A meta-analysis of math performance in Turner syndrome

JOSEPH M BAKER[1] | ALLAN L REISS[1,2]

1 Division of Brain Sciences, Center for Interdisciplinary Brain Sciences Research, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA; 2 Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA.

Correspondence to Joseph M Baker at Division of Brain Sciences, Center for Interdisciplinary Brain Sciences Research, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Main Administration, 401 Quarry Road, Stanford, CA 94305-5717, USA. E-mail: jbaker2@stanford.edu

AIM Studies investigating the relationship between Turner syndrome and math learning disability have used a wide variation of tasks designed to test various aspects of mathematical competencies. Although these studies have revealed much about the math deficits common to Turner syndrome, their diversity makes comparisons between individual studies difficult. As a result, the consistency of outcomes among these diverse measures remains unknown. The overarching aim of this review is to provide a systematic meta-analysis of the differences in math and number performance between females with Turner syndrome and age-matched neurotypical peers.

METHOD We provide a meta-analysis of behavioral performance in Turner syndrome relative to age-matched neurotypical populations on assessments of math and number aptitude. In total, 112 comparisons collected across 17 studies were included.

RESULTS Although 54% of all statistical comparisons in our analyses failed to reject the null hypothesis, our results indicate that meaningful group differences exist on all comparisons except those that do not require explicit calculation.

INTERPRETATION Taken together, these results help elucidate our current understanding of math and number weaknesses in Turner syndrome, while highlighting specific topics that require further investigation.

The capacity to represent number is an adaptive evolutionary ability shared by humans and animals alike.[1] At the heart of this ability, an inherent number sense is thought to underlie our preverbal representation of number and ultimately facilitate higher-order numerical processes such as mathematics.[2,3] Deficits in such processes (i.e. dyscalculia) lead to an impaired number sense that manifests as poor behavioral performance on tests of math and number aptitude. Given the importance of math and number processing in our daily lives, understanding the cause of dyscalculia has important educational and social implications. Notably, dyscalculia has been associated with specific chromosomal abnormalities such as Turner syndrome.[4] Given the high rates of dyscalculia among females with Turner syndrome, this group represents an ideal population in which to study math learning disabilities.[5,6]

Turner syndrome is a relatively common genetic disorder that affects roughly 1 in 1900 live female births and is characterized by partial or complete monosomy X.[7] Symptoms of Turner syndrome often include observable physical features (e.g. short stature and webbed neck), as well as cardiovascular and endocrine problems related to ovarian dysgenesis. As mentioned, Turner syndrome is commonly associated with performance deficits in mathematics, although overall IQ generally remains intact.[8–10] Impor-

tantly, compared with neurotypical peers, the known cognitive deficits related to Turner syndrome provide valuable contrasts that help elucidate the complex behavioral, genetic, and neural etiology of specific learning disorders such as dyscalculia.[11]

Multiple studies have identified a relationship between Turner syndrome and poor math performance in populations ranging from children to adults.[12–19] Importantly, this trend has emerged despite a wide variety of tools used to assess math performance across each study. However, as discussed by Murphy et al.,[18,19] the diverse range of assessment tools across studies has also led to inconsistent findings. For instance, studies conducted by Rovet,[12] Mazzocco,[15] Collaer et al.,[20] and Molko et al.[4] each provided Turner syndrome and age-matched neurotypical peers with basic arithmetic assessments (e.g. addition, subtraction). Rovet[12] and Mazzocco[15] both used standardized measures (Wide Range Achievement Test-Revised and KeyMath respectively). Rovet identified a significant statistical difference between Turner syndrome and neurotypical females, whereas the study by Mazzocco did not. Conversely, the studies by Collaer et al.[20] and Molko et al.[4] both used non-standardized arithmetic measures that were similar in scope to the standardized measures mentioned above. In a pattern similar to that of Rovet[12] and

Mazzocco,[15] a significant statistical difference was only identified in the study by Collaer et al.[20] Thus, despite each comparison measuring arithmetic performance within age-matched cohorts, these studies do not identify a consistent trend in outcomes.

A closer examination of the effect sizes for each comparison identifies moderate to large values of Cohen's d, ranging between 0.44 and 0.84. As reported by Cohen,[21] effect sizes in this range indicate that the mean performance of the neurotypical group was between the 66th and 79th centiles of the performance of the group with Turner syndrome. That is, the average neurotypical performance was better than the top 66% to 79% of all performance ranges in the group with Turner syndrome. Thus, these values are large enough to infer that meaningful performance differences may exist between Turner syndrome and neurotypical populations on each measure, but that measurement errors resulting from insufficient sample size and statistical power might have resulted in a failure to reject the null hypothesis (i.e. type II error). Importantly, comparison of values of Cohen's d for effect size allows an assessment of differences between Turner syndrome and neurotypical peers across different math assessment tools that are less biased than the interpretation of p values alone.[21]

