

# Factors Predicting Sexual Risk Behaviors of Adolescents in North-Eastern Thailand

Athiwat BUTDABUT<sup>1</sup> and Pissamai HOMCHAMPA

*Health System Science Department, Faculty of Medicine, Mahasarakham University, Maha Sarakham, Thailand*

**Abstract.** The number of young people who have had sex at an early age increases in proportion, it concerns unsafe sexual behaviors, teenage pregnancy, HIV aids and sexually transmitted infections (STIs). This study examines the health behaviors and factors predicting sexual risk behaviors pertaining to teenage pregnancy among adolescents in Thailand. Adolescents consulted the reproductive health center about problems with the same gender. The factors of adolescent reproductive behaviors were significantly associated with age, education level, and the perception of peer norms. Receiving social support from media information also significantly correlated with those behaviors. The results recommend that to prevent premature pregnancy, adolescents should protect themselves. Parents should take the issue of social media use by their teenagers very seriously.

**Keywords:** Teen pregnancy, teen sexual health, pregnancy prevention, reproductive health

## 1. Introduction

Adolescence is a unique age of transition from childhood to adulthood in many aspects [7]. The subjects change many varieties including physical, mental, and social shifts. These changes bring both risks and opportunities that influence the development into adulthood. Teenagers today are growing up in societies with developed economies, social systems and technologies, especially communication technology. Advancements such as social media with mobile phones, tablets, and other wireless devices have revolutionized communication styles and the way of accessing knowledge. According to WHO, approximately 12 million girls aged 15–19 years and at least 777,000 girls under 15 years give birth each year in developing regions [1]. At least 10 million unintended pregnancies occur each year among adolescent girls aged 15–19 years in the developing world [5]. Complications during pregnancy and childbirth are the leading cause of death for 15–19-year-old girls globally [6]. Adolescent mothers (ages 10–19 years) face higher risks of eclampsia, puerperal endometritis, and systemic infections, compared to women aged 20 to 24 years, and their babies face higher risks of low birth weight, preterm delivery and severe neonatal conditions [10]. This study investigates the behaviors that may play a role in the reproductive health of teenagers in Health Region 7 in Northeast of Thailand, aiming at determining which behaviors enhance reproductive health

---

<sup>1</sup> Corresponding Author: Athiwat Butdabut, Faculty of Medicine, Mahasarakham University, MahaSarakhm, Thailand. Email: athiwatbutdabut@gmail.com

promotion among teens with the goal of further developing, supporting, and encouraging them.

## 2. Methods

### 2.1. Study Design and Participants

The sample comprised 15–19 years old secondary school teenagers from 4 provinces under the responsibility of the 7th Regional Health Promotion Centre including Khon Kean, Roi-Et, Mahasarakham, and Kalasin. Selected by a randomization method, we included 389 teenagers in our study.

### 2.2. The Questionnaire Ethical Concern and Statistical Analysis

The questionnaire consisted of 137 questions on following topics: 1) Socio-economic and characteristic, 2) behaviors related to reproductive health, 3) attitudes towards self-control ability, 4) gender, 5) peer norms perception, 6) social support in media information, confirmed by Index of Item-Objective Congruency (IOC) scores  $\geq 0.5$  Cronbach's  $\alpha \geq 0.7$ . The participants were protected, being very aware of the individual rights and taking action to prevent inadvertent negative effects based on ethical research principles, accredited by the Human Research Ethics Committee of the Mahasarakham University (No. 78/2018). Data analysis used descriptive statistics (Pearson Chi-square), then testing correlation of each categorical data by  $\chi^2$  test. The level of statistical significance was  $p < 0.05$ . All data were analyzed by SPSS Statistics version 18.

## 3. Results

**Table 1.** Correlation between other Factors and Teen Reproductive Health Behavior (n = 389)

Factors	Adolescent Reproductive Behaviors			Chi-Square	p-value
	Poor	Medium	Good		
Gender				8.489 <sup>a</sup>	0.075
- Male	17(25.6)	28(35.4)	34(43.0)		
- Female	56(19.6)	139(48.8)	90(31.6)		
- LGBT	7(28.0)	14(56.0)	4(16.0)		
Age				27.217 <sup>a</sup>	<0.001 <sup>**</sup>
- Less than 16 years	12(13.3)	28(31.1)	50(55.6)		
- Over 16 years	68(22.7)	153(51.2)	78(26.1)		
Education Level				47.366 <sup>b</sup>	<0.001 <sup>**</sup>
- High school	22(12.0)	71(38.8)	90(49.2)		
- Voc. Cert.	55(28.1)	103(52.6)	38(19.3)		
- High Voc. Cert.	3(30.0)	7(70.0)	0(0.0)		
GPA				3.544 <sup>a</sup>	0.170
- Less than 3.00	12(32.4)	15(40.6)	10(27.0)		
- More than or equal to 3.00	68(19.3)	166(47.2)	118(33.5)		
Monthly Expense				1.737 <sup>a</sup>	0.420
- Less than 2,000 baht	24(24.7)	45(46.4)	28(28.9)		
- More than 2,000 baht	56(19.2)	136(46.6)	100(34.2)		
Peer Norms Perception				80.233 <sup>a</sup>	<0.001 <sup>**</sup>
- Very risky level	69(30.3)	115(50.4)	44(19.3)		
- Moderate risk level	7(16.3)	26(60.5)	10(23.2)		
- Low risk level	4(3.4)	40(33.9)	74(62.7)		
Receiving Social Support Information from Media				11.468 <sup>b</sup>	0.013 <sup>*</sup>
- Internet/TV programme	68(18.8)	170(47.1)	123(34.1)		
- Printed media	10(50.0)	8(40.0)	2(10.0)		
- Health personnel	2(25.0)	3(37.5)	3(37.5)		

