



## RESEARCH ARTICLE

# REVISED Language phenotypes in children with sex chromosome trisomies [version 2; referees: 3 approved]

Dorothy V. M. Bishop , Annie Brookman-Byrne , Nikki Gratton, Elaine Gray, Georgina Holt , Louise Morgan, Sarah Morris , Eleanor Paine, Holly Thornton, Paul A. Thompson

Department of Experimental Psychology, University of Oxford, Oxford, Oxon, OX2 6GG, UK

**v2** First published: 08 Nov 2018, 3:143 (<https://doi.org/10.12688/wellcomeopenres.14904.1>)  
 Latest published: 28 Jan 2019, 3:143 (<https://doi.org/10.12688/wellcomeopenres.14904.2>)

## Abstract

**Background** Sex chromosome trisomies (47,XXX, 47,XXY and 47,XYY) are known to be a risk factor for language disorder, but it is hard to predict outcomes, because many cases are identified only when problems are found.

**Methods** We recruited children aged 5-16 years with all three types of trisomy, and divided them into a High Bias group, identified in the course of investigations for neurodevelopmental problems, and a Low Bias group, identified via prenatal screening or other medical investigations. Children from a twin sample were used to compare pattern and severity of language problems: they were subdivided according to parental concerns about language/history of speech-language therapy into a No Concerns group (N = 118) and a Language Concerns group (N = 57). Children were assessed on a psychometric battery and a standardized parent checklist. After excluding children with intellectual disability, autism or hearing problems, the sample included 28 XXX, 18 XXY and 14 XYY Low Bias cases and 7 XXX, 13 XXY and 17 XYY High Bias cases.

**Results** Variation within each trisomy group was substantial: within the Low Bias group, overall language scores were depressed relative to normative data, but around one third had no evidence of problems. There was no effect of trisomy type, and the test profile was similar to the Language Concerns comparison group. The rate of problems was much greater in the High Bias children with trisomies.

**Conclusions** When advising parents after discovery of a trisomy, it is important to emphasise that, though there is an increased risk of language problems, there is a very wide range of outcomes. Severe language problems are more common in those identified via genetic testing for neurodevelopmental problems but these are not characteristic of children identified on prenatal screening.

## Keywords

Trisomy X, Klinefelter syndrome, XYY syndrome, ascertainment bias, variability, psychometric testing, parent report

## Open Peer Review

Referee Status:

	Invited Referees		
	1	2	3
<b>version 2</b> published 28 Jan 2019	 report	 report	 report
<b>version 1</b> published 08 Nov 2018	 report		

- 1 **Kate Baker** , University of Cambridge, UK
- 2 **Laura Zampini** , University of Milano-Bicocca, Italy
- 3 **Nancy Raitano Lee** , Drexel University, USA

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Dorothy V. M. Bishop ([dorothy.bishop@psy.ox.ac.uk](mailto:dorothy.bishop@psy.ox.ac.uk))

**Author roles:** **Bishop DVM:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Brookman-Byrne A:** Investigation, Project Administration, Writing – Review & Editing; **Gratton N:** Investigation, Writing – Review & Editing; **Gray E:** Investigation, Project Administration, Writing – Review & Editing; **Holt G:** Investigation, Methodology, Writing – Review & Editing; **Morgan L:** Investigation, Project Administration, Writing – Review & Editing; **Morris S:** Investigation, Methodology, Project Administration, Writing – Review & Editing; **Paine E:** Investigation, Project Administration; **Thornton H:** Investigation, Project Administration; **Thompson PA:** Formal Analysis, Validation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was funded by Wellcome Trust Programme Grant 082498.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2019 Bishop DVM *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Bishop DVM, Brookman-Byrne A, Gratton N *et al.* **Language phenotypes in children with sex chromosome trisomies [version 2; referees: 3 approved]** Wellcome Open Research 2019, 3:143 (<https://doi.org/10.12688/wellcomeopenres.14904.2>)

**First published:** 08 Nov 2018, 3:143 (<https://doi.org/10.12688/wellcomeopenres.14904.1>)

**REVISED Amendments from Version 1**

**Overview:** We discovered a scripting error that exclusionary criteria were based on the wrong variable and so Ns were misreported. There is no substantive impact on conclusions, but all results are affected. These are now corrected in text and deposited scripts/data. We now include plots showing results for children who are excluded because of autism spectrum disorder or intellectual disability. Other changes have been made in response to the review by Kate Baker.

**Abstract:** Corrected Ns reported: Twins: No Concerns group (N = 118), Language Concerns group (N = 57). SCT cases: 'the sample included 28 XXX, 18 XXY and 14 XYY Low Bias cases and 7 XXX, 13 XXY and 17 XYY High Bias cases'.

'Characteristic' used instead of 'typical' in final line.

**Introduction,** para 2. Revised to include quote from [Linden et al. \(2002\)](#).

**Participants:** as well as presenting corrected Ns, added a paragraph stating numbers with different reasons for prenatal testing

**Figure 1–Figure 6, Figure S1–Figure S3:** revised to show corrected Ns

**Comparison group:** minor changes in wording to improve clarity to this section; Ns corrected in text.

**Table 2–Table 5 and Table S1:** updated with correct Ns

**Figure 3–Figure 6:** redrawn with correct Ns

**Results** Sections A, B and C: to reduce amount of missing data, MANOVA now restricted to three composites, with separate ANOVA for the 4<sup>th</sup> measure.

End of results section A, and end of CCC-2 section: added analysis to test for variance differences between those with extra X vs extra Y.

**Figure 4&Figure 6:** Figure redrawn to include cases with ASD or intellectual disability. These are *not* included in the analysis, but are shown to indicate the impact of their exclusion.

**Discussion** :minor rewording for clarification.

End of para 1, added statement on statistical power.

Para 4, clarification of Ns with missing data.

Final paragraph: more about clinical implications.

**See referee reports**

Chromosome trisomies arise when an error of cell division leads to an egg or sperm that contains two copies rather than one copy of the chromosome. Trisomies that affect one of the autosomes are usually lethal, and in survivors they lead to marked physical and mental abnormalities. Trisomies of the sex chromosomes, however, have milder impacts, and often go undetected ([Printzlaue et al., 2017](#)). This complicates the study of the consequences of sex chromosome trisomies in two ways; first, it can be difficult to recruit large samples of cases, and second, those that are studied may not be representative of the population. In particular, there is a danger of overestimating the severity of impairment if we include cases where discovery of the trisomy was prompted by genetic testing to investigate developmental abnormalities.

In the 1960s, a multicentre project was initiated with the aim of evaluating the impact of sex chromosome trisomies in samples identified on newborn screening that were free from ascertainment bias ([Robinson et al., 1979](#)). The three kinds of trisomy - trisomy X (47,XXX), Klinefelter's syndrome (47,XXY) and 47,XYY karyotypes were all found to be associated with neurodevelopmental problems, particularly affecting language and motor functions, see [Leggett et al. \(2010\)](#) for review. Nevertheless, in summarising implications for genetic counselling, [Linden et al. \(2002\)](#) concluded: 'It is now known that these individuals are at an increased risk for developmental problems, but that most are in the normal range of development, and marked abnormality is not usually seen.' (p 4)

To our knowledge, there have been no more newborn screening studies initiated in the past 50 years. One reason is practical: each of the three trisomy types has a prevalence ranging from 1 in 600 to 1 in 1000, so many thousands of cases need to be screened to identify even a small sample. Second, ethical issues are raised when a screening procedure identifies a trisomy in a newborn: while knowledge of the trisomy could be useful in helping parents take steps to ensure early intervention, it can also be damaging by creating anxiety ([Valentine, 1979](#)).

It is now possible to detect sex chromosome trisomies in the foetus in the course of prenatal screening. Prenatal screening for Down syndrome was developed in the 1980s and led to cases of sex chromosome trisomy being detected as an incidental finding ([Cuckle & Maymon, 2016](#)). Because prenatal screening is offered to older mothers, there is some bias in samples identified this way, but detection of the trisomy is not dependent on the child's developmental outcome.

A study of neurodevelopmental outcomes of prenatally-identified children was reported by [Bishop et al. \(2011\)](#). Outcome measures were based on standardized parental report, as it was deemed unethical to conduct direct testing of children who might not be aware of their trisomy. This study was consistent with the earlier newborn screening studies in finding a high rate of language difficulties in all three trisomies, but in addition, there was an elevated risk of autism spectrum disorder (ASD) diagnosis in boys with both XXY and XYY karyotypes. This had not been reported in the earlier prospective studies of newborns, possibly because diagnostic criteria for autism were far more stringent in the 20th century. A second group of children with trisomies detected postnatally was found to have a similar profile, but levels of all neurodevelopmental problems were higher, as would be expected in a group where there was ascertainment bias. In the course of this study, we established that 43% of children had been told about the trisomy. This suggested that a study that involved direct neurodevelopmental assessment of children would be feasible.

The association with language problems is intriguing because there are few genetic aetiologies that selectively impair language development. There are two reasons why it is of interest to compare the language phenotype of children with sex

chromosome trisomies with that seen in children with developmental language disorder (DLD) of unknown origin. First, a comparative study could also help determine whether interventions that have been devised for children with DLD are also likely to be effective for children with a sex chromosome trisomy: (Bishop *et al.*, 2017). Second, if the similarities are not just superficial, then studying how an extra chromosome affects language development could help us understand aetiology of DLD (Bishop & Scerif, 2011). Studies to date have suggested that there may be different profiles of impairment in the three types of trisomy, which could be related to the different genetic and hormonal impacts of an extra X or Y chromosome (Bender *et al.*, 1983), but we still have only a very limited body of data from samples free from ascertainment bias. It has also been suggested that an additional X chromosome could have more variable effects than an additional Y chromosome (Skuse, 2018).

In 2011 we embarked on data collection for a new study of children with sex chromosome trisomies, including both pre- and post-natally identified cases. The goals were, first, to document the nature and range of language and behavioural difficulties in the three types of trisomy, and second, to test whether individual variation in outcomes could be predicted by variation in specific genetic variants. Here we report descriptive data relating to the first question, with a particular emphasis on language characteristics. Behavioural and psychiatric findings will be covered in a companion paper. To date we have not found any associations of phenotypes with variants in *CNTNAP2* and *NRXN1* genes; these analyses are reported elsewhere (Newbury *et al.*, 2018).

Our focus here is both on documenting whether there are reliable differences in language and cognitive outcomes for the three types of trisomy, and how far the pattern of language problems resembles that seen in children with developmental language disorder (DLD) who do not have any known genetic abnormalities. To minimize ascertainment bias, these analyses were restricted to children whose trisomy was either detected prenatally, or as an incidental finding during other medical investigations (referred to as a 'Low Bias' group). To address the latter question, we compared the results of trisomic children with those of children who had participated in a twin study of language development and disorders, and who had been given the same test battery. Finally, we considered how far trisomy outcomes are influenced by ascertainment bias, by comparing the Low Bias group with a High Bias group whose trisomies were discovered in the course of investigation of behavioural or neurodevelopmental problems.

The specific questions we considered were:

1. Does severity or variation in language profile differ across the three types of trisomy?
2. Is there a distinctive language and cognitive phenotype of children with sex chromosome trisomies, or do they resemble children with language problems who have the usual XY or XX karyotype?
3. How much does ascertainment bias affect the language and cognitive profiles of children with sex chromosome trisomies?

## Procedure

Ethical approval was obtained for the study in 2011 from the Berkshire NHS Research Ethics Committee (reference 11/SC/0096), and data collection started in August of that year, finishing in October 2016. Information sheets, consent forms and ethics approval documents are available on [Open Science Framework](#).

## Participants

### Sex chromosome trisomy cases

We recruited 45 girls with XXX, 46 boys with XXY and 51 boys with XYY to the study, in the age range from 5 years 0 months to 15 years 11 months. 28% of the children had participated in the Bishop *et al.* (2011) study. Recruitment was via NHS Clinical Genetics centres, two support groups (Unique: the Rare Chromosome Support Group, and the Klinefelter Syndrome Association) and self-referral through social media. To be included in the study, the child had to be aware of their trisomy status and assent to take part after viewing an information video that explained what was involved in the study.

