Detection and characterization of male sex chromosome abnormalities in the UK Biobank study

Yajie Zhao1, Eugene J. Gardner1, Marcus A. Tuke2, Huairen Zhang2, Maik Pietzner1,3, Mine Koprulu1, Raina Y. Jia1, Katherine S. Ruth2, Andrew R. Wood2, Robin N. Beaumont2, Jessica Tyrrell2, Samuel E. Jones2,4, Hana Lango Allen1, Felix R. Day1, Claudia Langenberg1,3, Timothy M. Frayling2, Michael N. Weedon2, John R.B. Perry1, Ken K. Ong1,*, Anna Murray2,*

ARTICLE INFO

Article history:
Received 31 January 2022
Received in revised form 15 May 2022
Accepted 16 May 2022
Available online xxxx

Keywords:
47.XXY
Disorders of sexual development
Endocrinology
Klinefelter syndrome
Thrombosis
Type 2 diabetes

ABSTRACT

Purpose: The study aimed to systematically ascertain male sex chromosome abnormalities, 47.XXY (Klinefelter syndrome [KS]) and 47.XYY, and characterize their risks of adverse health outcomes.

Methods: We analyzed genotyping array or exome sequence data in 207,067 men of European ancestry aged 40 to 70 years from the UK Biobank and related these to extensive routine health record data.

Results: Only 49 of 213 (23%) of men whom we identified with KS and only 1 of 143 (0.7%) with 47.XYY had a diagnosis of abnormal karyotype on their medical records or self-report. We observed expected associations for KS with reproductive dysfunction (late puberty: risk ratio [RR] = 2.7; childlessness: RR = 4.2; testosterone concentration: RR = −3.8 nmol/L, all P < 2 × 10⁻⁸), whereas XYY men appeared to have normal reproductive function. Despite this difference, we identified several higher disease risks shared across both KS and 47.XYY, including type 2 diabetes (RR = 3.0 and 2.6, respectively), venous thrombosis (RR = 6.4 and 7.4, respectively), pulmonary embolism (RR = 3.3 and 3.7, respectively), and chronic obstructive pulmonary disease (RR = 4.4 and 4.6, respectively) (all P < 7 × 10⁻⁶).

Conclusion: KS and 47.XYY were mostly unrecognized but conferred substantially higher risks for metabolic, vascular, and respiratory diseases, which were only partially explained by higher levels of body mass index, deprivation, and smoking.

Crown Copyright 2022 Published by Elsevier Inc. on behalf of American College of Medical Genetics and Genomics. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Introduction

The most common sex chromosome aneuploidies in men are 47,XXY (Klinefelter syndrome [KS]) and 47,XXX, with population prevalence estimates of 100 per 100,000 men and 18 to 100 per 100,000 men, respectively. Men with KS typically present during adolescence with delayed puberty or as adults with infertility. Other recognized features include tall adult stature; high body fat percentage; poor muscle tone; low bone mineral density (BMD); and increased risks of neurocognitive disability, psychoses, and disorders of personality. KS has also been associated with higher risks of type 2 diabetes and venous thromboembolism. By contrast, 47,XYY is less well-characterized because many of these individuals may not present to health services and thus are unaware of their karyotype. Reported features associated with 47,XXX may therefore be affected by sampling bias. These include tall stature; scoliosis; learning difficulties; poor muscle tone; increased central fat; and increased risks of seizures, asthma, and emotional and behavioral problems (eg, autism and attention deficit disorder). Although infertility has been reported in some men with XYY, most studies report normal sexual development and fertility.

Previous studies identified men with KS or 47,XXY from medical records, and therefore case ascertainment was based on recognition of their typical phenotypic features. Therefore, the reported penetrance of these features may have been biased and the full spectrum of clinical features are overlooked. A more robust alternative approach is to identify such individuals from large population-based studies using systematic measurements to produce unbiased estimates of the effects of sex chromosome aneuploidy on unselected diseases. We recently used this approach to show that mosaic X-chromosome aneuploidy in women (mosaic Turner syndrome, 45,X) conferred a lower penetrance of infertility than that reported by earlier clinic-based studies. In brief, this approach uses k-means-clustering to group individuals by the first 4 genetic principal components. In addition, we excluded individuals who were classified as White European by our k-means approach but self-identified as being of ancestry other than White European. We further excluded individuals whose samples failed UK Biobank genotyping quality control (QC) parameters and those who withdrew consent. Accordingly, 207,067 men were included in all association testing analyses. We were unable to incorporate non-European individuals when modeling the relationship between abnormal karyotypes and phenotypes outlined in this manuscript. We identified only 16 non-White European males with abnormal karyotypes: 9 with 47,XXX and 7 with 47,XXX.

Materials and Methods

Study population

The UK Biobank is a large prospective cohort that recruited approximately 500,000 participants aged 40 to 70 years across the island of Great Britain. A broad range of phenotypic and health-related information was collected from each participant, including physical measurements, lifestyle indicators, biomarkers in blood and urine, imaging, and routine health record data.