<sup>a</sup> Pearson Chi-square, <sup>b</sup> Fisher's Exact Test, <sup>\*</sup> Significance level 0.05, <sup>\*\*</sup> Significance level p-value < 0.001

The gender distribution among the adolescent showed females (73.3%), males (20.3%), and LGBT (Lesbian, Gay, Bisexual, and Transgender) (6.4%) with an average age of  $16.55 \pm 1.18$ . Most of the students studied at the first year vocational certificate level (27.0%). The most family received an income less than 15,000 baht per month (36.4%). Most teen caregiver's relationships were parents (76.9). Most people lived in parent's house (88.9%), mostly in adolescent's family with 4-6 people (68.6%). Adolescents and friends consulted on reproductive health problems mostly hold the same gender (88.9). The factors significantly associated with adolescent reproductive behaviors were: 1) age group ( $p$ -value  $<0.001$ ), 2) educational level ( $p$ -value  $<0.001$ ), 3) perception of peer group norms ( $p$ -value  $<0.001$ ), and 4) social support in media information ( $p$ -value = 0.013).

## 4. Discussion

### 4.1. Characteristics of the Adolescent and their Reproductive Behaviors

The result revealed that age group and educational level were significantly associated with adolescent reproductive behaviors. An earlier study reported that the age group is of relevance. Females who reported indirect aggression toward peers had earlier ages at first sexual intercourse, while females who were more victimized in adolescence experienced later ages at first sexual intercourse [9]. The education also affected several things for females in Pakistan. Literacy, for instance, is lower for women than for men. Only 20% of all females have attended primary school. Although most women know at least 1 contraceptive method, it is the urban educated woman who is twice as likely to know a source of supply and 5 times more likely to be a user [2]. Thai culture values virginity in the female gender, and marriage is the method that makes early pregnancy more acceptable. The research results eventually prompted health officials to advise teenagers on pregnancy prevention. Most of the peer norms perception are at high and medium risk level regarding adolescent reproductive behaviors. Van de Bongardt et al. [8] performed a meta-analysis to investigate the associations between three types of peer norms (descriptive norms (peer sexual behaviors), injunctive norms (peer sexual attitudes)), the peer pressure to have sex, and two adolescent sexual behavior outcomes (sexual activity and sexual risk behavior). Adolescent sexual activity was stronger associated with descriptive norms than with injunctive norms or peer pressure. Compared with the sexual activity outcome, the effect size of descriptive norms (peer sexual risk behavior) for sexual risk behavior was smaller. They also approved that age, gender, peer type, and socio-cultural context significantly moderated these associations. Another study explains that perceived peer norms supporting safer sex were inversely associated with recently having two or more sexual partners after controlling for demographic characteristics. Perceived peer norms around safer sexual behavior contribute to a lower likelihood of engaging in two HIV/STI risk behaviors: inconsistent condom use and multiple partnering [4].

### 4.2. Social Media Support Information

The adolescent's behaviors are also correlated to social support received by gathering information from any kinds of media. Regarding the use of information especially in the

context of health education such as safe sex, we recommend that adults should select supporting media. A study [3] conducted a systematic review of the literature to examine the effectiveness of social media among young adults aged 15 through 24 years, which indicated that social media and text messaging can increase knowledge regarding the prevention of STDs. These interventions may also affect behavior, such as screening/testing for STDs, sexual risk behaviors, and STD acquisition, but the evidence for effect is weak.

## 5. Conclusion

In conclusion, age group and educational level were significantly associated with adolescent reproductive behaviors. It also verified that the adolescent is influenced by peer norms perceptions in both the very risk level and medium level of adolescent reproductive behaviors. These findings emerge empirical factors for the risk behavior especially of peer norms and hold important implications for reproductive health of teenagers.

