International Journal of Advanced Science and Computer Applications (2025); 4(2)
DOI: 10.47679/ijasca.v4i2.79
https://ijasca.org/index.php/ijasca

RESEARCH ARTICLE

Knowledge Graph-based JingFang Drug Efficacy Analysis With a Supportive Randomized Controlled Influenza-like Illness Clinical Trial

Yuqing Li, Zhitao Jiang, Zhiyan Huang, Wenqiao Gong, Yanling Jiang, Guoliang Cheng

Published online: 4 September 2025

Abstract

This paper presents a novel methodology for drug efficacy analysis using a knowledge graph, validated by a randomized controlled clinical trial. To provide a comprehensive understanding of drug treatment effects, a learning-based workflow is developed to mine drug-disease entities and relations from literature. These relations build a knowledge graph used for clustering-based drug efficacy analysis. Our tool reports the learned relatedness between drugs and diseases, indicating efficacy levels. JingFang is identified as effective for flu and colds. To validate this, a clinical trial was conducted on Influenza-like illness. Between August 25 and October 12, 2020, 106 patients were randomly assigned in a 1:1 ratio to either the combined group (53) or the control group (53). Patients in the combined group received Xinkangtai Ke and JingFang, while the control group received Xinkangtai Ke only for 7 days. The combined group's cure rate was 92.5% (49) compared to 81.1% (43) in the control group (p=0.0852). The very effective rate was 98.1% (52) in the combined group versus 92.5% (49) in the control group (p=0.3692). For middle-aged and elderly participants, the combined group's recovery rate was significantly higher than the control group's (100% vs 78.4%, p=0.0059, 95% CI: 21.6 (8.3, 38.2)). No adverse effects were observed in either group. The results indicate that JingFang is effective for patients with Influenza-like illnesses, especially those over 34 years old. This study highlights the potential of knowledge graph-based analysis in drug efficacy research.

Keyword: Knowledge Graph, Clinical Trial, Influenza-like illness, JingFang, Drug Efficacy Analysis

INTRODUCTON

A biomedical network can be described by a knowledge graph (KG) [1], where a node can represent various types of bio-entities, including proteins, drugs, chemicals, diseases, and species, and an edge describes a relationship between an entity pair. A KG can be decomposed into a collection of <head entity, tail entity, predicate> triples, the predicate connects the head and tail items, showing a connection. For instance, <drug A, protein B, affect> can represent the regulation relationship between a drug and a protein. Also, each node and edge in a KG can be embedded with a set of attributes that offer additional information such as the sources of the research articles where the relationship is obtained from. Via literature mining and deep learning models, a wide range of KG has been built and applied to various notable domains in bio-science, including drug discovery and re-purposing [2], protein-protein interaction [3, 4], chemical-protein interaction [5], disease mechanism identification [6], and disease biomarkers network [7].

Drug efficacy prediction and analysis have been essential tasks in computational pharmacology [8]. Recent years have witnessed the development of a spectrum of KG-based methods for drug efficacy analytics [9, 10]. A core mission of these methods is to investigate the similarity between drugs as well as the treatment efficacy of drugs on diseases [11], with a basic assumption that two similar drugs may present similar efficacy for the same diseases. For pharmaceutical companies, however, it is also important to understand the efficacy of a certain drug on various diseases along the way ever since the drug was approved and consumed in the market, which can usually be found in clinical trials reported in research articles. To fulfill this need, it is expected to build a system that can track and compile the relevant clinical traits that have the drug involved and tested.

We present a novel methodology to demonstrate knowledge graph-based drug efficacy analysis, which is validated by a randomized controlled clinical trial conducted for JingFang [12]. To provide a comprehensive understanding of JingFang's treatment effect and function, a machine learning-based pipeline is developed to mine drug-disease entities and relations from the literature. These extracted relations are used to build a knowledge graph that is further used for a clustering-based drug efficacy analysis. With a given drug, our tool can report the learned relatedness between a drug and a

Zhangjiagang TCM Hospital Affiliated to Nanjing University of Chinese Medicine, 215600, China

*) corresponding author

Email: 18905391156@163.com

Received 15 May 2025 | Received in revised form 11 Juli 2025 | Accepted 04 August 2025
Available online 4 September 2025

disease, indicating the degree of efficacy between the drug-disease pair.