In the UK Biobank, 488,377 participants had DNA samples assayed using 1 of 2 genotyping arrays: UK Biobank Lung Exome Variant Evaluation (UK BILEVE study, N = 49,950) and Affymetrix Axiom UK Biobank array (UK Biobank Axiom [Affymetrix], N = 438,427). These 2 arrays tested 807,411 and 825,927 SNVs, respectively, with 95% overlap between arrays. We restricted our analysis to men of White European genetic ancestry as classified by the approach previously described by Thompson et al. In brief, this approach uses k-means-clustering to group individuals by the first 4 genetic principal components. In addition, we excluded individuals who were classified as White European by our k-means approach but self-identified as being of ancestry other than White European. We further excluded individuals whose samples failed UK Biobank genotyping quality control (QC) parameters and those who withdrew consent. Accordingly, 207,067 men were included in all association testing analyses. We were unable to incorporate non-European individuals when modeling the relationship between abnormal karyotypes and phenotypes outlined in this manuscript. We identified only 16 non-White European males with abnormal karyotypes: 9 with 47,XXX and 7 with 47,XXX.

Identification of male sex chromosome aneuploidy heterozygotes from SNV array data

To identify men with sex chromosome aneuploidy, we downloaded genotyping fluorescent signal intensity (log2 ratios, LRR) and QC information for all SNVs on the X chromosome (chrX) and Y chromosomes (chrY) from the UK Biobank data showcase (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=22431 and https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=1955). We excluded SNVs that (1) were located within pseudoautosomal regions (PAR), (2) did not have a calculable LRR on both arrays, (3) did not pass QC in all 106 batches, or (4) were flagged as failing QC by UK Biobank. After these steps, 16,599 chrX SNVs and 579 chrY SNVs remained. We then calculated the median LRR across all remaining SNVs on chrX and chrY to generate the values mLRR-X and mLRR-Y, respectively. These values represent the median fluorescence signal intensities across the entire X or Y chromosome. Using the thresholds described by Bycroft et al., men with −1 ≤ mLRR-Y < 0.23 and mLRR-X > −0.2 were categorized as having 47,XXX (KS) and men with mLRR-Y ≥ 0.23 and mLRR-X < −0.2 as 47,XXX (men with mLRR-Y ≥ 0.23 and mLRR-X > −0.2 were categorized as 48,XXX and were not included in further analyses).

Confirmation of male sex chromosome aneuploidy heterozygotes from exome sequencing data

To confirm sex chromosome aneuploidy status using an orthogonal approach, we used exome sequencing data
available for 83,104 White European men in UK Biobank. To estimate sex chromosome dosage, we calculated the average read depth of 3 target regions: (1) non-PAR regions on chrX, (2) X-degenerate regions (XDRs) on chrY as defined by Skov et al., and (3) autosomes. First, we used SAMtools (version: 1.9) to convert the provided CRAM files for each participant to Binary Alignment Map files on the basis of the GRCh38 reference sequence (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/GRCh38_reference_genome/GRCh38_full_analysis_set_plus_decoy_hla.fa). UK Biobank provided the GRCh38 coordinates of the targeted regions for its exome sequencing design with a Browser Extensible Data file (https://biobank.ndph.ox.ac.uk/showcase/ukb/auxdata/xgen_plus_spikein.GRCh38.bed). We created 3 subsets of this Browser Extensible Data file by extracting the overlap between the target regions and non-PAR regions on chrX, XDRs on chrY, and autosomes according to their GRCh38 coordinates. Then, they were converted to picard interval lists using the Picard (version: 2.21.6-SNAPSHOT) function BedToIntervalList, on the basis of the same reference sequence. Using these picard interval lists, the Binary Alignment Map file of each participant was inputted to calculate the average coverages of non-PAR regions on chrX, XDRs on chrY, and autosomes using the Picard function CollectHsMetrics. The relative read depth of non-PAR regions on chrX, XDRs on chrY, and autosomes according to their GRCh38 coordinates. The relative read depth of non-PAR regions on chrX and XDRs on chrY were defined as the average coverage in each of these regions divided by the average coverage across the autosomes. The relative read depth of non-PAR regions on chrX and XDRs on chrY multiplied by 2 were used as a proxy of chrX dosage and chrY dosage, respectively. Men with chrX dosage > 1.2 were categorized as having 47,XXY (KS) and men with chrY dosage > 1.5 as 47,XXX. Men with chrX dosage > 1.2 and chrY dosage > 1.5 were categorized as 48,XXYY.