## References

- [1] Darroch JE, Woog V, Bankole A, Ashford LS. Adding it up: costs and benefits of meeting the contraceptive needs of adolescents. New York: Guttmacher Institute, 2016, <https://www.guttmacher.org/report/adding-it-meeting-contraceptive-needs-....>
- [2] Greenspan A. Effects of education on reproductive behavior: lessons from Pakistan. *Asia Pac Pop Policy*. 1992 Dec;(23):1-4.
- [3] Jones K, Eathington P, Baldwin K, Sipsma H. The impact of health education transmitted via social media or text messaging on adolescent and young adult risky sexual behavior: a systematic review of the literature. *Sex Transm Dis*. 2014 Jul;41(7):413-9. doi: 10.1097/OLQ.000000000000146.
- [4] Kapadia F, Frye V, Bonner S, Emmanuel PJ, Samples CL, Latka MH. Perceived peer safer sex norms and sexual risk behaviors among substance-using Latino adolescents. *AIDS Education and Prevention* 01 Feb 2012, 24(1):27-40. DOI: [10.1521/aeap.2012.24.1.27](https://doi.org/10.1521/aeap.2012.24.1.27) PMID: 22339143
- [5] Kiani MA, Ghazanfarpour M, Saeidi M. Adolescent pregnancy: a health challenge. *International Journal of Pediatrics* 2019; 7(7): 9749-9752.
- [6] Neal S, Matthews Z, Frost M, Fogstad H, Camacho AV, Laski L. Childbearing in adolescents aged 12–15 years in low resource countries: a neglected issue. New estimates from demographic and household surveys in 42 countries. *Acta Obstet Gynecol Scand*. 2012 Sep;91(9):1114-8. doi: 10.1111/j.1600-0412.2012.01467.x.
- [7] Spear LP. Neurobehavioral changes in adolescence. *Current Directions in Psychological Science* 2000; 9(4): 111-114.
- [8] Van de Bongardt D, Reitz E, Sandfort T, Deković M. A meta-analysis of the relations between three types of peer norms and adolescent sexual behavior. *Pers Soc Psychol Rev*. 2015 Aug;19(3):203-34. doi: 10.1177/1088868314544223.
- [9] White DD, Gallup AC, Gallup Jr GG. Indirect peer aggression in adolescence and reproductive behavior. *Evolutionary Psychology* 2010; 8: 147470491000800106.
- [10] WHO U, Mathers C. Global strategy for women's, children's and adolescents' health (2016-2030). Organization 2016; 201: 4-103.

# Conceptual Model for Behaviour Change Progress – Instrument in Design Processes for Behaviour Change Systems

Helena LINDGREN<sup>1</sup> and Saskia WECK

*Department of Computing Science, Umeå University, Umeå, Sweden*

**Abstract.** The aim of the work presented in this article was to develop a conceptual model for behavior change progress, which could be used for automated assessment of reasons for progress or non-progress. The model was developed based on theories for behavior change, and evaluated by domain experts. The information models of two prototype systems of a digital coach under development for preventing cardiovascular diseases and stress respectively, were evaluated by comparing the content of the prototypes with concepts in the model. The conceptual model was found useful as instrument to evaluate to what extent the prototypes are based in theories for behavior change, whether some vital information is missing, and to identify mechanisms for short and long time goal setting. Moreover, the connection between the ontology underpinning the prototypes and the conceptual model could be defined. Future work includes the integration of the conceptual model to function as a meta-ontology, which could be used for capturing causal relationships between information collected by the applications at baseline and at runtime.

**Keywords.** behavior change, ontology, participatory design, knowledge engineering, personalisation

## 1. Introduction

Theories of behavior change (e.g. [1-5]), and design models for behavior change systems, partly based on these theories [6-7], are increasingly applied in research on persuasive technology for behavior change [8]. When applying a participatory design process to develop such systems, a central assumption is that the participating stakeholders are experts in the domain in which the persuasive technology is aimed to be used. The design process aims to translate this knowledge into digital tools, which an individual can use for health coaching. Moreover, the knowledge needs to be translated into formal, computational models, which the system can use for tailoring support and automatically and intelligently adapt its behavior to the individual (e.g. [9-10]).

The purpose of this study is to develop and evaluate a conceptual model of behavior change progress, which can be used in a participatory design process of behavior change systems, and for automated analysis of progress over time as a complement to everyday assessments at runtime.

---

<sup>1</sup> Corresponding Author, Helena Lindgren, Department of Computing Science, Umeå University, 901 87 Umeå, Sweden; E-mail: [helena.lindgren@umu.se](mailto:helena.lindgren@umu.se).

### 2. Methods

Theories on behavior change explain the determinants of human behavior, motivations for behavior, and how to guide behavior change. The conceptual model is based on the Theory of Planned Behavior [1], the Self-Determination Theory [2], Prochaska’s Transtheoretical Model of Change [3], the Goal Setting Theory [4], and the Fogg Behavior Model [6]. Concepts from the Health Belief Model [5] were also embedded.

The model was created in the following steps [11]. Relevant concepts were selected from behavior change theories, and defined. Next, based on the theoretical frameworks, the relationships between the concepts were defined and explained. In a third step, the initial model was validated by five experts through interviews and adjusted accordingly. Participating experts represented expertise in epidemiology, nutrition, sociology and nursing. The model was applied and further evaluated as part of two ongoing participatory design processes [10,12]. One aims at developing a digital coach application for preventing cardio-vascular diseases and promoting change of behavior to improve health [13], another to support stress rehabilitation. The preliminary information models of the two prototypes, which had emerged in focus group discussions and workshops, were evaluated by connecting the information to the conceptual model.

### 3. Results

The experts interviewees suggested changes of the conceptual model based on their knowledge of the concepts, theories and experience from clinical practice, gave elaborations on the concepts, and proposed additional concepts and relationships, partly based on the Health Belief Model (e.g., *Risk Perception, Previous Attempts*) [5,11].

