



Adopting Text Mining on Rehabilitation Therapy Repositioning for Stroke

Guilin Meng^{1,2}, Yong Huang^{2,3}, Qi Yu⁴, Ying Ding², David Wild², Yanxin Zhao¹, Xueyuan Liu^{1*} and Min Song^{5*}

¹Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai, China, ²School of Informatics Computing and Engineering, Indiana University, Bloomington, IN, United States, ³School of Information Management, Wuhan University, Wuhan, China, ⁴School of Management, Shanxi Medical University, Shanxi, China, ⁵School of Informatics, Yonsei University, Seoul, South Korea

OPEN ACCESS

Edited by:

Hans-Leo Teulings,
Neuroscript, United States

Reviewed by:

Junguk Hur,
University of North Dakota,
United States
Anis Yazidi,
OsloMet—Oslo Metropolitan
University, Norway

*Correspondence:

Xueyuan Liu
1510922@tongji.edu.cn
Min Song
min.song@yonsei.ac.kr

Received: 05 October 2018

Accepted: 05 March 2019

Published: 19 March 2019

Citation:

Meng G, Huang Y, Yu Q, Ding Y, Wild D, Zhao Y, Liu X and Song M (2019) Adopting Text Mining on Rehabilitation Therapy Repositioning for Stroke. *Front. Neuroinform.* 13:17. doi: 10.3389/fninf.2019.00017

Stroke is a common disabling disease that severely affects the daily life of patients. Accumulating evidence indicates that rehabilitation therapy can improve movement function. However, no clear guidelines have specific and effective rehabilitation therapy schemes, and the development of new rehabilitation techniques has been relatively slow. This study used a text mining approach, the ABC model, to identify an existing rehabilitation candidate therapy method that is most likely to be repositioned for stroke. In the model, we built the internal links of stroke (A), assessment scales (B), and rehabilitation therapies (C) in PubMed and the links were related to upper limb function measurements for patients with stroke. In the first step, using E-utility, we retrieved both stroke-related assessment scales and rehabilitation therapy records and then compiled two datasets, which were called Stroke_Scales and Stroke_Therapies, respectively. In the next step, we crawled all rehabilitation therapies co-occurring with the Stroke_Therapies and then named them as All_Therapies. Therapies that were already included in Stroke_Therapies were deleted from All_Therapies; therefore, the remaining therapies were the potential rehabilitation therapies, which could be repositioned for stroke after subsequent filtration by a manual check. We identified the top-ranked repositioning rehabilitation therapy and subsequently examined its clinical validation. Hand-arm bimanual intensive training (HABIT) was ranked the first in our repositioning rehabilitation therapies and had the most interaction links with Stroke_Scales. HABIT significantly improved clinical scores on assessment scales [Fugl-Meyer Assessment (FMA) and action research arm test (ARAT)] in the clinical validation study for acute stroke patients with upper limb dysfunction. Therefore, based on the ABC model and clinical validation, HABIT is a promising repositioned rehabilitation therapy for stroke, and the ABC model is an effective text mining approach for rehabilitation therapy repositioning. The findings in this study would be helpful in clinical knowledge discovery.

Keywords: text mining, ABC model, stroke, hand-arm bimanual intensive training, upper extremity

INTRODUCTION

Stroke is a common disabling health-care problem, and it is the second-leading cause of mortality and disability worldwide. In the United States, nearly 0.8 million people have stroke annually, and the estimated direct and indirect cost of stroke was \$95 billion in 2015 and is expected to rise to 185 billion in 2030 (Brainin and Zorowitz, 2013; Cramer et al., 2017). The symptoms of acute stroke include physical impairments and cognitive dysfunction, and physical impairments of the affected limbs include movement restriction, sensory loss, and muscle activation abnormalities (Kuehn, 2018). About 50% of acute stroke survivors suffer from dysfunction of the upper limbs in the chronic phase (Favre et al., 2014), severely affecting the daily life and the therapeutic effect of rehabilitation therapy and reducing the quality of life of patients after stroke.

Rehabilitation therapies offer a chance for an individual to improve recovery and adapt to the new situation following acute stroke. Effective rehabilitation can help remodel the residual cortex, establish synaptic connections, and improve neurological function (Silva et al., 2018). A number of studies have investigated methods of rehabilitation management, including task-oriented training (Carrico et al., 2016), impaired limb forced training (Kwakkel et al., 2015), movement science-based therapy, robotic-assisted movement, virtual reality (VR) training (Brunner et al., 2017), functional electrical stimulation (Kattenstroth et al., 2018), and skill acquisition training paired with impairment mitigation and motivational enhancement. Small clinical studies have demonstrated that these rehabilitation methods can improve post-stroke dysfunction; however, no rehabilitation management is currently used as a part of the routine guideline practice, and the relative effectiveness of existing rehabilitation strategies has not been sufficiently evaluated in large clinical trials (Pollock et al., 2014). In addition, the quality, precision and focus, and risk of bias of these methods in these studies have not been assessed, and little breakthrough has been achieved with respect to intervention timing, dosage, biomarkers, standard assessment time points, and measures of current clinical practices. Furthermore, the stagnant development of new competitive rehabilitation strategies impedes rehabilitation development.

Therefore, the study of stroke rehabilitation has a long way to go. With the advent of the era of big data, new ideas can be provided for current standardized and personalized stroke rehabilitation strategies, and collaborative and multidisciplinary research programs need to be developed to provide new intervention choices for stroke rehabilitation with a standard universal rationale. Rehabilitation therapy repositioning provides new insight into possible unfolding data sources, such as a large amount of literature in PubMed, to mine appropriate new therapy strategies for stroke. The principal advantages of rehabilitation therapy repositioning over new therapy development are that the approved therapy has already been tested for safety, and repositioning can eliminate the time and cost of developing new therapies. Text mining can automatically extract information (facts or data) by systematically scrutinizing a vast number of abstracts or

full-text versions of scientific publications (Percha and Altman, 2015; Meng et al., 2018a). Although there is no other published report on text mining in rehabilitation, some classic application examples can be found in the health science field (Westergaard et al., 2018). First, the publications in this field of text mining are growing at an exponential rate (Ding et al., 2013). Second, the medical literature is more standardized than other types of literature (He et al., 2011). Third, terms of the medical literature are relatively uniform and more convenient for mining. Typical text mining systems include the IBM Watson diagnosis system (Ahmed et al., 2017), DrugQuest (Papanikolaou et al., 2016), and disease-causing genes system (Sun et al., 2017).

