Studying development in Williams syndrome: Progress, prospects and challenges

Jo Van Herwegen[1]

Harry Purser[2]

Michael S.C. Thomas[3]

& WiSDom team

[1] Department of Psychology and Human Development, UCL Institute of Education

[2] Department of Psychology and Human Development, UCL Institute of Education

[3] School of Psychology, Birkbeck University of London

Abstract

Williams Syndrome (WS) is a rare neurodevelopmental disorder associated with a specific cognitive profile of strengths and impairments. It has been argued that studying cognitive development of this disorder would not only allow improved knowledge of WS but also provide insight into alternative pathways in development. However, due to the rarity and nature of the disorder, there are a number of challenges to collect longitudinal data. This letter describes a new multi-lab based approach to examine development in WS longitudinally and discusses some of the challenges and solutions that need to be taken into account when putting together either previously obtained or newly collected data from different labs.

.

Keywords: Williams Syndrome; longitudinal; cross-sectional; development; multi-lab based

Corresponding author: j.vanherwegen@ucl.ac.uk

Williams syndrome (WS) is a genetic disorder caused by a deletion on the long arm of chromosome 7 spanning 1.5 million to 1.8 million base pairs. The disorder has a prevalence of about 1 in 20,000 live births and includes a specific clinical profile, comprising cardiovascular difficulties, idiopathic hypercalcaemia, and dysmorphic facial features amongst others. Behaviourally, individuals with WS have been described as very sociable with high empathy and little fear of strangers. They do, however, show a number of repetitive behaviours and experience anxiety and social difficulties. Cognitively individuals with WS are characterized by mild to moderate intellectual deficits with IQ scores ranging from 40 to 100 and an average Full Scale IQ score of 55 (Martens, Wilson, & Reutens, 2008).

Although both being delayed from infancy onwards, language abilities in WS outperform non-verbal abilities such as visuo-spatial and number abilities (Mervis et al., 2000; Van Herwegen, Rundblad, Davelaar, & Annaz, 2011), and the discrepancy between verbal and non-verbal abilities increases with age (Jarrold, Baddeley, Hewes, & Phillips 2001).

Understanding the development of this uneven cognitive profile of individuals with WS provides a way to investigate the impact of impaired abilities present in infancy on the phenotypic outcome later in life but can also generate insight into alternative pathways in development (see Karmiloff-Smith et al., 2012 and Van Herwegen & Karmiloff-Smith, 2015 for a detailed discussion). In sum, a greater understanding of cognitive development in WS not only improves our understanding of WS itself, which would allow better educational and behavioural interventions being put in place, but can also improve our knowledge of development itself, especially our understanding of the impact of domain general abilities on domain specific outcomes and plasticity of alternative pathways.

Progress: Taking development seriously

One way to study changes in performance across development is the use of a cross-sectional design with developmental trajectories (Thomas et al., 2009). In cross-sectional studies a function is constructed, using regression methods, between age and performance on a task using groups with a wide-age range. In the first instance, a developmental trajectory for a typically developing (TD) group is built based on which it is possible to establish whether the performance of a participant with WS fits anywhere on the trajectory. Secondly, the trajectory allows assessment as to whether the performance of an individual with WS matches the point on the trajectory for their chronological age (CA). This allows one to investigate whether the performance on the task in the disorder group is different from the TD trajectory by using a linear regression model with one-between-groups factor or, when multiple dependent factors are included, a mixed-design linear regression model including within-participants factors. Next, it can be explored whether the mental age (MA) of an individual with WS, assessed by one or more cognitive tasks, fits on the trajectory of the TD group according to their MA. In addition, the developmental trajectory approach allows assessment of how performance on tasks can be predicted by performance on other tasks and assesses relations between different cognitive processes. Furthermore, it potentially allows distinguishing between different types of delay (see Thomas et al., 2009 for a detailed discussion) such as the difference between delayed and atypical development.

However, cross-sectional studies have a number of limitations and difficulties (see discussion in Doherty, Shimi, & Scerif, 2014). First of all, cross-sectional studies include snapshots of cognitive abilities across different age groups and the individual differences between these individuals might mask any real changes over time across an entire group. Thomas et al. (2009) advocated the validation of cross-sectional trajectories with longitudinal follow-up. Secondly, cross-sectional studies require participants across a wide age range and thus, tasks to be administered need to be sensitive across a wide age range in order to avoid

floor and ceiling effects between the two groups (see discussion Thomas, Purser, & Van Herwegen, 2012). Importantly, this design also assumes that individuals with the same disorder will follow the same developmental trajectory. However, seeing the heterogeneity in WS (Van Herwegen et al., 2011), this may not always be the case.

