Supplemental data appendix 1

Comparison of Cass Review to Kennedy Report (Learning from Bristol), Cumberlege Review (First do no harm), and National Maternity Review (Better births)

Independent Review of Gender Identity Services for Children and Young People	**Learning From Bristol: The Report of the Public Inquiry into Children's Heart Surgery at the Bristol Royal Infirmary**	**First Do No Harm: The Report of the Independent Medicines and Medical Devices Safety Review**	**Better Births: Improving Outcomes of Maternity Services in England**
Years: 2020-2024 **Chair**: Dr Hilary Cass, Paediatrician and former President of the Royal College of Paediatrics and Child Health	**Years:** 1998-2001 **Chair**: Professor Sir Ian Kennedy, Lawyer and Professor of Health Law, Ethics, and Policy	**Years:** 2018-2020 **Chair:** Baroness Julia Cumberlege, Former Health Minister and Healthcare Campaigner	**Years**: 2015-2016 **Chair:** Baroness Julia Cumberlege, Former Health Minister and Healthcare Campaigner
Number of Recommended Actions			
32 Recommendations	198 Recommendations	9 Recommendations 50 Actions for improvements	7 General Recommendations 28 Specific Recommendations
Terms of Reference			
Care Pathways: Evaluate pathways for less complex gender incongruence and specialist services, including referral criteria. **Clinical Models:** Examine clinical management from assessment to discharge, outlining objectives, benefits, outcomes, and interventions. **Complex Cases:** Identify best practices for complex presentations. **Medication Review:** Analyse hormone treatments and gender-affirming drugs, with NICE reviewing objectives, benefits, outcomes, and risks.	To investigate the management of complex cardiac surgical care for children at Bristol Royal Infirmary from 1984 to 1995, assess service adequacy, determine actions taken in response to concerns, identify any delays in addressing issues, and provide recommendations to ensure high-quality NHS care.	**Objective:** Recommend improvements in how the healthcare system responds to safety concerns regarding clinical interventions, including medicines and medical devices. **Assessment Focus:** Evaluate historical evidence on scientific knowledge, decision-making, and actions related to Primodos, sodium valproate, and pelvic mesh by manufacturers, regulators, clinicians, and policymakers. **Key Areas of Evaluation** 1. Use of scientific evidence 2. Patient consent practices 3. Response to Safety Concerns	**Evaluating Evidence:** Review UK and international evidence to recommend safe and efficient maternity service models, including midwife-led units. **Supporting Women's Choices:** Ensure the NHS enables women to make safe and appropriate maternity care choices. **Supporting NHS Staff:** Support NHS staff, particularly midwives, in providing responsive care. **Addressing Geographic Challenges:** Focus on achieving these objectives in geographically isolated areas.

Cheung CR, *et al. Arch Dis Child* 2024;0:1–5. doi: 10.1136/archdischild-2024-327994

Audit and Research: Recommend ongoing audits, long-term follow-up, data reporting, and research priorities. Workforce Needs: Assess current and future workforce requirements. Referral Trends: Explore reasons for increased referrals, especially among natal females, and their implications.		4. Action Timeliness and Effectiveness of actions taken by manufacturers, regulators, and policymakers	
Resources Drawn Upon to Inform Investigation			
The review drew on systematic reviews, qualitative and quantitative research, audits, professional input from clinicians and staff, focus groups, thematic roundtables, insights from service users and parents, and international perspectives through surveys, guideline appraisals, and meetings with global clinicians and policymakers.	Resources included examining the wider context of paediatric cardiac surgical services, gathering and independently analysing all available clinical data and case notes, commissioning expert groups for historical context and statistical analysis, and incorporating academic research, expert interactions, and parent testimonies.	The review gathered evidence from a diverse range of stakeholders, including clinicians, Royal Colleges, the pharmaceutical industry, medical device manufacturers, NHS and private sector providers, regulatory and professional bodies, and the Department of Health and Social Care.	The review's conclusions and recommendations are based on three key sources: Dr. Bill Kirkup's assessment of current maternity care quality, an independent review by the National Perinatal Epidemiology Unit at Oxford University, and the final report on care failings at Morecambe Bay NHS Trust. The review team also engaged extensively with the public, service users, staff, and other stakeholders over twelve months.
Conclusions			
Waiting Lists: Long waiting times for gender services impact individuals and NHS delivery. **Evolving Identity:** Young people's sense of identity can evolve; no hierarchy of gender expression should exist. **Transition Decisions:** While some may urgently seek transition, others might advise a slower approach. Outcomes vary; some may transition, de/retransition, or experience regret.	**Child-Centred Care:** Future care must be child-centred with trained staff and suitable facilities, guided by a national director for children's healthcare. **Safety:** The NHS must eliminate unsafe practices, promote openness, and enforce a duty of candour. Clinical negligence litigation should be abolished, and safety should be overseen by a non-executive board member. **Professional Competence:** Future	**Government Apology:** Issue a full apology for the impacts of Primodos, sodium valproate, and pelvic mesh. **Patient Safety Commissioner:** Appoint an independent Commissioner to oversee patient safety and user perspectives. **Redress Agency:** Create an independent agency to address systemic harm from medicines and devices. **Support Schemes:** Set up schemes to cover additional care costs for those harmed.	**Personalised Care**: Focus on women, babies, and families with genuine choice and unbiased information. **Continuity of Carer**: Ensure safe, trust-based care through consistent relationships aligned with women's decisions. **Safer Care**: Promote teamwork across boundaries, rapid referrals, and leadership in safety culture. Emphasise investigation and learning from mistakes. **Enhanced Mental Health Care**: Address historic underfunding in postnatal and

