

IN THE CLINIC

# Text Mining

## Gathering Evidence From Practice

Aimee Piller

Catherine Candler

**Using interventions for children with coordination delay as an example, five steps to analyzing information from electronic health records to determine intervention efficacy.**

**R**esearch concerning the efficacy of occupational therapy intervention has been identified as being of great importance for the profession, in line with the AOTA Centennial Vision, which emphasizes the importance of increasing research in the field of occupational therapy (2007).

This article describes a retrospective analysis of intervention within a single pediatric occupational therapy clinic. Our purpose was to explore how practitioners can contribute to the growing body of evidence for occupational therapy interventions through routine data collection, by using information gathered from the electronic health record (EHR) through the technique of text mining.

### EHR

The EHR is the accepted standard of medical documentations and files (U.S. Department of Health and Human Services, n.d.). The electronic format of the record stores information regarding

patients, treatment interventions, and outcomes in a readily accessible format. Within the EHR, narrative daily notes are a common form of occupational therapy documentation. The narrative format of treatment notes produces qualitative data in the form of free form text.

Text mining is a method used to examine and analyze free form text (Suominen, 2014). Retrospective studies in health care use text mining for classification of treatments (Zhou, Amundson, Yu, Kessler, Benzinger, & Wippold, 2014) and to gather evidence on frequency of interventions that guides decisions and research in health care (Suominen, 2014). It provides a method to gather and analyze data collected by practitioners during occupational therapy treatment.

Occupational therapists at our pediatric clinic used a five-step process of utilizing retrospective information from the EHR to provide information on treatment practices and outcomes within the daily practice of occupational therapy. A retrospective analysis of the EHR for

children with coordination delay provides an example of the process.

### Gathering Evidence Through Practice

#### STEP 1 Research Question

The first step in the process of analyzing information from the EHR is to determine what information to gather from the record. Determination of data begins with asking a research question. Questions should be specific and provide clear guidance for data collection. Questions regarding intervention require three categories: (1) the population of interest, (2) the intervention under examination, and (3) outcome measures. In our review, the following research questions were identified:

- What were the characteristics of children with coordination delay who received occupational therapy services at an outpatient clinic from 2011 to 2014?

- What interventions do occupational therapists employed in an outpatient pediatric clinic use most in treatment of kids with coordination delay?
- What were the therapeutic outcomes for the children with coordination delay who received occupational therapy services at an outpatient clinic from 2011 to 2014?

## STEP 2 Literature Review

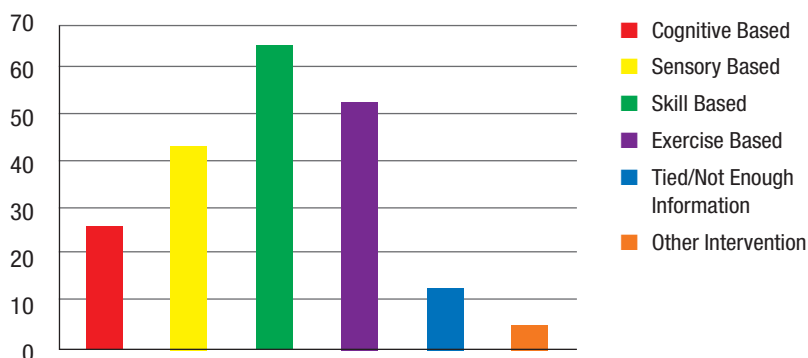
Step two is to perform a thorough literature review on the current treatment interventions for the population chosen. The literature review identifies the evidence, as well as gaps in the evidence, that are currently available about the most effective interventions for the target population. A review of the literature provides a summary of evidence and a summary of evidence gaps for treatment interventions for the target population.

The literature review revealed several recognized treatments for children with coordination delay. Researched treatments include cognitive orientation to daily occupational performance, sensory integration, neuro-motor task training, goal-oriented group interventions, exercise programs, and compensatory strategies. The interventions varied in effectiveness and were only evaluated as isolated treatments (Armstrong, 2012).

## STEP 3 Gathering the Data

Once the literature review is complete, step three is to set inclusion criteria and identify the records that meet the criteria. The records must keep the identifying patient information hidden to ensure confidentiality. It is always best to include demographic information regarding clients. Demographic information provides data on the population receiving the interventions. Gather any pretest or posttest data available in the record from evaluations, reevaluations, and discharge summaries. Compile the section of the daily treatment notes that describes the treatment interventions (i.e., the objective portion of the SOAP [subjective, objective, assessment, and plan] note) into a word processing document.