Disease association testing

To test for the disease burden associated with male sex chromosome aneuploidies, we performed logistic regression models with KS or 47,XXXY (coded 1) compared with the normal male karyotype 46,XY (coded 0) as the exposure. Outcomes comprised 875 International Classification of Diseases (ICD)-10 coded diseases amalgamated from death registries, hospital episode statistics, primary care records (in a subset, n = 94,959), and self-reported conditions (from the first occurrence of disease data set released by UK Biobank). The data set contains further 19 case definitions from dedicated working groups that used multiple sources for case identifications, such as for chronic obstructive pulmonary disease (COPD) or end stage renal disease. For each participant, all events from all sources were mapped to an ICD-10 code and the date of the first disease occurrence from any source was taken as the event date. From this data set, we filtered out likely erroneous disease events if the disease occurrence date: (1) was unknown or missing, (2) matched or preceded the date or year of birth, or (3) occurred after the data set release date. We performed logistic regression models in R (version: 3.6.0) among unrelated men of White European genetic ancestry (maximum N = 162,322) and adjusted for age at study baseline, test center, and the first 10 genetically derived principal components. Resulting odds ratios were converted to risk ratios (RR) using the formula described by Zhang and Yu. We applied a stringent Bonferroni corrected P value threshold of P < .05/875 = 5.7 × 10−5 to define statistical significance (Supplemental Table 1 and Figure 1).

Study phenotype association testing

To test the association of male sex chromosome aneuploidy status against selected anthropometric, reproductive, metabolic, cardiovascular, learning/memory, and behavioral study-measured traits (Supplemental Table 2), we used a linear mixed model implemented in BOLT-LMM (version: 2.3.2). The outcome childlessness was derived from the response 0 to the question “How many children have you fathered?” among men aged 55 and older. The Townsend deprivation index was used as an indicator of socioeconomic status, on the basis of participants’ home postcode. The 2 binary exposure variables described earlier were converted to BGEN file format using plink2 (version: 2.00-alpha) and inputted to BOLT-LMM via the bgenFile flag. A genetic relationship matrix was generated on the basis of all autosomal variants that had minor allele frequency of >1%, passed QC in all 106 batches, and were present on both genotyping arrays. Genotyping chip, age at baseline, and the first 10 genetically derived principal components were included as covariates. For binary outcomes, we also performed logistic regression and calculated the RR from the odds ratio as described earlier.

Nuclear magnetic resonance metabolic biomarkers association testing

We analyzed 168 circulating metabolic traits measured by proton nuclear magnetic resonance spectroscopy (Nightingale Health Plc) in nonfasting plasma samples in UK Biobank men with 46,XY (n = 49,806), KS (N = 48), or 47,XXXY (n = 38). For each metabolic traits, we first performed adjustment for technical variations using the R package ukbnmr (unpublished data: Ritchie SC et al 2021. https://www.medrxiv.org/content/10.1101/2021.09.24.21264079v2.), then performed inverse rank normalization, and then further adjusted for sex, age at the first study visit, body mass index (BMI), and the first 10 genetically derived principal components. Associations between abnormal karyotype and each metabolic trait were tested in separate linear regression models (Supplemental Figure 1).
Figure 1  Circos plot summarizing phenome-wide disease association tests for KS and 47,XXY compared with 46,XY. Each segment represents each International Classification of Diseases (ICD)-10 chapter in lexicographical order. P values (on a negative logarithmic scale) were from logistic regression models for KS (outer circle) and 47,XXY (inner circle) with each of 875 ICD-10 coded disease outcomes, adjusted for age and 10 principal genetic components. Outcomes reaching the multiple testing corrected statistical significance threshold ($P < 0.05/875 = 5.7 \times 10^{-5}$; dashed line) are indicated by large circles (for positive associations) and diamonds (for negative associations). AIODCE, arthropathies in other diseases classified elsewhere; CAFAC, cutaneous abscess, furuncle and carbuncle; DOAADCIDCE, disorders of arteries, arterioles and capillaries in diseases classified elsewhere; MABDDTUCOT, mental and behavioral disorders due to use of tobacco; NEC, not elsewhere classified; OBAATCODCTOC, other bacterial agents as the cause of diseases classified to other chapters; OCOPD, other chronic obstructive pulmonary disease; OEOAM, other extrapyramidal and movement disorders; ONDOLVALN, other noninfective disorders of lymphatic vessels and lymph nodes; ONGAC, other noninfective gastroenteritis and colitis; OLIOSAST, other local infections of skin and subcutaneous tissue; OSCAMPN, other sex chromosome abnormalities, male phenotype; NEC; OWPF, osteoporosis without pathological fracture; PIDCE, polyneuropathy in diseases classified elsewhere; SASATCODCTOC, streptococcus and staphylococcus as the cause of diseases classified to other chapters; SDDOSS, specific developmental disorders of scholastic skills; UALRI, unspecified acute lower respiratory infection.
Results

Prevalence of male sex chromosome aneuploidy in a population scale biobank

Using genotyping array data on 207,067 men of European ancestry, we identified 213 men with 47,XXY (Klinefelter syndrome), 143 with 47,XYY, and 2 with 48,XXYY. Of these cases, who also had exome sequencing data, we observed 100% confirmation of aneuploidy status (62/62 men with KS and 54/54 men with 47,XYY) (Figure 2B and C).

