PharmaSUG China 2022 - Paper RW-146

Real World Data Standardization for an Ophthalmology Real World Study in China and Exploration of New eSource Solutions

Junkai Lai, Hangzhou LionMed Medical Information Technology Co., Ltd; Peking University Clinical Research Institute
Chen Yao, Peking University Clinical Research Institute, Peking University First Hospital, Hainan Institute of Real World Data;
Bin Wang, Department of Infectious Diseases, the Second Affiliated Hospital Zhejiang University School of Medicine;
Kai Wang, Peking University First Hospital;
Xiwen Liao, Peking University First Hospital

ABSTRACT

Boao Lecheng International Medical Tourism pilot area is an area that allows the usage of foreign innovative medical products in China without obtaining domestic market approval, the potential to use real world data (RWD) collected within this region as evidence for regulatory approval has attracted much attention. Collecting and properly analyzing the RWD generated by patients after using medical products can generate evidence sources for obtaining domestic approval for listing. However, RWD collected needs additional data standardization before it can be used to fill in electronic case report form (eCRF) and methods used to standardize source data are not clear. Therefore, it is necessary to further explore current methods used to standardize data and the possible problems and their causes in an actual case study. Through the CATALYS femtosecond case study, this paper will introduce and evaluate a standardization method for transforming source data into eCRF data. The study included 36 participants with the same inclusion and exclusion criteria as the CATALYS femtosecond project. The study will explain and examine the standardization methods used and measure the accuracy of the data standardization methods. The results show that the accuracy of the data standardization method used in this study is 98.6% but eCRF data completeness remains at 23.9%. Through the case study, we summarized three key problems in the process of data transformation from real-world source data to eCRF data, namely, the lack of research relevant source data and the complexity associated with the standardization of unstructured source data. Finally, we also explored some feasible solutions developed in China for the above key problems.

INTRODUCTION

Boao Lecheng International Medical Tourism pilot area is an area that enables domestic citizens to first use foreign innovative products in China without first obtaining domestic market approval, therefore has attracted much attention [1-2]. Collecting and properly analyzing the real world data (RWD) generated from patient visit data in Boao after using innovative medical products can help generate real world evidence that can be used for further domestic market approval [3]. However, due to the lack of data standardization that meets the requirements of electronic case report form (eCRF), the above real-world data cannot be directly used in clinical research eCRF at present. Existing technical barriers include: the prevalence of unstructured text data, the usage of non-standard terminology, and low transparency in the process of data standardization [4-7]. Therefore, it is necessary to further explore the explicit problems and causes of these problems when standardizing real-world data for clinical research through a real world study.

Catalyst femtosecond laser medical device is the second medical device approved using RWD obtained in Boao. This study was originally a prospective, single group observational real-world study, and its efficacy is mainly evaluated by comparing with a previous single arm objective performance criteria (OPC) [8]. The source data for the Catalyst study was initially recorded in the hospital information system, specifically the electronic medical record (EMR) system, and then transcribed to the research eCRF by manual entry. The regulatory department of Lecheng hopes to improve the current workflow by developing standardized methods that directly utilize hospital source data to fill in eCRFs and improve

their capability to conduct source data verification. This paper will demonstrate and evaluate the standardization method currently used to transform hospital source data to eCRF data. In this case, one of the challenging aspects of data standardization is the prevalence of unstructured text data that requires the usage of natural language processing (NLP). The study will aim to present a framework to document such methods of handling data.

STUDY DESIGN

This study will include 36 participants with the same inclusion and exclusion criteria as the catalyst femtosecond study. The data comes directly from the hospital source data at Boao Super Hospital. The data standardization process will be documented using a newly developed framework to organize the methods used and also record the China data standards associated with the source data. The study will first develop the standardization method and then test the robustness of the extraction method. In the development phase, 50% of the participant data will be random sampled for development and the remaining 50% of the participant data will be used to validate the methods by data integrity and accuracy. 3 clinical research coordinators will evaluate the integrity and accuracy of the final eCRF data after data standardization. The integrity of the final research data set, filled eCRF, will evaluate the completeness of the filled eCRF and eCRF fields that are unable to be filled due to missing source data will be flagged as a s0f0 problem. The accuracy of the filled eCRF will exclude data fields experiencing a s0f0 problem and inaccurate data will be categorized into three situations: unable to extract the correct source data (s1f0: did not extract existing source data), incorrectly extracting the source data (s0f1: extracting the wrong source data), extracting the correct source data but producing the wrong answer (s1f1: conversion error). A summary of the above s0f0, s0f1, s1f0 and s1f1 problems will be calculated using a percentage which will help diagnose precise problems and the size of these problems experienced in the standardization process. Table 1 lists the eCRF data modules that will be filled in using hospital source data and the corresponding visit time points.