**Figure 1. Number of Words Used Within Narrative Data That Are Associated With Particular Categories of Interventions for Coordination Delay**



Inclusion criteria include information regarding age, diagnosis or area of dysfunction, and setting of treatment. It is by the research question and subsequent review of current evidence on the topic. In addition, the practitioner may bring in clinical knowledge to determine inclusion criteria. Inclusion criteria for our study were:

- Children ages 6 years 0 months to 12 years, 11 months at the time of initial evaluation
- Received outpatient occupational therapy services at the clinic sometime between January 2011 and September 2014
- Demonstrated delay in coordination as measured on standardized motor assessment

The particular EHR template utilized in this study had a separate field to pull data without providing identifying client information. Thirty-two client records met the inclusion criteria. The study used more than 1,900 narrative notes from the 32 records. Data pulled for analysis included the objective portion of the SOAP note. The information from the free form text of the treatment notes was compiled into a word processing document. We gathered additional data from the EHR into a separate research field that included pretest and posttest data.

## STEP 4 Data Analysis

Step four is to analyze the data collected. The original research question guides the data analysis.

### Population

The population describes the demographics of the identified records. Our study included six females and 26 males. Eleven of the participants also received speech therapy at the facility. The mean number of treatment sessions was 58.56. (See Table 1 on p. 11 for the types of diagnoses.)

### Intervention

For our study, we needed to identify the interventions used from the qualitative data in the SOAP notes. Text mining was our tool of choice to determine the frequency of words used in free form text. Several commercial software programs are available to analyze free form narrative data.

We used NVivo software, which is available for download for a nominal fee ([www.qsrinternational.com/products\\_nvivo.aspx](http://www.qsrinternational.com/products_nvivo.aspx)). A word query within NVivo counted word frequency from the compiled narrative notes. The query included stem words in the count of the base word and excluded stop words (e.g., “the,” “a,” “on,” etc.). Following the query, qualifying words and words that defined the treatment session were excluded so that only treatment approaches were used in the search. We identified the top 200 words from the final list. We then grouped the words into categories based on intervention methodology.

Six pediatric occupational therapists analyzed the top 200 words by placing them in one of four main categories of treatments for coordination delay. These categories consisted of treatment approaches identified in the literature.

**Table 1. Diagnoses of Clients**

Medical Diagnosis	Number of Records	Treatment Diagnosis	Number of Records
Autism	3	Autism	4
Attention deficit hyperactivity disorder	3	Muscle weakness	12
Speech delay	2	Lack of coordination	12
Developmental Coordination disorder	1	Sensory integrative dysfunction	1
Sensory integrative dysfunction	2	Lack of expected development	2
Dysphagia	1	No diagnosis	1
No diagnosis	20	<b>Total</b>	<b>32</b>
<b>Total</b>	<b>32</b>		

**Text mining allows for analysis of narrative data collected on a daily basis within any therapy setting.**

We designed the categories to classify the intervention techniques described by each word. The first category consisted of cognitive approaches, including behavioral approaches, specific cognitive approaches, and other cognitive-based strategies. The second category included sensory integration interventions and sensory strategies. The third category consisted of skill-based interventions, such as direct practice of specific developmental skills. The final category, exercise, incorporated biomechanical interventions, such as interventions addressing strength, postural control, and balance. A miscellaneous category provided a place for interventions that did not fit into any of the four provided categories. A final category allowed for the placement of words that did not have enough information to make a determination as to the category of intervention.

We used a rule of majority to identify the category of treatment for each word. A seventh therapist provided an additional category to place the word in the event of equal frequency from the six raters. Figure 1 on page 10 shows the frequency of the categories of treatment interventions used in this setting.

We also wanted to determine whether a treatment session used multiple intervention types. We randomly chose five notes from each participant and

compiled them into a text document. We compiled the top three words by frequency from each of the four intervention categories. We ran a query of the 160 random notes using 12 words to reveal the frequency of interventions within each note. One note did not include any interventions. However, this was identified as a progress report note. No notes utilized only one type of intervention, and eight notes utilized two types of interventions. One hundred and fifty-one notes utilized three or more intervention types within one documented session.

**Outcome measures**

In the outpatient setting, it is routine to use standardized measures of skills and functional abilities. Pretest scores from the initial evaluation and posttest scores from reevaluation or discharge summaries are essential aspects for measuring client outcomes. Statistical analysis of scores provides information on significance of improvement.

Outcome measures gathered from the records included pretest and posttest scores from the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (Bruininks & Bruininks, 2005). However, only three of the records had both initial evaluation and reevaluation scores. There was not enough information to make a conclusion or analysis of outcomes.