We propose a literature-based measure between "drug composition and efficacy impact". The rising expense of drug research, combined with a significant decrease in the number of new pharmaceutical approvals, has increased the demand for innovative target identification and effectiveness prediction tools. Here, we introduce a literature-based measure of the "efficacy of a drug component on a disease" that quantifies the interaction between a drug component target and a disease. By adjusting for known biases in the interaction group, proximity aids in detecting a drug's therapeutic impact and distinguishing between unsuccessful and effective therapies. Based on the analytical result, we identify that JingFang has been reported to be effective in treating flu and colds. To further validate the result, a randomized controlled clinical trial is conducted to evaluate JingFang's efficacy on Influenza-like illness, a subtype of cold.

Influenza-like illness refers to symptoms similar to the common cold, such as chills, fever, limb aches, nasal congestion, runny nose, headache, and cough, when exposed to air conditioning for a long time, and is also called Influenza-like illness syndrome. The treatment of Influenza-like illness is based on symptomatic treatment and symptom relief while paying attention to rest, proper hydration, and keeping indoor air circulation.

The purpose of this study was to establish a method for measuring pharmacological effectiveness using knowledge graphs, integrating data from the literature, and validating by a randomized controlled clinical trial conducted on Jing Fang granules, by combining the results of knowledge graphs and clinical trials, We can more accurately assess the efficacy of Jingfang's drugs against influenza-like illness (Figure 1).

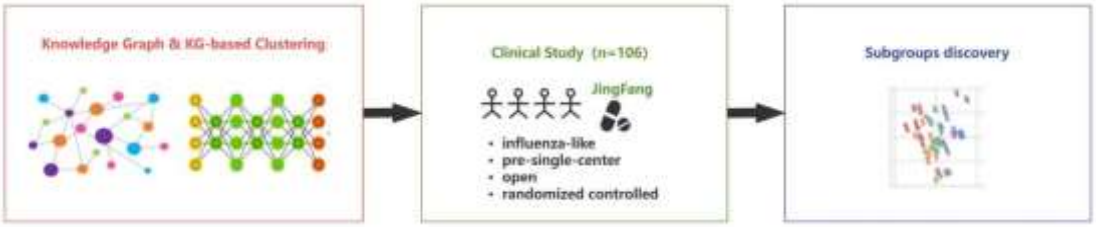

Fig 1. The evaluation framework of knowledge graph and influenza-like illness clinical trials. Combining the results of the knowledge graph and clinical trials, JingFang's efficacy is accurately evaluated.

METHODS

Knowledge Graph-based Analytics

Literature Scraping

We utilized a self-developed tool for web scraping. As shown in Table 1, a total of 19,053 paper abstracts were collected using four different keywords, including "JingFang", "JingFang" (Chinese), "Flu", and "Influenza-like illness". After an initial screening, 4,429 relevant abstracts were kept in the dataset for knowledge extraction. The fields used for literature scraping include the following: paper type, the title of the paper, author list, author affiliation, source, keywords, abstract, publication time, funding, volume, issue, page, URL, and DOI.

Table 1. Stats of literature collection

Keyword	# abstracts	# Abstracts after cleaning
JingFang	642	221
JingFang (Chinese)	2,324	578
Flu	8,592	1,327
Influenza-like illness	7,495	2,303
Total	19,053	4,429

Building a Knowledge Graph

Each abstract scraped from the Internet is semi-structured, in which the structured information such as the author list, the year of publication, the affiliations, and so on, and the unstructured title and abstract text. Our knowledge graph contains three entity types, including abstract, drug, and disease. The relation between a drug and a disease can be either "treat" or "cause". As shown in Figure 2, an abstract text is fed into a MacBERT pre-trained model for the extraction of entities and relationships. Each extracted relation is a three-tuple $< e_1, e_2, r >$, where e_1 and e_2 are the entities of the head and tail. Here, they are a drug and a disease. r, on the other hand, is the relation connecting the drug and disease. Other structured attributes, along with the extracted drugs and diseases, can be used to build a knowledge graph. To support further analysis, the knowledge graph is processed to generate an adjacency matrix to encode the interaction between drugs and diseases. Specifically, if a drug can treat a disease, and the relation has appeared in n abstracts, then the value of the corresponding cell of the drug and disease in the matrix should be n.