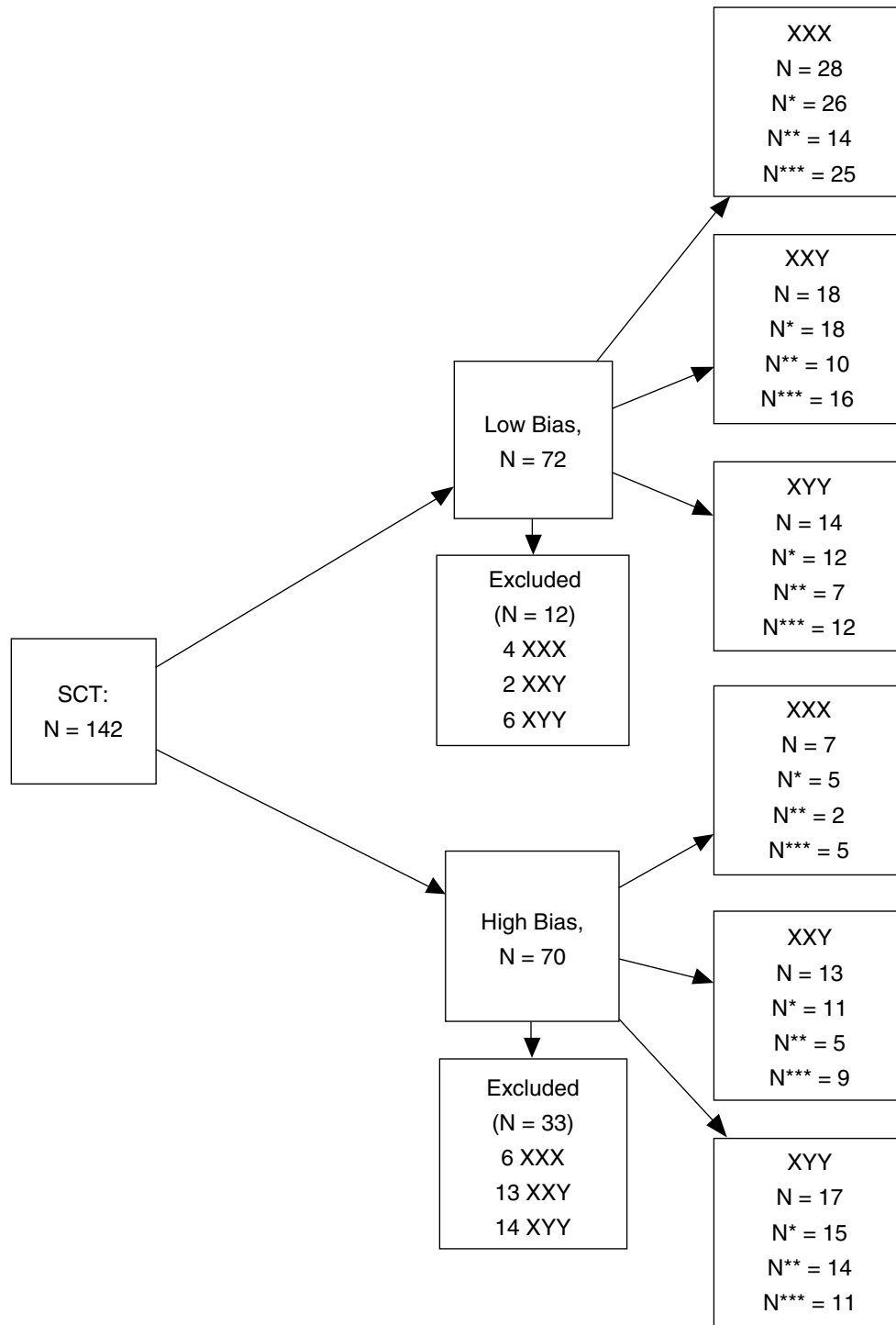
During the initial telephone interviews, caregivers were asked how their child was diagnosed, in particular whether this followed postnatal testing motivated by neurodevelopmental/behavioural problems. The phenotype of such children may be more severe, potentially biasing the sample, and so, as shown in [Figure 1](#), we grouped them prospectively as a High-risk-of-bias (High Bias) subgroup, comprised of 13 XXX, 26 XXY and 31 XYY cases. All other children formed the low-risk-of-bias (Low Bias) subgroup, with 32 XXX, 20 XXY and 20 XYY cases.

This latter group included 49 families who underwent amniocentesis: all but 4 parents provided a reason: 30 because of maternal age, 11 because nuchal screening or ultrasound scan indicated risk of a chromosome disorder, and 4 because of a family history of a chromosome disorder. A further 23 children had genetic testing for a range of medical conditions, including growth problems, failure to thrive, and failure to go through puberty (in the case of boys with XXY). Note that in referring to these as a 'Low Bias' group, we do not imply they are a representative sample of children with an extra sex chromosome: we make the more limited claim that the trisomy was not discovered as a result of the child manifesting language or behaviour difficulties. All genetic testing was done via the National Health Service, so the sample was not limited by parents' ability to pay for a test.

There was an unanticipated imbalance between the three trisomy types in terms of the proportion of cases in the Low Bias group: 80% for XXX girls, 58% for XXY boys, and 45% for XYY boys; chi-square (2) = 8.73,  $p = 0.01$ .

### Exclusionary criteria

Given our current focus on language disorder, we excluded from the main analysis children who had a nonverbal ability scaled score (PIQ) more than 2 SD below the mean (2 Low Bias and 9 High Bias), those with a diagnosis of Autism Spectrum Disorder (ASD) (9 Low Bias and 23 High Bias), and those who



\* With complete language data,  
 \*\* Complete language: aged 6 to 11 yr,  
 \*\*\* With useable CCC-2 data

Ascertainment  
bias

Trisomy

**Figure 1.** Numbers of children with sex chromosome trisomies included in different analyses.

failed a hearing screen (1 Low Bias and 1 High Bias). This enabled us to make a direct comparison of language scores with the twin sample, which also excluded children on this basis, and to see how far language problems were seen in children who did not have accompanying conditions that are likely to affect language function. However, note that this entails that our conclusions are based on a selected sample of trisomy cases. Where children who were excluded because of low IQ or ASD had completed the test battery, their individual data are shown in the plotted results, so the reader can form an impression of the impact of the exclusions on summary results. ASD in children with sex chromosome trisomies will be the focus of a companion paper.

### Missing data

Where a child had missing data on just one of the language/cognitive tests, that value was prorated from the mean of that child on the other measures. Likewise, when there was just one missing measure on reading-related tests, the value was prorated from other reading measures. 14 children had more missing data than this, and were excluded from the current analyses. This included 7 five-year-olds, for whom standard scores were not computable as they fell below the age range of norms. All but one case of missing data in children aged over 5 years came from the High Bias group, reflecting refusal or inability to attempt some tests. Norms for the CCC-2 extend across the full age range covered here, but 19 children had missing data because parents either did not complete the CCC-2 ( $N = 12$ ), or because they failed the consistency check ( $N = 7$ ), which is suggestive of invalid responses.

The numbers included in different analyses are shown in the flowchart in [Figure 1](#). When comparing test profiles of trisomy cases with those of twins with language concerns (see below), we restricted consideration to children covering the same age range, i.e. 6 yr 0 months to 11 yr 11 months.

### Comparison group

Our second question involved comparing the language profile of children with trisomies with that of children with language disorders of unknown cause. The comparison sample consisted of twins recruited via fliers sent to primary schools around the UK, advertisements on our group's website and via twins' clubs. The age range for this sample was narrower than for the SCT cases. We aimed to recruit families with twin children aged between 6 years 0 months and 11 years 11 months, with over-representation of those where one or both twins had language or literacy problems. Further details are provided in Wilson and Bishop ([Wilson & Bishop, 2018](#))

Usually, language disorder would be diagnosed on the basis of language test scores, but if we took that approach, the result would be a foregone conclusion, in that we would use the same measures to define the independent variable (language status) and the dependent variables. To avoid this circularity, twin children were subdivided according to parental concern about language and history of speech and language therapy (SALT), rather

than by language test scores, which were treated as dependent variables. Parental concern was coded from the initial interview, and used to divide the twin sample into a Language Concerns group with ongoing parental concerns about oral language (mild or severe) and the remainder (which included some cases where there had been transient concerns in preschool that had resolved, or where the concern affected only reading). The latter is referred to as the No Concerns group. In addition, children who had received speech and language therapy after the age of 4 years were included in the Language Concerns group.

To be included in the twin sample, neither twin could have a pre-existing diagnosis of autism spectrum disorder (ASD), or a serious long-term illness. In addition, children were excluded if an ASD diagnosis was confirmed after the study was completed ( $N = 1$ ), if the child had a performance IQ score below 70 ( $N = 1$ ) or failed a hearing screen ( $N = 4$ ). We followed the CATALISE criteria for DLD ([Bishop \*et al.\*, 2017](#)), which meant that other diagnoses, such as dyslexia, attentional deficit hyperactivity disorder (ADHD), or dyspraxia, were not grounds for exclusion.

One twin from each pair was selected at random to avoid dependencies in the data. The final sample consisted of 57 children in the language concerns group, and 118 children with no concerns. The latter group were used to derive a normative range against which to evaluate the other groups on language measures. [Figure 2](#) shows the number of twin children included in the study, and the number for whom useable CCC-2 data were available.

## Methods

### Language, literacy and cognitive assessments

The assessment battery administered to the child is shown in [Table 1](#).

Parents were asked to complete the Children's Communication Checklist – 2 ([Bishop, 2003](#)) and Social Responsiveness Scales ([Constantino, 2005](#)) and return them by mail; the response rate was 86% for the trisomy families and 83% for the twin families. The SRS data and results from an online interview covering behavioural and psychiatric characteristics will be described in a companion paper.

Results from the assessments in [Table 1](#) were converted to age-scaled scores on a common scale with mean 100 and SD 15. All tests had published norms covering the age range of the twin sample (6 to 11 years), but some did not extend outside this range. As explained in [Newbury \*et al.\* \(2018\)](#), where feasible, norms were extrapolated based on data from other samples encompassing the age range (see Appendix 2 on Open Science Framework project: <https://osf.io/ae8yn/>). For Oromotor Skills, however, extrapolated norms gave scaled scores that were well below the range of other tests. For the current analyses, therefore the No Concerns twin group was used to derive norms, using the regression of total correct on age to compute a standardized residual, which was then scaled to mean 100 and SD 15.



**Table 1.** Psychometric assessment battery.Page 7 of 29

Our original intention was to use MANOVAs to test group differences that related to our research questions, but preliminary analysis showed that the MANOVA assumption of multivariate normality was not met when all 14 language/cognitive measures were entered into the analysis. Accordingly, we created four composites for the psychometric tests, by averaging scores as follows: Nonverbal ability (Matrices and Block Design), Core language (Vocabulary and Woodcock-Johnson Comprehension), Verbal production/memory (Sentence repetition, Nonword repetition, Oromotor sequences) and Literacy skills (all subtests from Neale Analysis of Reading Ability, Test of Word Reading Efficiency, and PhAB Rapid Naming). These composites mostly met criteria for multivariate normality within subgroups. Distributions of scores on the component tests for the different groups are shown in Supplementary Material, [Figure S1](#). Multivariate normality was also an issue for the CCC-2 test data, so composites were formed based on three sets of scales, identified as Structural Language (scales A-C), Pragmatics (scales D-G), and Autistic Features (scales H-J). The means and SEs for the original 10 subscales are shown in Supplementary Material, [Figure S2](#).

### Covariates

Two measures of socio-economic background that have been associated with language status were included in the analysis. These were:

- a) Educational level of mother, transformed into an ordinal scale based on age at leaving full-time education/qualifications obtained, with points of 0 (prior to age 16 years), 1 (16 years/did GCSE or O-levels), 2 (18 years/ did A-levels), 3 (21 years, degree), 4 (postgraduate study).
- b) An index of multiple deprivation based on postcode was obtained for those living in England from the website <http://imd-by-postcode.opendatacommunities.org/>. This uses local statistics from the Department for Communities and Local Government to rank 32,844 postcodes on the basis of a weighted sum based on income, employment, education, health, crime, housing and living environment. The rank score was converted to a z-score to give a normally-distributed variable, termed Neighbourhood Advantage index, by dividing by 32,844, before applying the qnorm function in R. A Neighbourhood Advantage index of zero (i.e. average) was assigned to 13 SCT cases and 4 twin pairs from Wales, Scotland and Northern Ireland, where postcode rankings were not available.

