Pediatric Neurology ■■ (2018) ■■–■■

Contents lists available at ScienceDirect

Pediatric Neurology

journal homepage: www.elsevier.com/locate/pnu

Original Article

Is It Feasible to Identify Natural Clusters of TSC-Associated Neuropsychiatric Disorders (TAND)?

Loren Leclezio [a], Sugnet Gardner-Lubbe [b], Petrus J. de Vries [a,*]

[a] Division of Child and Adolescent Psychiatry, University of Cape Town, South Africa
[b] Department of Statistics and Actuarial Science, Stellenbosch University, South Africa

ARTICLE INFO

Article history:
Received 19 July 2017
Accepted 7 December 2017

Keywords:
Tuberous sclerosis complex
TAND
Natural TAND clusters
Neuropsychiatric
ASD
Cluster analysis
Factor analysis

ABSTRACT

Background: Tuberous sclerosis complex (TSC) is a genetic disorder with multisystem involvement. The lifetime prevalence of TSC-Associated Neuropsychiatric Disorders (TAND) is in the region of 90% in an apparently unique, individual pattern. This "uniqueness" poses significant challenges for diagnosis, psychoeducation, and intervention planning. To date, no studies have explored whether there may be natural clusters of TAND. The purpose of this feasibility study was (1) to investigate the practicability of identifying natural TAND clusters, and (2) to identify appropriate multivariate data analysis techniques for larger-scale studies.
Methods: TAND Checklist data were collected from 56 individuals with a clinical diagnosis of TSC (n = 20 from South Africa; n = 36 from Australia). Using R, the open-source statistical platform, mean squared contingency coefficients were calculated to produce a correlation matrix, and various cluster analyses and exploratory factor analysis were examined.
Results: Ward's method rendered six TAND clusters with good face validity and significant convergence with a six-factor exploratory factor analysis solution. The "bottom-up" data-driven strategies identified a "scholastic" cluster of TAND manifestations, an "autism spectrum disorder-like" cluster, a "dysregulated behavior" cluster, a "neuropsychological" cluster, a "hyperactive/impulsive" cluster, and a "mixed/mood" cluster.
Conclusions: These feasibility results suggest that a combination of cluster analysis and exploratory factor analysis methods may be able to identify clinically meaningful natural TAND clusters. Findings require replication and expansion in larger dataset, and could include quantification of cluster or factor scores at an individual level.

Background

Tuberous sclerosis complex (TSC) is an autosomal dominant genetic disorder caused by mutations in either of two genes, the *TSC1* gene on chromosome 9q34[1,2] or the *TSC2* gene on chromosome 16p13.3.[2,3] It is a multisystem disorder and can affect nearly any organ system, with most common physical manifestations being benign tumors in the central nervous system, skin, kidneys, heart, and lungs. It is also very common to see a vast and variable range of TSC-Associated Neuropsychiatric Disorders (TAND) that can present in infancy, childhood, adolescence, or adulthood. The multilevel manifestations of TAND include behavioral, psychiatric, intellectual, academic, neuropsychological, and psychosocial difficulties.[4–10] Table shows the various aspects of TAND, indicating the specific variables and domains examined in this study.

In a recent conceptual review of TAND, we identified three clinical and research challenges.[10] The first challenge was the "assessment and treatment gap" of TAND. That is, despite the fact that the lifetime prevalence of TAND is in the region of 90%,[11] there appears to be a global lack of assessment for these problems. A study conducted in the United Kingdom in 2010 showed that only 18% of all families had ever received any of the evaluations or treatments recommended in the 2005 TSC guidelines.[11,12] Similarly, in the international Tuberous Sclerosis Registry to Increase Disease Awareness (TOSCA) study of over 2000 patients from 31 countries, fewer than 40% of patients had ever had an intellectual ability assessment, and very high rates of missing TAND data were reported.[13] These findings suggested that even in expert TSC centers, TAND are likely to be under-identified and under-treated.

A second challenge for TAND is the perceived "uniqueness" of individual TAND profiles.[10] It appears that each person has his or her own set and combination of TAND features, and no two

* Corresponding author.
E-mail address: petrus.devries@uct.ac.za (P.J. de Vries).

https://doi.org/10.1016/j.pediatrneurol.2017.12.006
0887-8994/

TABLE.
Multiple Levels of TSC-Associated Neuropsychiatric Disorders (TAND)