Only 49 of 213 (23.0%) men with KS and 1 of 143 (0.7%) with 47,XYY had a diagnosis of sex chromosome abnormality on routine medical records or self-reported data (ICD10: Q98 other sex chromosome abnormalities, male phenotype, not elsewhere classified). Similar proportions were found in the subsample of men who had primary care data: only 24 of 89 (27.0%) with KS and 1 of 76 (1.3%) with 47,XYY had known sex chromosome abnormality. Conversely, of the men with a diagnosis of sex chromosome abnormality on their health record, from our analysis we classified 4 as 46,XX (mLRR-Y < -1) and a further 8 as having a normal male karyotype.

Quantification of typical features of 47,XXY and 47,XYY

Compared with men with a normal karyotype (46,XY), men with KS and 47,XYY had taller adult height, by 2.7 cm ($P = 7 \times 10^{-15}$) and 7.9 cm ($P = 8 \times 10^{-77}$), respectively, and were more likely to have had taller than average childhood height (RR = 1.3, $P = .01$ and RR = 1.7, $P = 5 \times 10^{-10}$, respectively) (Table 1). Men with KS and 47,XYY were more likely to be childless (RR = 4.2, $P = 4 \times 10^{-17}$ and RR = 2.4, $P = 2 \times 10^{-17}$, respectively), but only those
with KS (not 47,XYY) were more likely to report later than average puberty timing (RR = 2.7, P = 2 × 10⁻⁸).

In addition, men with KS and 47,XYY were less likely to have a university or college degree (RR = 0.39, P = 5 × 10⁻⁸ and RR = 0.50, P = 9 × 10⁻⁸, respectively) and had lower fluid intelligence test scores (beta = −2.1, P = 3 × 10⁻¹⁵ and beta = −1.6, P = 5 × 10⁻⁸, respectively), were more likely to be smokers (RR = 1.1, P = 0.28 and RR = 1.2, P = 8 × 10⁻³, respectively), report depressive episodes (RR = 2.4, P = 6 × 10⁻¹⁰ and RR = 2.7, P = 3 × 10⁻⁹, respectively), and live in areas with higher deprivation index (beta = 1.6, P = 1 × 10⁻¹⁵ and beta = 1.6, P = 1 × 10⁻¹⁰, respectively). Men with KS and 47,XYY were also more likely to live without a partner (RR = 2.2, P = 4 × 10⁻²⁰ and RR = 2.1, P = 1 × 10⁻¹², respectively) and report loneliness and isolation (RR = 2.2, P = 4 × 10⁻¹³ and RR = 2.5, P = 3 × 10⁻¹³, respectively) and poor overall health (RR = 4.2, P = 2 × 10⁻²⁶ and RR = 3.8, P = 2 × 10⁻¹⁴, respectively).

**Anthropometric features of men with 47,XYY and 47,XY**

Compared with 46,XY men, those with KS and 47,XYY had higher BMI (beta = 1.8 kg/m², P = 8 × 10⁻¹¹ and beta = 2.2 kg/m², P = 5 × 10⁻¹¹, respectively), higher percentage of total body fat (beta = 4.8%, P = 8 × 10⁻⁸ and beta = 2.2%, P = 3 × 10⁻⁷, respectively), and weaker hand grip strength (beta = −7.1 kg, P = 1 × 10⁻³⁶ and beta = −2.6 kg, P = 3 × 10⁻⁴, respectively) (Table 2). Other features differed between the groups. Even after accounting for their taller adult heights, men with 47,XYY (but not KS) had higher total fat-free mass (beta = 3.8 Kg, P = 4 × 10⁻¹⁵). Furthermore, men with 47,XYY had slightly higher BMD (beta = 0.04 g/cm², P = 2 × 10⁻²), whereas men with KS had lower BMD (beta = −0.05 g/cm², P = 5 × 10⁻⁶) and higher likelihood of osteoporosis with pathological fractures (RR = 10.8, P = 2 × 10⁻⁷).

**Hormonal, metabolic, and vascular features of men with 47,XYY and 47,XY**

Compared with 46,XY men, those with KS (but not 47,XYY) had lower plasma total testosterone concentration (beta = −3.8 nmol/L, P = 2 × 10⁻⁵⁰). Both KS and 47,XYY men had lower plasma IGF-1 concentrations (beta = −1.7 nmol/L, P = 1 × 10⁻⁶ and beta = −2.3, P = 3 × 10⁻⁸, respectively).