This study aimed to examine the indirect relationship in the literature through the ABC model (Swanson and Smalheiser, 1997) to discover a promising rehabilitation strategy (rehabilitation repositioning). The ABC model has an internal connection among disease (A), assessment scale (B), and rehabilitation therapy (C). For stroke, most assessment scales of upper limb impairment assess a person's ability to manage daily activities that require the use of the upper limbs; in addition, the rehabilitation therapies for functional improvement of upper limb share the same set of assessment scales, regardless of whether they are applicable for stroke or not. Therefore, we planned to identify undiscovered rehabilitation therapies for stroke through shared assessment scales and ultimately validate the most promising candidate in clinical settings.

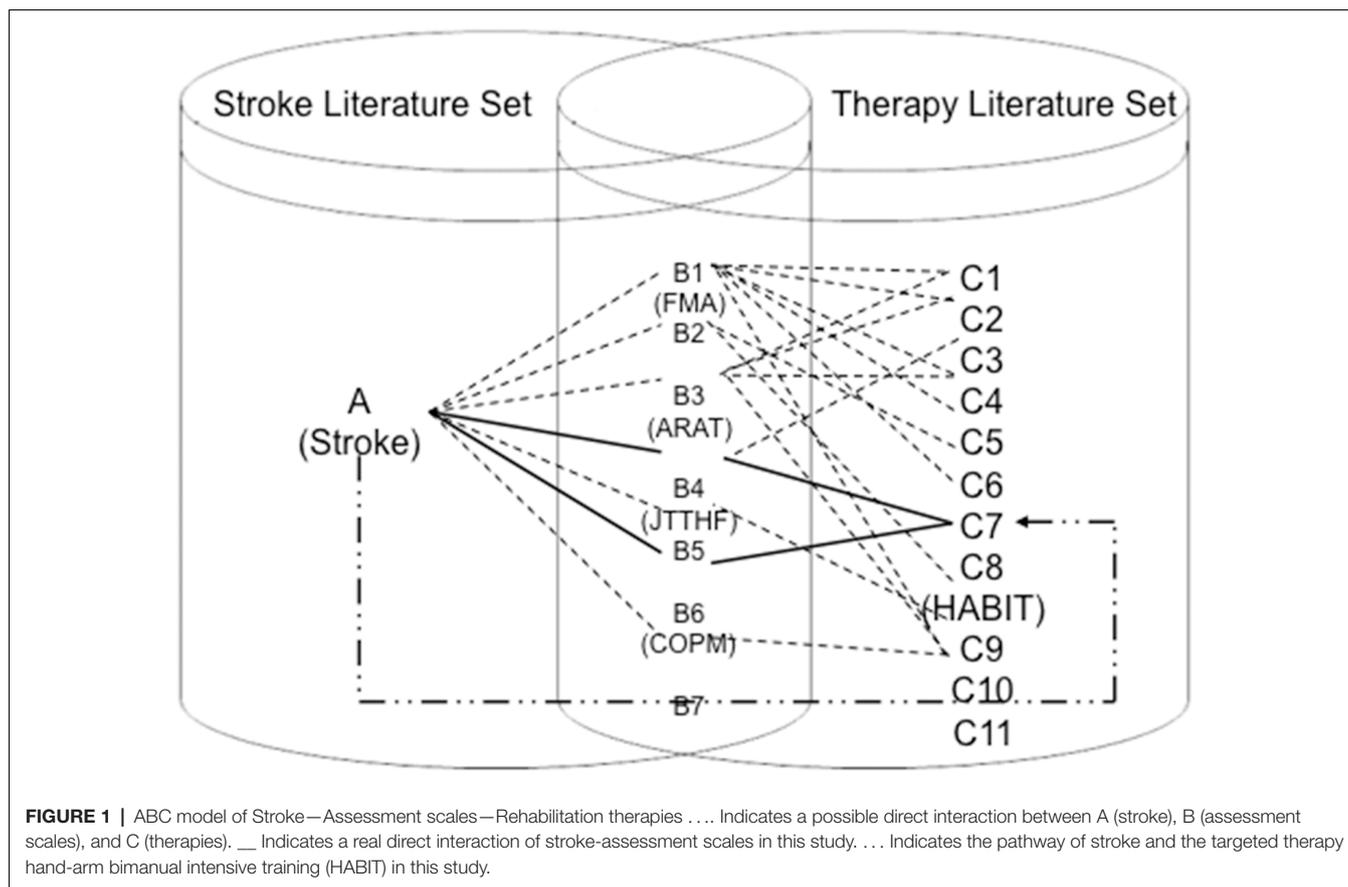
MATERIALS AND METHODS

Our first focus was to find repositioning rehabilitation therapy candidates from an extensive collection of articles in PubMed. An early version of this study was published at Meng et al. (2018b). The present study describes our method in more detail. The codes and stroke-related PubMed data of these articles can be found in https://github.com/hyyc116/Stroke_findings/tree/master/DSTN. DSTN is the abbreviation of "disease scale test network."