In addition, following a suggestion by [Boada et al. \(2009\)](#), we considered whether report of a positive family history of language problems was related to the language phenotype in the child. In an initial telephone interview with a parent, details of the mother, father and any siblings were recorded, and the informant was asked if each relative had any problems with hearing, speech, language or reading, and asked to elaborate if so. Family history was coded as positive if a relative was recorded as having definite evidence of language or literacy problems that had led to them requiring speech-language therapy or additional support at school. A three-point scale was used, with score of 0 for no

family history, 1 for one affected relative, and 2 for two or more affected relatives.

### Procedure

Children were seen for an individual assessment at home or school. The language, reading and nonverbal ability tests ([Table 1](#)) were given in an initial session lasting around 90 minutes, followed by an assessment of laterality, the results of which are described elsewhere ([Wilson & Bishop, 2018](#)).

### Analysis

[Table 2](#) shows the characteristics of the trisomy and twin groups on age, covariates and family history, for children included in the analysis of language test data.

### Statistical comparisons

For each type of variable (Psychometric tests and parent report on CCC-2) three comparisons were conducted, corresponding to the three research questions, A) Comparison of the three trisomies for Low Bias group only; B) Comparison of all Low Bias trisomy cases aged 6-11 yr with the Language Concerns twin comparison group; C) Comparison of Low Bias and High Bias trisomy groups. A Bonferroni-corrected alpha level of  $.05/6 = .008$  was adopted.

### Psychometric tests

#### A. Comparison of three trisomy groups (Low Bias group only).

As a first step, data from the psychometric composites were plotted to see the range and distribution of scores for the three Low Bias trisomy groups, as shown in the red beeswarm plots in [Figure 3](#). (These plots do not show scores for children who were excluded because of intellectual disability or ASD, but note that scores for those children are shown below in [Figure 4](#)). The dotted horizontal bars show the mean for each group. The yellow shaded area shows the range covered by mean  $\pm 1$  SD for the No Concerns twins. The distribution of blue points, corresponding to the Language Concerns twins will be discussed below (analysis B).

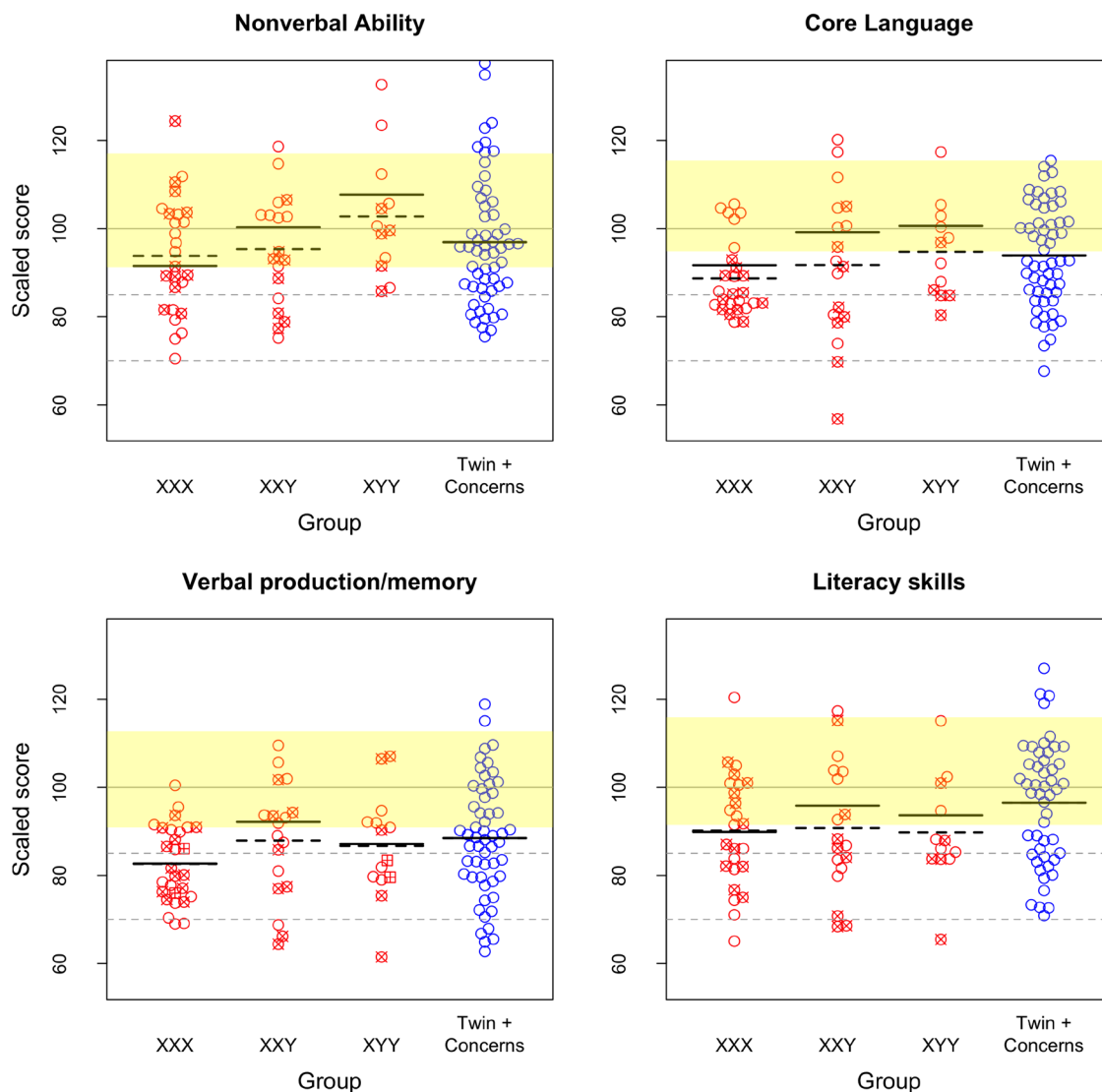
An initial impression from these plots is that there is variation from test to test in severity of impairment, but on all measures, the most striking feature is a wide spread of scores, with some children scoring above average and others well below. Although the mean test scores vary from trisomy to trisomy, the within-group variation is much greater than between group variation. This pattern is also evident in data from the individual tests making up the composites (Supplementary Material, [Figure S1](#)).

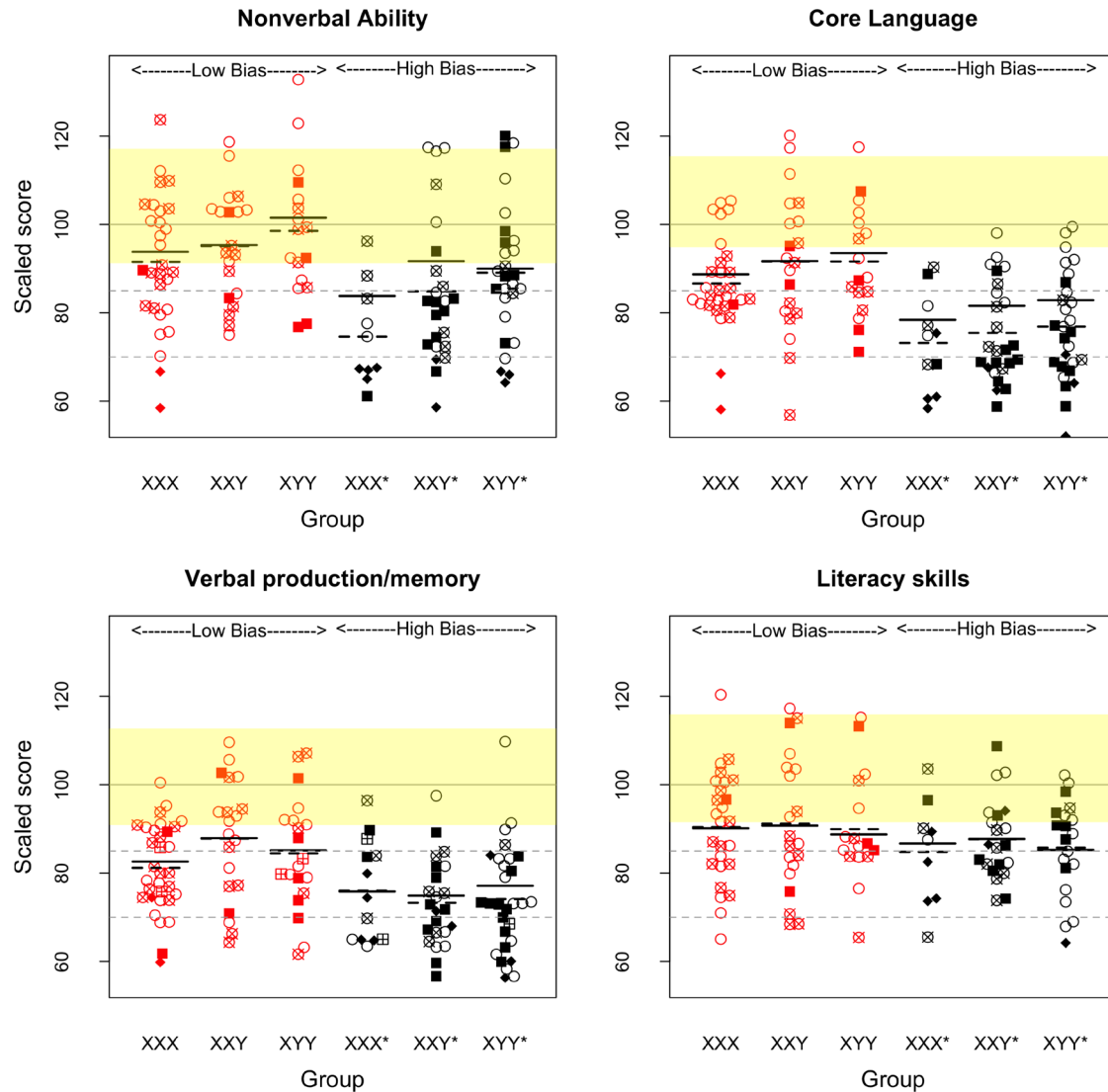
MANOVA was used to test whether there were reliable differences between the three trisomy groups on the first three composite measures, after taking covariates (Mother's educational level and Neighbourhood Advantage index) into account. These tests are based on the means shown in dotted lines, including all ages. No effect of trisomy type was found (see [Table 3](#), row A). In addition, to test the prediction by [Skuse \(2018\)](#) of smaller variance in the XYY group compared with the groups with an extra X chromosome, F-tests were conducted, but on no measure was there any support (all p-values above .05).



**Table 2.** Characteristics of the trisomy and twin groups on background variables: age, mother's educational level and Neighbourhood Advantage index.

Group	N	age (mo)	Mo. educ.	Neighbourhood Advantage	Family history
XXX: Low Bias	28	136 (46.7)	2 (0.96)	0.6 (0.82)	0.5 (0.7)
XXY: Low Bias	18	140 (43.4)	1.9 (1.06)	0.5 (0.86)	0.4 (0.61)
XYX: Low Bias	14	123 (42.4)	2.2 (0.89)	0.4 (0.77)	0.4 (0.51)
XXX: High Bias	7	117 (53.5)	1.4 (0.53)	-0.3 (1.14)	0.7 (0.49)
XXY: High Bias	13	132 (41.2)	1.8 (1.07)	0.6 (0.79)	0.6 (0.77)
XYX: High Bias	17	112 (33.5)	1.6 (1)	-0.2 (0.86)	0.6 (0.71)
Twin: No concerns	118	106 (19.3)	2.4 (0.93)	0.6 (0.95)	0.4 (0.64)
Twin: Language concerns	57	108 (18.9)	2.1 (0.9)	0.4 (1.17)	0.5 (0.68)

**Figure 3.** Distributions of scores on four clusters of psychometric tests for the Low Bias trisomy groups and the Language Concerns Comparison group. Open circles show cases in age range 6 to 11 yr. Solid line is mean for 6–11 yr olds, dashed line is mean for whole sample including those outside 6–11 yr age range. Yellow band is mean +/- 1 SD for No Concern comparison group.



**Figure 4.** Distributions of scores on four clusters of psychometric tests for the Low Bias vs High Bias trisomy groups. Unfilled circles show cases in age range 6 to 11 yr, unfilled circles with a cross show older children, unfilled squares with cross show 5-year-olds. Cases with ASD (filled squares) and intellectual disability (filled diamonds) are included in the plot, but were excluded from the analysis comparing bias groups. Asterisk denotes High Bias group. Solid lines are group means for children included in MANOVA; dashed lines show means with ASD/ID cases included. Yellow band is mean  $\pm$  1 SD for No Concern comparison group.