Most math and number studies included in this review (61%) used standardized tests to quantify differences in math aptitude between those with Turner syndrome and neurotypical age-matched comparison groups. These tests included the arithmetic subscales of the Wechsler intelligence tests,[8,12,22–26] the calculations subtest of the Woodcock–Johnson Test of Cognitive Ability,[19,27,28] the Test of Early Mathematics Ability—Second Edition,[15,19,28] the arithmetic subscale of the Wide Range Achievement Test,[12,25,29] the KeyMath Diagnostic Assessment,[15] the quantitative reasoning subscale of the Stanford–Binet Intelligence Scale,[15] and the numerical ability subscale of the Differential Aptitude Test.[20] Additionally, a subset of these studies also included non-standardized tests that measure various aspects of math ability.[6,18–20,28] Finally, 27% of studies comparing Turner syndrome and neurotypical populations relied solely on non-standardized measures to quantify math performance.[14,16,17,30,31]

In an effort to determine whether the combined effect of each comparison differs significantly from zero, we conducted a meta-analysis of each math and number aptitude comparison made between females with Turner syndrome and neurotypical comparisons. Furthermore, closer inspection of each measure identified three dichotomous categories into which each comparison could be classified. These categories included assessment type (i.e. standardized vs non-standardized), primary outcome variable (i.e. accuracy vs response time), and question type (i.e. calculation vs non-calculation). In an effort to elucidate the effect of each category on math comparisons effects, we conducted individual three-level meta-analyses using each category as a covariate in the model.

What this paper adds
- Meta-analysis provides a detailed description of math-related deficits in females with Turner syndrome.
- Math performance in Turner syndrome is severely affected by questions requiring a speeded response.
- Performance on questions that do not require explicit calculation is similar to neurotypical age-matched peers.

METHOD

Identification of comparisons

Several approaches were used to identify the relevant comparisons that have been published in peer-reviewed journals. Most comparisons were identified by a computer-based search of the PsychINFO, MEDline, ERIC, PubMed, Embase, Web of Science, and Google Scholar databases. The search used the following terms: 'Turner syndrome math', 'Turner syndrome math deficits', 'Turner syndrome math number', 'Turner syndrome cognitive deficits', and 'aneuploidy math number'. Collectively, these methods identified 26 articles reporting a total of 147 relevant comparisons. The pool of comparisons was then narrowed to 112 comparisons across 17 individual articles that fitted the criteria below (see Fig. 1).

Comparison exclusion criteria

The primary focus of this review is to facilitate a better understanding of the differences in the observed math and number performances between Turner syndrome and neurotypical groups. As a result, studies reporting reviews of Turner syndrome math and number processing,[8,9,32–36] comparisons between Turner syndrome and other atypical populations,[37,38] and within-group Turner syndrome comparisons[39–45] were not included. Furthermore, comparisons were excluded if they did not provide data necessary to calculate both Cohen's d[46–50] and the pooled variance needed for the effect size weighting procedure conducted within the random-effects model.[4,22,31]

Data subset procedure

As discussed above, each assessment included in the meta-analysis was dichotomously classified across three categories including assessment type (standardized vs non-standardized), primary outcome variable (accuracy vs response time), and question type (calculation vs non-calculation). Table I provides a breakdown of the number of each category drawn from all 17 studies included in the meta-analysis.

Assessment type

All comparisons that used an age-normed test of math aptitude were classified as 'standardized' ($n=37$) (see Fig. 1 for complete list of standardized assessments and corresponding studies). All remaining comparisons were classified as 'non-standardized' ($n=75$).

Primary outcome variable

Each assessment was further categorized on the basis of its use of accuracy ($n=84$) or response time ($n=28$) as a primary outcome variable. All standardized assessments used

(a) Databases searched

PsychInfo | MEDline | ERIC | PubMed | Embase | Web of Science | Google Scholar

(b) Search results

Total articles identified = 26

Total comparisons identified = 147

(c) Search reduction

Articles excluded = 9
Non-Turner syndrome/control comparisons = 15
Within Turner syndrome comparisons = 9
Insufficient data for effect size calculation = 11

Articles included = 17

Relevant comparisons meta-analyzed = 112

(d) Meta-analyses

Assessment type
Standardized = 37
Non-standardized = 75

Primary outcome variable
Accuracy = 84
Response time = 28

Comparison type
Calculation = 73
Other = 39

Figure 1: Article/comparison selection and reduction.

accuracy as a primary outcome variable. The non-standardized assessments were reviewed to determine the outcome variable used. In total, 24.1% of all comparisons used response time as a primary outcome variable.

Question type
Each assessment was then categorized on the basis of whether or not the questions asked required explicit calculation to complete. Calculation questions (n=73) included addition, subtraction, multiplication, division, and geometry. Non-calculation questions (n=39) included bisection, counting, copying digits, digit comparisons, magnitude judgments, number comparisons, problem verification, procedural involvement, quantitative reasoning, reading numbers, subitizing, transcoding numbers, and writing numbers.

