Sequence Mining of Comorbid Neurodevelopmental Disorders Using the SPADE Algorithm*

Inna Pimus1; Mor Peleg1; Mitchell Schertz2

1Department of Information Systems, University of Haifa, Haifa, Israel; 2Institute for Child Development, Meuhedet, North, Haifa, Israel

Keywords
Sequence mining, SPADE, neurodevelopmental disorders, comorbidity

Summary
Objectives: Understanding the progression of comorbid neurodevelopmental disorders (NDD) during different critical time periods may contribute to our comprehension of the underlying pathophysiology of NDDs. The objective of our study was to identify frequent temporal sequences of developmental diagnoses in noisy patient data.

Methods: We used a data set of 2810 patients, documenting NDD diagnoses given to them by an NDD expert at a child development center during multiple visits at different ages. Extensive preprocessing steps were developed in order to allow the data set to be processed by an efficient sequence mining algorithm (SPADE).

Results: The discovered sequences were validated by cross validation for 10 iterations; all correlation coefficients for support, confidence and lift measures were above 0.75 and their proportions were similar. No significant differences between the distributions of sequences were found using Kolmogorov-Smirnov test.

Conclusions: We have demonstrated the feasibility of using the SPADE algorithm for discovery of valid temporal sequences of comorbid disorders in children with NDDs. The identification of such sequences would be beneficial from clinical and research perspectives. Moreover, these sequences could serve as features for developing a full-fledged temporal predictive model.

1. Introduction

Neurodevelopmental Disorders (NDD) are chronic brain disorders characterized by onset during the developmental period, often starting in infancy, and with a range of developmental deficits that impair normal functioning [1]. Examples include attention deficit hyperactivity disorder (ADHD), developmental coordination disorder (DCD), and developmental language disorders [1]. The prevalence of NDDs among pre-school children is about 10% [2].

While timely and accurate detection of NDDs is very important, it is a difficult task, due to missing or ambiguous diagnostic definitions, which allow different interpretation of symptoms according to various approaches, unclear terminology, and confusing symptoms which may be attributed to different disorders [3].

While the Diagnostic and Statistical Manual of Mental Disorders (DSM) and diagnostic instruments for NDDs strive to present clear, practical guidelines for diagnosis of NDDs, the correct and immediate detection and interpretation of these disorders is influenced by professional experience with these types of diagnoses. Primary care pediatricians acknowledge their difficulty with tackling this problem and report a need for some practical aid and professional enrichment which will make the diagnosis procedure more effective [4].

1.1 The Importance of Identifying Comorbid NDDs

The causes or mechanisms of development of NDDs are not yet understood and are conjectured to be due to combinations of genetic and environmental effects [5]. The phenomenon of comorbidity in NDDs is extremely common; some researchers report up to 10 times higher rates of co-occurring disorders than of single NDDs [5]. Clearly, it would be helpful to find patterns of co-occurring NDDs, which may help elucidate neurological mechanisms involved in NDDs. In addition, discovery of comorbidities allows physicians to understand the extent and nature of the problems more clearly and to plan treatment in a more effective manner [6]. Importantly, age or development period determines how different situations combine with the biological predisposition and lead to different outcomes [7]. Knowledge of temporal patterns of NDDs can help physicians and patients to prepare in advance for additional symptoms that may arise [8].

© Schattauer 2016

http://dx.doi.org/10.3414/ME15-01-0142
received: October 27, 2015
accepted: December 21, 2015
epub ahead of print: February 5, 2016

* Supplementary material published on our website http://dx.doi.org/10.3414/ME15-01-0142
In previous research, our group has used Self-Organizing Maps (SOM) to identify comorbid NDDs [9], guided by an ontology of NDD groups that are known in the medical literature. The ontology is used to direct the clustering process, validate its results, and correct some of the clusters. However, this previous work has focused on analyzing NDD patterns at a single time point – disregarding the temporal progression of those patterns. In our current work we extend our previous research by focusing on discovering the temporal progression of comorbid NDDs.

1.2 Comorbid NDD Artifacts and their Reflection in Patient Data

Because machine learning methods are based on statistical probabilities of patterns that exist in data, it is important to understand which patterns reflect real existing phenomena of comorbid disorders and which describe artifacts. Artifacts can exist due to the following reasons:

a) Work habits: The majority of physical, occupational and speech/language therapy treatments for children older than six are not provided at the main clinic but at external institutions; and as such, the doctor writing the diagnosis may only focus on the diagnosis that he is managing.

b) Disorders come in and out of focus: At certain developmental stages, the medical expert may focus on different problems and disregard other persisting diagnoses. For example, certain clinical concerns (e.g., regarding specific learning disorders, Specific language impairment (SLI)), raised either by parents of kindergarteners or by school teachers, are usually noticed at particular ages [10].

c) Medical experts may focus on the more severe NDDs and sometimes do not record less severe disorders.

