No information given on consent rates.	0.5	Low follow-up rates at each timepoint.	0	Data collected from medical records.	1	N/A		Age and sex matched z scores were calculated and summarised descriptively.	1	0	Follow-ups to 24 months post-treatment	1	Laboratory, medication, and anthropometric data abstracted from the medical record. Creatinine measurements performed at clinical laboratories of study sites and at outside facilities.	1	N/A		
Mullins 2021	Retrospective pre-post single group study	4	Single clinic that included those initiating treatment and age 13 to 24 years at initiation of GAHT.	0.5	All included in analysis.	1	Data collected from medical records.	1	N/A		Unadjusted analyses were used, most descriptive summaries combined the data from gender groups.	0	0	Median duration of treatment 574 days (IQR: 283-962) - minimum duration unclear	0.5	Data collected as part of routine clinical practice.	1	No indication patients were tested for thrombosis before inclusion.	0	
Olson 2014	Prospective pre-post single group study	4.5	Single-clinic study. No information provided on consent rates.	0	One participant was missing information on testosterone levels and excluded from the analysis.	1	Information on average dose provided - confident that clinic data used.	1	N/A		Only birth-registered females were included in the study. No other covariates were adjusted for.	0.5	0	6 months after starting treatment	1	Hormone levels, lab values, anthropometric measurements and menstrual history were obtained via chart review.	1	N/A		
Olson-Kennedy 2018	Prospective pre-post single group study	4.5	Single-clinic study. No information provided on consent rate. Authors state that 101 participants were evaluated at baseline, but only 59 presented in study.	0	High follow-up.	1	Detailed information on treatment provided.	1	N/A		Gender groups analysed separately. Unadjusted analyses were used.	0.5	0	Follow-up carried out 21-31 months after initiation of hormones	1	Baseline and follow-up physiologic data were abstracted from the medical charts of the participants.	1	N/A		
Perl 2020	Retrospective pre-post single group study	5	Participants recruited from a national clinic, only 3 participants were excluded due to missing data.	1	Those with missing data were excluded from the study.	1	Medical records data used to identify those on treatment.	1	N/A		Only birth-registered females were included in study. No information is given on adjustment for baseline variables.	0.5	0	Follow-up average 4 months (SD 2) - some but not all participants sufficient follow-up	0.5	Data extracted from medical records. BP measured during a clinic visit within 1–4 weeks of GnRHa and testosterone injections using a Welch Allyn Vital Signs Monitor VSM 300 (Welch Allyn, Inc., Beaverton, OR).	1	N/A		
Perl 2021	Retrospective pre-post single group study	5	Participants were recruited from a national clinic, only 1 participant was excluded due to missing data.	1	Those with missing data were excluded from the study.	1	Data extracted from medical records.	1	N/A		Only birth-registered males were included in the study. Unadjusted analyses were used.	0.5	0	Follow-up average 18.5 months, range 3-63 months - some not all participants sufficient follow-up	0.5	Data extracted from medical records. BP measured using a Welch Allyn Vital Signs Monitor VSM 300 (Welch Allyn, Inc., Beaverton, OR).	1	N/A		
Schagen 2018	Prospective pre-post single group study	4	National clinic. No information on consent rates but selected from all eligible.	0.5	No information given on follow-up rates.	0	Details on duration of treatment provided.	1	N/A		Separate analyses carried out for males and females. No other covariates were adjusted for.	0.5	0	Analyses used data up to 2 years post-treatment	1	Detailed information given on laboratory investigations.	1	N/A		
Schagen 2020	Prospective pre-post single group study	5	National clinic. Small number of all treated excluded due to DXA scans not being available at the start of GnRHa.	1	No information given on missing data rates at follow-up.	0	Detailed treatment protocol provided.	1	N/A		Analyses adjusted for pubertal stage and sex. No other covariates were adjusted for.	1	0	Analyses presented up to 36 months of treatment	1	Dual-energy x-ray absorptiometry (DXA) performed using Hologic QDR 4500. Markers of bone formation and resorption determined in fasting blood samples, drawn on day of scans.	1	N/A		
Segev-Becker 2020	Retrospective pre-post single group study	3	National clinic. Consecutive participants recruited.	1	No information given on follow-up rates.	0	Information on treatment delivery presented.	1	N/A		Some but not all descriptive summaries stratified by gender/sex. Participants were split into pre-pubertal and pubertal groups.	0	0	No information given on follow-up period	0	Data collected retrospectively from clinical records.	1	N/A		
Sequeira 2019	Retrospective pre-post single group study	5	Single-clinic study. No information given on number excluded.	0	All included in model.	1	Information in paper provides confidence that clinic data used.	1	N/A		Only birth-registered females were included. The analysis controlled for age, hormone dose rate and baseline BMI z-score.	1	0	Follow-up carried out to 12 months post-initiation of treatment	1	Height and weight documented in each clinical encounter were used to calculate BMI in kg/m2.	1	N/A		
Stoffers 2019	Retrospective pre-post single group study	4	Single-clinic study. Only 2/64 participants declined to participate.	0.5	High rates of follow-up at 6 months post-treatment, but low rates at 12 and 24 months post-treatment.	0.5	Information in paper provides confidence that clinic data used.	1	N/A		Only birth-registered females were included. Unadjusted analyses were used.	0.5	0	The median duration of follow-up was 12 months (range 5-33 months)	0.5	Data collected via chart review.	1	N/A		
Tack 2016	Retrospective pre-post single group study	4	Single-centre study in country with three clinics. Small number (5 out of 43) excluded due to missing data.	0.5	No information given on follow-up rates at each timepoint.	0	Information in paper provides confidence that clinic data used.	1	N/A		Only birth-registered female adolescents were included. Unadjusted analyses were used.	0.5	0	Baseline, 6 and 12 month follow-up	1	Data collected as part of clinical follow-up.	1	N/A		