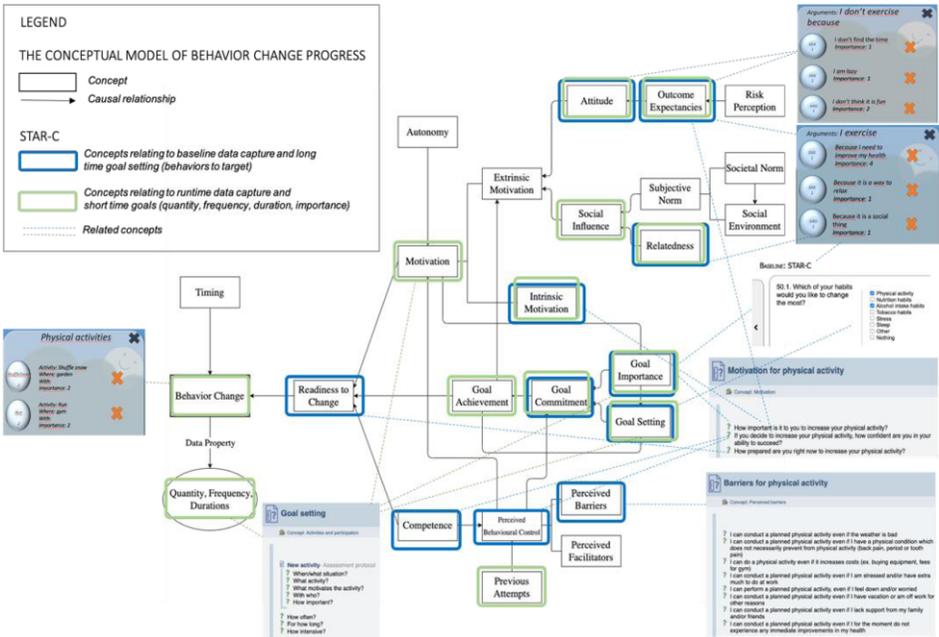


Figure 1. STAR-C information mapped to the conceptual model of behavior change progress.

The baseline questions for the cardio-vascular intervention cover most of the factors relating to a person’s own attitude and intention to change, while less information is collected on how the social environment may influence the factors for behavior change. The runtime version contains mechanisms that may capture also conflicting motives for activities and more on perceived barriers, which are expected to illuminate reasons for non-adherence to goals specified, and be used for explaining and motivating changes in behaviors.

When mapping the contents of the baseline and ecological momentary assessment (EMA) questions for the coaching system for managing stress (Jonglera) to the conceptual model, it is clear that stress management requires a different approach to design than the previous example, in particular, since mobilizing the physical and mental effort required to changing behavior is part of the behavior change progress, and targeted as part of the digital intervention (Figure 2). In this case, the EMA questions are posed to assess daily engagement and experiences, forming a basis for the system to translate this into motivating trajectories of positive trends, which can be communicated to the person as an extrinsic motivator. As such, it will target also the factors in the behavior change progress model such as *outcome expectancies, attitudes, barriers and facilitators*.

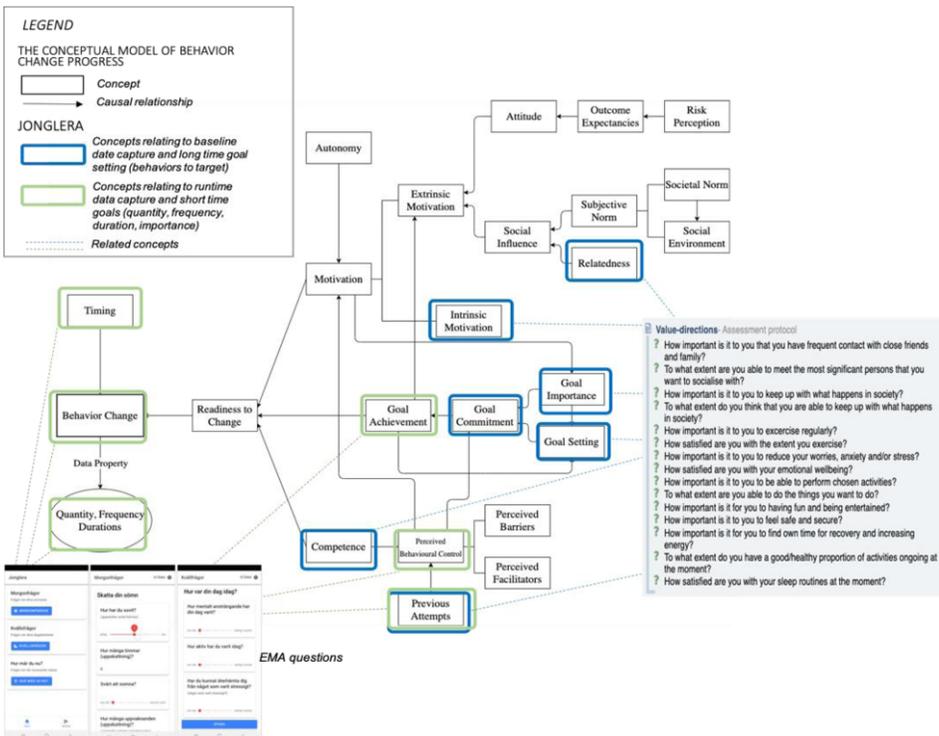


Figure 2. Jonglera information mapped to the conceptual model of behavior change progress.