Besides artifacts, there is difficulty of diagnosing certain disorders at specific ages, as different diagnoses are challenging patients in terms of daily functionality at different periods of time. This situation is termed by some medical experts as “floating diagnosis” and describes the clinical scenario whereby a child presents symptoms consistent with NDD-X at time 1 but different symptoms, consistent with NDD-Y, at time 2. For example, in the preschool years, milder forms of hyperactivity and inattention may be present but do not cause significant dysfunction [11]. Consequently, the concerns may be noted but a diagnosis of ADHD will not be made. However with entrance into grade school, even mild forms of hyperactivity will be disruptive and poor attentional skills will significantly interfere with learning. At that time, a diagnosis of ADHD will be made. The reverse can occur as well, when certain preschoolers are diagnosed with ADHD that resolves over time. Similar issues of diagnostic stability are present with DCD [12] and language disorders [13].

2. Materials and Methods

We describe below the SPADE algorithm, justify its choice, and specify the data set and the preprocessing steps needed for executing the SPADE algorithm on the data set. We then describe the statistical measures for assessing the quality of the temporal sequences discovered and the evaluation methods.

2.1 The SPADE Sequence Mining Algorithm

Sequence mining techniques discover item-sets where the time variable is discrete (i.e., when time is divided into intervals) [14, 15]. Sequence mining techniques were used over clinical data in a variety of scenarios: recommender systems, decision support systems and theoretical research [14–18]. Because of the size of the data set that was not very large, and in line with our emphasis on discovering the temporal progression of NDDs along expert-defined meaningful time intervals, we have chosen to use a sequential pattern mining algorithm rather than considering a method that takes into account time intervals with a duration.

The SPADE (Sequential PAttern Discovery using Equivalence classes) [19] algorithm is a particular kind of sequence mining algorithm, based on the classical frequent item-set detection algorithm – Apriori [20]. The SPADE algorithm works over a set of distinct items \( I = \{i_1, i_2, \ldots, i_m\} \) (e.g., NDDs). An event is a sub-set of \( I \). In our case, an event is a non-empty unordered group of (co-occurring) NDDs at time \( T_j \) (age group, in our case). A sequence is denoted as \( s = a_1 \rightarrow a_2 \rightarrow \ldots \rightarrow a_m \) where each \( a_k \) is an event. The algorithm discovers frequent sequences, i.e., sequences with support (proportion in which the sequence is found in the database) that is higher than pre-defined user-selected minimal value.

We have selected SPADE for its ability to work with noisy empirical data and integrate domain knowledge [19], its efficiency [20], and its ability to use data that is asymmetrically distributed between categories and with complex multivariate data, including various types of temporal data and different definitions of uniqueness (for example, whether each row in the dataset describes a different patient or a different visit).

For practical implementation of SPADE in our work, the “rulessequences package” [21] for R statistical programming language (version 2.15.1) was used.

2.2 Data Set

We used a data set of de-identified clinical data from the Institute for Child Development, Kupat Holim Meuhedet at Herzeliya, Israel. Ethical approval for use of these was obtained from the Ethics Committee at the University of Haifa. The data set includes records of 56,033 visits of 9796 patients (ages 0–18 years) who were diagnosed at the institute from October 12, 1999 to June 30, 2011. The original dataset contained the following information: randomly generated ID, gender, birth date, clinical exam date and diagnostic terms (Appendix A for an example). The diagnostic terms include mostly NDD diagnoses but also few findings (e.g., Average intellect). The intervals between follow-up visits of the patients in the data set range from 30 days at the minimum to 804 days at the maximum; the median is 234 days.
2.3 Data Preprocessing

The original format of the data sets did not allow easy statistical processing for several main reasons: 1) each observation in the dataset (a row) refers to a single patient visit. Thus, the same patient can appear in numerous rows; 2) standard terminology was not used (several terms were associated with a single diagnosis); 3) the data contains records with typing errors and missing values. Consequently, data preparation (preprocessing) was performed in four steps, described below, using python scripts.

2.3.1 Establishing Data Integrity: Handling Missing or Irrelevant Values

The original dataset contains records with typing errors in NDDs and date of birth, and missing values. Obvious typing errors were corrected. For example, the date of birth 26/5/2100 was changed to 26/5/2001 – notice that the year 2010 is not possible, since examination dates were documented before 2010, hence the birth year cannot be 2010. Rows containing empty values in columns ID, date of birth, clinical exam date or diagnosis were removed.

A medical expert (co-author MS) defined impossible combinations of disorders – NDDs that cannot co-occur with “Normal psychomotor development”: DCD, Gross Motor Developmental Delay, and Specific Language Impairment. Patient with impossible combinations were removed from dataset.

2.3.2 Data Standardization

The diagnostic terminology that is used in the original dataset does not use controlled vocabulary codes. In order to eliminate duplicate terminology, we prepared a consistent standardized concept list that contains the terms which are used in original dataset and their mapping to standard vocabulary terms (Appendix B). The list of consistent concepts was generated using a graph-based algorithm [22] that collects the relevant terms from the most extensive clinical vocabulary within UMLS version 2012AA (SNOMED-CT, which covers 86% of the original terms in the dataset), uses its hierarchy as the hierarchy of the final concept list, and adds into the hierarchy terms found only in complementing vocabularies from UMLS (six additional vocabularies were used to cover all terms found in the dataset, including mostly DSM IV and ICD-10 terms). Following the procedure described in [22], we mapped the 568 original terms to an equivalent set of 197 consistent concepts, which were approved by the medical expert. We automatically replaced all of the diagnostic terms with the concepts from the consistent standardized concept list.