Therefore, although cross-sectional studies can provide some insight into the development of cognitive functions in WS, these studies should ideally be followed up by longitudinal studies. Yet, longitudinal studies are costly and include a number of practical difficulties due to the rarity of WS. For example, in order to obtain a reasonable sample size, participants from a wide geographical area need to be recruited but this in turn means the project includes high travel costs and it is time consuming to revisit participants repeatedly due to longer travel times. Therefore, longitudinal studies assessing cognitive development in individuals with WS are rare (N=11) and they often include small sample sizes (average sample size N= 19.09, range 1 to 47).

Prospect: multi-lab based approaches

In our most recent project, named WiSDom, we examined cognitive development in WS longitudinally by collating data that has been gathered during the past fifteen years across different labs around Greater London. The WiSDom project includes principle investigators from across six different universities and incorporates data from standardised cognitive ability tasks that are often used as background measures in studies with WS, including Raven's Coloured Progressive Matrices (Raven, 1990), British Picture Vocabulary task, British Ability Scales (BPVS; Dunn, Dunn, Whetton & Burley, 1997), Pattern Construction from British Ability Scale (BAS; Ellioth, Smith & McCullock, 1997) and Test Reception of Grammar (Bishop, 2003). Given the rarity of the syndrome, the close proximity of the universities, and the fact that participants were recruited with help from the Williams

Syndrome Foundation (WSF), these six units have often assessed the same individuals with WS across a large number of different projects, allowing the creation of a longitudinal dataset that includes a much larger participant number (for example for BPVS; N= 192 in total for one data point, N= 90 for three data points, and N= 37 for six data points) compared to previous studies, even those who have examined only one data point cross-sectionally.

However, the fact that this multi-lab based approach was not planned when the studies were first carried out by the individual units has resulted in a number of challenges. One of the biggest challenges included the need to maintain participant anonymity, yet ensure that data from the different units was matched appropriately to the right individual with WS. Therefore, based on the database list from the WSF, a list of participant names with a corresponding anonymised code was generated by the data manager on the project. This master list did not include any data and was managed by the data manager only. This list was then sent out to the units so that the units could anonymise their datafiles using the appropriate code for each participant. Each unit then deleted the master list and sent their anonymised datafile to the lead researcher on the WiSDom project who collated all of the data. Although sharing even non-anonymised personal data for research or statistical purposes is permitted by the Data Sharing Code of Practice, the General Data Protection Regulation (GDPR) recommends anonymising where possible, with pseudo-anonymisation an acceptable second-choice. It is also worth noting that within GDPR, scientific research and statistical analyses are explicitly not considered to be incompatible with any initial purposes of data collection, whatever those initial purposes may have been. Going back further, the 1998 Data Protection Act has a specific exemption whereby data may be used for research purposes even if those purposes were not the initial reason for collecting the data. In sum, current data protection regulations do not present any obvious impediment to this sharing approach.

A second challenge included the fact that participants completed the tasks at different times and thus, the time between the different assessment periods differed between the participants. However, a linear mixed modelling approach readily accommodates not only different numbers of measurements across participants (and groups) but also irregular intervals between those measurements, given that it is based on line- or curve-fitting in essentially the same way that simple regression is performed. In addition, due to the time lapse participants were sometimes administered different version of the same standardised task so that the use of raw scores was not possible. However, for some tasks raw scores have been found to be more sensitive than mental age or standardised scores, for example for pattern construction abilities in WS (see discussion Purser & Van Herwegen, 2016), especially when measuring improvements over time. Given that a primary reason for the lack of sensitivity of mental age scores is the lumping together of very low or very high scores into single minimum or maximum scores, respectively, one solution is extrapolate the relationship between raw score and mental below and above the minimum and maximum, in order to estimate mental ages for such raw scores, thereby reducing the impact of floor and ceiling effects.

Another difficulty with a multi-lab based approach is the fact that numerous researchers have contributed to the data samples and should be acknowledged in the authorship of any published materials. One solution is the establishment of a consortium identity which can serve as a co-author and acknowledges the multiple contributors to the study.

Finally, although the participants were assessed on standardised tasks which should by definition be carried out in a standardised way so that scores can be easily comparable, participants were assessed by different researchers and thus, the small personal differences in which the tasks were administered may either add general noise to the data or generate some

systematic differences between cohorts and studies. Either of these could exaggerate or minimise the progress an individual with WS might make with increasing chronological age. However, owing to the fact that our sample size included a large number of data points for each researcher, we have the opportunity to examine the impact of the researcher on the outcome, again using linear mixed models. For example, examination of BPVS data of 37 participants who were assessed at least 6 times by 4 different experimenters, showed that the experimenter explained an additional 2% of the variance in the data. One application of this specific kind of investigation would be to shed light on a possible mechanism behind the replicability crisis: although the researcher effect here was modest, it might be far larger elsewhere. It also suggests a potential approach to replication, whereby linear mixed models, or some other hierarchical analysis, might be used to analyse composite datasets from different research groups, with the effects of those groups (and perhaps also their individual researchers) accounted for by the model. Such an approach has been used by the PING study to accumulate large brain imaging data sets (http://pingstudy.ucsd.edu/welcome.html).