TAND					
Behavior	Psychiatric	Intellectual	Academic	Neuropsychological	Psychosocial
Aggression*	Autism spectrum disorder	Intellectual disability	Reading*	Sustained attention*	Self-esteem
Temper tantrums*	ADHD	Uneven intellectual profiles	Writing*	Dual-tasking*	Self-efficacy
Anxiety*	Anxiety disorder		Mathematics*	Attentional switching	Parental stress
Depressed mood*	Depressive disorder		Spelling*	Memory recall*	Relationship difficulties
Self-injury*				Spatial working memory*	
Inattention*				Cognitive flexibility	
Hyperactivity*					
Impulsivity*					
Language delay*					
Poor eye contact*					
Repetitive behaviors*					
Sleep problems*					

Abbreviation:
ADHD = Attention deficit/hyperactivity disorder.
Items indicated with * are TAND variables included in this feasibility study.

individuals have the same TAND "signature." This uniqueness poses significant challenges for diagnostic workup, psycho-education, and intervention planning, particularly given the rarity of TSC and the multilevel nature of TAND. To date, there have been no obvious indications of natural clustering, that is, natural, predictable grouping of specific neuropsychiatric characteristics across levels of TAND.[10] We suggested that if we could identify such natural TAND clusters, it may represent a strategy not only to manage the assessment or treatment gap, but also to reduce the third challenge ("treatment paralysis"), where clinical teams believe they are unable to implement appropriate treatments given the complexity of the TAND manifestations in a particular individual.[10] For further detail, please see reference 10.

We investigated the feasibility of identifying natural TAND clusters and tried to identify suitable multivariate data analysis techniques for larger-scale studies. We proposed that data reduction may identify a small number of TAND clusters that could help to reduce and make manageable the complexity of TAND.

Methods

Subjects

Participants were part of the "Pilot Validation of the Tuberous Sclerosis-Associated Neuropsychiatric Disorders (TAND) Checklist" study.[14] Twenty participants from South Africa and 37 from Australia were recruited. Participants had to meet criteria for TSC, had to be able to complete the TAND Checklist[11] with the researcher, or had to have a parent or caregiver who could complete the TAND Checklist with the researcher. For informed consent purposes, TAND Checklists were completed only by individuals over age 18 years or by their primary caregivers. Data were deliberately included on a wide age (four to 42 years) and ability range (normal intellectual ability—profound intellectual disability) of participants.

Procedures

The TAND Checklist was administered to parents and caregivers of individuals with TSC by one of the researchers (LL or PJdV). The Checklist follows the neuropsychiatric levels of investigation outlined previously[11] and contains the following 12 sections: (1) Basic developmental milestones; (2) Current level of functioning; (3) Behavioral concerns; (4) Psychiatric disorders diagnosed; (5) Intellectual ability; (6) Academic skills; (7) Neuropsychological skills; (8) Psychosocial functioning; (9) Parent, caregiver, or self-rating of the impact of TAND; (10) Prioritization list; (11) Additional concerns; and (12) Health-care professional rating of the impact of TAND. Questions required simple "Yes" or "No" responses to most sections.[12]

Data analysis

All analyses were performed with the R software package (R Core Team, Vienna, Austria).[15] For detail of R code see Appendix S1. The following Sections of the TAND Checklist were included in the analysis: Section 3, behavioral challenges (19 questions/variables); Section 5, academic skills (four variables); and Section 7, neuropsychological skills (six variables). Given that all variables were binary (Yes/No), model-based clustering assuming an underlying Gaussian distribution was not considered suitable. Instead, the mean squared contingency coefficient[16] was used to compute a correlation matrix for the 29 variables selected from the TAND Checklist. Where missing values were present, these were omitted pairwise in correlation computations.

Several clustering solutions were compared. Hierarchical clustering methods provide a clustering tree visually representing the merging of TAND variables and suggesting a suitable number of clusters. Complete linkage, average linkage, Ward's method, and McQuitty's methods were applied with the hclust() R function. Although hierarchical clustering has often been used with great success, the algorithm is fairly naïve and some more recent methods in the R package cluster[17] were therefore also investigated. PAM (partitioning around medoids) is an extension of the popular k-means clustering method. However, k-means could not be applied, given that it required a numerical data matrix rather than a dissimilarity matrix based on the square contingency coefficient correlation matrix. The FANNY (fuzzy clustering) method allocates a probability for belonging to each cluster rather than simply allocating each item to a single cluster. DIANA (divisive analysis), also a hierarchical clustering method, was also used. In contrast to the other methods where larger clusters are formed by merging smaller clusters, this method forms smaller clusters by dividing larger clusters.

After a suitable clustering solution was obtained, exploratory factor analysis was employed with the function fa() from the R package psych.[18] The factor analysis was also performed on the means squared contingency coefficient correlation matrix. All the different options of factor extraction and rotation available in the fa() R function were investigated. These combinations were