ABC Model

In order to identify new rehabilitation therapies for stroke, we developed a relation extraction method based on ABC model, with which new knowledge could be discovered from sets of disjointed scientific entities, as shown in **Figure 1**. **Figure 1** is a graphic representation of the ABC model adopted to discover rehabilitation therapy repositioning for Stroke. It illustrates that there are three different term sets: (1) stroke related terms extracted from Stroke Literature Set; (2) therapy related terms extracted from Therapy Literature Set; and (3) intermediate terms that occur in both literature sets. The connection among these three term sets is from Stroke to Therapy *via* intermediate terms (a.k.a. B terms). In particular, In **Figure 1**, dotted lines and solid lines represent a connection or a path from source to target terms. The solid line implies that it is a target discovery linked from Stroke to Therapy *via* B terms whereas the dotted line is not of interest from the new discovery perspective.

The ABC model embodies the stages of retrieval, interpretation, and analysis. The retrieval phase investigates



unstructured stroke-related information, designs text queries, and extracts the entity documents; the interpretation phase constructs a connected graph to represent A-B and B-C relationships between entities; in the analysis phase, annotation information between transitivity diffusion entities is used to sort the best entity that should be further verified. The ABC model starts with the theme “A” (stroke) in PubMed that collects scientific questions. The phrase “B” (assessment scale) is then listed, and a separate search is used for each “B” term; then, the phrase “C” (rehabilitation therapy) appearing in the code of the B is compiled; finally, with frequency criteria, the C terms are ranked such that a high ranked C term is used to represent the most promising hypothesis. For example, the term C may be the name of a therapeutic strategy that has not been tested for A (stroke) but has been demonstrated in other situations (e.g., in other forms of physical injury model or experimental animal model) with curative effects, suggesting that C may be explored as a new therapy.

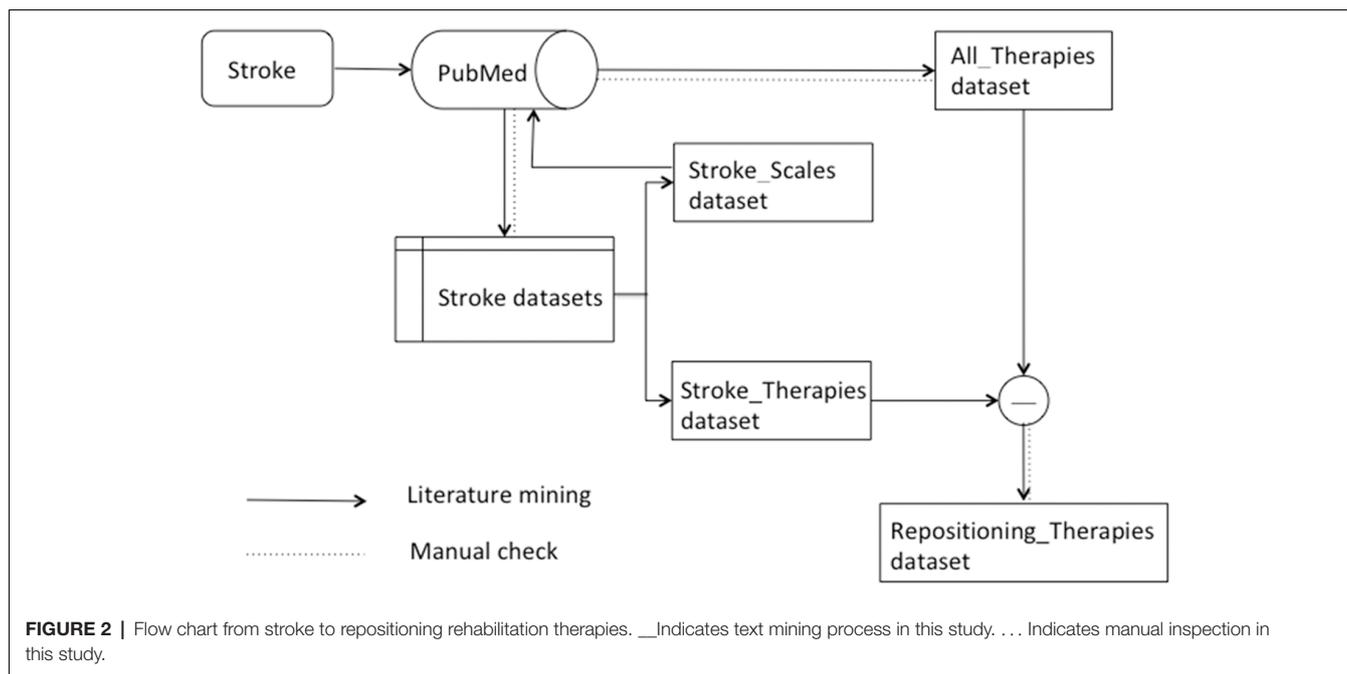
Study Procedure

The overall flow chart is shown in **Figure 2**.

Stroke-Related Assessment Scales and the Creation of Rehabilitation Therapy Datasets

To collect articles related to stroke with upper limb impairment, we searched PubMed with stroke-related keywords (“stroke”

OR “cerebral infarction” OR “brain ischemia” OR “cerebral hemorrhagic” OR “subarachnoid hemorrhage”) and (“hand” OR “arm” OR “upper extremity” OR “upper limb”) in April 2016. We did not use Medical Subject Headings (MeSHs) terms because most specific assessment scales and therapies do not directly belong to MeSH terms. When we searched stroke-related rehabilitation items, the MeSH terms that we obtained were only the common words, which should be deleted as confounding factors. The NCBI provides the E-utilities API for data query and download; therefore, we utilized python to call the API to fetch all query-related data in PubMed of NCBI through the HTTP protocol, and the data were then stored in plain text files (French et al., 2015). Scales and rehabilitation therapies are always used as noun phrases (NPs) in scientific articles. Thus, we applied a shallow chunk analyzer to extract NPs. The NPs ending with “test,” “scale,” “assessment,” “measure,” “score,” or “index” with frequency as the number of articles more than five were considered possible stroke assessment scale candidates. In addition, NPs ending with “training,” “therapy,” “treatment,” “treatments,” “practice,” “program,” “practise,” or “simulation” with frequency more than five were considered possible stroke therapy candidates. Besides frequency, the term frequency-inverse document frequency (TF-IDF) was also used as an index in this study. TF-IDF, a commonly used weighting technique for information retrieval and data mining, can be used to



evaluate the importance of a phrase in a file; the importance of a phrase increases in proportion to the frequency it appears in the file. Because relying only on text mining could lead to false negatives, two senior neurologists (Liu and Zhao) examined the whole set independently and compared against each other to identify true positives as accurately as possible. The agreement rate of the two neurologists was as high as 96%, and the reconfirmation of the individual set with doubts reached a unified conclusion. In addition, they compared the results with the latest stroke rehabilitation guidelines of American Heart Association (Winstein et al., 2016) before stroke-related scale dataset (Stroke_Scales) and rehabilitation therapy dataset (Stroke_Therapies) were established.