To summarize, the conceptual model illuminates the complexity of behavior change for improving emotional wellbeing relating to stress and exhaustion syndrome, and the importance to design mechanisms, which can translate small efforts into positive outcome expectancies and attitudes, which can in turn, reinforce sense of competence and self-efficacy.

#### 4. Discussion and Conclusions

A conceptual model of behavior change progress was developed based on theories of behavior change, evaluated by experts on behavior change and applied to evaluate the information models of two prototype systems for supporting behavior change. It was concluded that the model is useful for assessing to what extent an information model is capturing vital factors for behavior change, and how personalized intervention needs to be designed in order to promote behavior change.

Ontologies have been developed for different purposes, e.g., behavior change interventions, or identifying barriers for change [14-15]. The proposed conceptual model for behavior change progress aims to capture the vital concepts of the relevant theories, and their relationships in the development of persuasive systems. As such, the model complements existing models.

Further evaluation studies need to be done, in particular for exploring how the model can be used as instrument by the participating domain experts and stakeholders in a design process. Future work includes also the integration of the conceptual model to function as a meta-ontology in the system architectures of the two prototype systems, which could be used for capturing causal relationships between information collected by an application at baseline and at runtime.

#### References

- [1] Ajzen I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 1991;50:179-211.
- [2] Ryan RM, Deci LM. Self-Determination Theory and the facilitation of intrinsic motivation, social development and well-being. *American Psychologist* 2000;55:68-78.
- [3] Prochaska JO, Redding CA, Evers KE, The transtheoretical model and stages of change. In: Glanz A, Rimer BK, Viswanath K (Edrs.) *Health behavior and health education: Theory, research, and practice*. pp. 97–121. Jossey-Bass; 2008.
- [4] Locke EA, G.P. Latham GP. Goal setting theory. In: *Motivation: Theory and Research* 2012:23-40.
- [5] Janz NK, Becker MH. The Health Belief Model: A Decade Later. *Health Education Quarterly* 1984;11,1: 1-47.
- [6] Fogg BJ. A behaviour model of persuasive design, In *Proc. of the 4<sup>th</sup> International Conference on Persuasive Technology (Persuasive'09)*. Association for Computing Machinery, New York, NY, USA, 2009:1-7.
- [7] Oinas-Kukkonen H, Harjuma M. Persuasive Systems Design: Key Issues, Process Model, and System Features. *Communications for the Association for Information Systems* 2009;24,28: 485-500.
- [8] Orji R, Moffatt K. Persuasive technology for health and wellness: State-of-the-art and emerging trends. *Health Informatics Journal* March 2018;66-91.
- [9] Paramythis A, Weibelzahl S, Masthoff J. Layered evaluation of interactive adaptive systems: framework and formative methods. *User Modeling and User-Adapted Interaction* 2010;20:383-453.
- [10] Lindgren H, Kampik T, Guerrero E, Blusi M, Nieves JC. Argumentation-based Health Information Systems : A Design Methodology. *IEEE Intelligent Systems* 2021;36,2:72–80.
- [11] Weck S. A Conceptual Model of Behavior Change Progress for the Application within Coaching Systems to Support Sustainable Lifestyle Changes. [Dissertation]. Umeå University; 2021.
- [12] Janols R, Lindgren H. A method for co-designing theory-based behaviour change systems for health promotion. *Studies in Health Technology and Informatics* 2017;235:368–72.
- [13] Lindgren H, Guerrero E, Jingar M, Lindvall K, Ng N, Richter Sundberg L, et al. The STAR-C Intelligent Coach : a Cross- Disciplinary Design Process of a Behaviour Change Intervention in Primary Care. *Studies in Health Technology and Informatics* 2020;273:203-208.
- [14] Wright AJ, Norris E, Finnerty AN et al. Ontologies relevant to behaviour change interventions: a method for their development. *Wellcome Open Res* 2020;5:126
- [15] Alfaiifi Y. An Ontology to Support Behaviour Change Interventions: Barriers to Activity. Doctoral dissertation, University of Liverpool; 2020.

# Automatic Extraction and Decryption of Abbreviations from Domain-Specific Texts

Michil EGOROV <sup>a,1</sup> and Anastasia FUNKNER <sup>a,2</sup>

<sup>a</sup>*ITMO University, Saint Petersburg, Russia*

**Abstract.** This paper explores the problems of extraction and decryption of abbreviations from domain-specific texts in Russian. The main focus are unstructured electronic medical records which pose specific preprocessing problems. The major challenge is that there is no uniform way to write medical histories. The aim of the paper is to generalize the way of decrypting abbreviations from any variant of text. A dataset of nearly three million medical records was collected. A classifier model was trained in order to extract and decrypt abbreviations. After testing the proposed method with 224,307 records, the model showed an F1 score of 93.7% on a valid dataset.

**Keywords.** Clinical text, medical records, natural language processing, abbreviations

## 1. Introduction

Electronic health records (EHR) are widely used to build models for predicting the process of healthcare provision [1]. Such texts contain terms, specific abbreviations, and acronyms, whose decryption depends on the field of usage, particular medical institution, or even particular specialist [2]. These factors make research of the task of extraction complicated. This paper presents a method of automatic detection and decryption of abbreviations.