**STEP 5  
Conclusions**

The final step in a retrospective study of treatment notes is to draw conclusions based on the data analysis. Conclusions guide future practice, provide information for researchers, and provide information to improve therapy practice and documentation within the setting.

The study on treating coordination delay in children provided valuable information. First, interventions provided in this setting, for the most part, matched the interventions defined in the literature. The therapists in this setting were using evidence-based practice in treating coordination delay. Second, the information regarding frequency of interventions indicated that practitioners used a variety of approaches. The literature only provides evidence for interventions performed in isolation. This disparity highlighted the need to explore the efficacy of combined interventions. Finally, not enough information was available to make a determination of therapeutic outcomes. This lack of data illuminated a hole in the documentation practices of the clinic. The retrospective study resulted in a modification to documentation and reevaluation procedures in this setting.

**Application of Evidence From Practice**

The advent of the EHR provides an easy way for the occupational therapy practitioner to perform a retrospective study of treatment practices and interventions. The material gathered from the EHR provides information regarding the demographics of patients seen, duration of intervention, and frequency of treatment. In addition, the EHR provides a forum to access free-form text in the daily treatment notes of therapists. Text mining allows for analysis of narrative data collected on a daily basis within any therapy setting. The information gathered from a retrospective analysis defines the treatments utilized in the occupational therapy treatment process. Historically, clinicians have been more consumers of research, and less



# Winning partnerships

**Lucinda Dale, EdD, OTR, CHI, FAOTA**

*UIndy Professor*

**Research Interest:**

Use of tools for outcomes measurement in clinical practice

**Recent Project:**

Partnering with a local hospital to collaboratively develop and evaluate an intervention for tennis elbow sufferers.

**Recent Honor:**

Named a Fellow of the American Occupational Therapy Association

“Universities need clients to provide experience for students. Hospitals need researchers. When the two work together, students do research that’s clinically relevant and guides OT practice. It’s research that really matters—a win for everyone.”

[uindy.edu/otp](http://uindy.edu/otp)

UNIVERSITY of  
**INDIANAPOLIS.**  
*School of Occupational Therapy*

*Inspiring Excellence.®*



of contributors to the field of evidence. Treatment interventions defined from the perspective of the clinician provides information to the researcher for further study of the effectiveness of interventions. Text mining is a feasible way for all practitioners to contribute to the body of evidence for occupational therapy treatments. It provides a way to begin to delve into therapy interventions for the world to see, use, and study. 📄

## References

American Occupational Therapy Association. (2007). AOTA's Centennial Vision and executive summary. *American Journal of Occupational Therapy*, 61, 613–614. <http://dx.doi.org/10.5014/ajot.61.6.613>

Armstrong, D. (2012). Examining the evidence for interventions with children with developmental coordination disorder (report). *British Journal of Occupational Therapy*, 75, 532.

Bruininks, R. H., & Bruininks, B. D. (2005). *Bruininks-Oseretsky Test of Motor Proficiency: Second edition (BOT-2)*. Bloomington, MN: Pearson.

Suominen, H. (2014). Text mining and information analysis of health documents. *Artificial Intelligence in Medicine*, 61(3), 127. <http://dx.doi.org/10.1016/j.artmed.2014.06.001>

U.S. Department of Health and Human Services. (n.d.). *Key features of the affordable health care act by year*. Retrieved from <http://www.hhs.gov/healthcare/facts/timeline/timeline-text.html>

Zhou, Y., Amundson, P., Yu, F., Kessler, M., Benzinger, T., & Wippold, F. (2014). Automated classification of radiology reports to facilitate retrospective study in radiology. *Journal of Digital Imaging*, 27, 730–736. <http://dx.doi.org/10.1007/s10278-014-9708-x>

**Aimee Piller**, MOTR/L, is owner of Piller Child Development, an outpatient pediatric clinic located in Phoenix, Arizona. She is currently pursuing her PhD at Texas Woman's University.

**Catherine Candler**, OTR, PhD, BCP, is professor at Abilene Christian University. At the time of this project she was director of the School of Occupational Therapy at Texas Woman's University.

## For More Information



### Evidence-Based Practice and Research Resources

[www.aota.org/practice/researchers](http://www.aota.org/practice/researchers)



### Evaluation in Occupational Therapy: Obtaining and Interpreting Data, 4th Edition

By J. Hinojosa & P. Kramer, 2014. Bethesda, MD: AOTA Press. (\$69 for members, \$98 for nonmembers. To order, call toll free 877-404-AOTA (2682) or shop online at <http://store.aota.org>, and enter order #900356.