We then balanced the amount of patients in each category. If this step is omitted, disorders with smaller number of instances will not be discovered by the algorithm, since their support will be too small. Hence we associated the 197 concepts with 79 more abstract terms, by unifying several terms into the most-specific higher-level parent term from the UMLS hierarchy, by going one or more levels up the UMLS hierarchy. When several higher-level terms were available, the expert chose the most relevant one (e.g., Triplegic cerebral palsy and Athetoid cerebral palsy have several mutual parents in the SNOMED-CT hierarchy including disorder of brain and cerebral palsy; the expert chose cerebral palsy as the most relevant parent). Non-informative rows which did not include diagnoses but only listed non-specific terms (“observation”, “other diagnosis”, “to be determined”, and “syndrome”) were removed.

2.3.3 Partitioning of the Age Feature into Clinically-meaningful Intervals

Our medical expert was interested in finding temporal sequences that relate to clinically-meaningful age intervals (i.e., given a child with current set of NDDs and current age interval, what are patterns of progression to future age intervals). Hence, we partitioned the age variable into five time intervals (in years), which are clinically meaningful: 0–2, 2–3, 3–6, 6–9 and >9 years [23]. The periods were defined independently of the data, based on theoretical knowledge of the progress of NDDs and practical considerations (e.g., school age, developmental services are covered by national health insurance till age nine).

Patients with data in only one age category (meaning that there are no correlations to be considered) were removed from the dataset. 2810 patients (28%) remained. Most of the patient visits were concentrated in the age groups: 2–3, 3–6, 6–9 (Appendix C for age group distribution).

In some cases, a certain disorder was documented for a patient at some age point, at the next age point the disorder was not mentioned, but was mentioned again at some later age points. In such situations we assumed, after discussion with the medical expert, that the disorder was not resolved but was simply not documented, and we added it to the rows of the database. The completed NDDs (completed for 143 patients) are shown in Appendix D.

In order to conserve statistical power, disorders which appeared in less than 3% of patients (minimum 85 patients) were associated with the "Other" category (code 100). As a result, the database contained 19 most frequent disorders, shown in Appendix E.

2.3.4 Format Adjustment

The original data format that stores NDDs per visit of a child was converted into a format suitable for processing by the arulesSequences R package implementation of the SPADE algorithm (Table 1 for an example). Each child was given a fake patient ID (column 1). All of the NDDs (items) given to a child during a particular age category are represented as a single event. The event ID (column 2) is represented by the age category, as in [19]. We coded the time interval and the concept (consistent abstract term (diagnosis), as in Appendix B) as one variable according to the format: concept_ageCategory. For example 10_1 means concept ten in the first age category (time interval). Thus the ageCategory serves as a time stamp for the SPADE algorithm. This representation is meaningful, since developmental disorders behave differently at various ages. Thus, for
example, ADHD at age interval 1 and age interval 2 has different implications. Column 3 shows the number of disorders (comorbidities) in the event and Column 4 provides the disorder codes.

2.4 Measures

The output of the SPADE algorithm shows the sequences and their support, from which we calculate lift and confidence.

2.4.1 Support

Support (supp) is output by the arulesSequences R package implementation of SPADE for events and for sequences of events. Support is the proportion in which the event/sequence is found in the database. Note that each patient is counted only once, no matter how many times he was diagnosed with a disorder.

Sequences with low support were filtered out in the pre-processing step, as they represent sequences that occur too scarcely. The choice of the cutoff value was determined empirically as 0.03 by examining the distribution of support for sequences and finding the level of support in which there is a significant change in distribution density (Appendix F).

2.4.2 Confidence

Rules were induced from a set of frequent sequences, which satisfy the minimum confidence threshold. Using the notation of [19], the rules have the form \((E_1 \rightarrow \cdot \cdot \cdot \rightarrow E_k \rightarrow E_l)\), where the left and right side of the rule are sequences, and \(E_l\) is an event that occurs earlier than event \(E_k\).

The confidence of a sequence rule is the conditional probability of the existence of a certain sequence of NDD or NDD-group, given the existence of an earlier sequence.

Confidence is defined as:

\[
\text{conf}(E_1 \rightarrow \cdot \cdot \cdot \rightarrow E_k \rightarrow E_l) = \frac{\text{supp}(E_1 \rightarrow \cdot \cdot \cdot \rightarrow E_k \rightarrow E_l)}{\text{supp}(E_1 \rightarrow \cdot \cdot \cdot \rightarrow E_k)},
\]

which we notate as \(\text{conf}(E_1 \rightarrow \cdot \cdot \cdot \rightarrow E_l)\).

For example, \(\text{conf}(\text{ASD}[2-3] \rightarrow \text{ASD}[3-6]) = 0.73\) means that in 0.73 of the cases where ASD appears at age interval 2–3 it also appears later in age interval 3–6.

A threshold of confidence \(\geq 0.3\) for filtering output results was defined with our medical expert, who advised that practically interesting patterns with high conditional probability should be considered.