Entire Rehabilitation Therapy Dataset Creation

With a direct stroke—assessment scales and therapies—assessment scales relationship, we crawled all therapies-related NP in PubMed co-occurring with Stroke_Scales *via* E-Utilities. The proceeding entire rehabilitation therapy dataset, named All_Therapies building, was similar to Stroke_Scales building mentioned above, and the NPs ending with “training,” “therapy,” “treatment,” “treatments,” “practice,” “program,” “practice,” or “simulation” with frequency more than five were kept to be possible therapy candidates, in which the assessment items were in Stroke_Scales. Manual inspection was the same as described above in Stroke_Therapies.

Potential Repositioning Rehabilitation Therapy Dataset

Therapies already included in Stroke_Therapies were deleted from All_Therapies so that the remaining therapies were not stroke-applied therapies, which could be repurposed for stroke. Manual inspection was then carried out again.

Hypothesis Validation

Validation of the Retrieved Repositioning_Therapy Dataset in PubMed

The articles related to potential repositioning rehabilitation therapies and stroke-related keywords were retrieved from PubMed to ensure that no articles involve associations of therapies with stroke.

Further Rehabilitation Theory of Potential Candidate Exploration

The knowledge discovery is full of uncertainty and complicated. In the knowledge discovery, algorithms and methods could be perfect in theory, while the precision, recall, or some other metrics could be meaningless to some extended theories. This study aimed to find potential candidates. Therefore, clinical values of the potential candidates should be further comprehensively explored in the mechanism and principles of the rehabilitation theory.

Validation in Clinical Trial

A pre-clinical trial of adult acute stroke patients was carried out to test the rehabilitation effect by analyzing several sorts of data to clarify the potential advantage of rehabilitation therapy and determine the optimal rehabilitation approach for stroke patients.

RESULTS

Text Mining Based on the ABC Model

We retrieved 11,418 records from PubMed with stroke-related keywords, and there were 10,992 records with abstracts. From this dataset, 241,044 unique NPs were extracted, including 81 different scales and 215 different rehabilitation therapies.

In the potential scale list, the common phrases, such as “pre-test,” “post-test,” and “outcome assessment,” as well as other unrelated scales, such as “body mass index,” “depression score,” and “MMSE score,” were deleted after a manual check from the stroke-related scale dataset (Stroke_Scales), and 26 scales were finally obtained (Table 1). We ranked these assessment scales with frequency. The Fugl-Meyer assessment (FMA) had the highest frequency of 22.7%, followed by the Ashworth scale (AS), Barthel index (BI), action research arm test (ARAT), wolf motor function test (WMFT), Rankin scale (RS), functional independence measure (FIM), and primary outcome measure (POM) main outcome measure (MOM). Besides, the TF-IDF ranking was almost consistent with frequency. These scales are the most widely used assessment scales in stroke-related studies.

Accordingly, in the potential therapy list, the following common words were deleted: “clinical practice,” “conventional therapy,” “medical therapy,” “physical therapy,” “specific training,” and “combined therapy”; in addition, the following drug therapies were deleted in the manual inspection process: “antiplatelet therapy,” “anticoagulant therapy,” “antihypertensive therapy,” and “antithrombotic therapy.” Subsequently, we obtained the stroke-related rehabilitation therapy dataset, Stroke_Therapies, which comprised 47 rehabilitation therapies (Table 2). Two senior neurologists conducted the manual inspection independently, and we also compared our results with those in one recent review,

TABLE 1 | Stroke_Scales dataset items with frequencies and term frequency-inverse document frequency (TF-IDF).

Abbreviations	Full name	Frequencies	TF-IDF
FMA	Fugl-meyer assessment	248	8.967532517
Ashworth	Ashworth scale	138	14.86100317
BI	Barthel index	115	21.46757397
ARAT	Action research arm test	81	12.07425711
WMFT	Wolf motor function test	75	13.60741523
RS	Rankin scale	55	19.33877467
FIM	Functional independence measure	42	15.96626141
MOM	Main outcome measure	41	5.37136245
MI	Motricity index	40	7.990027148
SIAS	Stroke impact scale	25	11.18520612
BBS	Berg balance scale	23	24.26706778
MAS	Motor assessment scale	22	14.7664421
RMA	Rivermead motor assessment	20	17.80426457
AAI	Arm index	14	12.81320984
BBT	Box and block test	14	9.018969872
FAT	Frenchay arm test	13	18.36081387
NIHSS	Nihss score	12	15.94629694
POM	Primary outcome measure	11	5.529408988
JTTHF	Jebsen-Taylor Test of Hand Function	8	15.01043442
DAS	Disability assessment scale	8	14.62921899
BRUNSTROM	Brunnstrom scale	7	9.218846945
COPM	Canadian occupational performance measure	7	13.53841027
NHPT	Nine hole peg test	7	16.42409311
VAS	Visual analog scale	7	7.112188973
RS	Rankin scale	7	19.33877467
AMAT	Arm motor ability test	6	10.48555474

TABLE 2 | Stroke_Therapies dataset items with frequencies.

