RESEARCH ARTICLE

Language phenotypes in children with sex chromosome trisomies [version 1; referees: awaiting peer review]

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

Abstract

Background: Sex chromosome trisomies (47,XXX, 47,XXY and 47,XYY) are known to be a risk factor for language disorder, but typical outcomes are hard to estimate, 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 the 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 = 132) and a Language Concerns group (N = 41). Children were individually assessed on a psychometric battery, and a standardized parent checklist. After excluding children with intellectual disability, autism or hearing problems, psychometric data were available for 31 XXX, 20 XXY and 19 XYY Low Bias cases and 13 XXX, 25 XXY and 32 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 typical of children identified on prenatal screening.

Keywords
Trisomy X, Klinefelter syndrome, XYY syndrome, ascertainment bias, variability, psychometric testing, parent report
Corresponding author: Dorothy V. M. Bishop (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: © 2018 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 1; referees: awaiting peer review] Wellcome Open Research 2018, 3:143 (https://doi.org/10.12688/wellcomeopenres.14904.1)

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

Chromosome trisomies arise when an error of cell division leads to an egg or sperm that contains two copies rather than one copy of a 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 (Printzlau 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 multi-centre 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,XXY—were all found to be associated with neurodevelopmental problems, particularly affecting language and motor functions, see (Leggett et al., 2010) for review.

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. (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, but it seems likely that this was 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 of interest because there are few genetic aetiologies that selectively impair language development. There are two reasons why it is worth comparing 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 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 the 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 differences 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 inCNTNAP2 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 establishing 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?

Methods
Participants
Sex chromosome trisomy cases. We recruited 44 girls with XXX, 45 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. 22% 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. The sample was the same as that reported by Newbury et al. (2018), except that some children were excluded from that analysis because DNA samples were not available.

Ascertainment bias. 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

![Figure 1](image-url)

**Figure 1.** Numbers of children with sex chromosome trisomies included in different analyses.
a high-risk-of-bias (High Bias) subgroup, comprised of 13 XXX, 25 XXY and 32 XYY cases. All other children formed the low-risk-of-bias (Low Bias) subgroup, comprised of 31 XXX, 20 XXY and 19 XYY cases.

**Exclusionary criteria.** Given our current focus on language disorder, we excluded children who had a nonverbal ability scaled score more than 2 SD below the mean (2 Low Bias and 11 High Bias), those who met criteria for ASD (6 Low Bias and 13 High Bias) and those who failed a hearing screen (1 Low Bias and 1 High Bias). This enabled us to make a direct comparison 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 means that our conclusions are based on a selected sample of trisomy cases. ASD and other diagnoses in children with sex chromosome trisomies will be the focus of a companion paper.

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

**Comparison group**
The comparison sample consisted of twins recruited via flyers 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 sex chromosome trisomy 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 & Bishop (2018).

Our second question involved comparing the language profile of children with trisomies with that of children with language disorders of unknown cause. 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, 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.

We followed the CATALISE criteria for DLD (Bishop et al., 2017) and excluded twin pairs if either twin had a pre-existing diagnosis of ASD, failed the hearing screen, had a performance IQ score below 70, or had a serious long-term illness; 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 41 children in the language concerns group, and 132 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 Children’s Communication Checklist–2 (CCC-2) data were available.

**Ethical approval and consent to participate**
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 the Open Science Framework.

**Language, literacy and cognitive assessments**
The assessment battery administered to the child is shown in Table 1. Parents were asked to complete the CCC–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 by 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.