In 2019, the Center for Drug Evaluation of the China National Medical Products Administration (NMPA CDE) put forward core suggestions on the data management plan of real-world studies in the "guidelines for real world data used to generate real world evidence" guidance [3]. The guide recommends that data extracted from real-world data sources should provide a description of the source data, sufficient traceability, document the conversion process, and relevant data standards used in the conversion process. The guide further points out that reliable NLP algorithm can be used to improve the conversion efficiency on the premise of ensuring the accuracy of data conversion and traceability of the source data for the eCRF. Therefore, the study will give an example of the documentation framework for conversion methods.

eCRF Modules	**Pre-Operative**	**Operative**
Visit Date	x	X
Informed Consent Form		
Medical History	x	
Demographics（人口学）	x	
UCDVA（裸眼远视力）	x	
BCDVA（最佳矫正远视力）	x	
Manifest Refraction	x	
Slit Lamp Exam	x	
Medical Findings		
Dilated Fundus Exam	x	

eCRF Modules	Pre-Operative	Operative
Intraocular Pressure Screen	x	
Biometry	x	
Corneal Topography	x	
Cataract Status		
Ocular Symptoms	x	
Surgery Operation		x
Day of Surgery Medication Regimen		x
Other Surgical Procedures		x
Surgical Complications		x

Table 2. eCRF Data Modules

SOURCE DATA METADATA

Based on China's guidance, each eCRF field should specify the metadata of the source data. The description of metadata should include the source system, the name of the original data element, and the corresponding China EMR standard used. The EMR standard used to encode the source data can leverage the most widely used standards at the moment including the "2011 data element dictionary code" and "the 2014 EMR basic data set internal code" [9-11]. For example, the following source data descriptions will appear in the eCRF data field "sex": HIS (source data system), Sex (source data element name), hdsd00.11.110 (standard data element code), Sex (standard data element name) and WS 445-2014 (standard name) (see Table 2).

eCRF Field	Source System	Source DE Name	Standard DE Code	Standard DE Name	Standard Name
Is there Medical History?	HIS	Discharge Diagnosis	HDSD00.11.024	Primary Discharge Diagnosis-Name	WS 445-2014
Race	HIS	Ethnicity	HDSD00.11.077	Ethnicity	WS 445-2014
Sex	HIS	Sex	HDSD00.11.110	Sex	WS 445-2014

Table 2. eCRF 字段源数据表述

DATA CONVERSION METADATA

The data conversion from source data to CRF fields includes the conversion of both structured and unstructured text data. A description of the types of methods used and the potential routes that source data may take to get to research data can is described in Table 3.

The detailed implementation of the conversion method for the CRF fields can be seen in Appendix A. L1 conversions leverage existing structured data and map the data to a specific value domain including a data format requirement or research terminology. For instance, the “Operation Time” CRF field, takes the conversion route S>L1>T, which only utilizes a single L1 conversion. Correspondingly, the L1 column of Appendix A shows the value domain of “18 Digit Standard Date Time”. L1 conversions commonly require specific terminology standards of controlled vocabulary such as using ICD-10 codes for a diagnosis related CRF field.

L2 conversions start with unstructured text data for a given narrative section of text. For eye exam related CRF fields, the source data may look like “Eye Exam: UCDVA: OD: 0.6, OS: 0.08; BCDVA: OD: 0.6, OS: 0.4;” where individual exams are split by a semicolon implying an end to a sentence. Related eye exam CRF fields including “UCDVA Left Eye Result” and “Has UCDVA been Performed” have different conversion routes with the prior taking “S>L2>T” and the latter taking “S>L2>L3>T”. L2 conversion methods have three core aspects including the entity extraction, entity relation compliance, and entity output. For “UCDVA Left Eye Result”, the first task is to extract needed entities, including [UCDVA], [Left Eye], [Result] entity types, that ensures the relevance of the text data to the corresponding CRF field. Without [Left Eye] for instance, will lead to the extraction of both eyes result. The methods of extracting the entities may vary, including the usage of a term lists (OS, Left Eye) for the [Left Eye] entity and the usage of regular expression (/^[-+]?[0-9]+[.]?[0-9]*([eE][-+]?[0-9]+)?$/) to extract float numbers for the [Result] entity. The second task is to decipher which entities are linked together and filter irrelevant information through entity relationship rules. In the above example, the [Result] entity type has extracted all the different numbers in the source data including 0.6 for OD, 0.08 for OS, and similar numbers for the BCDVA exam. Therefore, the entities must meet the entity relationships, [UCDVA] in front of [Left Eye] within a sentence and [Left Eye] in front of [Result] in a phrase, to establish that the [Result] corresponds to the left eye UCDVA examination. Finally, the last task in L2 conversion is to output the entity of interest to fill in the CRF field, which in the above example is the [Result].