Stroke_Therapy items	Frequencies	Stroke_Therapy items	Frequencies
Transcranial magnetic stimulation	495	Treadmill training	15
Electrical stimulation	296	Peripheral nerve stimulation	14
Induced movement therapy	251	Bimanual training	13
Robot therapy	125	Gait training	13
Mental practice	97	Motor imagery training	13
Mirror therapy	97	Computer interface training	12
Intensive occupational therapy	95	Smart arm training	12
Motor training	74	Massed practice	10
Somatosensory stimulation	60	Transcutaneous electrical nerve stimulation	10
Repetitive practice	48	Virtual reality training	10
Intensive training	44	Physical and occupational therapy	9
Neuromuscular electrical stimulation	40	Active neuromuscular stimulation	8
Bilateral training	32	Deep brain stimulation	7
Bilateral arm training	29	Functional strength training	7
Noninvasive brain stimulation	27	Functional task practice	7
Cortical stimulation	24	Motor cortex stimulation	7
Tactile stimulation	24	Music therapy	7
Upper extremity training	24	Wrist training	7
Hand training	23	Based mental practice training	6
Neuromuscular stimulation	22	Constraint-induced movement therapy	6
Median nerve stimulation	20	Forced use therapy	6
Task practice therapy	19	Paired associative stimulation	6
Aerobic exercise training	17	Surface neuromuscular electrical stimulation	6
Unilateral training	17		

which focused on stroke rehabilitation assessment (Santisteban et al., 2016). It was found that our data of Stroke_Scales and Stroke_Therapies were consistent with those in the previous review.

Using the extracted Stroke_Scales, we searched again in PubMed and retrieved 60,307 records, in which 60,202 records had abstracts. From these records, we extracted the rehabilitation therapies (All_Therapies dataset), removed those applied for stroke and listed in Table 2 (Stroke_Therapies dataset), and obtained the potential repositioning rehabilitation therapies for

TABLE 3 | Potential repositioning rehabilitation therapies.

Unknown rehabilitation therapies	Frequencies	Unknown rehabilitation therapies	Frequencies
Cognitive behavior therapy	146	Acupuncture treatments	12
Massage therapy	30	Hand arm bimanual intensive training	7
Homeopathic treatment	16		

stroke (Table 3). The interactions between Stroke_Scales and All_Therapies datasets are shown in Figure 3.

As shown in the upper left quadrant in Figure 3, Stroke_Scales were densely distributed and were interacted with the All_Therapies dataset from the other three quadrants. Among All_Therapies, the potential repositioning rehabilitation therapies were marked in red and rose, while the existing Stroke_Therapies dataset items were indicated in other colors. Among the potential repositioning rehabilitation therapies, “hand arm bimanual intensive training” had the most interactions with Stroke_Scales, including Jebsen-Taylor Test of Hand Function (JTTHF), canadian occupational, performance measure (COPM), Assisting Hand Assessment (AHA), Pediatric Evaluation of Disability Inventory (PEDI), Box and Blocks Test (BBT), Six-Minute Walk Test (SMWT), and Goal Attainment Scale (GAS), none of which was the commonly used stroke-related assessment scale according to targeted populations and discipline difference between adult stroke and pediatric cerebral palsy.

We used these potential repositioning rehabilitation therapies with “stroke” in PubMed search to exclude the records that include the associations of those rehabilitation therapies with stroke. We found that except for “hand arm bimanual intensive training,” other NPs all co-occur with stroke. Thus, the “hand arm bimanual intensive training” was our target validation candidate.

Hand-Arm Bimanual Intensive Training (HABIT)

Hand arm bimanual intensive training (HABIT) is a bimanual intervention addressing the specific upper extremity impairments in pediatric congenital hemiplegia, which is the most common physical disability in childhood (Charles and Gordon, 2006; Gordon et al., 2007), typically with impairments of spasticity, sensation, and reduced strength. HABIT has been found to improve the pediatric patients’ bimanual hand coordination and the space control of actions. Furthermore, HABIT has been demonstrated to be the prioritized optimal approach to improve bimanual hand use and activity performance for children with hemiplegia (Green et al., 2013), and its principles include motor learning (practice specificity, types of practice, and feedback) and improved neuroplasticity (practice-induced brain changes arising from repetition, increased movement complexity, motivation, and reward), which are also the critical contents of stroke rehabilitation functional goals.

