Transl Clin Pharmacol. 2024 Dec;32(4):177-186
https://doi.org/10.12793/tcp.2024.32.e17
pISSN 2289-0882·eISSN 2383-5427

Translational and Clinical Pharmacology TCP

Original Article

Check for updates

Analyzing collaborations in clinical trials in Korea using association rule mining

Ki Young Huh [1,2] **and Ildae Song** [3,*]

[1]Department of Clinical Pharmacology and Therapeutics, Seoul National University College of Medicine and Hospital, Seoul 03080, Korea
[2]Clinical Trials Center, Seoul National University Hospital, Seoul 03080, Korea
[3]Department of Pharmaceutical Science and Technology, Kyungsung University, Busan 48434, Korea

OPEN ACCESS

Received: Oct 16, 2024
Revised: Nov 21, 2024
Accepted: Dec 8, 2024
Published online: Dec 16, 2024

***Correspondence to**
Ildae Song
Department of Pharmaceutical Science and Technology, Kyungsung University, 309 Suyeong-ro, Nam-gu, Busan 48434, Korea.
Email: cianni77@kyungsung.ac.kr

ORCID iDs
Ki Young Huh
https://orcid.org/0000-0002-1872-9954
Ildae Song
https://orcid.org/0000-0002-3904-4735

Funding
This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE: 2023RIS-007).

Conflict of Interest
- Authors: Nothing to declare
- Reviewers: Nothing to declare
- Editors: Nothing to declare

Reviewer
This article was reviewed by peer experts who are not TCP editors.

ABSTRACT

Identifying how trial sites collaborate is essential for multicenter trials. The ways in which collaboration among trial sites is established can vary according to study phase and clinical trial domains. In this study, we employed association rule mining to reveal trial collaboration. We used trial approval data provided by the Ministry of Food and Drug Safety in Korea and organized the trial sites. We collected trial information from 2012 to 2023 and categorized the trials according to study phase and clinical trial domain. We performed association rule mining based on study phase and clinical trial domain. We identified 209 valid trial sites and analyzed 11,107 clinical trials conducted during this period. By study phase, phase 1 trials accounted for the largest number (5,451), followed by phase 3 (2,492), others (1,826), and phase 2 (1,338). We found that phase 1 clinical trials had the highest lift metrics. The mean lift for phase 1 trials was 5.40, which was significantly greater than that of phase 2 (1.68) and phase 3 trials (1.72). Additionally, the network structure for trial collaboration in phase 1 trials was highly condensed, with several trial sites located in Seoul and Gyeonggi-do. Different trial collaboration characteristics were noted among clinical trial domains, with mean and variability of the lift metrics for pediatrics being the highest. In conclusion, association rule mining can identify collaborations among trial sites. Collaboration in phase 1 trials is relatively more exclusive than in other phases, and aspects of collaboration differ among clinical trial domains.

Keywords: Machine Learning; Clinical Trial; Multicenter Study

INTRODUCTION

Collaboration among trial sites plays a key role in multicenter clinical trials. A review of clinical trials registered in *ClinicalTrials.gov* reported that 65.6% of industry-sponsored trials were conducted as multicenter clinical trials [1]. It was noted that results from a single-center trial can be prone to potential biases and tend to overestimate treatment effects compared to multicenter trials [2,3]. Although multicenter trials can provide a concrete evidence for treatment effect, preparing and conducting multicenter trials needs considerable efforts, especially for collaboration among trial sites [3]. Various initiatives have been made to increase the efficiency of multicenter clinical trials [4]. Therefore, research networks for

Author Contributions

Conceptualization: Huh KY, Song I; Data curation: Huh KY; Formal analysis: Huh KY; Funding acquisition: Song I; Investigation: Huh KY; Methodology: Huh KY; Project administration: Huh KY; Resources: Song I; Software: Huh KY; Supervision: Song I; Validation: Song I; Visualization: Huh KY; Writing - original draft: Huh KY; Writing - review & editing: Song I.

multicenter clinical trials have been established in various fields, including oncology [5] and rare diseases [6].

The ways in which collaboration among trial sites is established can vary according to study phase and clinical trial domain. For example, in oncology, international approaches to harmonizing regulatory requirements and establishing collaborative frameworks are highly active [5,7]. In contrast, a social network analysis of pediatric research co-authorships revealed that the research enterprise could be clustered and fragmented [8]. Additionally, early-stage clinical trials have been less accessible to patients than late-stage trials, making them concentrated in several specialized sites [9]. Therefore, it is necessary to evaluate the current status of trial collaboration to establish an optimal collaboration network for multicenter trials.