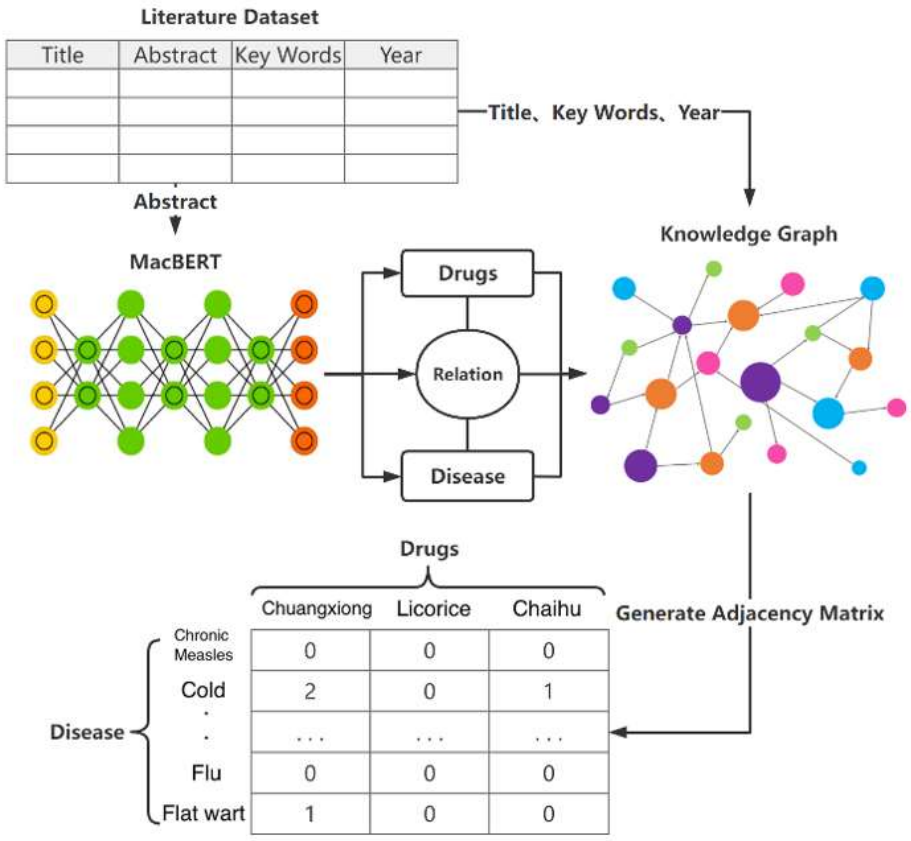

Fig 2. Workflow of building the drug-condition knowledge graph

KG-based Clustering for Drug Efficacy Analysis

The adjacency matrix generated from the previous step can be normalized and used to train a Graph Convolutional Network (GCN) [13] such that each graph node and edge may be represented as a numerical vector. To encode the semantics embedded in the abstract text, we pass the text through the MacBERT [14] model for a word vector mapping that converts each word to a vector. However, not all word vectors are utilized since most of the word tokens are not related to the drug efficacy analysis task. In our case, we only keep the word vectors for drugs and diseases. Each drug entity now has two representations, one from the GCN and one from the word vector mapping. Both vectors are fed into the T5-small [15] model, which is utilized as a feature-fusion module to combine the two representations. The output of T5-small is sent through a K-means [16] algorithm for clustering analysis. Essentially, the drugs have two types, a drug product, and its composition chemicals. For our KG, the extracted drug can be either. The idea of this analysis is that the closer the drug is to the cluster centroid, the stronger the positive correlation with the current disease. The process is described in Figure 3.