Meta-analytical modeling
Meta-analysis outcomes are influenced by a combination of within-study factors, such as the type of comparison that yields each effect size included in the analysis, and between-studies correlations (i.e. dependence) that arise when multiple effect sizes are drawn from a single study cohort.[51,52] For example, in cases such as ours, where relatively few studies (n=17) report a large number of individual comparisons (n=112), it is likely that high levels of dependence exist between effect sizes because many were estimated on the basis of the same group of participants. Thus, when interpreting a meta-analytical model, it is important that the influence of between-comparison dependence is known. To accomplish this, we used a three-level structural equation modeling approach to all parameter estimations. As described in detail by Cheung,[51]

Table I: Breakdown of the number of comparison categories drawn from all 17 studies included in the meta-analysis

References	Assessment type		Dependent variable		Question type		Total comparisons
	Standardized	Non-standardized	Accuracy	Response time	Calculation	Other	
Bender et al.[27]	1	—	1	—	1	—	1
Bruandet et al.[16]	—	16	7	9	8	8	16
Collaer et al.[20]	1	3	4	—	3	1	4
Downey et al.[24]	2	—	2	—	2	—	2
Kesler et al.[17]	—	4	2	2	4	—	4
Mazzocco[15]	5	—	5	—	2	3	5
Mazzocco et al.[28]	4	2	6	—	2	4	6
McCauley et al.[23]	1	—	1	—	1	—	1
Molko et al.[4]	1	—	1	—	1	—	1
Murphy and Mazzocco[19]	1	2	2	1	1	2	3
Romans et al.[25]	4	—	4	—	4	—	4
Ross et al.[26]	1	—	1	—	1	—	1
Rovet[12]	2	—	2	—	2	—	2
Rovet et al.[13]	13	—	13	—	13	—	13
Siegel et al.[29]	1	—	1	—	1	—	1
Temple and Marriott[14]	—	29	18	11	20	9	29
Temple and Sherwood[30]	—	19	14	5	7	12	19

this approach allows an estimation of the amount of heterogeneity within each comparison effect size (i.e. level 2), and between each study included in the meta-analysis (i.e. level 3).

When the intraclass correlation between studies is zero, a three-level model reduces to a two-level random-effects model ($y_i=\beta_0+u_i+e_i$). Here, $\mathrm{Var}(e_i)$ is the known sampling variance in the ith comparison, β_0 is the average population effect, and $\mathrm{Var}(u_i)$ is the study-specific heterogeneity (τ^2) that must be estimated.[53] Like traditional random-effects models, letting our estimation of τ^2 vary from zero allows our model inferences to be generalized beyond the studies included in our analyses.[51,53,54] Notably, a large τ^2 value indicates that the population effect sizes are likely to be heterogeneous, suggesting that the magnitude of effect sizes varies across different comparisons included in the analysis.

To explain such heterogeneity using individual comparison characteristics (e.g. assessment type, primary outcome variable, question type) as predictors, the standard random-effects model may be written as a mixed-effects model ($y_i=\beta_0+\beta_i x_i+u_i+e_i$), wherein β_0 and β_1 are the intercept and regression coefficients respectively, and $\mathrm{Var}(u_i)$ is the residual heterogeneity after controlling for the covariate x_1. When applied to meta-analyses, this standard mixed-effects model is written as $\lambda_i=\beta_0+\beta_i x_i+u_i$, where λ_i is the 'true' effect size in the ith study. Importantly, the two-level meta-analytical model may be extended to a three-level model by adding a cluster variable, which serves to group dependent effect sizes together during parameter estimation.[51] Thus, the level 2 model becomes $\lambda_{ij}=k_j+u_{(2)ij}$, where λ_{ij} is the 'true' ith effect size in the jth cluster (i.e. study cohort), k_j is the average effect in the jth cluster, and $\mathrm{Var}(u_{(2)ij})$ is the study-specific heterogeneity for level 2 of the model. Finally, the level 3 model may be written as $k_j=\beta_0+u_{(3)ij}$, where $\mathrm{Var}(u_{(3)ij})$ is the study-specific heterogeneity for level 3 of the model.

The significance of each three-level model is determined in multiple ways. First, the Q statistic[55] tests the null hypothesis that all effect sizes included in the model ($k=112$) are homogenous. The Q statistic is dependent on the entire population of effect sizes included in the meta-analysis, and is equal to two-level meta-analysis models.[51] Second, estimates of effect size and corresponding p values are calculated for each category-specific subgroup pair (e.g. standardized/non-standardized, accuracy/response time, or calculation/non-calculation) that is included as a covariate in each model. Third, the regression coefficient (β_1) and corresponding p value describe the degree to which the subgroups' estimated effect sizes differ. A significant regression coefficient indicates that the estimated effect sizes for both subgroups are significantly different from each other. Fourth, estimates of heterogeneity (τ^2) at levels 2 and 3 are made, along with corresponding p values. Significant τ^2 values indicate that the distribution of effect sizes is non-normal across individual comparisons (level 2) or individual study cohorts (level 3) included in the meta-analysis. Finally, the amount of variance in heterogeneity estimations accounted for by the category-specific subgroups (i.e. R^2) at levels 2 and 3 of the model are provided.