L3 conversions use structured data to derive the CRF field. A simple example would be using a structured “Birth Date” data element in the source data to derive the “Age” CRF field. L3 conversion could also be a part of a conversion route using unstructured source data. In the previous example, the “UCDVA Left Eye Result” was able to use a L2 conversion to output the structured [Number] entity. Using the output, the “Has UCDVA been performed” CRF field, can be derived from the [Result] output using an if-then condition: If “Number” is not null, then “Yes”, else “No”.

Method Type	**Description**	**Potential Route**
Structured（L1)	The source data field is already structured data. Only need to match the standardize to a value domain.	Source(S)>L1>Target(T) S>L2>L1>T
Unstructured（L2)	The source data field is text data. It is necessary to use NLP to extract relevant research data and use extraction rules to output structured data.	S>L2>T
Derived（L3)	Use of structured data to derive new data.	S>L1>L3>T S>L2>L3>T

		S>L2>L1>L3>T

Table 3. Data Conversion Method Category Explanation

RESULT

In Table 4, the summary statistics shows an data accuracy of 98.6% and a data completeness of 23.9%. In addition, the data conversion errors by category include S0F0 (74.7%), S0F1 (0.2%), S1F0 (0.9%), and S1F1 (0.3%).

The completeness of the CRF data is strongly affected by the S0F0 (no source data to be extracted) problem. The completeness of the data was affected by research modules with large numbers of prespecified medical observation fields. For example, the Ocular Symptoms research module has 92 CRF fields asking whether the physician perceived specific symptoms for the left and right eyes. The source data only states the specific symptoms observed and rarely state which eye was observed to minimize the amount of documentation. Making it uncertain whether the other symptoms were observed. Therefore, it was decided to not derive an output of “no” for observations that were not explicitly stated in the source data and classify the field as a S0F0 problem. The same problem was seen for other research modules including Day of Surgery Medication Regimen, Other Surgical Procedures, and Surgical Complications.

The accuracy of the CRF data had minor errors from S0F1 (extracted the wrong source data), S1F0 (source data was there but not extracted), and S1F1 (correct source data extracted but produced wrong results) problems. S0F1 and S1F0 problems were often due to the entity and entity relationship definitions, where the established definitions were either too restrictive or too lenient. S0F1 leaned on the lenient side. In Appendix B, the source data only expresses the symptoms related to the left eye, but the CRF data filled in symptoms for the right eye as well. In this case, the L2 entity relationship rule written for the CRF field did not add a rule using the [Right Eye] entity to restrict the results. S1F0 leaned on the restrictive side, often starting from being restrictive in the L2 entity definition. In Appendix C, the source data expresses that the surgical method used was a “watertight incision” which implies but does not explicitly state that no additional sutures were needed. Therefore, the clinical research coordinator considered the source data to be able to answer the Closure Type CRF field, but the data was not able to be extracted. The reason it was not extracted was that the CRF field expected “suture” or “sutureless” as the standard answer. The term lists used in the [Closure Type] entity definition did not incorporate terms beyond the standard answers, therefore was unable to extract the source data. In Appendix D, similar problems existed for the BCDVA Result CRF field, where the source data for BCDVA results were expected to use a calculation format to produce the result. The [Result] entity only incorporated a regular expression for the calculation format but did not expect situations where the physician only documented the result. There was an additional case of the S1F0 problem that was due to the source data having intraocular pressure for both eyes, but the CRF field only had a single field for Intraocular pressure that did not specify the laterality of the eye. Finally, S1F1 in the study were due to L1 mapping errors that mapped the source data terminology to the wrong CRF terminology. In this case, the source data for the Ethnicity CRF field was mapped incorrectly to the Race CRF field value specifications.