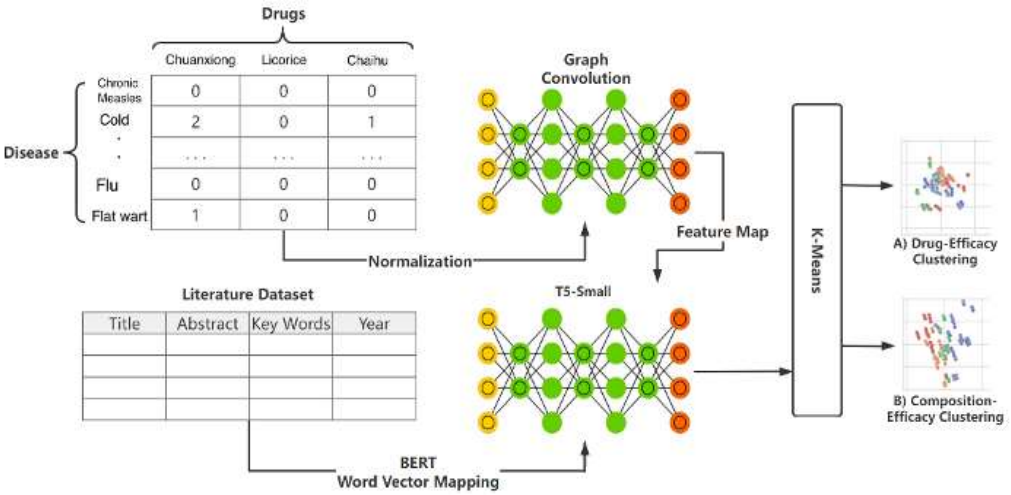

Fig 3. Workflow of KG-based Clustering for Drug Efficacy Analysis

A Randomized Controlled Clinical Study

Design of Study

In the second half of 2020, we conducted a single-center, open, randomized controlled clinical study from 25 August 2020 to 12 October 2020, with 108 patients participating in the study. The diagnostic criteria for Influenza-like illness were: onset on a hot day (June-October) and exposure to air conditioning or frequent entry and exit from an air-conditioned room for at least 3 days before onset, as well as meeting the following Western medical diagnostic criteria.

Western diagnostic criteria for the common cold are met: sneezing, nasal congestion, runny nose, cough, sore throat, and other local symptoms, and lacrimation predominates, and systemic symptoms such as chills and fever, general malaise, dizziness, and headache may or may not be evident; white blood cell count is normal or low.

A central randomization system (web-based Interactive Web Response System, IWRS) was used to achieve randomization groups in this study. Subjects were randomly divided into test and control groups in a 1:1 ratio and met the inclusion criteria. Subjects in the control group took only Neocontrol (Blue) (Sino-Medical), while subjects in the trial group took Neocontrol (Blue) plus JingFang (Shandong New Age Pharmaceutical Co., Ltd.). The study used a block randomization grouping method with a block length of 4. The randomization process was set up by a statistical and computer professional who set up the randomization grouping procedure.

This study has been approved by the Ethics Committee of Zhangjiagang City Hospital of Traditional Chinese Medicine and has been registered with the China Clinical Trials Registry (chictr.org.cn) under the registration number ChiCTR2000036543.

Inclusion and Exclusion Criteria

Males and females between the ages of 18 and 70 are eligible; onset of illness on hot days (June to October) and experience of exposure to an air-conditioned environment or frequent entry and exit from air-conditioned rooms for at least 3 days before onset; meeting the Western medical diagnostic criteria for the common cold; within 48 hours of onset; not having taken JingFang, Neocontrol (blue), Tylenol cold tablets, Neocontrol (red) or Day & Night Pepcid (night tablets) within 2 weeks before enrollment; subjects willingly engaged in the study and signed an informed consent form.

Subjects who met any of the following criteria were barred from participating: those with wind-heat colds (manifested by heavy fever, slight wind aversion, sweating, thirst, runny nose, red, swollen and hot throat, coughing and spitting yellow sputum, etc.); those with pharyngoconjunctivitis, acute attacks of chronic bronchitis, purulent tonsillitis, infectious upper respiratory tract infection; those with uncontrolled cardiovascular disease, diabetes, hypertension, thyroid disease, asthma, glaucoma, emphysema, chronic lung disease, dyspnea or prostatic hypertrophy; patients with pneumonia diagnosed by chest imaging; those who have used drugs for the treatment of this disease since the onset of the disease; those with active liver disease or uncontrollable liver disease; those with uncontrollable kidney disease or on kidney dialysis; axillary temperature ≥40 degrees Celsius, white blood cell total count 10 ×109/L or neutrophil classification > 80%; those who are allergic to the drugs used in this study; those who have mental or neurological disorders and cannot express their will correctly; pregnant women, lactating women and women of childbearing age who are not using contraception; those who are currently participating in clinical trials of other drugs or medical devices; and those who are considered by the investigator to be unsuitable for inclusion.

Medication

JingFang is produced by Shandong New Times Pharmaceutical Co., Ltd. The main ingredients are Bupleurum, Chuanxiong, Duhuo, Fangfeng, Poria, Licorice, Nepeta, Platycodon grandiflorum, Qianhu, Qianghuo, and Citrus aurantium. New Contac (Blue Pack), take 1 capsule every 12 hours after meals; do not exceed 2 capsules in 24 hours.JingFang, take 1 bag at a time, 3 times a day, with boiling water. The therapy lasts seven days. The subjects in the two groups received the same non-drug intervention program, that is, diet control and lifestyle improvement, mainly including: avoiding greasy and spicy food, avoiding tobacco and alcohol, avoiding overwork and overeating, and maintaining a good attitude.

Evaluation Criteria

The primary endpoint was the rate of healing within 7 days. Clinical cure: clinical symptoms and signs vanished or almost vanished, and the symptom score was decreased by 95%; efficacy: clinical symptoms and signs considerably improved, and the symptom score was lowered by 70%;

Clinical symptoms and indicators improved, and the symptom score was lowered by more than 30%. Clinical symptoms and indicators did not improve considerably, if at all, and the symptom score was lowered by less than 30%. Healing rate (%) = (number of clinically healed cases + number of apparent effect cases) ÷ total cases ×100%. The secondary endpoint was the incidence of adverse events.