RESULTS

All analyses were conducted using the metaSEM package[56] in *R* (R Foundation for Statistical Computing, Vienna Austria).[57] A value of α of 0.05 was used for each analysis. Correction for inflated type I error due to repeat testing was accomplished using the false discovery rate correction procedure.[58]

Complete data set

The results of the complete data-set analysis ($n=112$) identified a significant Q statistic ($Q=159.99$, $p=0.003$), indicating that the distribution of all effect sizes, without respect to category-specific covariates, is heterogeneous. This

model identified a moderate combined effect of 0.67, indicating that the average neurotypical performance is at the 74th centile of the group with Turner syndrome. This analysis yielded a significant z-value of 3.933 ($p<0.001$, 95% Wald confidence interval [CI] 0.332–0.992). These results indicate that the combined effect of all math and number aptitude comparisons is significantly greater than zero, suggesting that meaningful differences exist between the Turner syndrome and neurotypical groups. This result is important, given that over half of the total comparisons included in the meta-analysis (54%) failed to reject the null hypothesis. Notably, the test of heterogeneity of level 2 (comparisons) was significant ($\tau^2=0.418$, $p=0.004$, 95% Wald CI 0.131–0.707), indicating that the distribution of effect sizes across all 112 comparisons was non-normal. However, the test of level 3 heterogeneity was not significant ($\tau^2=0.031$, $p=0.694$, 95% Wald CI −0.125 to 0.187), indicating that variations in effect sizes were normally distributed across each study included in the meta-analysis.

Standardized versus non-standardized comparisons

To test for differences in effect size across standardized and non-standardized assessments, 'assessment type' was entered as a covariate into the same model reported above. This model identified significant combined-effect size estimates for both standardized (estimated effect size=1.221, $z=2.980$, $p=0.003$, CI 0.417–2.023) and non-standardized (estimated effect size=0.562, $z=3.107$, $p=0.002$, CI 0.207–0.917) comparisons (see Fig. 2) These values indicate that
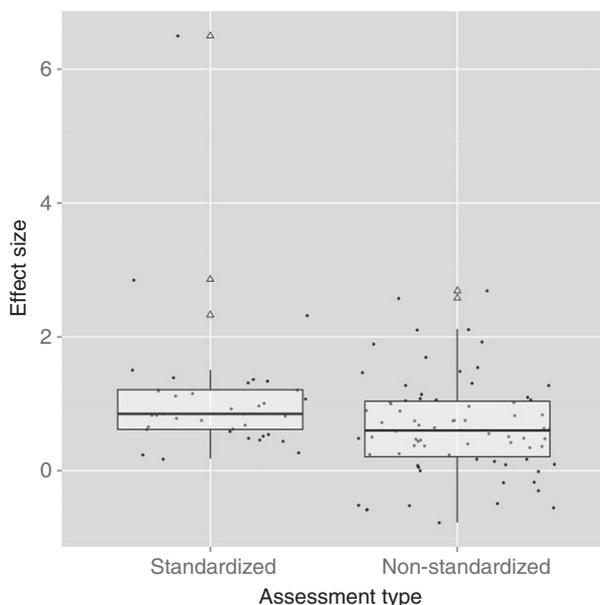
Figure 2: Distribution of effect sizes for all standardized and non-standardized comparisons. The boxplots provide the interquartile range of each effect size for both assessment types. The bold horizontal line in the center of each box provides the median effect size. Hollow triangles represent outlying effect sizes and correspond to neighboring points on the plot.

the average neurotypical control performance was at the 88th and 70th centiles of the group with Turner syndrome performance on standardized and non-standardized assessments respectively. Positive z-values indicate that the control group outperformed the group with Turner syndrome on both standardized and non-standardized comparisons. A non-significant regression coefficient ($\beta_1=0.658$, $z=1.481$, $p=0.138$, 95% Wald CI −0.213 to 1.529) indicates that the estimated effect sizes for standardized and non-standardized comparisons did not differ.

The test of heterogeneity at level 2 was significant ($\tau^2=0.401$, $z=2.819$, $p=0.004$, 95% Wald CI 0.122–0.679), indicating a non-normal distribution of effect sizes across standardized and non-standardized comparisons. However, a small R^2 value of 0.043 indicates that only 4.3% of the estimated heterogeneity at level 2 was accounted for by the comparison type. The test of heterogeneity at level 3 failed to reject the null hypothesis ($\tau^2=0.022$, $z=0.347$, $p=0.729$, 95% Wald CI −0.100 to 0.143), indicating that the distribution of effect sizes across individual studies was normal. Contrary to level 2, the R^2 value for level 3 accounted for 31.01% of estimated heterogeneity in effect sizes, indicating that the study from which each comparison was derived explained a significant proportion of the variance between standardized and non-standardized effect sizes.