S0F0	S0F1	S1F0	S1F1	准确率	完整率
74.7%	0.2%	0.9%	0.3%	98.6%	23.9%

Table 4. Summary Table of Data Integrity and Accuracy

DISCUSSION

The data conversion process from real world source data to CRF data in the study has discovered some core issues, including the lack of explicitly stated source data, lack of standardized terminology usage in the source data, and complexity in the data conversion of unstructured data. In the study, a framework was developed to diagnose the problems in the conversion process that would benefit the efficiency of correcting errors in the CRF data. The study also provided a documentation method that would help meet the requirement of China's data management protocol for real world studies. The results show that the current NLP method used can meet a high level of accuracy but has yet been able to increase the completeness of the CRF data. Without an increase in data completeness, the adoption of methods using electronic sources (eSources) directly to fill in CRFs will remain low since the methods only minimally help increase the efficiency of the clinical trial process as compared to the traditional manual method of data collection. In addition, the clinical trial process would have to consider the time it will take to construct a platform to utilize the data.

To raise efficiency and data quality when utilizing real world source data for clinical research, the standardization of source data according to clinical research data requirements is needed. Increasing data completeness would first require influencing the physician's behavior when documenting research source data. It may require physicians to explicitly state relevant research data for the NLP methods to be able to extract them for the CRF fields. In the study, it was found that physicians would only state their observed finding but will not state other prespecified observations regarding the Ocular Symptoms research module. Furthermore, physicians will not state the implied meaning of certain source data as it relates to the CRF field, such as for the Closure Type CRF field. Requiring physicians to explicitly state these observations would mean an increase in the time it will take to document the source data and increase their cognitive burden. Potential solutions to the problem may have to ensure that the physician can complete the documentation of research source data within the limits of their routine documentation time. Currently, these considerations may translate to tool that can reduce the time of documentation through a more efficient methods of input along with the incorporation of strategic notifications for the documentation of research source data in their EMR [12].

To raise the efficiency of the data conversion process will require influencing the workflow of the data engineering team. The current data conversion process requires many steps which contribute to its complexity and lack of immediate transparency. Transparency is influenced when data quality feedback from research medical experts are not immediately made known and changes made to data conversion process are not adequately documented. Changes in the process would require frequent communication between a data engineering team and research medical experts as well as the documentation of an audit trail of the changes in the conversion process and its effect on data quality. In terms of the management of complexity in the development process, members of the engineering team should find ways of reusing the existing entity related definitions either defined by a standard or through the reuse of other members work to limit the inconsistency in the data conversion process. Potential solutions to target both the transparency and complexity in the current conversion process may need to consider the usage of a clinical research data model, more automatic methods of data conversion, and immediate methods of

communicating data quality issues. First, the usage of the CDISC CDASH/SDTM clinical research data model would help create consistency in the definition of important entity types and limit the definition variations between the engineering team members [13]. Currently, the entity types are named without a standard such as [Left Eye], by using CDISC CDASH properties such as laterality [LAT] and location [LOC], the engineering team would reduce inconsistencies in entity definition and add to a central list of terminologies tied to the entity. By doing so, it would enable the adoption of terminology standards to reduce the work of manually adding terminology. It would also eliminate additional mapping of the output to the CRF fields that are already annotated with the CDISC standard. Second, NLP methods have progressed to using AI methods that handle the extraction of entities without the usage of a terminology list as well as the deciphering of entity relationships without the using explicit rules [14-17]. The limitation in these methods however is that they are considered black box methods that are hard to interpret and diagnosis if a problem in the data occurs [16,18]. However, progress is being made to make these algorithms more interpretable. For example, using the data conversion description framework, the data output from the AI algorithms can be used to deduce the rules and explicitly state the rules used for entity relationships and the collection of terminology that defined the entity used for the CRF fields. Finally, better data quality feedback would require the development of an interface that would help medical experts compare the source data with CRF data and flag data quality problems that would be sent to the engineering team.

A joint solution may be needed to increase the efficiency of research source data documentation and the conversion process of source data. It would not be ideal for the burden of standardizing source data to be given to physicians, therefore a balance between standardization at the point of documentation and after the point of documentation may be needed [19-21]. Several limitations warrant discussion for this study, the real-world data collected had previously been influenced by the prospective CATALYST study meaning that the documentation is more consistent and specific than the average routine care documentation in China. Therefore, the data conversion process developed in this study may not be able to translate well to a new data source with much more variations. In addition, the data collected in the study is restricted to data collected at the site hospital which means that CRF fields used for participant follow-up collected at a different site is incomplete and may reduce the data completeness of the CRF data.