Statistical methods

The key assessment criterion for this study is the drug's therapeutic effectiveness rate after 7 days of treatment. This study adopts the hypothesis of superiority. According to previous literature and preliminary test results, the control group's treatment effectiveness rate was 63.3%, and 63.3% in the experimental group. The treatment effective rate was 90%, and the superiority margin of the two groups was 3%, with α= 0.025 (one-sided), β= 0.2, and the sample size ratio of the two groups was 1:1. 45 patients were included, and after taking into account the 15% loss to follow-up rate, 53 patients were finally included in each group, there were a total of 106 patients in the two groups. SAS 9.4 statistical analysis software was used for statistical processing, and the obtained results measures were reported as mean, ±standard deviation, or median (upper and lower quartiles). Comparisons of measurement data were first tested for normality, and if they conformed to a normal distribution, parametric tests were used; otherwise, Wilcoxon rank sum tests were performed. The frequency (composition ratio) was used to statistically describe count data. To compare the count data, the chi-square test or Fisher's test was utilized. p<0.05 was chosen as the criterion of significance.

Subgroup analysis was performed on the cure rate and cure rate of patients in different age groups. Patients aged ≤34 years were young people and patients >34 years old were middle-aged and elderly people.

RESULTS

The experiments for this inquiry were carried out using Python 3.7.0. PyCaret was used to implement the learning algorithms[17]. Microsoft Office 365 Excel, Matplotlib 3.4.2, and Seaborn 0.11 were utilized to make the charts. BAIX (https://github.com/aibaix, accessed June 9th, 2022) was a self-developed Python tool used for data purification and exploratory data analysis.

Results Of Knowledge Graph Analysis

Discovery of Knowledge Graph-based Drug-Disease Relationships

We utilize Neo4J to store knowledge graph data, where the underlying storage structure is optimized for graph data attributes, delivering superior performance in processing relational data compared to other databases. Figure 4 presents an extracted portion of the knowledge graph, visualizing multiple node entities and the relationships graphically connecting them.

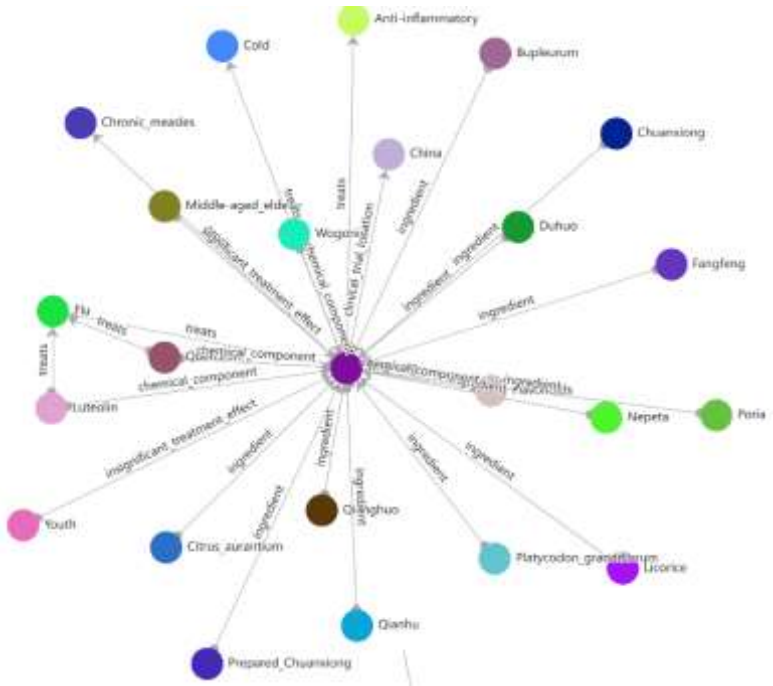

Fig 4. An example of the generated drug-condition knowledge graph

In the graph, node entities are represented by colored circles, with blue denoting diseases or symptoms, green representing herbal medicines or treatment methods, and purple symbolizing chemical components, among others. The node entities are linked through labeled edges that describe the semantic relationships between them, such as "treats," "ingredient," and "significant treatment effect."