We also identified several adverse metabolic and vascular health outcomes shared across both KS and 47,XYY, including higher risks for type 2 diabetes (RR = 3.0, P = 2 × 10⁻²⁰ and RR = 2.6, P = 3 × 10⁻¹⁰, respectively), albuminuria (RR = 1.9, P = 5 × 10⁻⁴ and RR = 2.4, P = 5 × 10⁻⁶, respectively), venous thrombosis (RR = 6.4, P = 3 × 10⁻²³ and RR = 7.4, P = 7 × 10⁻²², respectively), pulmonary embolism (RR = 3.3, P = 2 × 10⁻⁶ and RR = 3.7, P = 7 × 10⁻⁶, respectively), and atherosclerosis (RR = 3.1, P = 6 × 10⁻³ and RR = 5.5, P = 8 × 10⁻⁶, respectively). These disease associations were only partially attenuated after adjustments for BMI, household deprivation, and smoking (Table 3).

Exploration of red blood cell and platelet traits showed lower hematocrit and hemoglobin concentrations in KS (but not 47,XYY), was lower plasma IGF-1 concentrations (beta = 0.39, P = 8 × 10⁻³), lower levels of most HDL-related traits. Furthermore, men with KS and 47,XYY had lower levels of high-density lipoprotein (HDL) cholesterol (beta = −0.11 mmol/L, P = 2 × 10⁻⁷ and beta = −0.17 mmol/L, P = 3 × 10⁻¹¹, respectively) and men with 47,XYY (but not KS) had higher triglycerides (beta = 0.32 mmol/L, P = 1 × 10⁻⁶). In the subgroup with nuclear magnetic resonance metabolic data, we observed that KS (n = 48) and 47,XYY (n = 38) men had lower levels of most HDL-related traits. Furthermore,
Table 2  Anthropometric characteristics of Klinefelter syndrome and 47,XXY compared with men with normal (46,XY) karyotypes

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>46,XY</th>
<th>KS</th>
<th>47,XXY</th>
<th>KS vs 46,XY</th>
<th>47,XXY vs 46,XY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>27.9 (4.2)</td>
<td>29.7 (5.7)</td>
<td>30.2 (5.7)</td>
<td>1.8</td>
<td>1.2-2.3</td>
</tr>
<tr>
<td>Body fat percentage, %</td>
<td>29.5 (5.4)</td>
<td>34.3 (6.0)</td>
<td>31.6 (5.7)</td>
<td>4.8</td>
<td>4.1-5.6</td>
</tr>
<tr>
<td>Hand grip strength, kg</td>
<td>41.9 (8.9)</td>
<td>35.2 (8.9)</td>
<td>39.9 (9.7)</td>
<td>-7.1</td>
<td>-8.2 to -6.0</td>
</tr>
<tr>
<td>Fat-free mass, adj. height, kg</td>
<td>63.9 (7.7)</td>
<td>66.1 (9.4)</td>
<td>73.5 (9.2)</td>
<td>0.2</td>
<td>0.6-0.9</td>
</tr>
<tr>
<td>Bone mineral density, g/cm²</td>
<td>0.6 (0.2)</td>
<td>0.5 (0.2)</td>
<td>0.6 (0.2)</td>
<td>-0.05</td>
<td>-0.08 to -0.03</td>
</tr>
<tr>
<td>Binary traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteoporosis with pathological fracture</td>
<td>0.3%</td>
<td>2.4%</td>
<td>0.7%</td>
<td>10.8</td>
<td>4.5-25.3</td>
</tr>
<tr>
<td>Osteoporosis without pathological fracture</td>
<td>1.6%</td>
<td>8.9%</td>
<td>0.7%</td>
<td>6.2</td>
<td>4.0-8.9</td>
</tr>
</tbody>
</table>

*Beta*, regression coefficient from linear regression models; BMI, body mass index; KS, Klinefelter syndrome; RR, relative risk from logistic regression models.

| 47,XXY (but not KS) men showed lower levels across all lipid classes, apart from triglycerides (Supplemental Figure 1). |

Respiratory features of men with 47,XXY and 47,XY

Compared with 46,XY men, those with KS and 47,XXY had lower forced expiratory volume (beta = -0.68 L, P = 1 × 10⁻⁴⁶ and beta = -0.26 L, P = 2 × 10⁻⁵, respectively), peak expiratory flow (beta = -125 L/min, P = 2 × 10⁻³⁸ and beta = -68 L/min, P = 4 × 10⁻⁸, respectively) and vital capacity (beta = -0.93 L, P = 2 × 10⁻⁵⁰ and beta = -0.25 L, P = 3 × 10⁻³, respectively), and higher risks for COPD (RR = 4.4, P = 5 × 10⁻¹⁸ and RR = 4.6, P = 2 × 10⁻¹³, respectively) and asthma (RR = 2.0, P = 9 × 10⁻⁹ and RR = 1.7, P = 4 × 10⁻³, respectively). Again, these disease associations were only partially attenuated after adjustments for BMI, household deprivation, and smoking (Table 3).