Our original intention was to use multivariate analysis of variance (MANOVAs) to test group differences that related to the three research questions, but preliminary analysis showed that the MANOVA assumption of multivariate normality was not met when all 14 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...
Table 1. Psychometric assessment battery.

| Instrument | Measure |
|------------|---------|
| Audiometry: screen at 25 dB (HL) for frequencies 500, 1000, 2000 and 4000 Hz. | Passed if average threshold in better ear was 30 dB or less |
| Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999) | Matrices (nonverbal reasoning) Block Design (visuospatial skills) Vocabulary |
| Woodcock Johnson III Tests of Cognitive Abilities (Woodcock et al., 2007) | Verbal Comprehension |
| NEPSY: A Developmental Neuropsychological Assessment (Korkman et al., 1998) | Sentence Repetition Repetition of Nonsense Words Oromotor Sequences |
| Neale Analysis of Reading Ability – 2nd British edition (NARA-II) (Neale, 1999) | Passage reading, giving measures of accuracy, comprehension and rate |
| Test of Word Reading Efficiency (TOWRE) (Torgesen et al., 1999) | Speeded reading of words (Sight Word Efficiency) and nonwords (Phonetic Decoding Efficiency) |
| Phonological Assessment Battery (PhAB) (Frederickson et al., 1997) | Picture Naming Test Digit Naming Test |

Figure 2. Numbers of twin children with and without language concerns.
(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 Figure 1. Multivariate normality was again 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 Figure 2.

Covariates

Two measures of socio-economic variables 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 sex chromosome trisomy cases and 4 twin pairs from Wales, Scotland and Northern Ireland, where postcode rankings were not available.

In addition, following a suggestion by Bouda 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).

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 as prorated from other reading measures. In total, 29 children had more missing data than this, and were excluded from the current analyses. This included 11 5-year-olds, for whom standard scores were not computable as they fell below the age range of norms. All but four cases 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 31 children had missing data because parents either did not complete the CCC-2 (N = 20), or because they failed the consistency check (N = 11), which is suggestive of invalid responses.

Statistical comparisons

For each type of variable (psychometric tests and parent report on CCC-2) three comparisons were conducted using MANOVA, 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 0.05/6 = 0.008 was adopted. Other comparisons were analysed using chi square tests. All data analysis was conducted in the R programming language (R Core Team, 2018) version 3.5.0. Data were visualised using beeswarm plots (Eklund, 2016).

Results

Complete raw demographic data, alongside results of all psychometric tests and scripts used in this study, are available on Open Science Framework (Bishop, 2018).

Characteristics

Table 2 shows the characteristics of the trisomy and twin groups on age, covariates and family history. There was an unanticipated imbalance between the three trisomy types in terms of the proportion of cases in the Low Bias group: 70% for XXX girls, 44% for XXY boys, and 37% for XYY boys (p = 0.004 on chi square test).

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 exclude children with missing data on any composite. The continuous horizontal bars show the mean for each group. The yellow shaded area shows the range covered by mean ± 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
Table 2. Characteristics of the trisomy and twin groups on background variables: age (months), mother’s educational level, Neighbourhood Advantage index and family history.

| Group                  | N  | Age (mo) | Mo. educ. | Neighbourhood Advantage | Family history |
|------------------------|----|----------|-----------|--------------------------|----------------|
| XXX: Low Bias          | 31 | 135 (47.8)| 1.9 (0.93)| 0.6 (0.79)               | 0.6 (0.77)     |
| XXY: Low Bias          | 20 | 137 (42.0)| 2.0 (1.05)| 0.4 (0.85)               | 0.4 (0.59)     |
| XYY: Low Bias          | 19 | 125 (40.3)| 2.3 (0.89)| 0.4 (0.92)               | 0.4 (0.61)     |
| XXX: High Bias         | 13 | 138 (45.1)| 1.5 (0.67)| -0.5 (0.9)               | 0.6 (0.65)     |
| XXY: High Bias         | 25 | 146 (43.0)| 1.7 (0.93)| 0.2 (0.99)               | 0.9 (0.83)     |
| XYY: High Bias         | 32 | 115 (40.9)| 1.6 (0.87)| -0.1 (0.92)              | 0.7 (0.79)     |
| Twin: No concerns      | 116| 106 (19.5)| 2.4 (0.94)| 0.6 (0.96)               | 0.4 (0.65)     |
| Twin: Language concerns| 60 | 108 (19.0)| 2.1 (0.92)| 0.4 (1.15)               | 0.5 (0.70)     |

Figure 3. Distributions of scores on four clusters of psychometric tests for the Low Bias trisomy groups and the Language Concerns Comparison group. Filled circles show cases in age range 6 to 11 years. Dotted line is mean for 6–11-year-olds, solid line is mean for whole sample. Yellow band is mean ± 1 SD for No Concern comparison group.
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 Figure 1).