Drug Efficacy Analysis

Figures 5 and 6 show the two results that demonstrate the KG-based clustering analysis, in which Figure 5 displays the relatedness scores of JingFang and the commonly related conditions. It is shown that flu, chronic measles, anti-inflammatory, and cold are the top conditions that can be treated by JingFang. Specifically, flu has the highest score of 0.9374, meaning that in existing literature, JingFang is found to be the most effective drug to treat flu compared to other conditions. Figure 6, on the other hand, shows the pairwise relatedness between the composed chemicals of JingFang and conditions to identify how each chemical component takes effect on a certain condition. As shown in the figure, we intercepted only the top 8 ranked chemical components, quercetin, luteolin, kaempferol, wogonin, beta-sitosterol, naringenin, acacetin, and tanshinone IIIA, whose degree of influence varies in different diseases. The picture shows quercetin, and luteolin may be the key most effective ingredients in the treatment of influenza with JingFang.

Condition	Relatedness
Flu	0.9374
Chronic measles	0.6818
Anti-inflammatory	0.5539
Cold	0.3409
Anti-allergy	0.2983
Upper respiratory tract infection	0.2983
Flat wart	0.2557
Acute lung injury	0.2131
Mumps	0.2131
Atopic dermatitis	0.2131
White fresh skin	0.1704
Eczema	0.1704
Psoriasis	0.1704
Allergic dermatitis	0.1278
Pruritus	0.1278
Diabetic nephropathy	0.1278

Fig 5. Relatedness scores of JingFang and the commonly-related conditions

	Cold	Chronic Measles	Upper Respiratory Tract Infection	Atopic Dermatitis
quercetin	0.4857	0.2429	0.7286	0.6072
luteolin	0.6072	0.3642	0.4857	0.4857
kaempferol	0.3642	0.3642	0.2429	0.1214
wogonin	0.3642	0	0.2429	0.1214
beta-sitosterol	0.2429	0.1214	0.4857	0.2429
naringenin	0.2429	0.2429	0.1214	0.2429
acacetin	0.2429	0.1214	0.2429	0.1214
tanshinone II IA	0.2429	0	0.2429	0.3642

Fig 6. Pair-wise relatedness between the composed chemicals of JingFang and conditions

Results of clinical trials

Analysis of baseline characteristics of enrolled patients

A total of 108 patients were recruited from August 25, 2020, to October 12, 2020, and finally, 106 patients were enrolled and randomized to receive JingFang and Neocontrol (53 patients in the treatment group) or Neocontrol only (53 patients in the control group). Patients' ages in the treatment and control groups were (41.8 ± 15.8) and (43.5 ± 13.75) years, respectively, without any statistically significant variations. It's statistically significant in terms of gender structure, ethnic structure, BMI, total symptom score, and physical findings score, and were comparable in the difference between the treatment and control groups(see Table 2).

Table 2. Baseline characteristics of enrolled patients

	Test group (N=53)	Control group (N=53)	*p*-value
Mean age (SD)	41.8 (15.18)	43.5 (13.75)	0.58
# male patients (%)	19 (35.8)	12 (22.6)	0.14
BMI (SD)	23.09 (2.999)	23.41 (3.258)	0.61
Overall symptom score (SD)	5.5 (2.11)	5.7 (2.24)	0.81
Physical examination score (SD)	0.8 (0.55)	0.9 (0.48)	0.67

Efficacy analysis

The healing rate within 7 days was 92.5% (49 cases) in the test group and 81.1% (43 cases) in the control group, which was higher than the healing rate in the control group, but, no statistically significant difference existed between the two groups (p=0.0852, 95% CI: 11.3 (-2.0, 25.3)). Within 7 days, the healing rate in the test group was 98.1% (52 cases) and 92.5%

(49 cases) in the control group, which was higher than the control group's healing rate, but, no statistically significant difference existed between the two groups (p=0.3692, 95% CI: 5.7 (-3.5, 16.5)) (Table 3).

Table 3. Efficacy analysis

	Test group (N=53)	Control group (N=53)	p-value
Cured	49 (92.5)	43 (81.1)	0.09
Very effective	3 (5.7)	6 (11.3)	-
Effective	1 (1.9)	4 (7.5)	-
Not effective	0	0	-
Cured+very effective (%)	52 (98.1)	49 (92.5)	0.36

Subgroup analysis

In middle-aged and elderly subjects, the healing rate in the test group was 100% (32 cases) and 78.4% (29 cases) in the control group, which was statistically significantly higher than the healing rate in the control group(0.0059, 95% CI: 21.6 (8.3, 38.2)) (Table 4). In the youth population, the healing rates were essentially the same in both groups.