2.4.3 Lift

The lift is the ratio between the probability of a sequence occurrence in the dataset and the probability of its elements occurring independently [24]. It is defined as the proportion:

\[
\text{lift}(E_1 \rightarrow \cdot \cdot \cdot \rightarrow E_l) = \frac{\text{supp}(E_1 \rightarrow \cdot \cdot \cdot \rightarrow E_l)}{\text{supp}(E_1) \times \text{supp}(E_l)} = \frac{\text{conf}(E_1 \rightarrow \cdot \cdot \cdot \rightarrow E_l)}{\text{supp}(E_l)}
\]

Lift values other than 1 indicate that the events are not random. Lift values higher than 1 represent positive dependency between the disorders, meaning that if disorder \(E_l\) occurs, then it is more likely than random that \(E_k\) will follow. Lift values lower than 1 mean that the co-occurrence rate of the disorders is lower than in the independent case, and that if \(E_k\) occurs then \(E_l\) is unlikely to follow.

2.5 Evaluation Methods

We used two methods to validate the results: cross validation and expert validation. We used leave-50-out cross-validation, repeated 10 times. At each iteration, the similarity of sequences’ measures (support, confidence and lift) was examined by calculating Pearson and Spearman correlation coefficients for all the measures in the training and validation sets. In addition, Kolmogorov-Smirnov tests for similarity of distributions were performed for each pair of sets. Finally, proportion tests were performed on support and confidence measures for all the discovered sequences. For each of the ten iterations, we compared the proportion of each discovered pattern, between the two halves of the data set. To summarize these many results, we report the proportion of the discovered patterns for each iterations for which p-Values were significant (above 0.05), and the mean p-value.

As a second method of evaluation, we used expert validation in order to evaluate the validity of patterns. The evaluation was based on the opinion of one medical expert. The medical expert provided his opinion regarding each discovered se-
sequence whether it is approved (for sequences supported by the medical literature, reflecting known or plausible phenomena that are present in children), an artifact due to incomplete data caused by practitioners’ work habits, or requires future research to determine approval.

### 2.6 Post-processing of SPADE Output for Generating the Visualization Model

Visualization is important for clarifying our results and making them more accessible for clinical experts. The validated sequences with lift ≥ 1 predict positive dependencies between NDDs or NDD comorbid groups. They were visualized as nodes connected by arcs specifying that if the prefix NDD is present at a defined time period, the suffix NDD should be suggested for the next time period with certain confidence. The visualization was done manually using PowerPoint but an algorithm could be developed for that purpose.

In addition, Allen’s temporal relations [25] were used as a general framework for describing sequence mining results because it is a simple, consistent, standard, and practical framework for representing relationships between intervals. Each sequence of events (i.e., appearance/disappearance of NDD diagnosis) was mapped to one of the seven Allen relations (before, meets, overlaps, starts, during, finishes, equals). To ease comprehension by clinical experts, we catarinated “X meets X” intervals into longer intervals upon which a relation we call “persists” (previously called connectable by Shahar [26], fold by Lorentzos [27], coalesce by Sarda [28] or compress by Navathe and Ahmed [29]) holds, for marking the sequences which contain the same disorder as antecedent and descendant with meeting time intervals. For example sequence ASD[2–3] → ASD[3–6] is defined as “ASD persists [2–6].”

### 3. Results

Using the SPADE algorithm, we found 98 temporal NDD sequences. Table 2 shows the 38 sequences with confidence over 0.3. The column “Existing disorder/comorbid group” and the column “Future disorder/comorbid group” describe the left and right sides of the temporal pattern. Age interval in years is shown in square brackets. The support, lift, and Allen’s pattern category are also provided. The medical expert’s opinion is reported in the last three columns of Table 2, listing the category and more specific details. The last column shows whether the expert accepted the pattern.

According to the last column, 27 (71%) sequences were approved; 7 (18%) were found to be artifacts and were rejected and 4 (11%) were left for future research to determine whether they could be approved or are artifacts.

The approved sequences and the sequences left for future research whose lift values are ≥ 1 (total of 29 sequences) are visualized in Figure 1.

The detailed results of cross-validation for each of the ten iterations are shown in Table 3. Each row contains the number of patterns discovered at the iteration and results of performed statistical tests: Kolmogorov-Smirnov, Pearson and Spearman correlation, and proportion tests on support and confidence. All the measures were found to be similar for all the discovered patterns at any level of significance (all p-values lower than 0.05) at each iteration. Specifically, all of the Pearson and Spearman correlation coefficients are high (above 0.75 [30]), meaning that the measures in the training and validation sets are similar. The p-values of Kolmogorov-Smirnov tests for similarity of distributions are all above 0.1, so no significant differences in the distributions are seen. Very few significant differences were found in proportions (an average of 1% for confidence and 0 for support, which may be attributed to random noise effects), so the proportions are very similar.

### 4. Discussion

The objective of this study was to identify temporal patterns of NDDs. The identification of such patterns is beneficial from both clinical and research perspectives. Diagnosis of NDDs is a difficult task for several reasons. First, developmental diagnoses still lack, to different degrees, diagnostic stability [31]. For example, some infants with motor delay frequently “catch up” to a significant extent. Others often present with one type of developmental problem that may evolve into one of a number of different developmental problems at a later time [32]. Finally, developmental diagnoses are rarely unitary but often multiple in nature [33]. Therefore, automatically identifying patterns of developmental delays can help diagnose NDDs [34]. This will allow for improved parent counseling regarding the findings and better prepared child development staff in the developing of more appropriate interventions.