**Table 3.** Results 3 MANOVAs (A, B and C) testing for overall group differences on three composites from psychometric tests (Nonverbal Ability, Core Language and Verbal Production/Memory), with covariates of mother's education and Neighbourhood Advantage index.

Comparison	Effect	Wilks $\lambda$	F	df1	df2	p
A. Low Bias: XXX (26) vs XXY (18) vs XYY (12)	Group	0.885	1.030	6	98	0.409
	Mo. education	0.912	1.580	3	49	0.207
	Neighbourhood Advantage	0.890	2.010	3	49	0.124
B. Trisomies (31) vs. Lang concern twins (55): aged 6–11	Group	0.967	0.900	3	79	0.446
	Mo. education	0.802	6.490	3	79	0.001
	Neighbourhood Advantage	0.990	0.270	3	79	0.847
C. Low Bias (56) vs. High Bias (31)	Group	0.814	6.170	3	81	0.001
	Mo. education	0.940	1.710	3	81	0.171
	Neighbourhood Advantage	0.953	1.330	3	81	0.270

The literacy skills score had more missing data than other scores, and so was analysed separately using ANOVA, again revealing no effect of trisomy type,  $F(1, 53) = 0.001$ ,  $p = 0.97$ .

**B. Comparison of combined trisomy cases aged 6–11 yr (Low Bias) with Language Concern twin comparison group.** For this analysis, we combined all trisomy Low Bias cases aged 6–11 years (open circles in [Figure 3](#)) to compare scores on the first three psychometric composites with the twin comparison group with language concerns. [Table 3](#), row B, shows the MANOVA result for this comparison, which concerns the means shown as solid lines. Again there was no overall effect of group. The only factor affecting scores was mother's education, with each additional point on the scale associated with an increase of 4.40 points in average score on psychometric tests.

An ANOVA on the Literacy composite revealed that the Trisomy and Language Concerns groups were comparable in levels of impairment,  $F(3, 74) = 0.74$ ,  $p = 0.53$ .

**C. Comparison of effect of bias group: trisomy cases only.** [Figure 4](#) shows the distributions of scores on the four composites for children in the two ascertainment risk groups. For this analysis, we considered the effect of bias group, combining across trisomy type, again entering the first three composites into a MANOVA. As shown in [Table 3](#), row C, there was a strong effect of ascertainment bias, with the High Bias group having lower test scores. It is evident from inspection of [Figure 4](#) that this effect would have been larger still if we had included the cases with ASD and intellectual disability, who were more common in the High Bias group and tended to have low test scores.

The same dataset was used to explore the relationship with family history of language problems, by running linear regressions for each of the four composite measures, with the family history score as a predictor. In no case did the association prove statistically robust (all  $p$ -values  $> .05$ , maximum proportion variance explained = 0.03).

### Children's Communication Checklist - 2

Completion rates for the CCC-2 showed some variation by group: the checklist was completed by parents of 94.40% of the Low Bias and 77.80% of the High Bias trisomy group, and by 81.90% of the parents of No Concern twins and 76.40% of parents of Language Concern twins. Data was not useable for 5.10% of checklists which failed the consistency check: this is a criterion based on comparing scores for items describing strength and difficulties: if raw means are similar for both, this suggests that the respondent has not appreciated the need to change the polarity of responses between the two sets.

To establish potential bias introduced by those with missing or unusable data, we conducted a logistic regression analysis, with CCC-2 data coded as 1 (useable data) vs 0 (missing or inconsistent), and with maternal educational level, single parent status, trisomy status, and a composite measure of the child's language status (language factor from [Newbury et al., 2018](#)) as predictors. Availability of CCC-2 data not predicted by parental

characteristics (mother's education or single parent status), but was predicted by whether the child came from the twin or trisomy group, and by severity of language problems. The lower response rate by parents of twins might be explained by the fact that checklist completion was more onerous for them, as they were asked to complete a checklist for each twin. As shown in [Supplementary Table S1](#) and [Supplementary Figure S3](#), the impact of child's language status reflected that CCC-2 was less likely to be completed when the child had more severe language difficulties. Thus the CCC-2 results may underestimate the extent of communication difficulties in both samples.

Mean scores on the three CCC-2 composites are shown for the Low Bias and Language Concerns groups in [Figure 5](#). As expected, the No Concerns comparison group (range  $\pm 1$  SD shown as yellow shading) has mean scores close to the normative mean of 10.

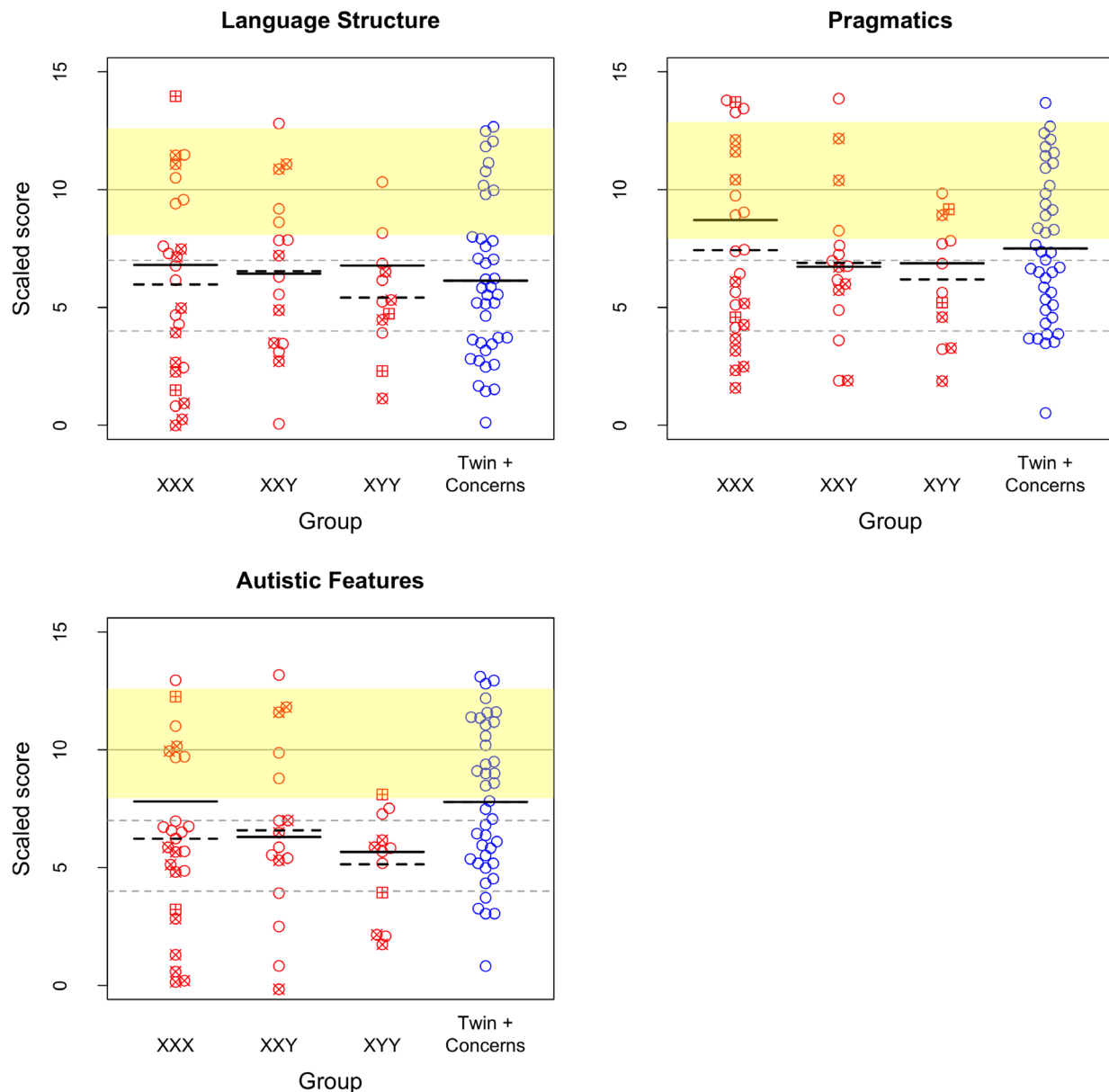
[Figure 6](#) contrasts CCC-2 composites for the Low Bias and High Bias trisomy groups. Here, the means are shown with and without cases of ASD and low IQ included: the MANOVA excluded these cases, focusing on differences in the means shown as solid lines. As with the psychometric tests, the scores were substantially lower for the High Bias groups, although there was a wide range in all three trisomies.

MANOVAs were conducted in the same way as for the psychometric test composites, again showing no effect of trisomy type within the Low Bias sample ([Table 4](#), row A), no difference between the combined trisomy group and the Language Concern twin comparison group (row B), but a substantial impairment in High Bias trisomy cases relative to the Low Bias cases (row C). For the Low Bias groups, the plot suggested there might be lower variance for the XYY children, in line with prediction by [Skuse \(2018\)](#), but  $p$ -values for F-tests comparing the variance of the XYY cases with the combined XXX and XXY cases were greater than .05 for all composites.

### Number of language tests below cutoff

In a final analysis, we focused on rates of impairment rather than mean scores. There have been various attempts to operationalise diagnostic criteria for language disorder, e.g. ([Tomblin et al., 1996](#)). Although an overall language test composite can be used for this purpose, this may miss cases who have an uneven profile, with striking deficits in just a few aspects of language function. To get a better impression of the nature of language deficits in the trisomy and comparison groups, we selected five language measures: Vocabulary, Comprehension, Sentence Repetition, Nonword Repetition, and Oromotor Sequences, and categorised each case according to whether their score was more than 1 SD below the population mean (i.e. 85 or less on the rescaled scores). This analysis did not include the cases meeting exclusionary criteria. The percentages of children in each group who scored this low on between 0 and 5 tests is shown in [Table 5](#).

Another indicator of language impairment is having a General Communication Composite (GCC) on the CCC-2 of 55 or less.



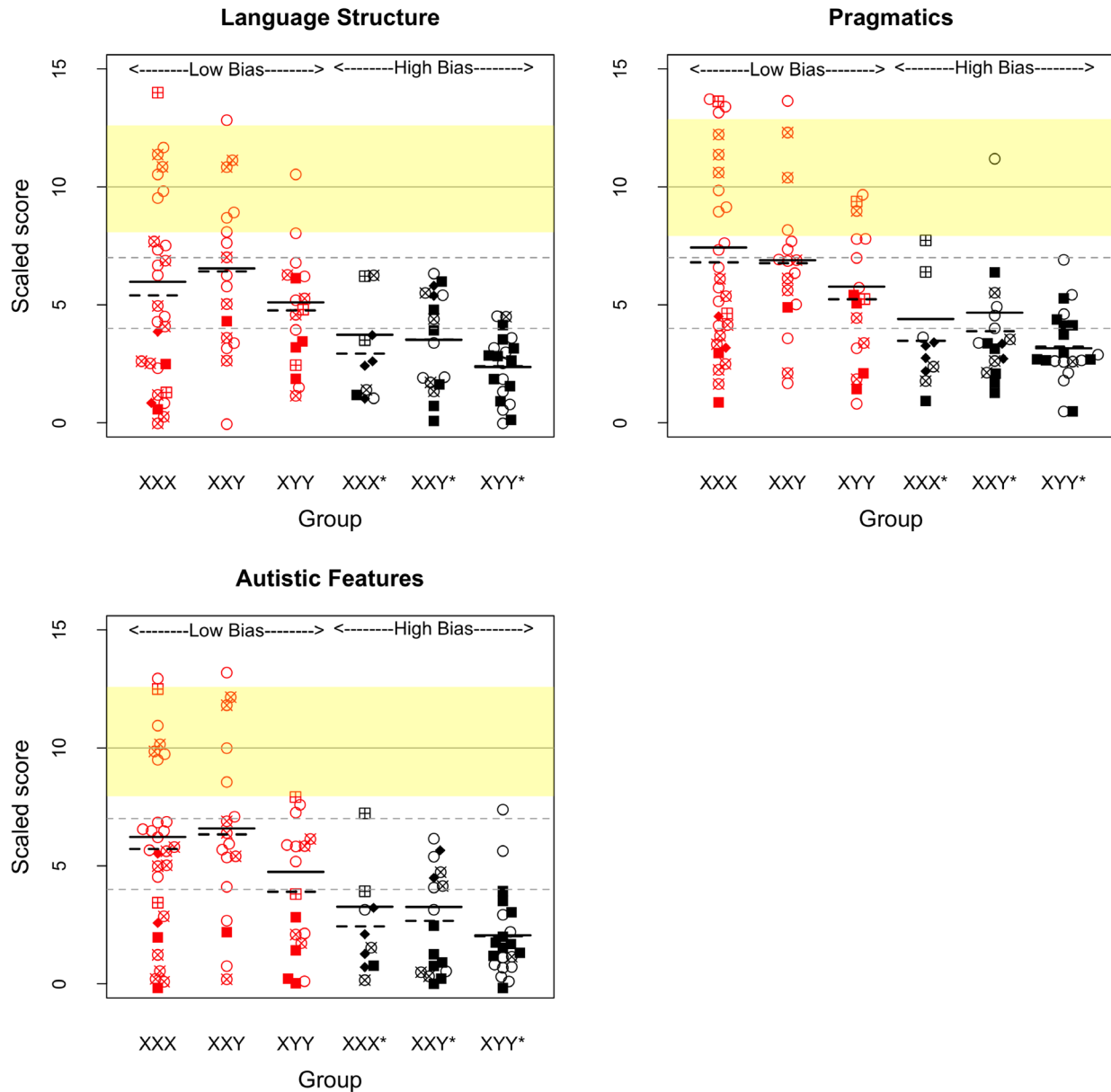
**Figure 5. Mean scores on the three CCC-2 composites for trisomy and comparison groups.** The yellow shaded band shows mean  $\pm$ 1 SD for the No Concerns twin group. For trisomies, dashed lines show mean for whole group, and solid lines show means for children aged 6 to 11 years.