Accuracy versus response time comparisons

Similar to the model described above, to test for differences in effect sizes between comparisons using accuracy or response time as primary outcome variables, 'primary outcome variable' was entered as a covariate in our model. This model identified significant effect sizes for assessments using both accuracy (estimated effect size=0.339, $z=2.041$, $p=0.041$, 95% Wald CI 0.014–0.666) and response time (estimated effect size=1.719, $z=6.047$, $p<0.001$, 95% Wald CI 1.162–2.276) outcomes (see Fig. 3). Thus, the control group significantly outperformed the group with Turner syndrome on comparisons using both outcome variables. However, a significant regression coefficient ($\beta_1=-1.379$, $z=-4.791$, $p<0.001$, 95% Wald CI −1.944 to −0.815) indicated that the effect size for response time comparisons was significantly greater than accuracy comparisons. This outcome is apparent in the centile standing of neurotypical compared with the group with Turner syndrome performance: average control group performance was at the 63rd centile of the group with Turner syndrome for accuracy-based comparisons, and was at the 96th centile for response time comparisons. This result is important, as it suggests that processing speed is significantly more impaired in females with Turner syndrome than processes needed for task accuracy.

The level 2 test of heterogeneity was not significant ($\tau^2=0.157$, $z=1.903$, $p=0.057$, 95% Wald CI −0.005 to 0.319), indicating that the effect size distribution across comparisons was normal. The corresponding R^2 value for level 2 of the model indicated that the primary outcome variable accounted for 62.44% of the estimated effect-size
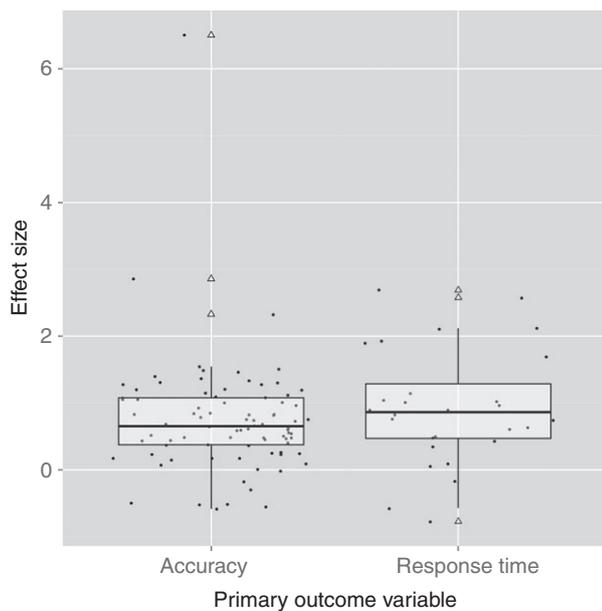
Figure 3: Distribution of effect sizes for all accuracy and response time comparisons. The boxplots provide the interquartile range of each effect size for both primary outcome variables. The bold horizontal line in the center of each box provides the median effect size. Hollow triangles represent outlying effect sizes and correspond to neighboring points on the plot.

Figure 4: Distribution of effect sizes for all calculation and non-calculation comparisons. The boxplots provide the interquartile range of each effect size for both question types. The bold horizontal line in the center of each box provides the median effect size. Hollow triangles represent outlying effect sizes and correspond to neighboring points on the plot.

heterogeneity. Moreover, the level 3 test of heterogeneity failed to reject the null hypothesis (τ^2=0.030, z=0.365, p=0.715, 95% Wald CI −0.131 to 0.192), highlighting a normal distribution of effect sizes across studies. The R^2 value for level 3 of the model only accounted for 3.47% of the estimated effect-size heterogeneity.

Calculation versus non-calculation comparisons

To test for differences in effect sizes between comparisons that did or did not require calculation, 'question type' was entered into our model as a covariate. Interestingly, this model highlighted a highly significant effect size for questions requiring calculation (estimated effect size=1.099, z=5.391, p<0.001, 95% CI 0.700–1.499), while the effect size estimation for non-calculation questions was not significant (estimated effect size=0.101, z=0.425, p=0.671, 95% CI −0.365 to 0.567) (see Fig. 4). Thus, the average comparison group was at the 85th centile of the group with Turner syndrome performance for calculation-based measures, but was nearly identical (i.e. 54th centile) as non-calculation-based comparisons. This finding is important as it suggests that the underlying cognitive processes that give rise to formal mathematics (e.g. abstract number discrimination)[59] may be preserved in females with Turner syndrome. This theory is supported by a significant regression coefficient (β_1=0.998, z=3.867, p<0.001, 95% Wald CI 0.492–1.505), which highlights a significantly greater effect size for comparisons requiring calculation than those that do not.

The tests for heterogeneity at levels 2 (τ^2=0.232, z=1.934, p=0.053, 95% Wald CI −0.003 to 0.467) and 3 (τ^2=0.028, z=0.329, p=0.741, 95% Wald CI −0.142 to 0.199) failed to reject the null hypothesis, indicating that the distribution of effect sizes across individual comparisons and studies was normal. R^2 values indicate that 44.6% of the variance in heterogeneity across comparisons was explained by question type, whereas 8.05% was explained by the study from which the comparison was drawn.