The research implications of the identification of such a predictive diagnostic model are even of greater importance. The identification of diagnostic clusters of developmental diagnoses, as was done in this research, could serve as a basis for developing a predictive model that could be tested. Such a model would allow identification of more homogeneous groups, each having a specific diagnostic outcome. This would lend itself to improved understanding of etiology and use of more homogeneous groups during intervention trials.

Our results – namely the identification of temporal patterns – support clinical thinking and previous research that examined particular developmental outcomes over time, given a particular initial developmental presentation [35]. Thus our results seem very much in line with the initial hypothesis that finding temporal patterns of NDDs has diagnostic and clinical benefits and the fact that most of the patterns found are clinically valid (see Section 4.1) supports SPADE as a useful mining algorithm for discovering temporal patterns of NDDs, given the child’s age and his history of NDDs. The SPADE algorithm is an efficient algorithm that could be used to find short and long sequences of events in large data sets which include a large number of potential items. The algorithm is mostly aimed at discovering sequences in episodic, time-stamped, point-based data. These items could be, in the medical domain, diagnostic codes, laboratory test results, procedures performed, claims data, etc. The preprocessing and post-processing ap-
## Table 2  Temporal sequences found in the data set

<table>
<thead>
<tr>
<th>ID</th>
<th>Existing disorder/comorbid group $E_i$</th>
<th>Future disorder/comorbid group $E_{i+1}$</th>
<th>Confidence</th>
<th>Support</th>
<th>Lift</th>
<th>Allen’s relation</th>
<th>Expert category</th>
<th>Expert comment</th>
<th>Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASD[2–3]</td>
<td>ASD[3–6]</td>
<td>73%</td>
<td>0.035</td>
<td>9.72</td>
<td>ASD persists</td>
<td>persists</td>
<td>Severe diagnosis tends to persist</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>(ADHD,SLI-C)[3–6]</td>
<td>ADHD[6–9]</td>
<td>67%</td>
<td>0.053</td>
<td>2.35</td>
<td>SLI-C starts ADHD</td>
<td>artifact of work habits</td>
<td>After age 6 kids tend to come only for ADHD medication</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>(DCD,ADHD)[3–6]</td>
<td>ADHD[6–9]</td>
<td>64%</td>
<td>0.131</td>
<td>2.25</td>
<td>DCD starts ADHD</td>
<td></td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>(DCD,ADHD,SLI-C)[3–6]</td>
<td>ADHD[6–9]</td>
<td>66%</td>
<td>0.038</td>
<td>2.34</td>
<td>DCD, SLI-C starts ADHD</td>
<td></td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>ADHD[3–6]</td>
<td>ADHD[6–9]</td>
<td>62%</td>
<td>0.192</td>
<td>2.17</td>
<td>ADHD persists</td>
<td>persists</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>SLI[2–3]</td>
<td>SLI[3–6]</td>
<td>62%</td>
<td>0.093</td>
<td>1.57</td>
<td>SLI persists</td>
<td>persists</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>(DCD,ADHD,SLI-C)[3–6]</td>
<td>ADHD[6–9]</td>
<td>61%</td>
<td>0.059</td>
<td>2.14</td>
<td>DCD, SLI starts ADHD</td>
<td>artifact of work habits</td>
<td>After age 6 kids tend to come only for ADHD medication</td>
<td>–</td>
</tr>
<tr>
<td>8</td>
<td>(ADHD,SLI)[3–6]</td>
<td>ADHD[6–9]</td>
<td>59%</td>
<td>0.076</td>
<td>2.06</td>
<td>SLI starts ADHD</td>
<td></td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>9</td>
<td>DD[2–3]</td>
<td>DD[3–6]</td>
<td>56%</td>
<td>0.055</td>
<td>4.03</td>
<td>DD persists</td>
<td>persists</td>
<td>The diagnosis is stable</td>
<td>+</td>
</tr>
<tr>
<td>10</td>
<td>SLI-C[2–3]</td>
<td>SLI[3–6]</td>
<td>49%</td>
<td>0.035</td>
<td>1.23</td>
<td>SLI-C meets SLI</td>
<td>persists</td>
<td>SLI-combined and SLI-expressive are similar</td>
<td>+</td>
</tr>
<tr>
<td>11</td>
<td>ASD[3–6]</td>
<td>ASD[6–9]</td>
<td>48%</td>
<td>0.036</td>
<td>10.54</td>
<td>ASD persists</td>
<td>persists</td>
<td>Severe diagnosis tends to persist. Clinically the expected stability of the disorder is around 60–80%.</td>
<td>+</td>
</tr>
<tr>
<td>12</td>
<td>DD[0–2]</td>
<td>DD[2–3]</td>
<td>46%</td>
<td>0.034</td>
<td>4.71</td>
<td>DD persists</td>
<td>persists</td>
<td>The diagnosis is stable</td>
<td>+</td>
</tr>
<tr>
<td>13</td>
<td>(SLI,SLI-C)[3–6]</td>
<td>SLI[6–9]</td>
<td>45%</td>
<td>0.035</td>
<td>2.58</td>
<td>SLI starts SLI</td>
<td>persists</td>
<td>SLI-combined and SLI-expressive are similar</td>
<td>+</td>
</tr>
<tr>
<td>14</td>
<td>(DCD,ADHD)[3–6]</td>
<td>DCD[6–9]</td>
<td>43%</td>
<td>0.088</td>
<td>1.99</td>
<td>ADHD starts DCD</td>
<td>persists and resolved</td>
<td>DCD can persist and ADHD resolved which can indeed occur – theoretically and clinically possible</td>
<td>+</td>
</tr>
<tr>
<td>15</td>
<td>(DCD,SLI-C)[3–6]</td>
<td>ADHD[6–9]</td>
<td>43%</td>
<td>0.047</td>
<td>1.51</td>
<td>DCD, SLI-C meets ADHD</td>
<td>floating Dx or artifact</td>
<td>ADHD was not present in preschool but becomes more significant afterwards</td>
<td>Fr</td>
</tr>
<tr>
<td>16</td>
<td>(DCD,ADHD,SLI-C)[3–6]</td>
<td>DCD[6–9]</td>
<td>43%</td>
<td>0.042</td>
<td>1.99</td>
<td>ADHD,SLI starts DCD</td>
<td>floating Dx</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>17</td>
<td>NPD[0–2]</td>
<td>NPD[2–3]</td>
<td>41%</td>
<td>0.032</td>
<td>4.69</td>
<td>NPD persists</td>
<td>persists</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>18</td>
<td>(DCD,ADHD,SLI-C)[3–6]</td>
<td>SLI[6–9]</td>
<td>40%</td>
<td>0.039</td>
<td>2.29</td>
<td>DCD, ADHD starts SLI</td>
<td>floating Dx</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>19</td>
<td>DCD[3–6]</td>
<td>ADHD[6–9]</td>
<td>39%</td>
<td>0.171</td>
<td>1.36</td>
<td>DCD meets ADHD</td>
<td>floating Dx or artifact</td>
<td>DCD can evolve into ADHD</td>
<td>Fr</td>
</tr>
<tr>
<td>20</td>
<td>(ADHD,SLI-C)[3–6]</td>
<td>SLI[6–9]</td>
<td>39%</td>
<td>0.051</td>
<td>2.26</td>
<td>ADHD starts SLI</td>
<td>floating Dx</td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>
Table 2  continued