Table 4. Efficacy analysis

Age group	Curative effect	Test group (N=53)	Control group (N=53)	p-value
Young	Cured	17(81.0)	14(87.5)	0.6796
	Very effective	3(14.3)	1(6.3)	-
	Effective	1(4.8)	1(6.3)	-
	Not effective	0	0	-
	Cured+very effective	20(95.2)	15(93.8)	1
Middle-aged and elderly	Cured	32 (100.0)	29(78.4)	0.0059
	Very effective	0	5(13.5)	-
	Effective	0	3(8.1)	-
	Not effective	0	0	-
	Cured+very effective	32 (100.0)	34(91.9)	0.243

Knowledge graph and clinical trial-based integration analysis

This study aims to evaluate the therapeutic efficacy of Jingfang for influenza-like illnesses by integrating knowledge graph technology with clinical trial data. We developed an innovative knowledge graph-based pharmacological analysis method and validated its effectiveness through a randomized controlled clinical trial.

First, we constructed a knowledge graph by extracting drug-disease entities and their relationships from literature using a machine learning workflow. Our tool can report drug-disease correlations that indicate the degree of efficacy between drug-disease pairs. Specifically, we collected 19,053 abstracts and used our in-house text-mining tool to extract relationship information between drugs and diseases. Each extracted relationship was encoded as an adjacency matrix for subsequent analysis. This knowledge graph not only contains drug and disease entities but also reflects the therapeutic or pathological associations between them.

To deeply mine the information embedded in the knowledge graph, we applied a graph convolutional network (GCN) to normalize the adjacency matrix and used a T5 mini-model to fuse the GCN-obtained representations with word vector graphs. Through this approach, we analyzed the association between Jingfang and various diseases and explored the potential therapeutic effects of Jingfang for influenza-like illnesses using the K-means clustering algorithm.

To validate the knowledge graph analysis results, we conducted a randomized controlled clinical trial in China. The trial enrolled 106 patients with influenza-like illnesses, and the results showed that the cure rate in the Jingfang combined treatment group (92.5%) was significantly higher than that in the control group (81.1%), especially among the middle-aged and elderly population. Subgroup analysis of the clinical data revealed that Jingfang had a more pronounced therapeutic effect on middle-aged and elderly patients aged 34 and above, which was consistent with the knowledge graph analysis results. However, the knowledge graph did not capture this age-related difference in efficacy, and future work may consider incorporating demographic information into knowledge representation and analysis.

The innovation of this study lies in proposing a novel framework for evaluating therapeutic efficacy by combining knowledge graphs with clinical trial results, thereby enhancing the understanding of drug treatment effects. This not only provides new analytical tools for similar drug development but also improves the efficiency and accuracy of drug development by systematically validating literature efficacy data and integrating it with actual clinical trial results. Additionally, applying a knowledge graph to evaluate the therapeutic effects of traditional Chinese medicines like Jingfang is an innovative and unique approach, bringing new perspectives to this under-explored field.

In terms of technical implementation, we constructed a multi-layered knowledge graph by extracting relevant data from a vast amount of biomedical literature and using automated text-mining tools to identify key drug and disease entities and their relationships. With the aid of graph convolutional network processing, we could capture complex associations between entities and discover drug combinations with similar therapeutic effects through clustering analysis. This multi-layered knowledge graph comprehensively presents the relationships between drug components and diseases, and reveals the potential therapeutic effects of different components on specific diseases, laying a theoretical foundation for clinical trials and drug development.

However, this study also has some limitations. First, the accuracy of clustering analysis depends on the quality and completeness of the literature data, and biases and omissions in the literature may affect the accuracy of the results. Second, the sample size of the clinical trial is relatively small, which may impact the stability and generalizability of the statistical results. Future work should expand the sample size and utilize more independent data sources to validate and optimize this integrated analysis method.

In conclusion, the knowledge graph-based method enhances our understanding of drug mechanisms of action and provides an effective tool for predicting and validating drug clinical efficacy. By integrating knowledge graph analysis with clinical trial results, we can more accurately evaluate the therapeutic efficacy of Jingfang for influenza-like illnesses and provide scientific evidence for its clinical application, promoting the modernization of traditional Chinese medicine evaluation.

CONCLUSION

This study introduces a novel approach to drug efficacy analysis using a knowledge graph (KG) methodology, complemented by a randomized controlled trial to validate the effectiveness of JingFang in treating Influenza-like illness. By extracting and analyzing drug-disease relationships from the literature, a comprehensive KG was constructed, serving as the foundation for the efficacy analysis. The trial results indicated a significantly higher cure rate for the JingFang group, especially among middle-aged and elderly patients, compared to the control group. This innovative approach not only provides a powerful tool for predicting drug efficacy but also combines traditional clinical trial results with advanced data analysis techniques, thereby enhancing the accuracy and reliability of drug efficacy evaluations. This method holds potential for broad application in drug development and repurposing, particularly in the context of Traditional Chinese Medicine.