Recognition of well-established abbreviations and acronyms is usually carried out with the help of dictionaries [3] and marked data [4], and mainly addresses data in the English language. MeDAL [3] contains medical texts with abbreviations and their possible decryptions. Models pre-trained on this dataset improve their metrics by 0.2-2%.

For the Russian language, as one of the low-resource languages, this area is not well researched. The author of [5] considers the problem of extraction and decryption of abbreviations from the Corpus of Legislative Acts of the Russian Federation. The paper considers approaches to topic modeling of texts to identify words that are similar in terms of use and contexts. As the author highlighted in their work, automatic decryption of abbreviations using such approaches is not accurate enough.

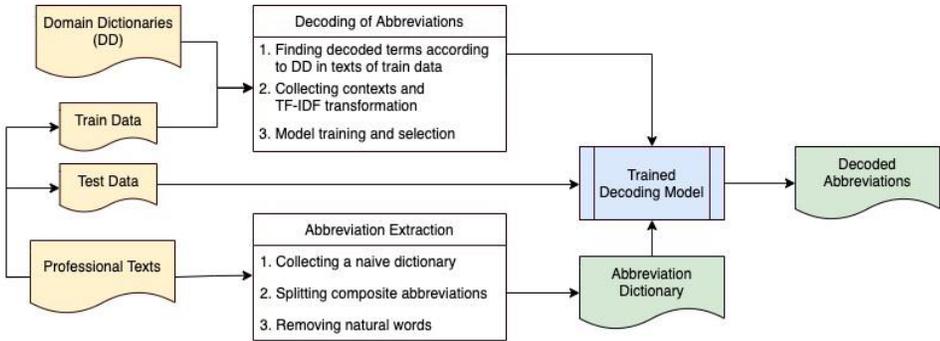
---

<sup>1</sup> Corresponding author. e-mail: egorovmichil9@gmail.com.

<sup>2</sup> Corresponding author. e-mail: funkner.anastasia@gmail.com.

## 2. Methods

This section sequentially describes the steps taken to extract and decrypt abbreviations.



**Figure 1.** Diagram of methods for extraction and decryption of abbreviations.

Firstly, a dictionary of abbreviations needs to be collected (Fig. 1). Keep in mind the assumption that a word is an abbreviation, if at least one half of its letters are upper-case. This allows to compile a naive abbreviation dictionary. However, this dictionary can include composite abbreviations, for example "BP.HR". Such abbreviations may be distinguished by punctuation. Moreover, the naive dictionary also contains natural words that were written in upper case. As an abbreviation often contains unnatural character pairs, the average entropy of character pairs in a word allows to remove natural words from the dictionary using the Shannon entropy formula. This means the more entropy, the more natural the word is. It remains to choose the threshold value of entropy.

The algorithm has two stages: training and applying. For the training stage, one should make a corpus with encoded abbreviations. Words in the corpus need to be normalized and filtered using a list of stop words. In the stage of applying, the trained model is used with an abbreviation and its context as input and a decoded term as output.

In order to train the decryption model, the input text is divided into three parts: context before and after an abbreviation, and its characters. Contexts are collected with a window and transformed into two Term Frequency-Inverse Document Frequency (TF-IDF) vectors. The stack of bag-of-characters vectors and two transformed TF-IDF contexts become features. Decryption models can only be trained on two TF-IDF vectorizer, since using bag-of-characters vectors often leads to model over-fitting. Any multiclassifier model can be used as a predicting model.

## 3. Result

For this research, we used medical recommendations from EHRs provided by the Almazov National Medical Research Centre, Saint Petersburg, Russia. The total volume of the dataset is approximately 3 million unstructured records with 440,000 unique words.

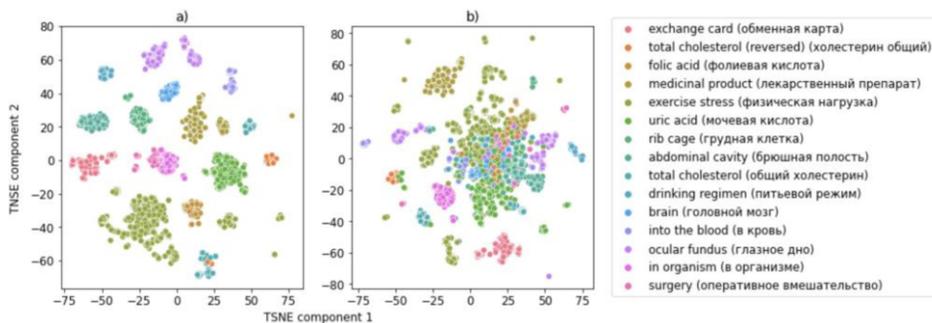
After the naive extraction, 36,856 unique abbreviated words were obtained and reduced to 26,989 after splitting by punctuation marks. To collect the statistics for entropy, the Leo Tolstoy novel "War and Peace" with 54,294 unique words was

processed. A dataset of 48,530 natural words and 23,346 abbreviations was collected to choose the threshold. The natural word classifier accuracy was 89% with 0.2 threshold.