<table>
<thead>
<tr>
<th>ID</th>
<th>Existing disorder/comorbid group ( E_2 )</th>
<th>Future disorder/comorbid group ( E_1 )</th>
<th>Confidence</th>
<th>Support</th>
<th>Lift</th>
<th>Allen’s relation</th>
<th>Expert category</th>
<th>Expert comment</th>
<th>Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>DCD[3–6]</td>
<td>DCD[6–9]</td>
<td>38%</td>
<td>0.17</td>
<td>1.77</td>
<td>DCD persists</td>
<td>persists</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>(DCD,SLI)[3–6]</td>
<td>SLI[6–9]</td>
<td>38%</td>
<td>0.086</td>
<td>2.2</td>
<td>DCD starts SLI</td>
<td>floating Dx</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>(DCD,SLI)[3–6]</td>
<td>DCD[6–9]</td>
<td>38%</td>
<td>0.085</td>
<td>1.73</td>
<td>SLI starts DCD</td>
<td>persists and resolved</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>SLI-C[3–6]</td>
<td>ADHD[6–9]</td>
<td>38%</td>
<td>0.067</td>
<td>1.35</td>
<td>SLI-C meets ADHD</td>
<td>artifact of work habits</td>
<td>After age 6 kids tend to come only for ADHD medication</td>
<td>–</td>
</tr>
<tr>
<td>25</td>
<td>(DCD,SLI)[3–6]</td>
<td>ADHD[6–9]</td>
<td>36%</td>
<td>0.081</td>
<td>1.26</td>
<td>DCD, SLI meets ADHD</td>
<td></td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>SLI[2–3]</td>
<td>DCD[3–6]</td>
<td>36%</td>
<td>0.053</td>
<td>0.81</td>
<td>SLI meets DCD</td>
<td>occurs less than expected by independent events</td>
<td>Artifact suspected</td>
<td>Fr</td>
</tr>
<tr>
<td>27</td>
<td>(ADHD,SLI)[3–6]</td>
<td>DCD[6–9]</td>
<td>35%</td>
<td>0.046</td>
<td>1.62</td>
<td>ADHD, SLI meets DCD</td>
<td>floating Dx</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>SLI[3–6]</td>
<td>SLI[6–9]</td>
<td>34%</td>
<td>0.133</td>
<td>1.93</td>
<td>SLI persists</td>
<td>persists</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>(DCD,ADHD)[3–6]</td>
<td>(DCD,ADHD)[6–9]</td>
<td>34%</td>
<td>0.07</td>
<td>2.85</td>
<td>DCD is equal to ADHD</td>
<td>persists</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>(DCD,SLI-C)[3–6]</td>
<td>DCD[6–9]</td>
<td>34%</td>
<td>0.037</td>
<td>1.56</td>
<td>SLI-C starts DCD</td>
<td>floating Dx</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>GMDD[0–2]</td>
<td>DCD[3–6]</td>
<td>34%</td>
<td>0.032</td>
<td>0.77</td>
<td>If GMDD occurs, then DCD is unlikely to replace it</td>
<td>Artifact suspected</td>
<td>At early age DCD cannot be diagnosed yet, it can be only be suspected and noted as GMDD</td>
<td>Fr</td>
</tr>
<tr>
<td>32</td>
<td>ADHD[3–6]</td>
<td>DCD[6–9]</td>
<td>33%</td>
<td>0.103</td>
<td>1.52</td>
<td>ADHD meets DCD</td>
<td>floating Dx</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>(DCD,ADHD,SLI)[3–6]</td>
<td>(ADHD,DCD)[6–9]</td>
<td>32%</td>
<td>0.031</td>
<td>2.71</td>
<td>SLI starts DCD, ADHD</td>
<td>floating Dx</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>(DCD,SLI-C)[3–6]</td>
<td>SLI[6–9]</td>
<td>32%</td>
<td>0.035</td>
<td>1.85</td>
<td>DCD, SLI-C meets SLI</td>
<td>floating Dx</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>(ADHD,SLI)[3–6]</td>
<td>(ADHD,SLI)[6–9]</td>
<td>31%</td>
<td>0.04</td>
<td>4.08</td>
<td>ADHD is equal to SLI</td>
<td>persists</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>(DCD,ADHD,SLI)[3–6]</td>
<td>(ADHD,SLI)[6–9]</td>
<td>31%</td>
<td>0.03</td>
<td>4.11</td>
<td>DCD starts ADHD, SLI</td>
<td>floating Dx</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>SLI-C[3–6]</td>
<td>SLI[6–9]</td>
<td>31%</td>
<td>0.054</td>
<td>1.79</td>
<td>SLI-C meets SLI</td>
<td>persists</td>
<td>The diagnosis is stable(SLI-combined and SLI are similar)</td>
<td>+</td>
</tr>
<tr>
<td>38</td>
<td>(DCD,SLI-C)[3–6]</td>
<td>LD[6–9]</td>
<td>31%</td>
<td>0.034</td>
<td>2.28</td>
<td>DCD, SLI-C meets LD</td>
<td>floating Dx</td>
<td>Reference to disorders becomes less specific in this age</td>
<td>+</td>
</tr>
</tbody>
</table>