Tack 2017	Retrospective pre-post single group study	4.5	Single-centre study in country with three clinics. All those who received CA for at least 6 months included.	1	No information given on follow-up rates at each timepoint.	0	Information in paper provides confidence that clinic data used.	1	N/A		Only birth-registered males were included in the study. Unadjusted analyses were used.	0.5	0	Baseline, 6 and 12 month follow-up	1	Data collected as part of clinical follow-up.	1	N/A		
van der Loos 2021	Retrospective pre-post study	4.5	Single-clinic study. 123 excluded due to DXA not being available.	0	Only participants who had a DXA were included.	1	Information in paper provides confidence that clinic data used.	1	N/A		Separate analyses were carried out for males and females.	0.5	0	Follow up >2 years after treatment initiation	1	Detailed information on DXA testing given.	1	N/A		
Vlot 2017	Retrospective pre-post single group study	4	Single-clinic study. A large number of eligible participants were excluded due to incomplete data.	0	Data indicates that more than 10% were missing data for outcomes.	0	Data collection took place at point of treatment.	1	N/A		Analysis stratified on sex and bone age. Unadjusted analyses were used.	1	0	Follow-up 24 months after start of hormone treatment	1	Detailed information on DXA testing provided.	1	N/A		
Cross-sectional																					
Arcelus 2016	Cross-sectional study with controls	3.5	299 eligible patients - 31 did not answer questions regarding NSSI and were excluded. National clinic.	1	More than 10% excluded from analysis - no information provided on those or explanation.	0	Self-reported data on treatments received prior to assessment at adult clinic.	0	Those not treated with hormones from same clinic sample.	1	Controlled for gender, self-esteem, transphobia, interpersonal problems, social support.	0.5	0	N/A		Validated assessment tools used.	1	N/A		
Burke 2020	Cross-sectional study with controls	5	Single clinic population - no information provided about recruitment and response, or number eligible.	0	All participants included in analysis.	1	Clinic data used to select / categorise treatment groups.	1	Treatment naïve from same source plus controls with no gender incongruence which was appropriate for outcome.	1	Controlled for puberty stage / age, sex assigned at birth but no other treatments. Cross-sectional so no baseline control.	1	0	N/A		Standard assessment - equipment and procedure explained in full - same applied to all participants (treatments and controls).	1	N/A		
Fontanari 2020	Cross-sectional study with controls	3	Self-selecting survey.	0	All participants who completed the survey were included in the analysis.	1	Self-report.	0	Non-exposed group from same survey sample.	1	No adjustment made for age, sex, co-interventions or socio-demographic confounders.	0	0	N/A		Validated scales used.	1	N/A		
Grannis 2021	Cross-sectional study with controls	4	Single-clinic population. Participants excluded on ability to participate in MRI based research, but no numbers are given on this.	0	Information on those included in analysis not provided.	0	Data collected from medical records.	1	Drawn from same source.	1	Age was adjusted for. Only birth-registered females were included.	1	0	N/A		Validated scales and detailed brain imaging protocol.	1	N/A		
Green 2022	Cross-sectional study with controls	4	Self-selecting survey	0	2895 participants missing data on treatment and excluded from analyses. Missingness could be related to the underlying value.	0	Self-report.	0	Drawn from same source as exposed group.	1	Analysis controlled for age, sex, socio-demographic characteristics, also controlled for parental support, victimisation, receipt of puberty blockers, receipt of gender identity conversion therapy.	1	1	N/A		Validated scale used to measure depression, items on suicidal thoughts and behaviours taken from Youth Risk Behaviour Survey.	1	N/A		
Millington 2021b	Cross-sectional study with controls	2.5	Single-clinic study with no information given on numbers excluded.	0	No information given on data collection rates.	0	Detailed information on treatment provided.	1	No information given on selection of non-exposed group.	0	Only birth-registered females were included. Mostly unadjusted analyses were carried out.	0.5	0	N/A		Results collected as part of clinical practice.	1	N/A		
Nokoff 2020	Cross-sectional study with controls	4	Single-clinic study. No information provided on consent rate.	0	All participants included in analysis according to table data.	1	Information in paper provides confidence that clinic data used.	1	Controls from Colorado RESistance to InSulin in Type 1 ANd Type 2 diabetes, and Health Influences in Puberty studies. Different ages in both and skewed clinical characteristics in one.	0	Separate analyses were carried out for sex, and analyses either matched on Tanner stage or adjusted for age. Analyses also matched on BMI.	1	0	N/A		Body composition measured using DXA and detailed information on laboratory assays provided.	1	N/A		
Strang 2022	Cross-sectional study with controls	4	Shared study protocol in two locations. No information given on consent rates.	0	Only those with complete report forms were included.	1	Collected through parent and self-report, and only verified through medical records when dates not recalled.	0	Drawn from same population as exposed group.	1	Analyses adjusted for assigned sex and age.	1	0	N/A		Validated scales and evaluations used.	1	N/A		

Turban 2022	Cross-sectional study with controls	3	National survey covering 50 states in collaboration with 400+ lesbian, gay, bisexual and transgender organisations. Population covers 18-36 year olds.	1	No information given on number of participants excluded from analyses due to missing data.	0	Self-reported by participants.	0	Drawn from same population as exposed group.	1	Separate analyses were carried out for participants receiving hormones in early adolescence, late adolescence and adulthood. Some but not all analyses adjusted for sex assigned at birth.	0.5	0	N/A		One validated scale used, the rest appear to be bespoke for the study.	0.5	N/A		