DISCUSSION

The results of our meta-analysis indicate that a meaningful group difference exists between females with Turner syndrome and age-matched neurotypical peers across all measures of math and number aptitude. This outcome emerges despite a majority of comparisons (54%) yielding non-significant statistical outcomes, indicating that high levels of false-negative outcomes (i.e. type II errors) have probably led to an underestimation of the severity of math and number performance deficits in the population with Turner syndrome. Thus, this review serves to highlight greater levels of consistency among measures of math and number aptitude that are missed when interpretations are made on the basis of individual comparisons alone.

A primary strength of our results lies in the identification of mathematical domains that pose the greatest difficulty to females with Turner syndrome. In particular, Turner syndrome performance was severely affected by questions that required a speeded response time. As discussed above, the combined effect size for all response time

comparisons was large (1.719) and indicated that the average neurotypical group performance was at the 95th centile of the group with Turner syndrome. Based on the statistical outcomes reported for these comparisons, of which only 29.1% were statistically significant, this result was unexpected and highlights a major discrepancy between meaningful effects and statistical outcomes. It is important to note that most of these comparisons (86.2%) were reported within only three studies. However, tests of heterogeneity of effect sizes across studies indicated that effect sizes were normally distributed. Thus, the females with Turner syndrome recruited for participation in these studies were representative of the greater population with Turner syndrome and probably did not artificially exaggerate the severity of response time deficits. Alternatively, non-significant findings may have been due to the authors' conservative corrections for type I error that arises with repeated testing. It is important that future studies address these issues directly by using response time measures across a diverse population of females with Turner syndrome, and by using a reasonable amount of individual assessments so that meaningful differences are not lost to statistical correction for inflated type I errors.

The role of processing speed as an underlying source of math and other cognitive deficits in Turner syndrome has previously been discussed elsewhere.[6,10,31] Inconsistent response time outcomes across various tests of executive function have led some researchers to reject the claim that a general slowing of processing speed in the population with Turner syndrome drives their performance differences.[6] However, as evidenced by the response time results above, meaningful response time differences may be masked by non-significant statistical outcomes. Thus, a comprehensive meta-analysis of all executive function task comparisons made between females with Turner syndrome and age-matched neurotypical peers is warranted to elucidate the role of processing speed on cognitive deficits. Importantly, clarification of the role of processing speed on cognitive deficits in the population with Turner syndrome may have important clinical implications.

Moreover, the identification of mathematical domains in which females with Turner syndrome perform equally well as their neurotypical peers may provide important insight into the etiology of the math deficits common within the population with Turner syndrome. On the basis of the current results, similar behavioral performance between females with Turner syndrome and neurotypical controls arises on math and number assessments that do not require explicit calculation. These results may suggest that the underlying cognitive processes that give rise to mathematics, such as abstract number processing,[59] may be intact in females with Turner syndrome, and that the observed deficits in mathematics may stem from higher-order cognitive processes related to executive functioning. Alternatively, non-calculation-based assessments may recruit verbal cognitive processes that are closely related to math and number processing.[3] As females with Turner syndrome are known to possess unimpaired verbal skills,[10] these results may be indicative of the reliance of the Turner syndrome population on verbal cognitive processes during certain mathematical tasks. Thus, females with Turner syndrome may benefit from mathematics instruction that provides sufficient time to answer questions, and which recruits sufficient levels of verbal cognitive processes.

As discussed above, inclusion of multiple comparisons drawn from a single study cohort may result in high levels of dependence across comparisons.[51,52] That is, individual study cohorts may have unique performance characteristics that influence their assessment outcomes. When one study cohort completes multiple assessments, the effect size of each outcome is similarly influenced by the group's unique performance characteristics. When compared with the entire population, these dependent comparisons may unfairly influence the distribution of effect sizes. Through multi-level modeling, we were able to model the dependence (i.e. heterogeneity) that was present across each comparison (level 2), as well as across each individual study (level 3) included in our analyses. Importantly, no comparison rejected the null hypothesis that variance in effect sizes across individual study cohorts was normal. Thus, these results indicate that inclusion of multiple comparisons from a single study population did not influence the outcomes of our meta-analyses.

REFERENCES

1. Brannon EM, Roitman JD. Nonverbal representations of time and number in animals and human infants. In: Meck WH, editor. Functional and Neural Mechanisms of Interval Timing. Boca Raton, FL: CRC Press, 2003: 143–82.

2. Nieder A, Dehaene S. Representation of number in the brain. *Annu Rev Neurosci* 2009; **32**: 185–208.

3. Dehaene S. The Number Sense: How the Mind Creates Mathematics. New York, NY: Oxford University Press, 2011.

4. Molko N, Cachia A, Rivière D, et al. Functional and structural alterations of the intraparietal sulcus in a developmental dyscalculia of genetic origin. *Neuron* 2003; **40**: 847–58.