+ approved sequences, – artifacts, Fr – future research; ADHD – Attention Deficit and Disruptive Behavior Disorders; ASD – Autistic Spectrum Disorder; DCD – Developmental Coordination Disorder; DD – Developmental Delay; Dx – Diagnosis; GMDD – Gross Motor Developmental Disorder; NPD – Normal Psychomotor Development; SLI – Specific Language Impairment; SLI-C – SLI-combined
Figure 1
Temporal comorbid disorders visualization model. The squares define comorbid groups and arrows define confidence. The numbers above arrows represent the number of the sequence (used in Table 2) and the numbers in the brackets represent the confidence.
ADHD – Attention Deficit and Disruptive Behavior Disorders; ASD – Autistic Spectrum Disorder; DCD – Developmental Coordination Disorder; DD – Developmental Delay; NPD – Normal Psychomotor Development; SLI – Specific Language Impairment; SLI-C – SLI-combined
In patterns with ID#: 1, 11, 5, 29, 33, 35, we have found that ADHD and ASD exhibit high persistence. This is a clinically relevant result and not just a trivial pattern. This result is reported in various articles such as [39][40] which note the persistence of ASD as the child ages and persists even into adulthood.

Additionally, we found numerous patterns with presence of comorbidities between various disorders:

- In the sequences: 29, 33, 35, 36 we have found that the patients with persistent ADHD tend to develop disorders in addition to ADHD – requiring constant vigilance on the part of the clinician during regular follow-up [41, 42].
- Comorbidities between DCD and SLI were found and supported in the literature [43, 44] – patterns: 16, 18, 22, 23, 33, 36.

Note that we have found the following co-morbidities: {ADHD,DCD}, {ADHD,SLI}, {DCD,SLI}, {ADHD,DCD,SLI}, which replicate the results that were found in previous work [9] which had used non-temporal NDD comorbidity discovery method. The fact that we have found similar insights by running an effective sequence mining algorithm – which may be performed in real time – emphasizes the usefulness of our approach in terms of correct and significant results, low cost, and high speed. These results yield additional value for clinicians who can see the progression of comorbid disorders over time in patient populations.

Two patterns with lift less than 1 (rows 26 and 31 in Table 2) were found. These patterns represent a type of “negative” co-morbidity; that is, events that have a lower than expected probability of appearing together. Those results were not shown in our

4.1 Important Results and Interesting Observations

The clinical expert was not expecting to find novel unexpected patterns but instead, he was particularly interested in automatically finding patterns that could be corroborated by formal large-scale long-term medical studies involving tens to hundreds of patients and spanning over several years. Specifically, some of the most important results are:

- In patterns with ID#: 1, 11, 5, 29, 33, 35, 36 we have found that ADHD and ASD exhibit high persistence. This is a clinically relevant result and not just a trivial pattern. This result is reported in various articles such as [39][40] which note the persistence of ASD as the child ages and persists even into adulthood.

Note that we have found the following co-morbidities: {ADHD,DCD}, {ADHD,SLI}, {DCD,SLI}, {ADHD,DCD,SLI}, which replicate the results that were found in previous work [9] which had used non-temporal NDD comorbidity discovery method. The fact that we have found similar insights by running an effective sequence mining algorithm – which may be performed in real time – emphasizes the usefulness of our approach in terms of correct and significant results, low cost, and high speed. These results yield additional value for clinicians who can see the progression of comorbid disorders over time in patient populations.