Conflict of Interest Statement

The authors declare no conflict of interest.

Author Contributions

Conceptualization and methodology, Y. L., Z. J., Z. H., W. G., G. C., and Y. J.; software, validation, and original draft preparation, Y. L., Z. J., Z. H., and W. G.; review and editing, G. C. and Y. J.. All authors have read and agreed to the published version of the manuscript.

Funding

This work was not supported by any funding.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Patrick Ernst, Amy Siu, and Gerhard Weikum. Knowlife: a versatile approach for constructing a large knowl- edge graph for biomedical sciences. BMC bioinformatics, 16(1):1– 13, 2015.

[2] Xiangxiang Zeng, Xinqi Tu, Yuansheng Liu, Xiangzheng Fu, and Yansen Su. Toward better drug discovery with knowledge graph. Current opinion in structural biology, 72:114– 126, 2022.

[3] Siyuan Cheng, Xiaozhuan Liang, Zhen Bi, Ningyu Zhang, and Huajun Chen. Proteinkg65: A knowledge graph for protein science. arXiv preprint arXiv:2207.10080, 2022.

[4] Sameh K Mohamed, Vít Nováek, and Aayah Nounu. Discovering protein drug targets using knowledge graph embeddings. Bioinformatics, 36(2):603–610, 2020.

[5] Xian Zhu, Yueming Gu, and Zhifeng Xiao. Herbkg: Constructing a herbal-molecular medicine knowledge graph using a two-stage framework based on deep transfer learning. Frontiers in Genetics, 13, 2022.

[6] Zhenfeng Lei, Yuan Sun, Yaser Ahangari Nanehkaran, Shuangyuan Yang, Md Saiful Islam, Huiqing Lei, and Defu Zhang. A novel data-driven robust framework based on machine learning and knowledge graph for disease classification. Future Generation Computer Systems, 102:534–548, 2020.

[7] Kun Yu, Weidong Xie, Linjie Wang, Shoujia Zhang, and Wei Li. Determination of biomarkers from microar- ray data using graph neural network and spectral clustering. Scientific reports, 11(1):1– 11, 2021.

[8] Jie Zhu, Jingxiang Wang, Xin Wang, Mingjing Gao, Bingbing Guo, Miaomiao Gao, Jiarui Liu, Yanqiu Yu, Liang Wang, Weikaixin Kong, et al. Prediction of drug efficacy from transcriptional profiles with deep learn- ing. Nature biotechnology, 39(11):1444– 1452, 2021.

[9] Wytze J Vlietstra, Rein Vos, Anneke M Sijbers, Erik M van Mulligen, and Jan A Kors. Using predicate and provenance information from a knowledge graph for drug efficacy screening. Journal of biomedical semantics, 9(1):1– 10, 2018.

[10] Yongjun Zhu, Chao Che, Bo Jin, Ningrui Zhang, Chang Su, and Fei Wang. Knowledge-driven drug repurpos- ing using a comprehensive drug knowledge graph. Health Informatics Journal, 26(4):2737–2750, 2020.

[11] Lan Huang, Huimin Luo, Suning Li, Fang-Xiang Wu, and Jianxin Wang. Drug – drug similarity measure and its applications. Briefings in Bioinformatics, 22(4):bbaa265, 2021.

[12] ShiRong Li, XiangZi Li, TianYe Yang, LiHong Pan, YuYu Xu, LiJuan Wang, MingMin Jiang, JiDong Zhou, ChengHong Sun, JingChun Yao, et al. Jingfang granules alleviate lps-induced mastitis by inhibiting inflam- mation, protecting the blood-milk barrier structure and regulating cell apoptosis. Pharmacological Research- Modern Chinese Medicine, 2:100072, 2022.

[13] Si Zhang, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. Graph convolutional networks: a comprehensive review. Computational Social Networks, 6(1):1–23, 2019.

[14] Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. Revisiting pre-trained models for chinese natural language processing. arXiv preprint arXiv:2004.13922, 2020.

[15] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res., 21(140):1–67, 2020.

[16] Hans-Hermann Bock. Clustering methods: a history of k-means algorithms. Selected contributions in data analysis and classification, pages 161– 172, 2007.
[17] Moez Ali. PyCaret: An open source, low-code machine learning library in Python, April 2020. PyCaret version 1.0