The resulting dictionary contains 4,658 unique abbreviations after discarding words whose entropy was above 0.2. Unfortunately, some abbreviations like "АД" (short for "blood pressure", but the word "ад" is translated as "hell") have very large entropy and cannot be divided from natural words, because it is already a separate word.

The recommendation records were normalized using pymorphy2 [6] and stop words were filtered using the NLTK Python module. Unencrypted medical terms were identified in the domain dictionary in the recommendation corpus. The result was 224,307 texts with 19,980 unique terms. The validation set contained 30% of data. The selected domain dictionary was the "Encyclopedic dictionary of medical terms". TF-IDF context vectors (context\_window=5, max\_features=200) were counted. The projections of the context feature vectors with t-distributed Stochastic Neighbor Embedding (t-SNE) [7] are shown in Fig. 2. As can be seen, TF-IDF vectorizers grouped by context together even if the abbreviation columns in the features were skipped.

After applying difference models to the validation dataset, the Support Vector Classifier emerges as the best model. Its accuracy is 94.5%, ROC-AUC score is 99.7%, and F1 score is 93.7% (Tab. 1).



**Figure 2.** Context feature maps for top 15 terms a) with abbreviation columns, b) without abbreviation columns.

**Table 1.** Results of abbreviations decryption.

Model	Accuracy	ROC-AUC score	F1 score
RandomForestClassifier	0.938	0.968	0.933
LogisticRegression	0.939	0.985	0.930
XGBClassifier	0.939	0.973	0.935
SGDClassifier	0.940	0.983	0.930
CatBoostClassifier	0.940	0.986	0.932
SVC	0.945	0.997	0.937

## Discussion & Conclusion

This work presents an algorithm for automatic extraction and decryption of abbreviations from medical records. To achieve this, we applied a naive hypothesis and then improved the algorithm via entropy of words. The abbreviation filtering algorithm is implemented. After testing the decryption method with randomly chosen records, the classifier shows high accuracy (94.5%).

However, this work does not consider the abbreviations that are written in lower case, for example "mmHg" (millimeter of mercury) and frequent abbreviations (they are virtually always written in their short form, so we were unable to find their context using a domain dictionary). We intend to develop more abbreviation rules and expand the naive dictionary.

In our future research, we intend to apply the algorithm to different corpora, especially where abbreviations are ambiguous. This module is aimed to help data scientists improve their models that use free-form records for predicting processes associated with healthcare.

## **Acknowledgments**

This research is financially supported by The Russian Science Foundation, Agreement #19-11-00326

## **References**

- [1] Gamal A, Barakat S, Rezk A. Standardized electronic health record data modeling and persistence: A comparative review. *Journal of Biomedical Informatics* 2021; 114.
- [2] Funkner AA, Egorov MP, Fokin SA et al. Citywide quality of health information system through text mining of electronic health records. *Applied Network Science* 2021; 6.
- [3] Wen Z, Lu XH, Reddy S. MeDAL: Medical Abbreviation Disambiguation Dataset for Natural Language Understanding Pretraining. In *Proceedings of the 3rd Clinical Natural Language Processing Workshop*. Association for Computational Linguistics; 2020.
- [4] Zweigenbaum P, Deleger L, Lavergne T, Neveol A, Bodnari A. A Supervised abbreviation resolution system for medical text. *CEUR Workshop Proceedings* 2013; 1179.
- [5] Shilov IM. Automatic identification and decryption of abbreviations in the text [dissertation on the Internet]. Saint-Petersburg: Saint-Petersburg State University; 2016. Available from: <https://nauchkor.ru/pubs/avtomaticheskoe-vyyavlenie-i-rasshifrovkaabbreviatur-i-sokrascheniy-v-tekste-587d36365f1be77c40d58984>
- [6] Korobov M. Morphological analyzer and generator for Russian and Ukrainian languages. In: Khachay MY, Konstantinova N, Panchenko A, Ignatov DI, Labunets VG (Edrs.) *Analysis of images, social networks and texts*. Berlin: Springer; 2015.
- [7] Laurens M. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* 2008; 15:3221-3245.

# Mapping of OpenEHR Archetypes to FHIR Resources in Use Case Oncology

Abdul Mateen RAJPUT <sup>a,1</sup> and Ina BRAKOLLARI <sup>a</sup>

<sup>a</sup> *Universitätsklinikum Köln, Kerpener Str. 62, 50937 Köln, Germany*

**Abstract.** Unambiguous data exchange among healthcare systems is essential for error-free reporting and improved patient care. Mapping of different standards plays a crucial role in making different systems communicate with each other and have an efficient healthcare systems. This work focuses on exploring the possibilities of semantic interoperability between two widely used clinical modelling standards, OpenEHR and FHIR (Fast Healthcare Interoperability Resources). A manually curated map is being developed where the same semantically meaning OpenEHR Archetypes are mapped to the relevant FHIR Resources.