The percentages of twins rated this low was 5.1% for the No Concerns group and 40.4% for the Language Concerns group. Among children with sex chromosome trisomies, 56.7% of the Low Bias group and 73.0% of the High Bias group were rated this poorly.

## Discussion

This study confirmed that there is a high rate of language problems among children with sex chromosome trisomies of all three kinds. In this relatively small sample, there were no consistent differences in the severity or profile of language problems between

those with XXX, XXY and XYY karyotypes. In their study of neonatally identified cases, [Bender \*et al.\* \(2001\)](#) noted that females with XXX ( $N = 10$ ) were more impaired than males with XXY ( $N = 11$ ), but the conclusion was not based on direct comparison of the two groups and the sample size was small. An unanticipated finding from our study was that girls with XXX were more likely than the XXY and XYY boys to come from the Low Bias group, suggesting that fewer of them had problems that brought them to clinical attention. [Ross \*et al.\* \(2009\)](#) found that boys with XYY had more pervasive and severe language problems than boys with XXY karyotype, but they noted that



**Figure 6.** Mean scores on the three CCC-2 composites for Low Bias vs High Bias trisomy groups, with the latter denoted by asterisks. For trisomies, solid lines show mean for children included in MANOVA, and dashed lines show means for with cases of ASD or low IQ included. The yellow shaded band shows mean +/- 1 SD for the No Concerns twin group.

**Table 4.** Results for three MANOVAs comparing groups on the CCC-2 composite scores.

Comparison	Effect	Wilks $\lambda$	F	df1	df2	p
A. XXX (29), XXY (17), XYY (17)	Group	0.862	1.260	6	98	0.282
	Mo. education	0.914	1.530	3	49	0.219
	Neighbourhood Advantage	0.963	0.630	3	49	0.602
B. All SCT vs Lang concern twins	Group	0.869	3.210	3	64	0.029
	Mo. education	0.867	3.260	3	64	0.027
	Neighbourhood Advantage	0.928	1.660	3	64	0.185
C. Bias	Bias	0.751	10.370	3	94	0.000
	Mo. education	0.980	0.650	3	94	0.583
	Neighbourhood Advantage	0.918	2.800	3	94	0.044

**Table 5. Percentages of children meeting criteria for impairment (score < 86) on a given number of tests (T0 to T5).**

Group	N	T0	T1	T2	T3	T4	T5
Comparison							
No concerns	118	60.2	26.3	7.6	5.9	0	0
Language concerns	57	22.8	24.6	14	19.3	8.8	10.5
Trisomy: Low Bias							
XXX	28	10.7	14.3	14.3	28.6	28.6	3.6
XXY	18	27.8	16.7	16.7	22.2	0	16.7
XYX	14	0	42.9	21.4	21.4	14.3	0
Trisomy: High Bias							
XXX	7	0	14.3	14.3	14.3	14.3	42.9
XXY	13	0	7.7	7.7	15.4	38.5	30.8
XYX	17	0	17.6	11.8	11.8	35.3	23.5

the former group included more postnatally identified cases, so ascertainment bias could affect findings. Although it is possible that karyotype differences would emerge with a larger sample, a striking feature of our data, and those of [Bender et al. \(2001\)](#), is the wide range of variation within each type of trisomy, which swamps any trisomy-specific effect. Indeed, for children selected in a manner that reduced ascertainment bias, we found that around one third of cases scored in the same range as a comparison group without problems, in terms of having at most one low score on a set of five language tests. A similar proportion scored within the normal range on the overall index from parental report, i.e. the General Communication Composite of the Children's Communication Checklist-2. Furthermore, for ethical reasons, we were only able to study children who knew about their sex chromosome trisomy: we know from our previous study that these tend to be children with more severe problems, where the trisomy may be disclosed in order to help the child understand about their difficulties ([Gratton et al., 2016](#)). It is therefore likely that even our Low Bias group may overestimate the extent of language difficulties in childhood, because those without any difficulties would be less likely to take part. On the other hand, the data reported here excluded 12 children from the Low Bias sample because of intellectual disability, ASD, or hearing problems, and so it needs to be borne in mind that had these cases been included, an even wider range of scores would have been observed. [Skuse \(2018\)](#) suggested that we might expect to see more variable expression of genes from an additional X chromosome than from the Y chromosome, in which case the language phenotype might have a narrower range in the XYX group; this prediction was not supported. Note, however, that detection of reliable differences in variance between groups would require substantial sample sizes. Even for detection of mean group differences, our sample size was too small to detect any

but large effects of karyotype: for comparisons of XXX+XXY vs XYY we had 80% power to detect an effect size on psychometric composites of .83, and on CCC-2 an effect size of .97. We cannot therefore conclude from the current study that there are no sex-chromosome-specific influences, only that if they do exist, they are not large, and are superimposed on substantial variation from other sources. Because of the difficulty of recruiting large samples, and the difficulty of controlling sampling biases, we will need to combine samples across different research groups to get a clear answer to the question of whether an additional X or Y has consistently different effects.

The twin sample was subdivided to provide a way of assessing how far the language problems in those with sex chromosome trisomies resembled those of children who had language problems in the absence of any known neurobiological condition. Allocation to the language concerns group was made purely on the basis of parental report: either there was ongoing concern about the child's language skills and/or that the child had speech and language therapy after the age of 4 years. As can be seen in [Figure 3](#) and [Figure 5](#), although the overall means were below average, many children selected this way did not have obvious language problems on the test battery used here. This mismatch between parental report and test performance is reminiscent of findings by [Broomfield & Dodd \(2004\)](#) who found that around 10% of children on the caseload of a speech-language therapy service had normal range performance on language assessment. This suggests that parents may be concerned about relatively minor problems that are not clinically important, or that are transient and have resolved by the time we assessed the child. It could also be that the battery used here was not sensitive to the kinds of difficulties that children experienced. For instance, some children receive speech and language therapy for problems with articulation, voice or fluency, and those problems would not necessarily be detected on our assessment battery. In the sample of [Broomfield & Dodd \(2004\)](#) speech difficulties were the most common type of problem.

Insofar as it was possible to compare the Language Concerns group with the trisomy cases, the profile and severity of their problems appeared similar. Thus, we did not detect any distinct phenotypic signature of a sex chromosome trisomy.

When we turn to consider children whose trisomy was discovered in the course of investigation for neurocognitive or behavioural difficulties (i.e. the High Bias group), we find a substantially higher rate of language problems. Furthermore, 33 children were excluded from the High Bias sample because of intellectual disability, ASD or hearing problems, and a further 3 of the remaining 37 children from this group were excluded from analysis because of missing language test data due to refusal or inability. Thus the language test scores shown here apply only to those who could complete the test battery and did not have additional problems. These results agree with those of [Wigby et al. \(2016\)](#), who found substantial differences in neurodevelopmental outcomes for girls with trisomy X, depending on whether the trisomy was identified prenatally or postnatally.



It is not surprising to find that children whose trisomy was identified in the course of investigations for developmental disorders should have high rates of problems, but this result reinforces the message that, when advising parents of likely outcomes of children in whom a sex chromosome trisomy is adventitiously discovered, it is important to focus on results from samples that minimise this kind of ascertainment bias (Linden *et al.*, 2002).

There are several possible factors that might account for the range of variation in phenotypes seen in sex chromosome trisomies. One possibility is that the impact of a sex chromosome trisomy might be influenced by environmental background. This idea has been raised by Bender *et al.* (1987). We included measures of environmental background - maternal years of education and neighbourhood advantage - in our analyses, and found evidence of an impact of maternal education on psychometric scores in one analysis (see Table 4, Analysis B), though this effect was not evident in an analysis that included only children with a sex chromosome trisomy. In a review of Klinefelter syndrome, Boada *et al.* (2009) proposed that family history of language problems might affect severity of language phenotypes, but we found no evidence for this in the current sample.

Elsewhere we have proposed that genetic variants on autosomes may interact with neurodevelopmental consequences of extra gene product from sex chromosomes, so there is amplification of impact from variants that usually have only a mild effect. To date we have not been able to identify such effects (Newbury *et al.*, 2018). This does not, of course, rule out the possibility that a mechanism of this kind does operate, but involving different genes. In future work, we plan to broaden the scope of our search for such genetic mechanisms. It is worth noting, however, the suggestion by Beach *et al.* (2017), that aneuploidy itself may be a cause of phenotypic heterogeneity. Most of their evidence comes from studies in yeast, where they found high variability in cell cycle progression among cells with identical aneuploidies, as well as variable response to environmental stress. They argued this tendency for gain or loss of a chromosome to cause genetic instability and increased variability in the phenotype may extend to mammals. They found wide variation in inbred mice with genetically engineered trisomy 19, despite a uniform genetic background and environment. If this applies to human sex chromosome trisomies, then a search for genetic or environmental correlates of phenotypic variation could prove fruitless.

Where a sex chromosome trisomy is discovered on prenatal screening, parents will be anxious to know the implications for the child's development. The results reported here illustrate the very wide range of outcomes that can be seen in children with an extra X or Y chromosome. Some children in our sample had no measurable language difficulties, while others were severely impaired. In interpreting these findings, it is also important to bear in mind that, even if we consider just the Low Bias group,

we may over-estimate the rate of impairment, if parents are more motivated to take part if their child has problems, and/or if the child's eligibility for the study (in terms of knowing about the trisomy) is affected by level of impairment. These considerations make advising parents challenging, as prediction of individual outcomes is impossible: the most that can be said in our current state of knowledge is that there is a clear risk that the child will have language problems that may interfere with daily life, social interaction and progress in school, but that such problems are by no means inevitable. Most children attended mainstream school and it was not uncommon to find language skills in the normal range. Where children had language deficits not accompanied by ASD or intellectual disability, they tended to be of a range and severity similar to those seen in other children who have developmental language disorder (i.e. in the absence of any known biological aetiology). This reinforces the advice given by Linden *et al.* (2002) that children with sex chromosome trisomies do not need special interventions, as their difficulties with speech, language and learning are similar to those of other children with typical chromosomes.