In our previous research, we comprehensively reviewed trends in clinical trials in Korea. We found that phase 1 clinical trials in Korea are increasing, with the oncology area accounting for a major proportion of these trials [10,11]. We also revealed that most clinical trials are conducted exclusively among trial sites in metropolitan areas [12]. However, the detailed characteristics of trial clusters that can describe the current status of trial collaboration were not evaluated.

Association rule mining is an unsupervised machine learning method used to investigate relationships between variables. This method explores frequent patterns in complex, high-dimensional datasets and can readily detect potential association rules [13]. It has been used in various ways in health informatics [14], such as finding associations between lifestyle, family medical history, and medical abnormalities [15], or identifying risk factors in atherosclerosis [16].

In this study, we employed association rule mining to reveal trial collaboration. We used the trial approval data provided by the Ministry of Food and Drug Safety in Korea and organized the trial sites. We detected potential associations among trial sites, which may imply potential trial collaborations. We analyzed the data according to study phase and clinical trial domain to unravel the detailed trial collaboration status in Korea.

METHODS

Data collection and categorization

Information on the approved clinical trials between 2012 and 2023 was obtained from the public database from the Ministry of the Food and Drug Safety [17]. The data consisted of the following six variables: investigational products (variable name: GOODS_NAME), sponsors (APPLY_ENTP_NAME), approval date (APPROVAL_TIME), trial sites separated by colons (LAB_NAME), study title (CLINIC_EXAM_TITLE), and study phase (CLINIC_STEP_NAME). It included all studies approved by the Ministry of Food and Drug Safety. Study phase was categorized into the following four categories (phase 1, 2, 3, and others) consistently as the previous study [12]. The number of clinical trials was counted once for multi-center clinical trials. For subgroup analysis, the following clinical trial domains were exploratorily extracted from the title of the study: lung cancer, breast cancer, pediatrics, and diabetes. These clinical trial domains were selected because they demonstrate different aspects of collaborations in clinical trials and can be readily extracted from the title using unique keywords.

The names of the trial sites were standardized, and when a trial site was not recorded, the trial was excluded from the analysis. In addition, since non-hospital organizations (i.e., clinical research organizations or laboratories) were included in the variable, we removed these non-hospital organizations from the list of trial sites in the dataset. For example, when trial site information was 'CRO A: Hospital A: Hospital B,' the information was cleaned to 'Hospital A: Hospital B' for analysis. Trial sites other than hospitals (e.g., clinical research organization or laboratories) were also excluded from the analysis. Trial sites were coded using a three-character province code (i.e., Busan, BUS; Chungcheongbuk-do, CHB; Chungcheongnam-do, CHN; Daegu, DAG; Daejeon, DAJ; Gangwon-do, GAN; Gwangju, GWA; Gyeonggi-do, GYE; Gyeongsangbuk-do, GYB; Gyeongsangnam-do, GYN; Incheon, INC; Jeju-do, JEJ; Jeollabuk-do, JEB; Jeollanam-do, JEN; Sejong, SEJ; Seoul, SEO; Ulsan, ULS) and a unique number, generated in alphabetical order and grouped by province. For example, the second trial site in Seoul was coded as 'SEO_2'.

Association rule mining

Also known as "market basket analysis," association rule mining explores frequent patterns and associations found in a dataset. Association rule mining first examines co-occurring patterns in a dataset. For example, in the following three example transaction data: (bread, milk), (bread, butter), and (bread, milk, butter), the itemset (bread, milk) occurs twice in the entire transaction set. In this case, the quality metric *support* is calculated as 2/3, which is defined as the fraction of transactions that contain an itemset. Another metric is *confidence*, which is defined as the percentage of data in an itemset. In mathematical notations, for items A and B, *support*(A→B) is defined as $P(\mathrm{A} \cap \mathrm{B})$, and *confidence*(A→B) is defined as $P(B|A) = \frac{P(A \cap B)}{P(A)}$. *Support* is typically used to identify frequent itemsets in a transaction set. *Confidence* refers to how likely item B is to occur when item A occurs. High support and confidence suggest that the association rule might be significant within the transaction set, and both metrics are commonly used as thresholds to identify association rules. For interestingness measure, *lift* is defined as the ratio of joint occurrence of A and B to the product of marginal occurrence, or $lift(\mathrm{A} \rightarrow \mathrm{B}) = \frac{P(A \cap B)}{P(A)P(B)}$ [13,18]. *Lift* can show the correlation between A and B; when *lift* > 1, it implies positive correlation; when *lift* < 1, negative correlation. A *lift* of 1 shows that item A and B is independent [13].