MANOVA was used to test whether there were reliable differences between the three trisomy groups on the four composite measures, after taking covariates (mother’s educational level and Neighbourhood Advantage index) into account. No effect of trisomy type was found (see Table 3, row A).

**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 cases aged 6–11 years (filled circles in Figure 3) with the twin comparison group with language concerns. Table 3, row B, shows the MANOVA result for this comparison, again revealing 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.

**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. As shown in Table 3, row C, there was a strong effect of ascertainment bias, with the High Bias group having lower 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 > 0.05, maximum proportion variance explained = 0.04).

**CCC-2**

Completion rates for the CCC-2 showed some variation by group: the checklist was completed by parents of 93.4% of the Low Bias and 80.3% of the High Bias trisomy group, and by 78.7% of the parents of No Concern twins and 75.4% of parents of Language Concern twins. Data was not useable for 4.7% 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 was 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 1 and Supplementary Figure 3, 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 ± 1 SD shown as yellow shading) has mean scores close to the population mean of 10.

Figure 6 contrasts CCC-2 composites for the Low Bias and High Bias trisomy groups. 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.

**Table 3.** Results of three MANOVAs (A, B and C), testing for overall group differences on four composites from psychometric tests, with covariates of mother’s education and Neighbourhood Advantage index.

| Comparison                                      | Effect                  | Wilks λ | F   | df1 | df2 | p      |
|-------------------------------------------------|-------------------------|---------|-----|-----|-----|--------|
| A. Low Bias: XXX (25) vs XXY (20) vs XYY (16)   | Group                   | 0.904   | 0.680 | 8   | 106  | 0.704  |
|                                                 | Mo. education           | 0.888   | 1.670 | 4   | 53   | 0.171  |
|                                                 | Neighbourhood Advantage | 0.902   | 1.450 | 4   | 53   | 0.232  |
| B. Trisomies (35) vs. Lang concern twins (51): aged 6–11 | Group                   | 0.956   | 0.910 | 4   | 79   | 0.465  |
|                                                 | Mo. education           | 0.814   | 4.500 | 4   | 79   | 0.003  |
|                                                 | Neighbourhood Advantage | 0.966   | 0.690 | 4   | 79   | 0.601  |
| C. Low Bias (61) vs. High Bias (50)             | Group                   | 0.715   | 10.060 | 4   | 101  | <0.001 |
|                                                 | Mo. education           | 0.927   | 1.990 | 4   | 101  | 0.102  |
|                                                 | Neighbourhood Advantage | 0.973   | 0.690 | 4   | 101  | 0.601  |

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).

Number of language tests below cut-off
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). 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. The percentages of twins rated this low was 5.2% for the No Concerns group and 41.7% for the Language Concerns group. Among children with sex chromosome trisomies, 61.4% of the Low Bias group and 75.7% 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 between those with XXX, XXY and XYY karyotypes in the severity or profile of language problems. 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 XXY had more pervasive and severe language problems than boys with XXY karyotype, but they noted that 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. Indeed, for children selected in a manner that reduced ascertainment bias, we found that around one-third of cases resembled 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 CCC-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

![Figure 5. Mean scores on the three Children’s Communication Checklist–2 composites for trisomy and comparison groups.](image)

The yellow shaded band shows mean ±1 SD for the No Concerns twin group. For trisomies, continuous lines show mean for whole group, and dotted lines show means for children aged 6 to 11 years.
Figure 6. Mean scores on the three Children’s Communication Checklist–2 composites for Low Bias vs High Bias trisomy groups, with the latter denoted by asterisks. For trisomies, continuous lines show mean for whole group, and dotted lines show means for children aged 6 to 11 years. The yellow shaded band shows mean ±1 SD for the No Concerns twin group.