EXPLORATION OF NEW ESOURCE METHODS

China is exploring ways to address the limitations of current research. An electronic source record (ESR) tool was developed and tested in Boao to improve the quality of source data and reduce the cost of clinical research [22]. The ESR tool is designed to help doctors record source data by extending the data collection mode and generating data collection tips, so as to improve the quality of source data. Data collection modes can include speech recognition, dialogue recognition, and optical character recognition (OCR).

First, regarding the Catalys study, source data of the initial study, including some data modules of surgical results, cataract status, and follow-up data, were not recorded in the hospital EMR, which made it impossible to standardize the data and fill in the eCRF. For example, some surgical results were collected during the operation, so they were directly recorded on the paper CRF by the research nurse, and the follow-up was not recorded in the EMR. The expansion of data collection mode can provide researchers with a recording method in different scenarios. For example, speech recognition can be used in surgical scenes, or pathological reports can be uploaded through OCR, so as to reduce repeated data entry.

Secondly, data collection tips are used to improve the integrity of research related data. In our case study, the integrity of research data is largely affected by the unclear record of research data. In order to improve the record integrity of research data, it is important to reduce the burden on doctors while

ensuring the quality. The ESR solves this problem by allowing doctors to either use the EMR template of the hospital or to customize the version suitable for their daily data collection process and configuring the relevant research data requirements as text prompts for each part of the EMR template. Compared with the traditional EMR system, the ESR recording method can reduce the time of recording the required research data by an average of 39%, with good system availability among clinicians.

The centralized storage of research source data can also improve the traceability of eCRF data and reduce the cost related to source data validation (SDV). For the traditional method of the conversion of research source data to eCRF, it requires a lot of work to compare the consistency between eCRF and source data, especially when the source data from different sources are not stored in one place. The ESR tool solves this problem by binding eCRF data fields to source data and creating an interface that can highlight the source data used by each field to facilitate the verification of source data. This interface can help trace the "location" of source data. In addition, the process of converting source data into eCRF should also record the "method" of data extraction. In the case study, the rules of extracting data are helpful to answer "how" to transform the source data, but it requires those who evaluate the method to understand the NLP method. Further research should be warranted to explore ways to improve the interpretability of NLP method, which the improvement of the transparency of the conversion process may be necessary.

Finally, researchers and regulators are interested in data from different studies with similar treatment or disease areas, providing a basis for producing more generalized clinical evidence and ultimately supporting analysis and decision-making. For regulators, the CDISC SDTM data model is a key data submission standard, which is conducive to the efficient inspection of research data. It is also an important trend to convert real-world data into clinical research data models. Organizing different data sources into data formats that can be used for secondary research can improve the repeatability of clinical research. In the case study, due to the non-standard definition of entity types and the different documentation habits of clinicians from different data sources, the NLP method of the project may not be applied repeatedly to projects in other fields. To solve the problem of reusability of NLP method, our research team has proposed a feasible solution [23]. This scheme uses NLP model based on artificial intelligence, in which the NLP model uses text data marked by CDISC standard for training [23]. The NLP model based on CDISC can reduce the development of data conversion algorithms in similar fields (such as laboratory data), and the extracted data can be directly filled into the eCRF annotated by CDISC.

CONCLUSION

This study created a framework to describe the eSource data conversion process, and evaluated the performance of the data conversion process in terms of eCRF data accuracy and data integrity. The results show that the data accuracy of this process is 98.6%, but the data integrity is only 23.9%. The limitations of this study include the use of a single real-world data source and the lack of follow-up data. This study also explores a feasible scheme to improve the quality and efficiency of data transformation.

REFERENCES

[1] 中华人民共和国中央人民政府. 四部门印发《关于支持建设博鳌乐城国际医疗旅游先行区的实施方案》[EB/OL]. (2019-09-10)[2022-07-04].http://www.gov.cn/xinwen/2019-09/17/5430452/files/7d1f580ed40b4853a061f1b5da0e23d2.pdf.

[2] 国家药品监督管理局. 中国药品监管科学行动计划第二批重点项目发布. [EB/OL]. (2021-06-24)[2022-07-04]. https://www.nmpa.gov.cn/directory/web/nmpa/xxgk/fgwj/gzwj/gzwjyp/20210628172854126.html

[3] 国家药品监督管理局. 国家药监局关于发布真实世界证据支持药物研发与审评的指导原则（试行）的通告(2020 年第 1 号)[EB/OL].(2020-01-03)[2021-07-05]. https://www.nmpa.gov.cn/xxgk/ggtg/qtggtg/20200107151901190.html.