Future recommendations

Studying development is no easy feat. However, a multi-lab based approach allows assessment of reproducibility of the smaller studies and provides a cost-effective way to gather data from a larger sample of participants over an extended period of time. In order for multi-lab based research to be effective, however, it is important to consider some of the practical aspects in advance. This includes, ensuring that protocols for task administration and scoring are shared, in order to make sure the same versions of the task are administered in the same way, as far as practicably possible, and not too closely in time (i.e., standardised tasks should not be repeated within a six month time frame). In addition, this approach would also require the sharing of the data to be considered within the consent form as this would

allow for the database to be made accessible to other researchers not part of the original study, which in turn would allow for the database to grow even further over time. One way of safeguarding data in such an arrangement would be to program an online database such that its users (aside from the data controller) could access only anonymised data, but the database would be linked to a secure file that stored participant dates of birth and full names. If users wished to add new data, they could check whether their participants were already in the file and either add a new participant, or else add further data for an existing participant. Finally, a planned repository would allow for the data to be entered by each unit itself, which would be more cost-effective in the long run.

Given that multi lab-based studies allow for a cost-effective way to replicate small-scale studies and obtain longitudinal data which allow a more in-depth understanding of development, especially in rare disorders, future studies should consider multi-lab approaches not only for standardised tasks but also for experimental tasks. This would require further standardisation of protocols across labs as well as an exchange of experimental tasks. Although a multi-lab based approach to experimental tasks requires more care and planning, the more studies that can validate cross-sectional trajectories via longitudinal studies, the more additional insight will be obtained into the individual variability present in a given disorder, as well as how compromised or approximate the previous literature using cross-sectional methods has been.

None of the data or materials for the experiments reported here is available, and none of the experiments was preregistered.

References

Bishop, D. (2003). *The Test for Reception of Grammar - Version 2*. London: Psychological Corporation.

Doherty, B., Shimi, A., & Scerif, G. (2014). Genetic disorders as models of high neurocognitive risk. In J. Van Herwegen & D. Riby (Eds.). *Neurodevelopmental disorders: Research challenges and solutions* (pp. 305-340). London: Psychology Press.

Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale II*. Windsor, Berkshire; NFER-Nelson Publishing Company.

Elliott, C.D., Smith, P., & McCulloch, K. (1997). *Technical manual British Ability Scales II*. Windsor, Berkshire: NFER-NELSON Publishing Company.

Jarrold, C., Baddeley, A.D, Hewes, A.K., & Phillips, C. (2001). A longitudinal assessment of diverging verbal and non-verbal abilities in the Williams Syndrome phenotype. *Cortex*, *37*(3), 423-431.

Karmiloff-Smith, A., D'Souza, D., Dekker, T., Van Herwegen, J., Xu, F., Rodic, M., & Ansari, D. (2012). Genetic and environmental vulnerabilities: The importance of cross-syndrome comparisons. *PNAS, 190(2),* 17261-17265.

Martens, M.A., Wilson, S. J., & Reutens, D.C. (2008). Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. *Journal of Child Psychology and Psychiatry*, *49*(6), 576-608.

Mervis, C.B., Robinson, B.F., Bertrand, J., Morris, C.A., Klein-Tasman, B.P. & Armstrong, S.C. (2000). The Williams Syndrome Cognitive Profile. *Brain and Cognition*, *44*(3), 604-628.

Purser, H. & Van Herwegen, J. (2016). Standardised and experimental psychological tasks: issues and solutions for research with children. In J. Prior. & J. Van Herwegen (Eds.). *Practical Research with Children* (pp. 105-131). London: Psychology Press.

Raven, J. (2007). *Coloured Progressive Matrices*. Oxford, UK: Oxford Psychologists Press Ltd.

Thomas, M.S.C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research*, *52*, 336–358. doi:10.1044/1092-4388(2009/07-0144)

Thomas, M.S.C., Purser, H., & Van Herwegen, J. (2012). The developmental trajectories approach to cognition. In E. K. Farran & A. Karmiloff-Smith (Eds). *Neurodevelopmental disorders across the lifespan: A Neuroconstructivist approach* (pp. 13-36). Oxford: Oxford University Press.

Van Herwegen, J. & Karmiloff-Smith, A (2015). Genetic developmental disorders and numerical competence across the lifespan. In R. Cohen Kadosh & A. Dowker (Eds.), *Oxford Handbook of Numerical Cognition* (pp. 721-731). Oxford: Oxford University Press.

Van Herwegen, J., Rundblad, G., Davelaar, E.J., & Annaz, D. (2011). Variability and standardised test profiles in typically developing children and children with Williams syndrome. *British Journal of Developmental Psychology, 29*, 883-894.