5. Mazzocco MM. Mathematical learning disability in girls with Turner syndrome: a challenge to defining MLD and its subtypes. *Dev Disabil Res Rev* 2009; **15**: 35–44.

6. Mazzocco MM, Hanich LB. Math achievement, numerical processing, and executive functions in girls with Turner syndrome: do girls with Turner syndrome have

math learning disability? *Learn Individ Diff* 2010; **20**: 70–81.

7. Davenport ML, Hooper SR, Zeger M. Turner syndrome in childhood. In: Mazzocco MM, Ross JL, editors. Neurogenetic Developmental Disorders: Variation of Manifestation in Childhood. Cambridge, MA: MIT Press, 2007: 3–45.

8. Rovet J, Ireland L. Behavioral phenotype in children with Turner syndrome. *J Pediatr Psychol* 1994; **19**: 779–90.

9. Rovet J. Turner syndrome: genetic and hormonal factors contributing to a specific learning disability profile. *Learn Disabil Res Pract* 2004; **19**: 133–45.

10. Hong D, Scaletta Kent J, Kesler S. Cognitive profile of Turner syndrome. *Dev Disabil Res Rev* 2009; **15**: 270–8.

11. Butterworth B, Kovas Y. Understanding neurocognitive developmental disorders can improve education for all. *Science* 2013; **340**: 300–5.

12. Rovet JF. The psychoeducational characteristics of children with Turner syndrome. *J Learn Disabil* 1993; **26**: 333–41.

13. Rovet J, Szekely C, Hockenberry MN. Specific arithmetic calculation deficits in children with Turner syndrome. *J Clin Exp Neuropsychol* 1994; **16**: 820–39.

14. Temple CM, Marriott AJ. Arithmetical ability and disability in Turner's syndrome: a cognitive neuropsychological analysis. *Dev Neuropsychol* 1998; **14**: 47–67.

15. Mazzocco MM. Math learning disability and math LD subtypes evidence from studies of Turner syndrome, fragile X syndrome, and neurofibromatosis type 1. *J Learn Disabil* 2001; **34**: 520–33.

16. Bruandet M, Molko N, Cohen L, Dehaene S. A cognitive characterization of dyscalculia in Turner syndrome. *Neuropsychologia* 2004; **42**: 288–98.

17. Kesler SR, Menon V, Reiss AL. Neurofunctional differences associated with arithmetic processing in Turner syndrome. *Cereb Cortex* 2006; **16**: 849–56.

18. Murphy MM, Mazzocco MM, Gerner G, Henry AE. Mathematics learning disability in girls with Turner syndrome or fragile X syndrome. *Brain Cogn* 2006; **61**: 195–210.

19. Murphy MM, Mazzocco MM. Mathematics learning disabilities in girls with fragile X or Turner syndrome during late elementary school. *J Learn Disabil* 2008; **41**: 29–46.

20. Collaer ML, Geffner ME, Kaufman FR, Buckingham B, Hines M. Cognitive and behavioral characteristics of turner syndrome: exploring a role for ovarian hormones in female sexual differentiation. *Horm Behav* 2002; **41**: 139–55.

21. Cohen J. A power primer. *Psychol Bull* 1992; **112**: 155.

22. Waber DP. Neuropsychological aspects of Turner's syndrome. *Dev Med Child Neurol* 1979; **21**: 58–70.

23. McCauley E, Kay T, Ito J, Treder R. The Turner syndrome: cognitive deficits, affective discrimination, and behavior problems. *Child Dev* 1987; **58**: 464–73.

24. Downey J, Elkin EJ, Ehrhardt AA, Meyerbahlburg HFL, Bell JJ, Morishima A. Cognitive-ability and everyday functioning in women with Turner syndrome. *J Learn Disabil* 1991; **24**: 32–9.

25. Romans SM, Stefanatos G, Roeltgen DP, Kushner H, Ross JL. Transition to young adulthood in Ullrich-Turner syndrome: neurodevelopmental changes. *Am J Med Genet* 1998; **79**: 140–7.

26. Ross J, Stefanatos G, Kushner H, Zinn A, Bondy C, Roeltgen D. Persistent cognitive deficits in adult women with Turner syndrome. *Neurology* 2002; **58**: 218–25.

27. Bender BG, Linden MG, Robinson A. Neuropsychological impairment in 42 adolescents with sex-chromosome abnormalities. *Am J Med Genet* 1993; **48**: 169–73.

28. Mazzocco MM, Singh Bhatia N, Lesniak-Karpiak K. Visuospatial skills and their association with math performance in girls with fragile X or Turner syndrome. *Child Neuropsychol* 2006; **12**: 87–110.

29. Siegel PT, Clopper R, Stabler B. The psychological consequences of Turner syndrome and review of the National Cooperative Growth Study psychological substudy. *Pediatrics* 1998; **102** (Suppl. 3): 488–91.

30. Temple CM, Sherwood S. Representation and retrieval of arithmetical facts: developmental difficulties. *Q J Exp Psychol A* 2002; **55**: 733–52.