Discussion

Using systematic case ascertainment in a large, unselected population of men of European ancestry aged 40 to 70 years, we report the prevalence of KS (103/100,000 men) and 47,XXY (69/100,000 men). Notably, only a small minority of these men had a diagnosis of sex chromosome abnormality on their medical records or by self-report (23% of KS and 0.7% of 47,XXY) and yet these conditions conferred substantially increased risks for multiple, potentially preventable diseases.

The underdiagnosis of KS and 47,XXY has been previously indirectly quantified in other settings, on the basis of the differences in their clinical prevalence compared with estimates from population-based cytogenetic surveys in newborn infants. Such studies estimated that only between 7% (in the United Kingdom) and 57% (in Australia) of expected KS cases were diagnosed on the basis of clinical presentation. For XYY, only between 3% (in the United Kingdom) and 18% (in Denmark) of expected cases were diagnosed.²

Our prevalence estimates in an adult study population are somewhat lower than those reported in those newborn infants (KS: 152/100,000 males and 47,XXY: 98/100,000 males).³ Although this could be interpreted as indicating higher mortality rates, it is recognized that UK Biobank comprises a more educated and healthier sample than the general population, likely owing to healthy volunteer bias.25 Similarly, the prevalence of other adverse genetic conditions is reportedly lower in UK Biobank than in other more representative studies.³⁶

Previous studies have highlighted higher disease risks in men with KS. Bojesen et al.²⁷ identified 832 men with KS from hospital records in Denmark and reported higher risks for venous thrombosis (hazard ratio = 5.3, 95% CI = 3.3-8.5), pulmonary embolism (hazard ratio = 3.6, 95% CI = 1.9-6.7), COPD (hazard ratio = 3.9, 95% CI = 2.5-6.1), type 2 diabetes (hazard ratio = 3.7, 95% CI = 2.1-6.4), and atherosclerosis (hazard ratio = 4.5, 95% CI = 2.8-7.1). Swedlow et al.²⁸ accessed data on 3518 UK patients with KS diagnosed since 1959 and followed up till mid-2003 and reported higher mortality from diabetes mellitus (standardized mortality ratio = 5.8, 95% CI = 3.4-9.3), pulmonary embolism (standardized mortality ratio = 5.7, 95% CI = 2.5-11.3), and chronic lower respiratory disease (standardized mortality ratio = 2.1, 95% CI = 1.4-3.0). Zöller et al.²⁹ reported a higher risk for venous thromboembolism (incidence rate ratio [IRR] = 6.4, 95% CI = 5.1-7.9) in 1085 men diagnosed with KS between 1969 and 2010 in Sweden. Our findings confirm these strong disease associations and also the reported higher risks of psychiatric illness and osteoporosis.