Table 4. Results for three MANOVAs comparing groups on the Children’s Communication Checklist–2 composite scores.

| Comparison          | Effect                  | Wilks λ | F     | df1 | df2 | p    |
|---------------------|-------------------------|---------|-------|-----|-----|------|
| A. XXX (24), XXY (17), XYY (11) | Group                   | 0.875   | 1.040 | 6   | 90  | 0.407|
|                     | Mo. education           | 0.877   | 2.100 | 3   | 45  | 0.113|
|                     | Neighbourhood Advantage | 0.936   | 1.030 | 3   | 45  | 0.388|
| B. All SCT vs Lang concern twins | Group                   | 0.912   | 1.990 | 3   | 62  | 0.125|
|                     | Mo. education           | 0.888   | 2.600 | 3   | 62  | 0.060|
|                     | Neighbourhood Advantage | 0.941   | 1.310 | 3   | 62  | 0.280|
| C. Bias             | Bias                    | 0.733   | 9.210 | 3   | 76  | 0.000|
|                     | Mo. education           | 0.937   | 1.690 | 3   | 76  | 0.176|
|                     | Neighbourhood Advantage | 0.930   | 1.910 | 3   | 76  | 0.136|

SCT, sex chromosome trisomy.
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 nine 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 XYY group. Visual inspection of the distribution of scores in the three trisomies (Figure 5 and Figure 6), however, does not support that prediction.

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 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 parent 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 (2014) 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, 25 children were excluded from the High Bias sample because of intellectual disability, ASD or hearing problems, and a further 25 children had missing data and were excluded from analysis. 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 it needs to be emphasised that one should not draw conclusions about typical outcomes for children with sex chromosome trisomies from samples that have this kind of ascertainment bias.

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, row B). In a review of

| Group | N  | T0 | T1 | T2 | T3 | T4 | T5 |
|-------|----|----|----|----|----|----|----|
| No concerns | 116 | 60.3 | 25.0 | 8.6 | 6.0 | 0 | 0 |
| Language concerns | 60 | 23.3 | 23.3 | 13.3 | 16.7 | 8.3 | 15.0 |
| Trisomy: Low Bias | XXX | 31 | 9.7 | 12.9 | 12.9 | 25.8 | 25.8 | 12.9 |
| | XXY | 20 | 30.0 | 15.0 | 15.0 | 20.0 | 5.0 | 15.0 |
| | XYY | 19 | 5.3 | 31.6 | 21.1 | 15.8 | 21.1 | 5.3 |
| Trisomy: High Bias | XXX | 13 | 0 | 7.7 | 15.4 | 23.1 | 23.1 | 30.8 |
| | XXY | 25 | 0 | 4.0 | 4.0 | 20.0 | 32.0 | 40.0 |
| | XYY | 32 | 0 | 9.4 | 6.2 | 12.5 | 34.4 | 37.5 |
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 came 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 operate. If such genetic instability 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. This makes 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. Where children have language deficits, they tend to be of a range and severity similar to those seen in other children who have language difficulties in the absence of any known biological aetiology.

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

Supplementary Figure 1. Distributions of scores on the 14 measures from psychometric tests for the Low Bias trisomy groups and the Language Concerns Comparison group.

Filled circles show cases in age range 6 to 11 years. Continuous line is mean for 6–11-year-olds, dotted line is mean for whole sample. Yellow band is mean ± 1 SD for No Concern comparison group. Note that on some tests there is a minimum scaled score that leads to a skewed distribution.

Click here to access the data

Supplementary Figure 2. Means for CCC-2 scales for trisomy groups subdivided by bias type (continuous lines = Low Bias, and dotted lines = High Bias) and for the two twin groups.

Error bars show standard errors. In the MANOVA for CCC-2, we used composites based on averaged subscales to minimise problems with non-normality of data; this figure summarises the data on the individual scales.

Click here to access the data
Supplementary Figure 3. Fitted regression lines for predicting CCC-2 completion from child status. Shown are fitted points for CCC-2 response rate in trisomy and twin samples, in relation to language status of the child (score on language factor, where low score indicates impairment). Note that checklist completion was lowest for parents whose child had more severe language problems.

Click here to access the data

Supplementary Table 1. Estimates of regression coefficients with CCC-2 completion rate as dependent variable, and parent and child variables as predictors.

Click here to access the data

References