31. Simon T, Takarae Y, DeBoer T, McDonald-McGinn D, Zackai E, Ross J. Overlapping numerical cognition impairments in children with chromosome 22q11.2 deletion or Turner syndromes. *Neuropsychologia* 2008; **46**: 82–94.

32. Geary DC. Mathematical disabilities: cognitive, neuropsychological, and genetic components. *Psychol Bull* 1993; **114**: 345.

33. Ansari D, Karmiloff-Smith A. Atypical trajectories of number development: a neuroconstructivist perspective. *Trends Cogn Sci* 2002; **6**: 511–6.

34. Mazzocco MM. Challenges in identifying target skills for math disability screening and intervention. *J Learn Disabil* 2005; **38**: 318–23.

35. Mazzocco MM, editor. The cognitive phenotype of turner syndrome: specific learning disabilities. *Int Congr Ser* 2006; **1298**: i–xiii, 83–92.

36. De Smedt B, Swillen A, Verschaffel L, Ghesquiere P. Mathematical learning disabilities in children with 22q11.2 deletion syndrome: a review. *Dev Disabil Res Rev* 2009; **15**: 4–10.

37. Alexander D, Money J. Turner's syndrome and Gerstmann's syndrome: neuropsychologic comparisons. *Neuropsychologia* 1966; **4**: 265–73.

38. Ross JL, Roeltgen D, Kushner H, Wei F, Zinn AR. The Turner syndrome-associated neurocognitive phenotype maps to distal Xp. *Am J Hum Genet* 2000; **67**: 672–81.

39. Money J, Alexander D. Turner's syndrome: further demonstration of the presence of specific cognitional deficiencies. *J Med Genet* 1966; **3**: 47.

40. Reske-Nielsen E, Christensen A-L, Nielsen J. A neuropathological and neuropsychological study of Turner's syndrome. *Cortex* 1982; **18**: 181–90.

41. Lahood BJ, Bacon GE. Cognitive abilities of adolescent Turner's syndrome patients. *J Adolesc Health* 1985; **6**: 358–64.

42. Ross J, Roeltgen D, Zinn A. Cognition and the sex chromosomes: studies in Turner syndrome. *Horm Res Paediatr* 2006; **65**: 47–56.

43. Ross JL, Mazzocco MM, Kushner H, Kowal K, Cutler GB, Roeltgen D. Effects of treatment with oxandrolone for 4 years on the frequency of severe arithmetic learning disability in girls with Turner syndrome. *J Pediatr* 2009; **155**: 714–20.

44. Kesler SR, Sheau K, Koovakkattu D, Reiss AL. Changes in frontal-parietal activation and math skills performance following adaptive number sense training: preliminary results from a pilot study. *Neuropsychol Rehabil* 2011; **21**: 433–54.

45. Erhan H, Belotserkovsky J. Neuropsychological impact of Turner syndrome: importance of parent education, early detection, and intervention: a case study. *Arch Clin Neuropsychol* 2014; **29**: 544–5.

46. Temple C, Carney R. Intellectual functioning of children with Turner syndrome: a comparison of behavioural phenotypes. *Dev Med Child Neurol* 1993; **35**: 691–8.

47. Mazzocco MM. A process approach to describing mathematics difficulties in girls with Turner syndrome. *Pediatrics* 1998; **102** (Suppl. 3): 492–6.

48. Butterworth B, Grana A, Piazza M, Girelli L, Price C, Skuse D. Language and the origins of number skills: karyotypic differences in Turner's syndrome. *Brain Lang* 1999; **69**: 486–8.

49. Thalheimer W, Cook S. 2002. How to Calculate Effect Sizes from Published Research Articles: A Simplified Methodology. Available from: http://www.work-learning.com/effect_sizes.
htm (accessed 20 April 2012).

50. Ross JL, Roeltgen D, Stefanatos GA, et al. Androgen-responsive aspects of cognition in girls with Turner syndrome. *J Clin Endocrinol Metab* 2003; **88**: 292–6.

51. Cheung MW-L. Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychol Methods* 2014; **19**: 211.

52. Scammacca N, Roberts G, Stuebing KK. Meta-analysis with complex research designs dealing with dependence from multiple measures and multiple group comparisons. *Rev Educ Res* 2014; **84**: 328–64.

53. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-analysis. Hoboken, NJ: John Wiley, 2011.

54. Hedges LV, Vevea JL. Fixed-and random-effects models in meta-analysis. *Psychol Methods* 1998; **3**: 486.

55. Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954; **10**: 101–29.

56. Cheung MW-L. metaSEM: an R package for meta-analysis using structural equation modeling. *Front Psychol* 2015; **5**: 1521.

57. Team RC.R: A Language and Environment for Statistical Computing. Vienna, Austria, 2015. Available from: http://www.R-project.org (accessed 16 June 2015).

58. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol* 1995; **57**: 289–300.

59. Halberda J, Mazzocco MM, Feigenson L. Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 2008; **455**: 665–8.