By contrast, few studies have reported on disease risks in men with 47,XXY. Berglund et al.²⁰ identified 255 men with 47,XXY from hospital records in Denmark and reported higher risks for venous thrombosis (IRR = 10.2, 95% CI = 4.6-22.6) and COPD (IRR = 5.8, 95% CI = 2.4-15.1). Our
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>46,XY Mean (SD)</th>
<th>KS Mean (SD)</th>
<th>47,XY Mean (SD)</th>
<th>KS vs 46,XY</th>
<th>Adj. BMI, TDI, and smoking</th>
<th>47,XY vs 46,XY</th>
<th>Adj. BMI, TDI, and smoking</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Continuous traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testosterone, nmol/L</td>
<td>12.0 (3.7)</td>
<td>8.2 (6.0)</td>
<td>11.5 (4.7)</td>
<td>-3.8</td>
<td>-4.3 to -3.3</td>
<td>2 x 10^-5</td>
<td>-3.3</td>
</tr>
<tr>
<td>SHBG, nmol/L</td>
<td>39.9 (16.8)</td>
<td>41.1 (22.7)</td>
<td>40.3 (20.4)</td>
<td>2.3</td>
<td>0.4-4.4</td>
<td>2 x 10^-7</td>
<td>4.2</td>
</tr>
<tr>
<td>IGF-1, nmol/L</td>
<td>21.9 (5.5)</td>
<td>20.4 (6.5)</td>
<td>20.0 (5.7)</td>
<td>-1.7</td>
<td>-2.3 to -1.0</td>
<td>1 x 10^-6</td>
<td>-1.2</td>
</tr>
<tr>
<td>HDL cholesterol, mmol/L</td>
<td>1.3 (0.3)</td>
<td>1.2 (0.3)</td>
<td>1.1 (0.3)</td>
<td>-0.11</td>
<td>-0.15 to -0.07</td>
<td>2 x 10^-7</td>
<td>-0.06</td>
</tr>
<tr>
<td>LDL cholesterol, mmol/L</td>
<td>3.5 (0.9)</td>
<td>3.4 (0.9)</td>
<td>3.4 (0.9)</td>
<td>-0.10</td>
<td>-0.22 to 0.01</td>
<td>6 x 10^-2</td>
<td>-0.06</td>
</tr>
<tr>
<td>Triglycerides, mmol/L</td>
<td>2.0 (1.2)</td>
<td>2.1 (1.0)</td>
<td>2.3 (1.4)</td>
<td>0.01</td>
<td>-0.05 to 0.25</td>
<td>2 x 10^-1</td>
<td>-0.01</td>
</tr>
<tr>
<td>Forced expiratory volume, L</td>
<td>3.3 (0.8)</td>
<td>2.7 (0.8)</td>
<td>3.2 (0.9)</td>
<td>-0.7</td>
<td>-0.8 to -0.6</td>
<td>1 x 10^-6</td>
<td>-0.6</td>
</tr>
<tr>
<td>Peak expiratory flow, L/min</td>
<td>469 (138)</td>
<td>348 (132)</td>
<td>408 (132)</td>
<td>-125</td>
<td>-144 to -106</td>
<td>2 x 10^-8</td>
<td>-117</td>
</tr>
<tr>
<td>Vital capacity, L</td>
<td>4.5 (1.0)</td>
<td>3.5 (0.9)</td>
<td>4.3 (1.0)</td>
<td>-0.9</td>
<td>-1.1 to -0.8</td>
<td>2 x 10^-5</td>
<td>-0.8</td>
</tr>
<tr>
<td><strong>Binary traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>10.3%</td>
<td>28.6%</td>
<td>25.0%</td>
<td>3.0</td>
<td>2.4-3.6</td>
<td>2 x 10^-20</td>
<td>2.3</td>
</tr>
<tr>
<td>Albuminuria</td>
<td>13.9%</td>
<td>26.6%</td>
<td>30.3%</td>
<td>1.9</td>
<td>1.3-2.6</td>
<td>5 x 10^-4</td>
<td>1.7</td>
</tr>
<tr>
<td>Venous thrombosis</td>
<td>1.8%</td>
<td>10.9%</td>
<td>12.6%</td>
<td>6.4</td>
<td>4.2-9.1</td>
<td>3 x 10^-23</td>
<td>5.4</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>2.5%</td>
<td>4.9%</td>
<td>7.5%</td>
<td>3.3</td>
<td>2.0-5.2</td>
<td>2 x 10^-6</td>
<td>2.8</td>
</tr>
<tr>
<td>Atherosclerosis</td>
<td>1.0%</td>
<td>2.8%</td>
<td>4.9%</td>
<td>3.1</td>
<td>1.4-6.8</td>
<td>6 x 10^-3</td>
<td>2.2</td>
</tr>
<tr>
<td>COPD</td>
<td>4.2%</td>
<td>16.4%</td>
<td>16.8%</td>
<td>4.4</td>
<td>3.2-6.0</td>
<td>5 x 10^-18</td>
<td>3.7</td>
</tr>
<tr>
<td>Asthma</td>
<td>11.8%</td>
<td>23.5%</td>
<td>19.6%</td>
<td>2.0</td>
<td>1.6-2.6</td>
<td>1 x 10^-7</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Adj. adjusted for: Beta, regression coefficient from linear regression models; BMI, body mass index; COPD, chronic obstructive pulmonary disease; HDL, high-density lipoprotein; KS, Klinefelter syndrome; LDL, low-density lipoprotein; RR, relative risk from logistic regression models; TDI, Townsend deprivation index.
findings confirm those observations and extend the list of diseases strongly associated with 47,XXY to also include type 2 diabetes, pulmonary embolism, and atherosclerosis.

We observed some notable differences between KS and 47,XXY. KS is a well-recognized cause of reproductive dysfunction and this was reflected in our data by their older age at puberty, lower testosterone levels, and high risk of being childless. Reproductive dysfunction likely also contributes to their tall stature (due to later pubertal growth completion), lower bone density and muscle strength, and greater adiposity. By contrast, men with 47,XXY appeared to have normal reproductive function, with no alteration in their puberty timing or testosterone levels and a more modestly higher risk of being childless, which could be explained by their similarly higher chance of living without a partner.

Hence, despite these marked differences in reproductive function, it is unclear why both KS and 47,XXY should show striking similarities in conferring substantially higher risks for many diseases in common—type 2 diabetes, atherosclerosis, venous thrombosis, pulmonary embolism, and COPD, which persisted after adjustments for several lifestyle behavioral-related traits (BMI, smoking, deprivation). Higher risks of type 2 diabetes, atherosclerosis, and microalbuminuria, with lower HDL cholesterol and higher adiposity, together indicate higher insulin resistance in both KS and 47,XXY men. Both conditions confer a triple dose of the PAR, containing the growth-promoting SHOX gene, which likely partially contributes to their tall stature, and there are case reports of insulin resistance in other conditions characterized by SHOX excess.29,30 However, the underlying mechanisms are yet unknown.