### Data availability

Data and analysis scripts are available on Open Science Framework: <https://osf.io/u2c7d/>, DOI: <https://doi.org/10.17605/OSF.IO/U2C7D> (Bishop, 2018). Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

### Grant information

This work was funded by Wellcome Trust Programme Grant 082498.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

### Acknowledgements

We offer warmest thanks to the families who took part in the study, and school staff who helped facilitate assessment arrangements. We greatly appreciate the help we had with recruitment from Unique (the Rare Chromosome Disorder Support Group), the Klinefelter's Syndrome Association, and staff from Clinical Genetics Centres in the following NHS Trusts: Royal Devon and Exeter, The Northwest London Hospitals, University Hospitals Bristol, Sheffield Children's, Liverpool Women's, University Hospital Southampton, Birmingham Women's Hospital, Great Ormond Street Hospital for Children, Leeds Teaching Hospitals, Newcastle upon Tyne Hospitals, Oxford University Hospitals, Central Manchester University Hospitals, University Hospitals of Leicester, Nottingham University Hospitals, Belfast Health and Social Care Trust and the NIHR Clinical Research Network. We extend special thanks to Dr Gillian Borthwick, National Research Coordinator NIHR CRN: Genetics, for helping us navigate the complexities of working with multiple NHS sites under the NIHR UK Rare Genetic Disease Research Consortium Agreement.

## Supplementary material

- Distributions of scores on individual psychometric tests (Figure S1)
- Means for CCC-2 scales (Figure S2)
- Prediction of CCC-2 checklist completion from child and parent variables (Figure S3 and Table S1)

[Click here to access the data](#)

## References

- Beach RR, Ricci-Tam C, Brennan CM, *et al.*: **Aneuploidy causes non-genetic individuality.** *Cell.* 2017; **169**(2): 229–242.e21.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bender B, Fry E, Pennington B, *et al.*: **Speech and language development in 41 children with sex chromosome anomalies.** *Pediatrics.* 1983; **71**(2): 262–267.  
[PubMed Abstract](#)
- Bender BG, Linden MG, Harmon RJ: **Neuropsychological and functional cognitive skills of 35 unselected adults with sex chromosome abnormalities.** *Am J Med Genet.* 2001; **102**(4): 309–313.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bender BG, Linden MG, Robinson A: **Environment and developmental risk in children with sex chromosome abnormalities.** *J Am Acad Child Adolesc Psychiatry.* 1987; **26**(4): 499–503.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bishop DVM: **Language and behavioural phenotypes in children with sex chromosome trisomies.** 2018.  
<http://www.doi.org/10.17605/OSF.IO/U2C7D>
- Bishop DVM: **The Children's Communication Checklist, version 2 (CCC-2).** London: Pearson. 2003.
- Bishop DV, Jacobs PA, Lachlan K, *et al.*: **Autism, language and communication in children with sex chromosome abnormalities.** *Arch Dis Child.* 2011; **96**(10): 954–959.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bishop DV, Scerif G: **Klinefelter syndrome as a window on the aetiology of language and communication impairments in children: the neurotrophin-neurexin hypothesis.** *Acta Paediatr.* 2011; **100**(6): 903–907.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bishop DVM, Snowling MJ, Thompson PA, *et al.*: **Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology.** *J Child Psychol Psychiatry.* 2017; **58**(10): 1068–1080.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Boada R, Janusz J, Hutaff-Lee C, *et al.*: **The cognitive phenotype in Klinefelter syndrome: a review of the literature including genetic and hormonal factors.** *Dev Disabil Res Rev.* 2009; **15**(4): 284–294.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Broomfield J, Dodd B: **Children with speech and language disability: caseload characteristics.** *Int J Lang Commun Disord.* 2004; **39**(3): 303–324.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Constantino JN: **The Social Responsiveness Scale.** Los Angeles: Western Psychological Services. 2005.
- Cuckle H, Maymon R: **Development of prenatal screening--A historical overview.** *Semin Perinatol.* 2016; **40**(1): 12–22.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Frederickson N, Frith U, Reason R: **Phonological assessment battery (PhAB).** Windsor: NFER-Nelson. 1997.  
[Reference Source](#)
- Gratton NC, Myring J, Middlemiss P, *et al.*: **Children with sex chromosome trisomies: parental disclosure of genetic status.** *Eur J Hum Genet.* 2016; **24**(5): 638–44.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Korkman M, Kirk U, Kemp SI: **NEPSY: A developmental neuropsychological assessment.** San Antonio: Psychological Corporation. 1998.  
[Reference Source](#)
- Leggett V, Jacobs P, Nation K, *et al.*: **Neurocognitive outcomes of individuals with a sex chromosome trisomy: XXX, XYY, or XXY: a systematic review.** *Dev Med Child Neurol.* 2010; **52**(2): 119–129.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Linden MG, Bender BG, Robinson A: **Genetic counseling for sex chromosome abnormalities.** *Am J Med Genet.* 2002; **110**(1): 3–10.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Neale MD: **Neale Analysis of Reading Ability Second Revised British Edition.** London: GL Assessment. 1999.
- Newbury DF, Simpson NH, Thompson PA, *et al.*: **Stage 2 Registered Report: Variation in neurodevelopmental outcomes in children with sex chromosome trisomies: testing the double hit hypothesis [version 1; referees: 2 approved].** *Wellcome Open Res.* 2018; **3**: 85.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Printzlau F, Wolstencroft J, Skuse DH: **Cognitive, behavioral, and neural consequences of sex chromosome aneuploidy.** *J Neurosci Res.* 2017; **95**(1–2): 311–319.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Robinson A, Puck M, Pennington B, *et al.*: **Abnormalities of the sex chromosomes: a prospective study on randomly identified newborns.** In A. Robinson, HA. Lubs, & D. Bergsma (Eds.) *Sex Chromosome Aneuploidy: prospective studies on children.* *Birth Defects Orig Artic Ser.* 1979; **15**(1): 203–241.  
[PubMed Abstract](#)
- Ross JL, Zeger MP, Kushner H, *et al.*: **An extra X or Y chromosome: contrasting the cognitive and motor phenotypes in childhood in boys with 47,XYY syndrome or 47, XXY Klinefelter syndrome.** *Dev Disabil Res Rev.* 2009; **15**(4): 309–317.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Skuse D: **Referee Report For: Stage 1 Registered Report: Variation in neurodevelopmental outcomes in children with sex chromosome trisomies: protocol for a test of the double hit hypothesis [version 1; referees: 1 approved, 2 approved with reservations].** *Wellcome Open Res.* 2018; **3**: 10.  
[Publisher Full Text](#)
- Tomblin JB, Records NL, Zhang X: **A system for the diagnosis of specific language impairment in kindergarten children.** *J Speech Hear Res.* 1996; **39**(6): 1284–1294.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Torgesen JK, Wagner R, Rashotte C: **Test of Word Reading Efficiency (TOWRE).** New York: Psychological Corporation. 1999.  
[Reference Source](#)
- Valentine GH: **The growth and development of six XYY children.** In A. Robinson, HA. Lubs, & D. Bergsma (Eds.): *Sex chromosomal aneuploidy: prospective studies on children.* *Birth Defects Orig Artic Ser.* New York: Alan R Liss. 1979; **15**(1): 175–190.  
[PubMed Abstract](#)
- Wechsler D: **Wechsler Abbreviated Scale of Intelligence.** San Antonio: Psychological Corporation. 1999.  
[Reference Source](#)
- Wigby K, D'Epagnier C, Howell S, *et al.*: **Expanding the phenotype of triple X syndrome: A comparison of prenatal versus postnatal diagnosis.** *Am J Med Genet A.* 2016; **170**(11): 2870–2881.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wilson AC, Bishop DVM: **Sex chromosome trisomies are not associated with atypical lateralization for language.** *Dev Med Child Neurol.* 2018; **60**(11): 1132–1139.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Woodcock RW, McGrew KS, Mather N: **Woodcock-Johnson III Tests of Cognitive Abilities.** Rolling Meadows, IL: Riverside Publishing. 2007.

# Open Peer Review

Current Referee Status:



Version 2

Referee Report 18 March 2019

<https://doi.org/10.21956/wellcomeopenres.16388.r34705>



**Nancy Raitano Lee** 

Department of Psychology, Drexel University, Philadelphia, PA, USA

I was delighted to review “Language phenotypes in children with sex chromosome trisomies” for Wellcome Open Research. Bishop and colleagues describe a carefully designed study of the language outcomes of children with sex chromosome trisomies (SCTs; 47,XXX, 47,XXY, 47,XYY) whose genetic diagnoses were ascertained either through (a) prenatal screening or other medical investigations (‘low bias group’) or (b) during the course of testing to identify the etiology of neurodevelopmental problems (‘high bias group’). The study sought to answer three questions. First, does the profile and severity of language difficulties experienced by children with SCTs differ across the trisomy variants? Second, does the language phenotype associated with SCTs differ from that observed in children without known chromosomal abnormalities whose parents expressed concerns about their language? Third, to what extent does ascertainment bias affect language and cognitive skills in youth with SCTs?

Findings can briefly be summarized as follows. First, the authors reported no statistically significant differences between SCT groups on language and cognitive testing. Second, the authors report similar language impairment severity and profiles for children with SCTs as compared to children from an existing twin sample whose parents expressed concerns about language development. Lastly, the authors noted that ascertainment bias appears to relate to the severity of language difficulties experienced by children with SCTs – that is, children from the high bias group showed greater language impairments than children from the low bias group.

This research has several noteworthy strengths. I will highlight three here. First, the authors took great care to classify participants based on how the child’s SCT was ascertained. Because SCTs are often not associated with frank dysmorphology, they can go undiagnosed<sup>1</sup>. Thus, a concern for studies examining language and learning problems in youth with SCTs is that they are only describing a subset of individuals whose diagnoses were discovered due to learning challenges. Second, the authors measured and covaried family background characteristics (mother’s education and multiple deprivation index) when evaluating study questions. As research suggests that family characteristics<sup>2</sup> and early life stress<sup>3</sup> relate to outcomes in children with SCTs, the inclusion of maternal education and adversity was an important methodological control implemented by these authors. Third, the authors included a comparison group of children with idiopathic developmental language concerns. In order to begin to understand whether children with SCTs and language impairments may benefit from similar interventions to those currently implemented for children with idiopathic language difficulties, direct comparisons of these groups are necessary.

My review of this methodologically rigorous research study also raised several questions that I hoped the authors could address in this or future research. I will briefly mention two of these. The first involves the question of ascertainment bias. While I wholeheartedly agree with the authors that the manner in which a child's SCT diagnosis is ascertained has the potential to bias study findings, I wonder if it is worth discussing the different types of bias involved with the low and high bias groups in greater detail. While the authors were careful to note that the low bias group was not free of bias, I had numerous questions about this group and whether they represent a less biased or rather differently biased group? In particular, I had questions about the subset of individuals in this group whose SCT was detected during prenatal testing. There appears to be multiple factors associated with who is offered and completes prenatal testing. The authors note that one factor is advanced maternal age. Based on this, I had the following questions. What were the ages of the mothers in the two groups? Were the mothers of the children in the low bias group older? Recent research suggests that older mothers experience greater (or perhaps more persistent) happiness after birth than younger mothers<sup>4</sup>. As there is research linking maternal mental health to child cognitive/linguistic development<sup>5</sup>, I wondered if the participants in the low bias group may have experienced advantages that could have influenced their language development favorably (above and beyond those that were controlled for statistically in the study's analyses). From my brief review of the literature, another factor that appears to relate to prenatal testing rates in the United Kingdom is ethnicity<sup>6</sup>. Based on this research, I hoped that the authors could address whether there was an even distribution of children from multilingual families across these groups? (I assume that this was the case, but I could not find specific mention of this in the Method). Lastly, I wondered about differences in the receipt of speech/language therapy services between the groups. Do the authors have information about any preventative or early speech-language therapy services that may have been sought for the participants in the low bias group who were diagnosed prenatally? I recognize that it is quite difficult to quantify the type and dose of treatment a child may have received retrospectively. However, I do wonder if some discussion of service provision for the different groups would be helpful to include.

My second set of questions relates to whether the authors have or will examine relations among different aspects of cognitive functioning (e.g., nonverbal IQ, verbal short-term memory) and language outcomes in the children with SCTs. In particular, I had questions about whether there were any differences in the strength of the relationships between different cognitive abilities and language skills in the youth with SCTs as compared to those with idiopathic language concerns. In recent work from my lab looking at autism spectrum disorder (ASD) symptoms in youth with idiopathic ASD and those with ASD within the context of Down syndrome (Godfrey et al., *in press*<sup>7</sup>), there was a suggestion (that we were not powered to evaluate statistically due to small sample sizes) of a differential relationship between verbal mental age and social communication impairments for children with ASD that was idiopathic in nature or occurred within the context of Down syndrome. Specifically, there appeared to be a tighter coupling between verbal mental age and social communication in children with Down syndrome and ASD than in the idiopathic ASD group. Have Bishop and colleagues examined relations between nonverbal IQ, verbal short-term memory, and language functioning in this sample? I raise this question, as (a) there may be differences in the cognitive underpinnings or correlates of language challenges in those with SCTs as compared to those with idiopathic language problems, and (b) differences in these relationships may be of use to consider when planning and implementing treatments for children with SCTs.