[4] U.S. Food and Drug Administration. (2018). Use of Electronic Health Record Data in Clinical Investigations. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-electronic-health-record-data-clinical-investigations-guidance-industry. Accessed Jun 23, 2022.

[5] F Jin, C Yao, X Yan, et al. Gap between real-world data and clinical research within hospitals in China: a qualitative study. Bmj Open 2020;10(12):e38375.

[6] AA Parab, P Mehta, A Vattikola, et al. Accelerating the Adoption of eSource in Clinical Research: A Transcelerate Point of View. Ther Innov Regul Sci 2020;54(5):1141-51.

[7] J Xie, EQ Wu, S Wang, et al. Real-World Data for Healthcare Research in China: Call for Actions. Value Health Reg Issues 2022;27:72-81.

[8] U.S. National Library of Medicine. (2020). A Real-World Evidence Study in China of the Catalys Precision Laser System. Available at https://clinicaltrials.gov/ct2/show/NCT04171518. Accessed Jun 23, 2022.

[9] 中华人民共和国国家卫生和计划生育委员会. 电子病历基本数据集. [EB/OL].(2014-06-20)[2022-07-07]. http://www.nhc.gov.cn/zwgkzt/s9497/201406/15349b9e45ff4aabac990fdd86332029.shtml

[10] 中华人民共和国卫生部. 卫生信息数据元目录. [EB/OL].(2011-08-02)[2022-07-07].http://www.nhc.gov.cn/zwgkzt/s9497/201108/52755/files/ef22d2039ed945fcb18f8f27bfd6f7fa.pdf

[11] 中华人民共和国卫生部. 卫生信息数据元值域代码. [EB/OL].(2011-08-02)[2022-07-07].http://www.nhc.gov.cn/wjw/s9497/201108/52759.shtml

[12] B Wang, X Hao, X Yan, et al. Evaluation of the clinical application effect of eSource record tools for clinical research. BMC Med Inform Decis Mak 2022;22(1):98.

[13] S Takahara, TI Saito, Y Imai, T Kawakami, T Murayama. A use-case analysis of Clinical Data Interchange Standards Consortium/Study Data Tabulation Model in academia in an investigator-initiated clinical trial. Nagoya J Med Sci 2022;84(1):120-32.

[14] T Chen, Y Hu. Entity relation extraction from electronic medical records based on improved annotation rules and BiLSTM-CRF. Ann Transl Med 2021;9(18):1415.

[15] H Liu, Z Zhang, Y Xu, et al. Use of BERT (Bidirectional Encoder Representations from Transformers)-Based Deep Learning Method for Extracting Evidences in Chinese Radiology Reports: Development of a Computer-Aided Liver Cancer Diagnosis Framework. J Med Internet Res 2021;23(1):e19689.

[16] S Wu, K Roberts, S Datta, et al. Deep learning in clinical natural language processing: a methodical review. J Am Med Inform Assoc 2020;27(3):457-70.

[17] N Perera, M Dehmer, F Emmert-Streib. Named Entity Recognition and Relation Detection for Biomedical Information Extraction. Front Cell Dev Biol 2020;8:673.

[18] R Daneshjou, MP Smith, MD Sun, V Rotemberg, J Zou. Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review. Jama Dermatol 2021;157(11):1362-69.

[19] Y Chen, D Hu, M Li, H Duan, X Lu. Automatic SNOMED CT coding of Chinese clinical terms via attention-based semantic matching. Int J Med Inform 2022;159:104676.

[20] H Forsvik, V Voipio, J Lamminen, et al. Literature Review of Patient Record Structures from the Physician's Perspective. J Med Syst 2017;41(2):29.

[21] K Hung, M Lau, V Fung. Successful Implementation of Terminology Binding in Hong Kong Hospital Authority. Stud Health Technol Inform 2019;264:1486-87.

[22] B Wang, X Hao, X Yan, et al. Evaluation of the clinical application effect of eSource record tools for clinical research. BMC Med Inform Decis Mak 2022;22(1):98.

[23] 赖俊恺, 王斌, 姚晨, 等. 从真实世界数据到临床研究数据的标准转化研究[J]. 中国食品药品监管, 2021, (11): 39-46.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Junkai Lai
Hangzhou LionMed Medical Information Technology Co., Ltd;
Peking University Clinical Research Institute
18600031841 (WeChatID: d8322489388)
junkai.daniel.lai@qq.com