Similarly, it is unclear why risks for venous thrombosis and pulmonary embolism are raised in both KS and 47,XXY to a similar substantial degree, around 6- to 7-fold higher risk for venous thrombosis. This is similar or even higher than that conferred by factor V Leiden, a genetic variant carried by around 5% of White and of European descent individuals.31 Hence, it might be considered to add sex chromosome aneuploidy to the screening for genetic causes of thrombophilia. Furthermore, because KS and 47,XXY confer higher risks for multiple potentially preventable diseases, future studies should explore the potential benefits of wider testing.

Strengths of our study include the use of systematic case ascertainment and 100% confirmation of genome-wide association study (GWAS) array-based categorization using exome sequencing data in a large subsample. Furthermore, our numbers of individuals with a male sex chromosome aneuploidy are similar to those reported by UK Biobank, which found 355 such individuals (99.2% of our 358 cases).12 GWAS array genotyping and exome sequencing are increasingly performed in clinical settings, however male sex chromosome aneuploidy status is not routinely derived. In addition, the wide range of traits and diseases available in UK Biobank allowed us to systematically quantify the disease and phenotypic effects of male sex chromosome aneuploidy.

Limitations include the healthy volunteer bias of the UK Biobank sample and the yet incomplete linkage to primary care health data. Hence it is likely that the true disease risks associated with KS and 47,XXY analyses are even higher than the substantial estimates that we observed.

In conclusion, our findings show that male sex chromosome aneuploidy can be reliably detected using GWAS or exome sequencing data. KS and 47,XXY were mostly unrecognized but conferred substantially higher risks of diverse potentially preventable diseases, including metabolic and vascular diseases, which were only partially explained by higher levels of BMI, deprivation, and smoking. Future studies should consider the utility of deriving male sex chromosome aneuploidy status when genetic testing is undertaken for existing clinical indications, eg, for thrombosis risk. Furthermore, our findings add significantly to ongoing debates regarding the potential benefits of wider population genetic screening.

Data Availability

All data are available via the UK Biobank Access Management System https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access.

Acknowledgments

Y.Z., E.J.G., M.P., M.K., R.Y.J., H.L.A., F.R.D., C.L., J.R.B.P. and K.K.O. are supported by the Medical Research Council, United Kingdom (Unit programs: MC_UU_00006/1 and MC_UU_00006/2). M.K. is supported by a Gates Fellowship. A.R.W. and J.T. hold Academy of Medical Sciences Springboard awards [SBF004\1079 and SBF006\1134], which are supported by the Academy of Medical Sciences, United Kingdom; the Wellcome Trust; the Government Department of Business, Energy and Industrial Strategy; the British Heart Foundation; and Diabetes UK, United Kingdom. R.N.B. and A.M. are supported by the Medical Research Council (MR/T00200X/1). K.S.R. is supported by Cancer Research UK, United Kingdom [C18281/A29019]. The authors acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work. The funders had no role in the design, analysis, and interpretation of data; writing of the report; or the decision to submit the article for publication.

Author Information

The online version of this article (https://doi.org/10.1016/j.xxx.2021.05.006) is available to authorized users.

**Ethics Declaration**

UK Biobank has approval from the North West Multi-centre Research Ethics Committee (REC reference 13/NW/0157, https://www.ukbiobank.ac.uk/media/lcvbdoik/21-nw-0157-favourable-opinion-with-conditions-18-06-2021.pdf) as a Research Tissue Bank (RTB) approval and informed consent (https://www.ukbiobank.ac.uk/media/t2hbbo35/consent-form.pdf) was provided by each participant. This approval means that researchers do not require separate ethical clearance and can operate under the RTB approval. This RTB approval was granted initially in 2011 and it is a renewal on a 5-yearly cycle; hence UK Biobank successfully applied to renew it in 2016 and 2021. This research was conducted using the UK Biobank resource under applications 9905 (https://www.ukbiobank.ac.uk/enable-your-research/approved-research/using-genetics-to-elucidate-the-relationship-between-socioeconomic-behavioural-traits-and-disease), 9072 (https://biobank.ctsu.ox.ac.uk/crystal/app.cgi?id=9072), and 49847 (https://www.ukbiobank.ac.uk/enable-your-research/approved-research/assessing-the-prevalence-penetrance-and-pathogenicity-of-rare-genomic-variants).

**Conflict of Interest**

The authors declare no conflicts of interest.

**Additional Information**

The online version of this article (https://doi.org/10.1016/j.gim.2022.05.011) contains supplementary material, which is available to authorized users.

**Affiliations**

1MRC Epidemiology Unit, Institute of Metabolic Science, School of Clinical Medicine, Cambridge University, Cambridge, United Kingdom; 2Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Royal Devon & Exeter Hospital, Exeter, United Kingdom; 3Computational Medicine, Berlin Institute of Health (BIH) at Charité, Universitätmedizin Berlin, Berlin, Germany; 4Institute for Molecular Medicine Finland (FiMM), Helsinki, Finland

**References**