Having noted my questions for the authors, I'd like to close by thanking Dr. Bishop and colleagues for inviting me to review this work. I enjoyed reviewing this manuscript and I look forward to the papers to come on this unique sample of children with SCTs.

## References

1. Boyd PA, Loane M, Garne E, Khoshnood B, Dolk H, EUROCAT working group: Sex chromosome trisomies in Europe: prevalence, prenatal detection and outcome of pregnancy. *Eur J Hum Genet.* 2011; **19** (2): 231-4 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Bender BG, Harmon RJ, Linden MG, Robinson A: Psychosocial adaptation of 39 adolescents with sex chromosome abnormalities. *Pediatrics.* 1995; **96** (2 Pt 1): 302-8 [PubMed Abstract](#)
3. van Rijn S, Barneveld P, Descheemaeker MJ, Giltay J, Swaab H: The effect of early life stress on the cognitive phenotype of children with an extra X chromosome (47,XXY/47,XXX). *Child Neuropsychol.* 2018; **24** (2): 277-286 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Myrskylä M, Barclay K, Goisis A: Advantages of later motherhood. *Gynakologe.* 2017; **50** (10): 767-772 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Ahun MN, Côté SM: Maternal depressive symptoms and early childhood cognitive development: a review of putative environmental mediators. *Arch Womens Ment Health.* 2019; **22** (1): 15-24 [PubMed Abstract](#) | [Publisher Full Text](#)
6. Rowe RE, Garcia J, Davidson LL: Social and ethnic inequalities in the offer and uptake of prenatal screening and diagnosis in the UK: a systematic review. *Public Health.* 2004; **118** (3): 177-89 [PubMed Abstract](#) | [Publisher Full Text](#)
7. Godfrey M, Hepburn S, Fidler D, Tapera T, Zhang F, Rosenberg C, Lee NR: Autism spectrum disorder (ASD) symptom profiles of children with comorbid Down syndrome (DS) and ASD: A comparison with children with DS only and ASD only. *Research in Developmental Disabilities.* (in press).

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Cognition and brain development in neurogenetic syndromes

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Referee Report 14 February 2019

<https://doi.org/10.21956/wellcomeopenres.16388.r34707>



**Laura Zampini** 

Department of Psychology, University of Milano-Bicocca, Milan, Italy

It is a great pleasure for me to read and comment on this interesting and important paper. I recently started studying language development in Italian children with sex chromosome trisomies and the study of Bishop *et al.* is an excellent model for investigating this topic.

In the present study, Bishop *et al.* analysed language development in children with sex chromosome trisomies comparing their language profile with that of children with language disorders of unknown cause. After excluding children with autism, intellectual disability or hearing problems, the authors found that children with sex chromosome trisomies identified via prenatal screening had an increased risk of language problems, but severe language disorders were not common in this specific group.

The study has many strengths, among which the wide number of participants with sex chromosome trisomies ( $N = 142$ ). The methods are appropriate. One of the merits is the division of children into two groups: those identified by prenatal screening (Low bias) and those diagnosed during postnatal testing for neurodevelopmental problems (High bias). Controlling for children's socio-economic background is also of great value.

Moreover, the study has improved after the first revision, in particular with the inclusion of information about the excluded cases (i.e., children with autism and intellectual disability) in the beeswarm plots. Data analyses are detailed, and the results are clearly reported.

My only concern is about the conclusions drawn by the study. The argumentation of the results, in the abstract and the discussion, is in my opinion too optimistic for the families of these children. It is true that around one-third of the children identified by prenatal screening does not show language disorders. However, about 17% of children in the Low bias group have been excluded due to a developmental disorder (i.e., autism, intellectual disability, hearing impairment). Furthermore, a high percentage of children with a prenatal diagnosis scored more than 1 SD below normative data in at least one area of language development (only 10.7% of girls with XXX and 27.8% of children with XXY did not show any significant problem).

Moreover, I wonder if this optimistic conclusion could also be related to the participant's age.

In our recent study<sup>1</sup>, although with a very limited number of participants (i.e., 15 24-month-old children), we found that 60% of children with sex chromosome trisomies (75% of children with XXY and 43% of children with XXX) were at risk for language impairments (i.e., they had a vocabulary size lower than 50 words at 24 months). The current follow-up on these children (data collection in progress) is showing that at 4 years almost all the late-talking children with sex chromosome trisomies meet the criteria for a diagnosis of language impairment.

Is it possible that some of the participants in the present study (in particular the oldest ones) have partially compensated for their linguistic problems?

One important missing information is the number of children with sex chromosome trisomies in both groups (Low and High bias) that had (or have) speech and language therapy. How many children undergo rehabilitation treatments? For how long? These data should be reported since it is possible that the involvement in speech and language therapy could influence the children's performance on standardised tests.

## References

1. Zampini L, Draghi L, Silibello G, Dall'Ara F, Rigamonti C, Suttora C, Zanchi P, Salerni N, Lalatta F,



Vizziello P: Vocal and gestural productions of 24-month-old children with sex chromosome trisomies. *Int J Lang Commun Disord*. **53** (1): 171-181 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** language development; children with genetic syndromes

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 15 Mar 2019

**Dorothy Bishop**, University of Oxford, UK

Many thanks to Dr Zampini for her comments on our paper, and for drawing our attention to the study she and colleagues have done, looking at language development in children with trisomy X and Klinefelter's syndrome in children around 24 months of age. Although their sample is small in absolute terms, (7 girls with trisomy X, and 8 boys with Klinefelter syndrome), it is, I believe, unique in having data on a number of these children all studied at the same age, and restricting consideration to cases identified on prenatal screening. Use of a combination of data from parental report and direct observation is a considerable strength of the study, and the follow-up results described in her review above suggest that a parent vocabulary checklist could be very useful in identifying those children with a sex chromosome trisomy who may benefit from early speech-language intervention.

I was surprised, though, that our conclusions were thought to be over-optimistic, as my impression was that our results are not so inconsistent with those of Zampini et al. In both studies, there is a wide range of outcomes, with significant numbers of children having problems, but others within normal limits. Zampini et al report that 60% of their sample had a vocabulary size below the 5<sup>th</sup> percentile at 24 months of age, but 27% scored above the 25<sup>th</sup> percentile. Our concern was that, in reporting our findings, we should not focus only on cases with poor language, but rather that we

should indicate the range of outcomes – which includes some who are relatively unaffected. It is precisely this lack of certainty about the severity of impact that makes it difficult for genetic counsellors to advise parents, but it seems in such a situation the only way forward is to give families information about the full range of possible outcomes, while making them aware that the child is at increased risk of difficulties.

I have now added [Table 1](#), showing children's experience of speech-language therapy, as reported by parents, in the Low Risk cases. I have included the cases with autism or intellectual disability, who were more likely to have had had therapy. As with our previous 2011 study, rates of therapy are indeed higher than in a comparison group, but there are still many children who had no contact with speech and language therapy services.

**Competing Interests:** No competing interests were disclosed.

Referee Report 29 January 2019

<https://doi.org/10.21956/wellcomeopenres.16388.r34701>



**Kate Baker** 

MRC Cognition and Brain Sciences Unit, and Department of Medical Genetics, University of Cambridge, Cambridge, UK

No new comments. I concur entirely with the balanced view detailed within the response to my initial comments. I am sincerely grateful for the opportunities to learn via this review process.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Genomic disorders and cognitive development

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

## Version 1

Referee Report 27 November 2018

<https://doi.org/10.21956/wellcomeopenres.16247.r34202>



**Kate Baker** 

MRC Cognition and Brain Sciences Unit, and Department of Medical Genetics, University of Cambridge, Cambridge, UK

Investigating the associations between rare genomic variants and developmental cognitive impairments is of ever-growing importance. The number of genomic tests being carried out is increasing rapidly, at any point from early prenatal to adulthood. The proportion of genomic tests reported as abnormal is also increasing, as the catalogue of disorder-associated genes grows week by week and the analysis of

chromosomal variants and sequence variants improves.

After a genetic diagnosis, three questions usually arise: To what extent does this diagnosis explain the individual's past and present problems (taking into account other risk factors they may also be exposed to)? To what extent does this diagnosis predict the types and severities of future problems for this individual? To what extent should this genetic diagnosis be taken into account when planning interventions for this individual? In other words, how much difference can or should knowledge of genetic cause really make?

At present we lack the evidence base to answer these questions, to make genetic testing count for patients and their families, and to insure that genetic diagnosis is helpful not harmful. The limited literature is rife with claims of specific associations between rare genotypes and developmental profiles, claims which lack robustness in multiple respects – theoretically unsound, underpowered studies, highly biased samples, lack of informative comparison groups, inadequate phenotyping measures and exploratory statistical analysis. The field has not yet embraced open science.

This paper by Bishop *et al* is of high importance, for our understanding of sex chromosome trisomies (SCTs) in particular, but also for genomic disorders more broadly. The study raises the bar for this research field, via transparent study design, reporting and open data. These results should inform post-diagnostic genetic counselling for SCTs, and influence the future research agenda for these groups and other genomic disorders.

The paper tackles three major theoretical and clinical problems – specificity of associations between genomic disorders (three different SCTs) and developmental outcomes (language impairments), specificity of associations between genomic and non-genomic risk factors (SCT vs twins with language concerns), and the influence of ascertainment bias on SCT data. To summarise the findings, the authors confirm that SCT is a risk factor for developmental language impairment but find no evidence for specificity of outcomes between genomic disorders, no evidence for specificity of outcomes between genomic and non-genomic disorders. They find further evidence that ascertainment bias predicts severity of impairments within SCTs.

However, there are some important limitations in drawing these conclusions, which I feel could be discussed more directly and fully in order to maximise the utility and impact of this paper:

1. Are the SCT groups reported in the paper truly representative?
  - The low bias group was mainly detected prenatally. Indications for prenatal testing include abnormalities on fetal ultrasound and abnormal first trimester screening (to a greater extent than advanced maternal age), which are early indicators of developmental risk.
  - For both low and high bias groups, only children aware of their genetic diagnosis could be invited to the study (an inclusion criterion likely to be predictive of developmental risk). Families volunteering for this research were likely to be motivated by concerns about their children's development, biasing both groups toward impairment.
  - I understand and appreciate the inclusion / exclusion criteria applied to compare the SCT groups to the control samples. However a very high proportion of the SCT groups were excluded due to intellectual disability and autism, which are deeply connected to the outcomes under investigation. For the high bias group in particular, these criteria led to more than one third of the sample being excluded, and a further third can unusable data. The full spectrum of impairments in the SCT

group is therefore heavily under-estimated. Would it be possible for the authors to provide more information within the paper about the excluded SCT participants, and discuss whether incorporating these individuals would be likely to alter their main findings?

2. Is the reported data sufficient for validity and interpretation of results?

- The biggest limitation here is the high proportion of missing data for both neuropsychological and questionnaire assessments, meaning that only mildly impaired individuals are included in statistical analysis. This is openly reported, but it is somewhat difficult to appraise the quantities of missing data across analyses and I wonder if the authors could summarise this more clearly. For example, Table 2 provides descriptive demographic data for all participants (with suggestion that the high bias SCT groups may be associated with lower maternal education, lower neighbourhood advantage and higher family history scores). However if this data was displayed in addition for the subset of participants (defined by age and data availability) included in statistical comparison of (a) neuropsychological assessments and (b) CCC-2, this might indicate whether there were potential non-genomic biases in the available datasets and results.
- Connected, the high rates of missing data emphasise the need to improve methods that can be applied to children with intellectual impairments. Put more simply, it is difficult to draw firm conclusions about the nature and severity of language impairments if a large proportion of individuals with the very communication deficits of interest cannot be investigated. Could the authors discuss whether this limitation challenges their conclusions? And could they make some suggestions for new approaches which could tackle this issue head-on?