Holistic Care: Care should be personal and holistic, including services outside NHS specialists. **Diverse Opinions:** There is no consensus on the best treatment approach; evidence is weak, and clinicians are uncertain about enduring trans identities. **Clinician Concerns:** Primary and secondary care clinicians worry about their ability to work with this population and fear social debate. **Long-Term Effects:** The long-term health impacts of hormone interventions are not well understood. **Transfer Vulnerability:** Young people are particularly vulnerable during the transition to adult services. **Increasing Demand:** The NHS faces rising requests for support for gender-related distress, requiring high-quality care. **Compassionate Support:** Every child and young person seeking help from the NHS should receive compassionate and appropriate support to thrive.	contracts must include appraisal, professional development, and revalidation to ensure competence. **Organisational Change:** All healthcare workers should have similar employment terms and clear accountability. **Care Standards:** Future care must follow published standards, with non-compliant hospitals barred from offering NHS services. **Openness:** There must be clear, accessible information on clinical performance for patients. **Monitoring:** Future systems should include internal checks and independent reviews.	**Specialist Networks:** Establish networks for care and advice on implanted mesh and medications during pregnancy. **MHRA Reforms:** Revise the MHRA for better adverse event reporting, device regulation, and patient engagement. **Patient Database:** Create a central database for tracking device implantation and outcomes. **Transparency in Payments:** Expand GMC registers and mandate reporting of payments by the pharmaceutical and device industries. **Implementation Task Force:** Form a task force to implement these recommendations with a set timeline.	perinatal mental health to improve well-being for women, babies, and families. **Multi-Professional Collaboration**: Break down barriers between midwives, obstetricians, and other professionals for safe, personalised care. **Cross-Boundary Coordination**: Support personalisation, safety, and choice in maternity services, with access to specialist care when needed. **Fair Payment System**: Ensure providers are fairly compensated for high-quality care, and support commissioners in promoting personalisation, safety, and choice.

Supplemental data appendix 2

Summary of criticism of the evidence appraisal process *

Criticism	Short Response	Long Response
York should not have introduced the additional evidence appraisal tools after the protocol was registered.	Modification to protocols are common in research. The York modifications were necessitated by study designs, and appropriately documented.	Methods adjustments, including modifying risk of bias appraisal tools to more closely meet study designs under review, are a common practice. The instrument in the registered protocol (MMAT) was unable to meet the need of systematic reviewers. The registered protocol provides this rationale, explains that after the literature search yielded a variety of study designs, so two more additional appraisal tools were added, the Newcastle Ottawa Scale (NOS) and AGREE II. Both represent well-established and widely used tools in evidence appraisal, and are appropriate choice for the included study design in the systematic review work.
York's 'single search strategy" by design misses 'many relevant studies'.	It is common to use a single search 'umbrella' strategy when evaluating evidence in related domains.	When doing systematic reviews related to one another, performing an umbrella search that is broad enough to cover the entire area is entirely appropriate. The comprehensive single search used two combined concepts: 'children', including all terms for children and adolescents; and 'gender dysphoria', including associated terms such as gender-related distress and gender incongruence, and gender identity terms including transgender, gender diverse and non-binary. The York search strategy yielded far more studies than any other reviews (237 studies from 18 countries covering 113,269 children and adolescents) and was the most comprehensive to date.
Important studies were ignored by the York reviews, including the 'highly impactful' study by Chen et al.,2023 and another by Tordoff et al., 2022	Both studies were considered by York. Neither change the conclusions of 'very low quality evidence.'	The York systematic reviews included low quality studies, but the systematic reviews did not summarize the outcome information from these low quality studies. This practice is entirely appropriate. When meta-analysis is feasible, systematic reviewers could conduct sensitivity analysis to examine the robustness and impact of including results from low quality studies. In this case of no meta-analysis, setting a threshold based on study quality (risk of bias), and providing summary of only studies meeting a certain method quality threshold is a sensible approach. Tordoff et al., 2022 was part of the systematic review and rated as low quality. Chen et al., 2023 was published after the search had been completed, but the York researchers did review it and determined that incorporating it would not change the systematic review conclusions. For clarity, we conducted an independent analysis of these two oft-quoted studies using NOS as well as the ROBINS-I scale which McNamara et al. state should have been used instead (Supplemental data Appendix 4). Both studies were rated 'critical risk of bias' and do not change York's conclusions. Notably, the NOS assessment yielded a higher rating than ROBINS-I, suggesting that York's choice of tool was likely more forgiving. See Supplemental data appendix 4.