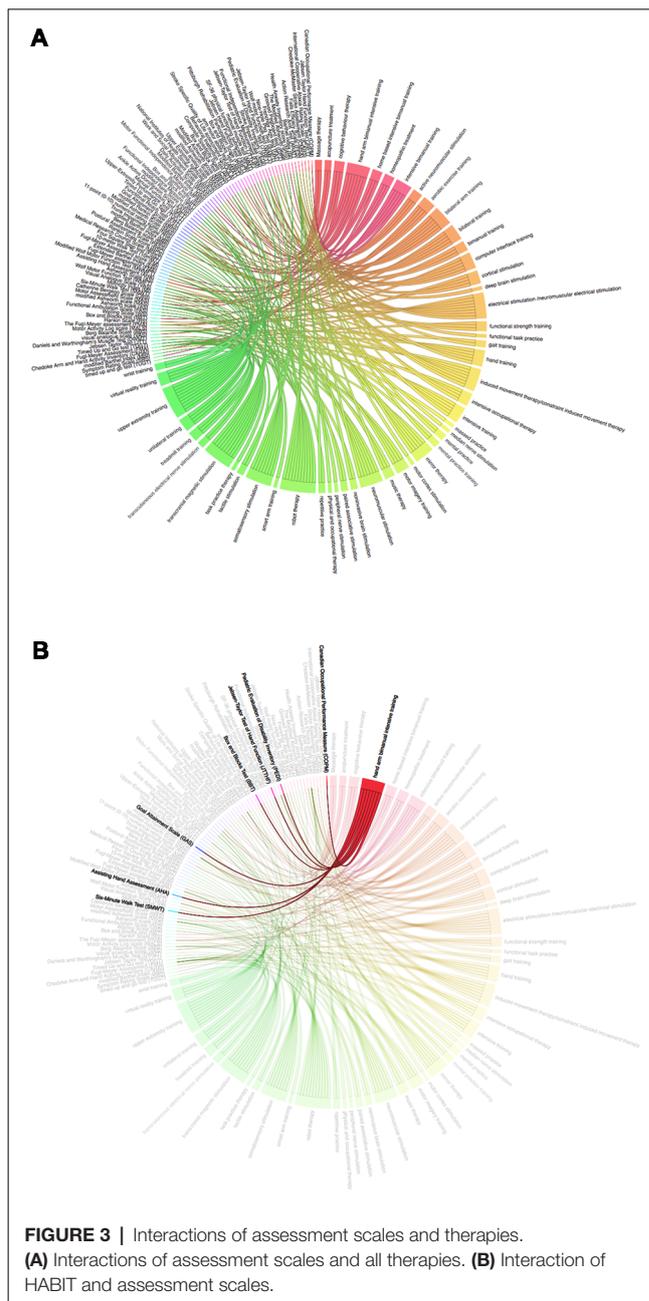


FIGURE 3 | Interactions of assessment scales and therapies. (A) Interactions of assessment scales and all therapies. (B) Interaction of HABIT and assessment scales.

We checked the frequency of different scales for stroke and found that compared with JTTHF and COPM, FMA was the more commonly used measure for the upper limb in the adult population, accounting for almost 30% of the total frequencies (Table 2). The findings are consistent with those of a systematic review of stroke rehabilitation (Santisteban et al., 2016). Based on the fact that FMA is usually applied in combination with ARAT in clinical studies, we decided to apply FMA combined with ARAT, not COPM or JTTHF, in our clinical validation study.

The validation of the potential candidate of repositioning rehabilitation therapies was examined in our little preclinical trial of stroke patients with upper limb impairment. The trial was approved by the local ethics committee of Shanghai Tenth

TABLE 4 | Clinical validation of HABILIT on patients with acute stroke.

No.	HABILIT group (n = 5)				No.	RRP group (n = 5)			
	FMA		ARAT			FMA		ARAT	
	Baseline	Post-training	Baseline	Post-training		Baseline	Post-training	Baseline	Post-training
#1	32	36	32	34	#1	30	33	31	33
#2	35	45	33	38	#2	32	33	31	34
#3	33	36	31	33	#3	32	34	29	30
#4	33	37	30	33	#4	37	39	33	34
#5	29	36	34	37	#5	33	34	32	34

RRR, the routine rehabilitation program; HABILIT, Hand-arm bimanual intensive training; FMA, Fugl-Meyer assessment; ARAT, action research arm test; $p(\text{FMA, baseline, HABILIT vs. RRP}) = 0.798815616$; $p(\text{FMA, post-training, HABILIT vs. RRP}) = 0.148803893$; $p(\text{ARAT, baseline, HABILIT vs. RRP}) = 0.433301901$; $p(\text{ARAT, post-training, HABILIT vs. RRP}) = 0.16682164$; $p(\text{HABILIT group, FMA, baseline vs. post-training}) = 0.030637606$; $p(\text{HABILIT group, ARAT, baseline vs. post-training}) = 0.049413148$.

People's Hospital. Patients adequately understood the purpose of the study, agreed to participate in this study, and signed the informed files. A total of 10 patients with acute stroke were recruited and randomly divided into two groups: the routine rehabilitation program (RRP) and the HABILIT groups. The same therapist who held national certification performed all treatments and training. The treatment duration was 2 weeks (10 working days), and the rehabilitation time was the same in both groups. In general, motor function and extremity activity were assessed using FMA and ARAT before and after therapy in these patients. Both groups showed statistically significant improvements of FMA and ARAT from baseline to post-treatment assessment. While more participates need to be enrolled to get a positive comparison between HABILIT group and RRP group (Table 4). More details were described in our follow-up full clinical trial study published in *Frontiers in Neurology* (Meng et al., 2018c).

DISCUSSION

Stroke is caused by the sudden rupture of blood vessels in the brain or blood vessels that cannot flow into the brain owing to vascular occlusion. Pediatric congenital hemiplegia mostly occurs because of poor nutrition during pregnancy, anemia during pregnancy, encephalopathy, birth asphyxia, pathological jaundice, and intrauterine distress, causing immature fetal brain and other organ tissues. Furthermore, pediatric congenital hemiplegia can further damage the central nervous system, leading to cerebral palsy. Although the pathogenesis and clinical manifestations of stroke and pediatric congenital hemiplegia are not the same, the HABILIT is effective for both conditions, indicating that the two diseases share similar rehabilitation mechanisms. Besides, patients who received HABILIT showed significant improvement, indicating that HABILIT has a positive impact on rehabilitation therapy for upper extremity impaired patients. The principle of HABILIT includes motor learning (task specificity, task type, and feedback), neuroplasticity, brain transformation upon increasingly difficult therapy and incentive reward (Nudo, 2003), and training cortical-spinal system reconstruction [manifesting as function recovery after injury (Eyre, 2003)] and is the theoretical foundation underlying the hypothesis that bimanual intensive therapy could improve the upper extremity function after acute stroke.

Although the repositioning of rehabilitation therapy is common in clinical practice, the discovery is often based on the clinician's personal experience or deduction from the medical community, without objective and systemic approaches with data mining application. For instance, VR, a rehabilitation intervention, was first applied to basic motor disability (Greenleaf and Tovar, 1994). Studies in the area of VR-based rehabilitation have gained growing recognition of the potential value of VR for other diseases with motor disorders, such as Parkinson's disease. VR is presently proposed as a new rehabilitation tool that potentially optimizes motor learning in a safe environment and replicates real situations to help improve functional activities in daily life such as gait, balance, and quality of life (Dockx et al., 2016). In our study, stroke (A) was associated with the assessment scales (B) in stroke literature, and in the rehabilitation therapy literature, assessment scales (B) represented the effect of rehabilitation therapy (C). It is highly likely that the retrieved rehabilitation therapies, which are unknown for stroke, have positive effects on stroke. In recent years, bioinformatics mining and omics studies have indirectly utilized the model (Smalheiser, 2017). To the best of our knowledge, this study is the first to use the ABC model to show the repositioning of rehabilitation therapy with positive validation. This model could be generalized as disease-assessment scale-rehabilitation therapy in future studies.