3. Are there other potential sources of false negative results to consider?

- Can the authors comment on the power that would be needed to detect significant differences within the analyses that they present? I appreciate the immense work that the team dedicated to this study and the impressive total sample size of rare disorders that they amassed over a five year period. The results should certainly be published. However, should interpretation of results dependent on a small subset of the total population be cautious?
- Inspecting the results in detail I note in Table 4, final row lists "Bias x Trisomy" but no results are given. This omission speaks to the question of whether there may be informative interactions yielding hypotheses for future studies, which could be explored with necessary caution. For example, Figure 4 suggests that bias within the SCT sample may have limited influence on core language skills or literacy skills, but impact on nonverbal ability and verbal production / memory. Furthermore, the impact of bias on these domains appears most marked for participants with XXX. On the other hand, participants with XYY demonstrate the most severe pragmatic and autistic features on CCC-2, in both the low and high bias groups. Would it therefore be wrong to conclude at this point that there are definitely no specific associations between SCTs and risk profiles, which may not be clinically predictive in view of high within-group variability but could nevertheless suggest differences in underlying mechanisms?

4. Is the concept of "typical" outcomes and profiles valid or useful?

- The abstract refers to typical outcomes in the Background and Conclusions. I wonder whether this concept is supported by the data, and whether either the abstract be reworded or the discussion extended to tackle this idea head-on.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Genomic disorders and cognitive development

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 05 Jan 2019

**Dorothy Bishop**, University of Oxford, UK

We thank the reviewer for her helpful comments, which are useful not just for this paper, but also for our forthcoming companion paper that focuses more on ASD and psychiatric aspects. In the course of addressing the comments, we discovered some errors in the reported numbers in different subsets of the sample, which had arisen because of confusion about the coding of variable names. This was the responsibility of DVMB, who is embarrassed at the errors, but glad of the opportunity to correct them – and grateful to the reviewer for noting inconsistencies that led to their discovery.

### **Sampling bias**

The first question raised by the reviewer is whether the SCT groups reported in the paper are truly representative. She notes that children diagnosed prenatally will have had indications for prenatal testing, that volunteer families may be atypical, and that children with ASD and intellectual disability were excluded.

We have now added more information about the Low Bias group, as follows:

*This latter group included 49 families who underwent amniocentesis: all but 4 parents provided a reason: 30 because of maternal age, 11 because nuchal screening or ultrasound scan indicated risk of a chromosome disorder, and 4 because of a family history of a chromosome disorder. A further 23 children had genetic testing for a range of medical conditions, including growth problems, failure to thrive, and failure to go through puberty (in the case of boys with XXY). Note that in referring to these as a 'Low Bias' group, we do not imply they are a representative sample of children with an extra sex chromosome: we make the more limited claim that the trisomy was not discovered as a result of the child manifesting language or behaviour difficulties. All genetic testing was done via the National Health Service, so the sample was not limited by parents' ability to pay for a test.*

The issue of volunteer bias is also very pertinent. This can cut both ways: as the reviewer notes, some parents may volunteer because they hope to obtain more information about their child's difficulties. On the other hand, some seemed keen to take part to emphasise how well their child was functioning, especially if they had been given an unrealistically poor prognosis at the time of diagnosis. The child's participation is also an issue. As we noted in our discussion, we previously found that parents are more likely to tell their child about the trisomy if the child is experiencing difficulties (Gratton et al, 2016), in which case the child can feel a sense of relief at having an explanation for why they are different from others. On the other hand, there is likely to be an opposing potential bias, which is that children who are anxious about social or communicative difficulties are less likely to assent to take part, even if parents want them to.

Could these problems be overcome? Potentially, one could conduct another study that involved neonatal screening, to complement the studies done in the 1960s, but even if ethics committees approved such a study, there would still be issues about bias among those who consented to take part. An interesting example comes from a study published by Tuke et al (2018), reporting on characteristics of women from the Biobank study who were discovered to have a sex chromosome trisomy. They were reported to be taller than average, to have earlier menopause, and to perform more poorly on a test of intellectual ability than other women of normal karyotype. Correspondence with the authors confirmed that none of these women were aware of the trisomy (or at least, none of them reported it on a medical history screen). At first glance, this looks like a perfect way of identifying a truly representative sample. However, there is volunteer bias in Biobank: only 5% of those invited to take part agreed to do so, and the response rate was more than twice as high for those from affluent areas than from areas of social disadvantage (Fry et al, 2018). Furthermore, Tuke et al (2018) identified 110 women with 47,XXX among 244,848 women in Biobank, which is less than half the number expected on the basis of estimated prevalence of around 1 in 1000 (Nielsen & Wohler, 1991). We know that adults with trisomy X tend to have relatively unskilled occupations and it seems likely that many may prefer not to participate in the Biobank study. These points are made not to criticise studies based on Biobank, but merely to note that recruitment of a truly representative sample is never possible, because those who consent will have certain characteristics, and one cannot compel people to take part in research. Given that is the case, the key issue is to be open about possible sources of bias and their likely impact.

It is also important to consider what question is being addressed, and how the information will be used. This is particularly important when families may be offered a termination of pregnancy if the trisomy is discovered prenatally, as we know that their decision can be heavily influenced by what they are told about outcomes (Jeon et al, 2012). For sex chromosome trisomies, there are enduring concerns similar to those expressed by Linden et al (2002), who note that parents seeking information on the internet may be confronted by information that either exaggerates the problems, or is factually wrong. Our interest is in language disorders, but we feel it is very important to draw attention to the fact that many children from the Low Bias sample had relatively mild language problems, with some scoring within normal limits. We noted the potential for bias toward impairment in our sample for this reason, but even taking that into account, there were many children who did well.

## Exclusions

In terms of the point about exclusions, it is important to recognise that our main focus is on the Low Bias group, where exclusions applied to 16% of cases (corrected data). The exclusion rate was high in the High Bias group, but our main purpose in including that group was to estimate the bias



that was introduced by including such cases. As we note in the paper, the differences between the two groups are substantial, even with ASD/ID cases excluded: clearly they would be even greater if we had included them. We make that point, but do not feel this compromises the main conclusion deriving from the Low Bias vs High Bias comparison – which is that it is very important to minimise bias as far as possible.

We accept the reviewer's point, that more information is needed about the excluded cases, rather than just referring to a future paper, and so in the current paper we have now included them in the beeswarm plots for the low-high bias comparison, distinctively marked, so readers can get an impression of the impact of the exclusion. We have not included them in the MANOVA, as the principal focus is the comparison of the SCT and twin samples in terms of language profile, and ASD/ID cases were not included in the twin sample.

One other note: since submitting this paper, we have been working on a manuscript documenting ASD and social anxiety in these children, and for consistency with that paper, there has been a change to the coding of ASD in the current paper. A subset of children were assessed using the Development and Wellbeing Assessment, and we had previously used one rater's diagnosis for ASD. We now have consensus data, which sometimes gave slightly different results, and this is what is used in the latest version of the paper.

The final paragraph, with clinical recommendations, has been modified to take into account these points about ascertainment bias.

### **Missing data**

The statement that 'only mildly impaired individuals are included in statistical analysis' represents a misunderstanding. It is evident from the beeswarm plots, that show the whole range of scores for all individuals, that the sample included children who scored very poorly – see, for instance, Figure 4, which shows many children who scored more than 2 SD below the mean. This was, however, confusingly reported, and in checking all the tables and figures, we realised there were some errors in the reporting. We thank the reviewer for picking this up, making it possible for us to make some corrections, as follows:

1. Figures 1 and 2 were intended to give a complete picture of the numbers of children at each point in the analysis for each sample, but a gremlin got into the script, which meant that some numbers were computed from the wrong subset of data. This has now been corrected.
2. Table 2 should have just shown the demographic data for the children included in the language test analysis, whereas in fact it showed data for the whole sample, including cases who had been excluded. This has now been corrected.
3. Because MANOVA required complete data for all children, any child with missing data on any of the 4 composites (PIQ, general language, verbal memory or literacy) was excluded from the analysis. Most missing data from the Low Risk group was simply because the children were too young for test norms. Indeed, our original goal had been to include only those aged 6 years and over, but given that we had some 5-year-olds volunteered for the study, and it would be feasible to collect a subset of data with them, including the CCC-2, we decided to reduce the lower age limit to 5 years towards the end of the study. So it would be wrong to conclude that these children could not attempt the tests. A second point is that for the majority of other children excluded from MANOVA, there was missing data on just one composite, and this was usually the Literacy

composite. To give a slightly larger sample on MANOVA, we decided to exclude literacy from that analysis; we now report a separate ANOVA for Literacy. In addition we note the numbers of cases with missing data, distinguishing between those out of age range for norms, and those who did not complete the tests.

We thank the reviewer for drawing our attention to this, because it is clear that by focussing attention only on children who completed the whole battery, we gave the misleading impression that many children could not be tested, which was not the case. In the High Bias group there were some children who were more challenging to assess and who were unwilling or unable to attempt certain tests. But as we stress throughout, the results from the High Bias group need to be treated with extreme caution: our main reason for including them is to show how ascertainment bias can distort results.

So in sum, the beeswarm plots have now been amended to include children who had partial data. We had previously included only those children who were in the MANOVA so that the beeswarm data would match the MANOVA data, and show the relevant means. We are still able to show that information, although it does make the plot a bit more complicated.

### Power

The reviewer raises an important point: all studies in this area have been beset by small sample sizes, and it is likely that this is why findings of karyotype-specific effects have been inconsistent. We have now added this comment in our discussion:

*Note, however, that detection of differences in variance between groups would require substantial sample sizes. Even for detection of mean group differences, our sample size was too small to detect any but large effects of karyotype: for comparisons of XXX+XXY vs XYY we had 80% power to detect an effect size on psychometric composites of .83, and on CCC-2 an effect size of .97. We cannot therefore conclude from the current study that there are no sex-chromosome-specific influences, only that if they do exist, they are not large, and are superimposed on substantial variation from other sources. Because of the difficulty of recruiting large samples, and the difficulty of controlling sampling biases, we will need to combine samples across different research groups to get a clear answer to the question of whether an additional X or Y has consistently different effects.*

### Table 4

Apologies: There was an error in Table 4 with a misleading row added. We did not look at the interaction between trisomy and bias group as this would not address the questions of interest. (And also the power would be abysmal even if there was interest in such an interaction). Our data suggested there could be differences between trisomies in terms of who gets into the High Bias group, but that was a post hoc observation, based on finding that the numbers depart from expectation, with relatively few XXX girls in the High Bias group. The chi square analysis we did on the Ns in the 2 x 3 table draw attention to this imbalance, which we think is the best way to flag this up.

### Is the concept of “typical” outcomes and profiles valid or useful?

This is a valid comment: our goal was really to try and distinguish what were the most common profiles seen in sex chromosome trisomies, though given the enormous heterogeneity it is perhaps unhelpful to talk of anything being 'typical'. We say a bit more about implications for genetic counselling: perhaps the message for parents is exactly this – there is no 'typical'. Nevertheless,

they need to be given some idea of the range of possibilities. We refer back to the 2002 paper of Linden et al on genetic counselling; we feel our paper supports their recommendations regarding a condition that is very challenging to advise on.

## References

- Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., . . . Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology*, 186(9), 1026-1034. doi:<https://doi.org/10.1093/aje/kwx246>
- Gratton, N. C., Myring, J., Middlemiss, P., Shears, D., Wellesley, D., Wynn, S., . . . Scerif, G. (2016). Children with sex chromosome trisomies: parental disclosure of genetic status. *European Journal of Human Genetics*, 24(5), 638-644. doi:10.1038/ejhg.2015.168
- Jeon, K. C., Chen, L.-S., & Goodson, P. (2012). Decision to abort after a prenatal diagnosis of sex chromosome abnormality: a systematic review of the literature. *Genetics in Medicine*, 14, 27-38.
- Linden, M. G., Bender, B. G., & Robinson, A. (2002). Genetic counseling for sex chromosome abnormalities. *American Journal of Medical genetics*, 110(1), 3-10. doi:10.1002/ajmg.10391
- Nielsen, J., & Wohler, M. (1991). Chromosome abnormalities found among 34910 newborn children: results from a 13-year incidence study in Århus, Denmark. *Human Genetics*, 87, 81-83.
- Tuke, M. A., Ruth, K. S., Wood, A. R., Beaumont, R. N., Tyrrell, J., Jones, S. E., . . . Murray, A. (2018). Mosaic Turner syndrome shows reduced penetrance in an adult population study. *Genetics in Medicine*. doi:<https://doi.org/10.1038/s41436-018-0271-6>

**Competing Interests:** None