AGREE II: Appraisal of Guidelines, Research and Evaluation

GRADE: The Grading of Recommendations Assessment, Development and Evaluation

MMAT: the Mixed Methods Assessment Tool

NOS: Newcastle-Ottawa Scale;

ROBIN-I: the Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool

Supplemental data appendix 3

Critical Appraisal of Chen et al. and Tordoff et al.

Chen D, Berona J, Chan YM, Ehrensaft D, Garofalo R, Hidalgo MA, Rosenthal SM, Tishelman AC, Olson Kennedy J. Psychosocial functioning in transgender youth after 2 years of hormones. *N Engl J Med*. 2023 Jan 19;388:240-250. doi: 10.1056/NEJMoa2206297. Correction *N Engl J Med*. 2023 Oct 19;389:1540. doi: 10.1056/NEJMx230007.

Tordoff DM, Wanta JW, Collin A, Stepney C, Inwards-Breland DJ, Ahrens K. Mental health outcomes in transgender and nonbinary youths receiving gender-affirming care. *JAMA Netw Open*. 2022 Feb 1;5:e220978. doi: 10.1001/jamanetworkopen.2022.0978. Correction *JAMA Netw Open*. 2022 Jul 1;5:e2229031. doi: 10.1001/jamanetworkopen.2022.29031

Table 1: Newcastle-Ottawa Scale (NOS)

Study	Study design	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-exposure group	NOS Q2 score†	NOS Q5. comparability of cohorts (part 1 - controls for key demographics; part 2 - controls for co-interventions)	NOS Q5 part 1 score †	NOS Q5 part 2 score †	NOS Q7 follow-up duration	NOS Q7 score†	NOS Q6 assessment of outcome	NOS Q6 score†	NOS Q4. Demonstration that outcome of interest not present at study start	NOS Q4 score†
Chen 2023	Prospective pre-post single group study	5	National/regional clinics, large cohort	1	Data were available for 81% of all possible observations; 75.2% participants (237 of 315) completed at least four study visits.	0	The intervention is clearly defined in this study: gender-affirming hormones (testosterone or estradiol) Participants initiating GAH as part of their clinical care were included in this cohort, and only those participants. This was ascertained through medical records	1	N/A		This study used mixed-effects model, which accommodated missing data and nonnormal distributions, age and natal sex were included in the model, but not other co-interventions	1	0	2 years	1	Validated scales	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	

* Higher scores indicate higher study quality. For each study, the rating is calculated as the percentage of the score this study has received (the total score for this study divided by the maximum score which this study can receive). Each study is classified as low (in shaded red, ≤ 50% of the maximum score), moderate (in shaded orange, >50 to 75%), and high (in shaded green, >75%). The rating for pre-post single group studies (Chen 2023) is calculated out of a total of 7 (because Q2 is not relevant); while the rating for Cohort studies (Tordoff 2022) is calculated out of a total score of 8.
†For each scoring item question, the score is colour-coded as shaded red (0), shaded orange (0.5), and shaded green (1). Not applicable items (N/A) did not generate a score and the respective cells are shaded grey. High scores indicate higher study quality on this item.
‡ The assessment for Tordoff is from published systematic reviews by the York systematic review team.[1]

[1] Taylor J, Mitchell A, Hall R, Langton T, Fraser L, Hewitt CE. Masculinising and feminising hormone interventions for adolescents experiencing gender dysphoria or incongruence: a systematic review. *Arch Dis Child*. 2024 Apr 9;archdischild-2023-326670.

Table 2. The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I)*

Study	Risk of bias due to confounding	Risk of bias due to selection of participants	Risk of bias in classification of interventions	Risk of bias due to deviations from intended interventions	Risk of bias due to missing data	Risk of bias in measurement of outcomes	Risk of bias in reported results	**Overall risk of bias**
Chen 2023†	Critical due to uncontrolled confoundings	Low	Low	Critical due to possible co-interventions	Low	Critical due to unblinded assessment	Low	**Critical**
Tordoff 2022	Critical due to uncontrolled confoundings	Low	Serious due to not well defined intervention	Critical due to possible co-interventions	Critical due to missing participant data	Critical due to unblinded assessment	Serious due to possible selective reporting	**Critical**

* The ROBINS-I instrument assessment tool classifies an individual study as with low, moderate, serious, and critical risk of bias for each risk of bias domain, and for the overall risk of bias of the study. For each risk of bias domain, and for the overall risk of bias, each was colour coded as shaded green (low risk of bias), shaded yellow (moderate risk of bias), shaded orange (serious risk of bias), and shaded red (critical risk of bias). Both studies are of critical risk of bias according to our assessment with the ROBINS-I tool.
†Our assessment shows that the assessment with the adapted NOS scale probably is more forgiving than the assessment with the ROBINS-I tool. Studies assessed as moderate quality would probably be of critical risk of bias due to its emphasis in confounding in non-randomized studies.

Cheung CR, *et al. Arch Dis Child* 2024;0:1–5. doi: 10.1136/archdischild-2024-327994

Cheung CR, *et al. Arch Dis Child* 2024;0:1–5. doi: 10.1136/archdischild-2024-327994