However, there is a major concern with the approach taken in the study. The HABILIT prototype had started earlier than text mining was adopted. In the beginning, the clinicians generated a hypothesis that HABILIT could be re-positioned in stroke rehabilitation and prepared to apply for the clinical ethics application. In order to avoid uncertainty in the trial, the clinicians sought cooperation from information scientists, hoping to find more supporting evidence from existing PubMed literatures using advanced text mining algorithms, and this is why the clinical research was not immediately carried out after the application of clinical ethics. Later, researchers in the two domains (clinical research and text mining) interacted with each other on numerous occasions, constantly trying to adjust the parameters, and finally provided effective evidence for HABILIT. Besides, we relied much on a manual check, more than half were deleted after the manual check.

We understand that knowledge discovery is a process of trial and error, and it is an iterative procedure that domain scientists

work closely with information scientists. The present study has shown the potential of applying text mining for new medical discoveries. Mining literatures to generate a hypothesis about rehabilitation therapy repositioning semi-automatically could be the *de facto* approach to inform clinicians who are trying to master the exponentially rapid expansion of publications and datasets. As demonstrated in this study, the ABC model can be applied to new hypothesis generation for rehabilitation therapy repositioning.

CONCLUSION

In the present study, we proposed a text mining approach to mine terms related to disease, rehabilitation therapy, and assessment scale from literature, and a subsequent ABC inference analysis was used to identify relationships of these terms across publications. The clinical validation study demonstrated that our approach could be used to identify potential repositioning rehabilitation therapy strategies for stroke. In particular, we identified a promising rehabilitation method, called HABIT, which was previously used in pediatric congenital hemiplegia. A subsequent full clinical trial confirmed the HABIT as a highly promising rehabilitation therapy for

stroke. Further clinical studies are needed to investigate the long-term effect of HABIT on stroke patients and identify optimal parameters of the therapy. In addition, further studies are needed to improve text mining and inference strategy and identify other clinical applications amenable to the technique.

AUTHOR CONTRIBUTIONS

GM, XL, and MS designed the study. YH and YD performed the experiments and prepared the figures. QY, DW, and YZ analyzed the data. XL and YZ performed the manual inspection as senior domain experts. GM wrote and discussed all sections of the manuscript. All authors reviewed and approved the manuscript.

FUNDING

This study was supported by the Excellent Young Physician Project of Shanghai Tenth People's Hospital (No. 2018SYPDRC038), Shanghai Youth Science and Technology Talents Sailing Project (No. 19YF1438400), and the Shanghai Science and Technology Commission Project (No. 18411961700, 17411950100, and 17411950101).