**Keywords.** OpenEHR, FHIR, OpenEHR Templates, Semantic Interoperability, HiGHmed

## 1. Introduction

OpenEHR is a technology for e-health, consisting of open specifications, clinical models, and software that can be used to create standards, and build information and interoperability solutions for healthcare [1]. It is an open-source healthcare information modeling standard that enables modelling of interoperable Electronic Health Records (EHRs). It started in response to the absence of an open-source platform for the exchange of clinical data. The standard is maintained by a community of healthcare professionals and software developers. Since its inception, it has been used widely across the globe [2–4].

On the other hand, FHIR (Fast Healthcare Interoperability Resources) is also an emerging open standard in healthcare. It is considered as “HTML” of healthcare [5], it facilitates extensive data modeling and data exchange, irrespective of the necessity of a common Electronic Health Record (EHR) system. Transitioning the patient data to FHIR would make the data able to connect various applications and increases interoperability significantly.

HiGHmed Consortium [6] uses OpenEHR as a standard for clinical data exchange. However, there are some use cases where usage of FHIR is foreseen and required. For that purpose mapping between both those emerging standards was needed.

---

<sup>1</sup> Abdul-Mateen RAJPUT, Universitätsklinikum Köln, Kerpener Str. 62, 50937 Köln, Germany ; E-mail: Abdul.mateen@uni-koeln.de.

## 2. Methods

The archetypes associated with the use case Oncology in OpenEHR Clinical Knowledge Manager (CKM) [7] have been identified and downloaded. The components of the archetypes were analyzed and extracted into an excel file. Each data element of OpenEHR archetypes was searched on the HL7 FHIR website [8] manually for the corresponding resource term. The FHIR resource elements, which had the same semantic meaning as OpenEHR archetypes elements, were mapped into the excel file accordingly.

## 3. Results

A mapping table was created that contains the archetype elements mapped to resource elements with the web address but also the archetype elements that can not be mapped when the subsequent resource elements were not found. Figure 1 shows the table of mapped terms.

Archetype	Archetype Element	FHIR Resource Element	Webpage
openEHR-EHR-EVALUATION.problem_diagnosis.v0	Date/time of onset	Condition.recordedDate	<a href="https://www.hl7.org/fhir...Condition.code">https://www.hl7.org/fhir...Condition.code</a>
openEHR-EHR-EVALUATION.problem_diagnosis.v0	Severity	Condition.severity	<a href="https://www.hl7.org/fhir...Condition.severity">https://www.hl7.org/fhir...Condition.severity</a>
openEHR-EHR-EVALUATION.problem_diagnosis.v0	Status	Condition.clinicalStatus	<a href="https://www.hl7.org/fhir...Condition.clinicalStatus">https://www.hl7.org/fhir...Condition.clinicalStatus</a>
openEHR-EHR-EVALUATION.problem_diagnosis.v0	Comment	Condition.note	<a href="https://www.hl7.org/fhir...Condition.note">https://www.hl7.org/fhir...Condition.note</a>

**Figure 1.** The table shows the OpenEHR archetype in first column, its captured elements in second column, corresponding FHIR resource elements and web address in third and fourth columns respectively.

Our findings show that archetypes were more specific as they were designed by clinical domain experts and they consist of a formal model. Their elements could only be mapped to only one FHIR resource element. On the contrary, FHIR resource elements could be mapped to more than one archetype element.

In addition, different elements of a single archetype can be mapped to multiple resources because a single FHIR resource does not cover all the information that a single archetype contains. It was also noticed that there are FHIR resources that overlap with each other and the boundaries are not strict in FHIR.

## 4. Discussion and Conclusion

A mapping between different standards is crucial for the efficient data exchange among different standards. It is a complex and time-consuming process. The emerging new health data standards, FHIR and OpenEHR, require to communicate as each of them has limitations. The manual mapping between both standards not only make it easier to understand the data representation in a different context but also streamline the data exchange. The mutual understanding of the limitations of each standard and the development of tools designed to facilitate communication among different standards is highly desirable.

## Acknowledgment

The project is funded by the German Federal Ministry of Education and Research (BMBF, grant id: 01ZZ1802U).

## References

- [1] [https://www.openehr.org/about/what\\_is\\_openehr](https://www.openehr.org/about/what_is_openehr) (Accessed 15 January 2021).
- [2] Maia TA, De Muylder CF, Reis ZSN. Archetype Development Process: A Case Study of Support Interoperability among Electronic Health Record in the State of Minas Gerais, Brazil. *J Med Syst* 2019; 43: 57
- [3] Min L, Tian Q, Lu X, et al. Modeling EHR with the openEHR approach: an exploratory study in China. *BMC Medical Informatics and Decision Making* 2018; 18: 75.
- [4] Mascia C, Frexia F, Uva P, et al. The openEHR Genomics Project. *Stud Health Technol Inform* 2020; 270: 443–447.
- [5] Munro D. Setting Healthcare Interop On Fire. *Forbes*, <https://www.forbes.com/sites/danmunro/2014/03/30/setting-healthcare-interop-on-fire/> (accessed 4 March 2020).
- [6] Home | HiGHmed, <https://www.highmed.org/> (accessed 14 January 2021).
- [7] Clinical Knowledge Manager, <https://ckm.openehr.org/ckm/> (accessed 29 July 2020).
- [8] Index - FHIR v4.0.1, <http://hl7.org/fhir/> (accessed 28 January 2021).