Two patterns with lift less than 1 (rows 26 and 31 in Table 2) were found. These patterns represent a type of “negative” co-morbidity; that is, events that have a lower than expected probability of appearing together. Those results were not shown in our

Table 3 Cross validation results

<table>
<thead>
<tr>
<th>Set</th>
<th>Iterations Num</th>
<th>Kolmogorov-Smirnov (P-values)</th>
<th>Pearson correlation</th>
<th>Spearman correlation</th>
<th>Proportion Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Supp</td>
<td>Conf</td>
<td>Lift</td>
<td>Supp</td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>0.957</td>
<td>0.161</td>
<td>0.996</td>
<td>0.979</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
<td>0.957</td>
<td>0.074</td>
<td>0.847</td>
<td>0.974</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>0.944</td>
<td>0.476</td>
<td>0.329</td>
<td>0.97</td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td>0.699</td>
<td>0.169</td>
<td>0.699</td>
<td>0.972</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>0.591</td>
<td>0.056</td>
<td>0.734</td>
<td>0.976</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>0.994</td>
<td>0.136</td>
<td>0.944</td>
<td>0.979</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>1</td>
<td>0.964</td>
<td>0.864</td>
<td>0.984</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>1</td>
<td>0.994</td>
<td>0.944</td>
<td>0.987</td>
</tr>
<tr>
<td>9</td>
<td>40</td>
<td>0.759</td>
<td>0.164</td>
<td>0.988</td>
<td>0.977</td>
</tr>
<tr>
<td>10</td>
<td>47</td>
<td>0.838</td>
<td>0.504</td>
<td>0.674</td>
<td>0.971</td>
</tr>
<tr>
<td>Avg</td>
<td>46.9</td>
<td>0.8739</td>
<td>0.3698</td>
<td>0.8019</td>
<td>0.9769</td>
</tr>
</tbody>
</table>

© Schattauer 2016
visualization, which shows comorbid disorders with increased probability of appearance at subsequent time intervals. However it would be interesting to explore why these disorders “protect” against the appearance of subsequent disorders.

In addition, we have found several patterns highlighting organizational behaviors (patterns marked as “artifact of work habits” in Table 2). These behaviors were known as characteristic work practices of the center, where not all NDDs were documented in focused visits that concerned followup regarding treatment of particular NDDs. Naturally, data completeness could limit pattern discovery. Discovery of these patterns provided us with ideas to improve the current center’s information system so that clinicians could visualize the child’s history with respect to NDDs and could easily mark which previous NDDs are also present in the current visit. NDDs that are predicted to co-occur in high confidence are also suggested and a warning may be issued to indicate that an inconsistent diagnosis was entered (see Section 2.3.1 – impossible combinations). The physician could also add additional NDDs from the standardized concept list (see Appendix G).

4.2 Study Limitations

The results of this study have several limitations. First, as noted, our data reflects in part patterns stemming from physician’s work practices at a particular location with the administrative need to focus only on certain NDDs (ADHD). As such, the data does not fully reflect the children’s development. Therefore, although we found temporal patterns of progression of NDDs we could not use them for predicting a child’s prognosis within acceptable certainty, mainly due to data incompleteness. For example we found that ASD persists in ages 3–6 to age 6–9 in 48% of children, but according to our medical expert, the expected presence at subsequent time intervals. However it would be interesting to explore why these disorders “protect” against the appearance of subsequent disorders.

In addition, we have found several patterns highlighting organizational behaviors (patterns marked as “artifact of work habits” in ▶Table 2). These behaviors were known as characteristic work practices of the center, where not all NDDs were documented in focused visits that concerned followup regarding treatment of particular NDDs. Naturally, data completeness could limit pattern discovery. Discovery of these patterns provided us with ideas to improve the current center’s information system so that clinicians could visualize the child’s history with respect to NDDs and could easily mark which previous NDDs are also present in the current visit. NDDs that are predicted to co-occur in high confidence are also suggested and a warning may be issued to indicate that an inconsistent diagnosis was entered (see Section 2.3.1 – impossible combinations). The physician could also add additional NDDs from the standardized concept list (see ▶Appendix G).

5. Conclusions

We have demonstrated the feasibility of applying a patterns discovery methodology for NDD comorbidity progression with the results from the automated approach accurately reflecting clinical research data. Discovering temporal patterns of comorbid NDDs can have both theoretical and practical value. Theoretically, comorbidities may result from a mutual event or mechanisms by which the comorbid NDDs develop. Practically, knowledge of such patterns may improve prognosis and diagnosis of NDDs based on previously existing NDDs and thus help medical professionals to devise better treatment [46].

Our work supplies supporting evidence for the feasibility of using sequence mining techniques for research of comorbidity of disorders. In future work, elaborating our current approach over a bigger and more representative dataset will allow developing a predictive model which will predict highly probable disorders based on discovered temporal patterns. In addition, the patterns that we have discovered could be used as features by various interval-based temporal data mining techniques [14, 37] as the next step for future studies, to better handle co-occurrence of interval-based events and states.

References


I. Pimus et al.: Sequence Mining of Comorbid Neurodevelopmental Disorders Using the SPADE Algorithm