REFERENCES

- Ahmed, M. N., Toor, A. S., O'neil, K., and Friedland, D. (2017). Cognitive computing and the future of health care cognitive computing and the future of healthcare: the cognitive power of IBM watson has the potential to transform global personalized medicine. *IEEE Pulse* 8, 4–9. doi: 10.1109/MPUL.2017.2678098
- Brainin, M., and Zorowitz, R. D. (2013). Advances in stroke: recovery and rehabilitation. *Stroke* 44, 311–313. doi: 10.1161/STROKEAHA.111.000342
- Brunner, I., Skouen, J. S., Hofstad, H., Aßmus, J., Becker, F., Sanders, A. M., et al. (2017). Virtual reality training for upper extremity in subacute stroke (VIRTUES): a multicenter RCT. *Neurology* 89, 2413–2421. doi: 10.1212/wnl.0000000000004744
- Carrico, C., Chelette, K. C., Westgate, P. M., Powell, E., Nichols, L., Fleischer, A., et al. (2016). Nerve stimulation enhances task-oriented training in chronic, severe motor deficit after stroke: a randomized trial. *Stroke* 47, 1879–1884. doi: 10.1161/strokeaha.116.012671
- Charles, J., and Gordon, A. M. (2006). Development of hand-arm bimanual intensive training (HABIT) for improving bimanual coordination in children with hemiplegic cerebral palsy. *Dev. Med. Child Neurol.* 48, 931–936. doi: 10.1017/s0012162206002039
- Cramer, S. C., Wolf, S. L., Adams, H. P., Chen, D., Dromerick, A. W., Dunning, K., et al. (2017). Stroke recovery and rehabilitation research: issues, opportunities and the national institutes of health strokeNet. *Stroke* 48, 813–819. doi: 10.1161/STROKEAHA.116.015501
- Ding, Y., Liu, X. Z., Guo, C., and Cronin, B. (2013). The distribution of references across texts: some implications for citation analysis. *J. Informetr.* 7, 583–592. doi: 10.1016/j.joi.2013.03.003
- Dockx, K., Bekkers, E. M., Van Den Bergh, V., Ginis, P., Rochester, L., Hausdorff, J. M., et al. (2016). Virtual reality for rehabilitation in Parkinson's disease. *Cochrane Database Syst. Rev.* 12:CD010760. doi: 10.1002/14651858.CD010760.pub2
- Eyre, J. A. (2003). Development and plasticity of the corticospinal system in man. *Neural Plast.* 10, 93–106. doi: 10.1155/np.2003.93
- Favre, I., Zeffiro, T. A., Detante, O., Krainik, A., Hommel, M., and Jaillard, A. (2014). Upper limb recovery after stroke is associated with ipsilesional primary motor cortical activity: a meta-analysis. *Stroke* 45, 1077–1083. doi: 10.1161/strokeaha.113.003168
- French, L., Liu, P., Marais, O., Koreman, T., Tseng, L., Lai, A., et al. (2015). Text mining for neuroanatomy using WhiteText with an updated corpus and a new web application. *Front. Neuroinform.* 9:13. doi: 10.3389/fninf.2015.00013
- Gordon, A. M., Schneider, J. A., Chinnan, A., and Charles, J. R. (2007). Efficacy of a hand-arm bimanual intensive therapy (HABIT) in children with hemiplegic cerebral palsy: a randomized control trial. *Dev. Med. Child Neurol.* 49, 830–838. doi: 10.1111/j.1469-8749.2007.00830.x
- Green, D., Schertz, M., Gordon, A. M., Moore, A., Schejter Margalit, T., Farquharson, Y., et al. (2013). A multi-site study of functional outcomes following a themed approach to hand-arm bimanual intensive therapy for children with hemiplegia. *Dev. Med. Child Neurol.* 55, 527–533. doi: 10.1111/dmcn.12113
- Greenleaf, W. J., and Tovar, M. A. (1994). Augmenting reality in rehabilitation medicine. *Artif. Intell. Med.* 6, 289–299. doi: 10.1016/0933-3657(94)90034-5
- He, B., Tang, J., Ding, Y., Wang, H., Sun, Y., Shin, J. H., et al. (2011). Mining relational paths in integrated biomedical data. *PLoS One* 6:e27506. doi: 10.1371/journal.pone.0027506
- Kattenstroth, J. C., Kalisch, T., Sczesny-Kaiser, M., Greulich, W., Tegenthoff, M., and Dinse, H. R. (2018). Daily repetitive sensory stimulation of the paretic hand for the treatment of sensorimotor deficits in patients with subacute stroke: RESET, a randomized, sham-controlled trial. *BMC Neurol.* 18:2. doi: 10.1186/s12883-017-1006-z
- Kuehn, B. (2018). Stroke Rehab Lacking. *JAMA* 320:128. doi: 10.1001/jama.2018.9047
- Kwakkel, G., Veerbeek, J. M., Van Wegen, E. E., and Wolf, S. L. (2015). Constraint-induced movement therapy after stroke. *Lancet Neurol.* 14, 224–234. doi: 10.1016/S1474-4422(14)70160-7
- Meng, G., Meng, X., Ma, X., Zhang, G., Hu, X., Jin, A., et al. (2018a). Application of ferulic acid for Alzheimer's disease: combination of text mining and experimental validation. *Front. Neuroinform.* 12:31. doi: 10.3389/fninf.2018.00031
- Meng, G., Huang, Y., Yu, Q., Ding, Y., Wild, D., Zhao, Y., et al. (2018b). Adopting literature-based discovery on rehabilitation therapy repositioning for stroke. *BioRxiv [Preprint]*. Available online at: <https://www.biorxiv.org/content/early/2018/09/19/422154> [accessed September 19, 2018].
- Meng, G., Meng, X., Tan, Y., Yu, J., Jin, A., Zhao, Y., et al. (2018c). Short-term efficacy of hand-arm bimanual intensive training on upper arm function

- in acute stroke patients: a randomized controlled trial. *Front. Neurol.* 8:726. doi: 10.3389/fneur.2017.00726
- Nudo, R. J. (2003). Adaptive plasticity in motor cortex: implications for rehabilitation after brain injury. *J. Rehabil. Med.* 35, 7–10. doi: 10.1080/16501960310010070
- Papanikolaou, N., Pavlopoulos, G. A., Theodosiou, T., Vizirianakis, I.S., and Iliopoulos, I. (2016). DrugQuest—a text mining workflow for drug association discovery. *BMC Bioinformatics* 17:182. doi: 10.1186/s12859-016-1041-6
- Percha, B., and Altman, R. B. (2015). Learning the structure of biomedical relationships from unstructured text. *PLoS Comput. Biol.* 11:e1004216. doi: 10.1371/journal.pcbi.1004216
- Pollock, A., Farmer, S. E., Brady, M. C., Langhorne, P., Mead, G. E., Mehrholz, J., et al. (2014). Interventions for improving upper limb function after stroke. *Cochrane Database Syst. Rev.* 11:CD010820. doi: 10.1002/14651858.CD010820.pub2
- Silva, A., Vaughan-Graham, J., Silva, C., Sousa, A., Cunha, C., Ferreira, R., et al. (2018). Stroke rehabilitation and research: consideration of the role of the cortico-reticulospinal system. *Somatosens. Mot. Res.* 35, 148–152. doi: 10.1080/08990220.2018.1500363
- Santisteban, L., Teremetz, M., Bleton, J. P., Baron, J.C., Maier, M. A., and Lindberg, P.G. (2016). Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review. *PLoS One* 11:e0154792. doi: 10.1371/journal.pone.0154792
- Smalheiser, N. R. (2017). Rediscovering don swanson: the past, present and future of literature-based discovery. *J. Data Inf. Sci.* 2, 43–64. doi: 10.1515/jdis-2017-0019
- Sun, P., Guo, J., Winnenburger, R., and Baumbach, J. (2017). Drug repurposing by integrated literature mining and drug-gene-disease triangulation. *Drug Discov. Today* 22, 615–619. doi: 10.1016/j.drudis.2016.10.008
- Swanson, D. R., and Smalheiser, N. R. (1997). An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artifi. Intelli.* 91, 183–203. doi: 10.1016/s0004-3702(97)00008-8
- Westergaard, D., Staerfeldt, H. H., Tonsberg, C., Jensen, L. J., and Brunak, S. (2018). A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLoS Comput. Biol.* 14:e1005962. doi: 10.1371/journal.pcbi.1005962
- Winstein, C. J., Stein, J., Arena, R., Bates, B., Cherney, L. R., Cramer, S. C., et al. (2016). Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals From the American Heart Association/American Stroke Association. *Stroke* 47, e98–e169. doi: 10.1161/STR.0000000000000098

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Meng, Huang, Yu, Ding, Wild, Zhao, Liu and Song. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.