As inspecting all combinations of association rules is computationally impossible, the *Apriori* algorithm suggested by Agrawal and Srikant was used [19]. The *Apriori* algorithm requires minimum *support* and *confidence* levels for investigating association rules and can minimize the number of evaluations [19]. We set a minimum empirical *support* level of 0.02 and a *confidence* level of 0.05 for investigating association rules across all clinical trials and according to study phase. Considering the smaller number of trials in the analysis by clinical trial domain, we set a higher *support* level of 0.05 to exclude rare cases. We calculated three quality metrics (*support*, *confidence*, and *lift*) for all analyses.

Statistical analysis and visualization

The quality metrics for each association rule mining analysis were summarized descriptively. Lift metrics were compared among study phases and clinical trial domains using the non-parametric Kruskal-Wallis test. A post-hoc analysis was performed using the Dunn test with Bonferroni correction. Support and confidence were not tested, given that the thresholds were set prior to association rule mining.

https://doi.org/10.12793/tcp.2024.32.e17

Quality metrics were plotted on a scatterplot, and normalized distributions were overlaid as data ellipses [20]. Given that the number of association rules is large and arbitrary thresholds for *support* and *confidence* were used to identify the important rules, the top 10 association rules were visualized as graphs using the interestingness measure, *lift*. We demonstrated the representative association rules using a graph where vertices represent trial site codes, while the size and color of the circles linking the vertices indicate the support and lift of the association rule, respectively. Association rule graphs were generated for overall clinical trials as well as by phases and clinical trial domains.

RESULTS

Trial collaboration by study phase

We identified 209 valid trial sites and analyzed 11,107 clinical trials conducted between 2012 and 2023. By study phase, phase 1 trials accounted for the largest number (5,451), followed by phase 3 (2,492), others (1,826), and phase 2 (1,338) (**Table 1**). The overall trial collaboration network was primarily formed within trial sites in metropolitan areas, specifically Seoul, Gyeonggi-do, and Incheon (**Fig. 1**).

We found that phase 1 clinical trials had the highest lift metrics. The mean lift for phase 1 trials was 5.40, which was significantly greater than that of phase 2 (1.68) and phase 3 trials (1.72) (**Table 1**, **Fig. 2**). Additionally, the network structure for trial collaboration in phase 1 trials was highly condensed, with several trial sites located in Seoul and Gyeonggi-do. In contrast, several clusters in phase 3 trials included trial sites outside of metropolitan areas (e.g., Daegu, Daejeon, Gwangju, and Gangwon-do). Lift metrics for phase 2 trials were not statistically different to those from phase 3 trials. In trials classified as "others" (mainly investigator-initiated trials), collaboration was similar to that in phase 1 trials, and high-lift collaborations were found (**Fig. 3**).

Exploratory analysis of trial collaboration by clinical trial domain

Different trial collaboration characteristics were noted among clinical trial domains. Lung cancer and breast cancer showed different trial collaboration statuses; trial collaboration in breast cancer was relatively more clustered and exhibited significantly higher lift metrics. Although the mean lift metrics for pediatrics were the highest, the variability was also the largest. There was no significant difference in trial collaboration between the breast cancer and pediatric domains.

Table 1. Summary of the quality metrics of the association rules

Variables	Study counts	Number of association rules	Support	Confidence	Lift	*p*-value	Post-hoc analysis		
Study phase						< 0.0001			
Overall	11,107	288	0.04 ± 0.02	0.67 ± 0.09	2.74 ± 0.44				
Phase 1	5,451	17	0.03 ± 0.01	0.61 ± 0.07	5.40 ± 1.22		Ref		
Phase 2	1,338	405	0.04 ± 0.04	0.67 ± 0.10	1.68 ± 0.36		< 0.0001	Ref	
Phase 3	2,492	6,552	0.03 ± 0.02	0.68 ± 0.10	1.72 ± 0.45		< 0.0001	1.0000	Ref
Others	1,826	335	0.03 ± 0.01	0.71 ± 0.12	3.64 ± 1.28		1.0000	< 0.0001	< 0.0001
Selected clinical trial domains						< 0.0001			
Lung cancer	453	262	0.09 ± 0.05	0.69 ± 0.11	1.64 ± 0.37		Ref		
Breast cancer	270	4,171	0.08 ± 0.05	0.84 ± 0.14	1.93 ± 0.69		< 0.0001	Ref	
Pediatrics	241	28	0.10 ± 0.05	0.66 ± 0.11	2.92 ± 3.46		0.6003	1.0000	Ref
Diabetes	247	2,295	0.07 ± 0.02	0.65 ± 0.11	2.00 ± 0.56		< 0.0001	< 0.0001	0.0288

Mean and standard deviations are presented for support, confidence, and lift. Overall *p*-values are calculated using the Kruskal-Wallis test, and *p*-values from the post-hoc Dunn test with Bonferroni correction are presented.

https://doi.org/10.12793/tcp.2024.32.e17

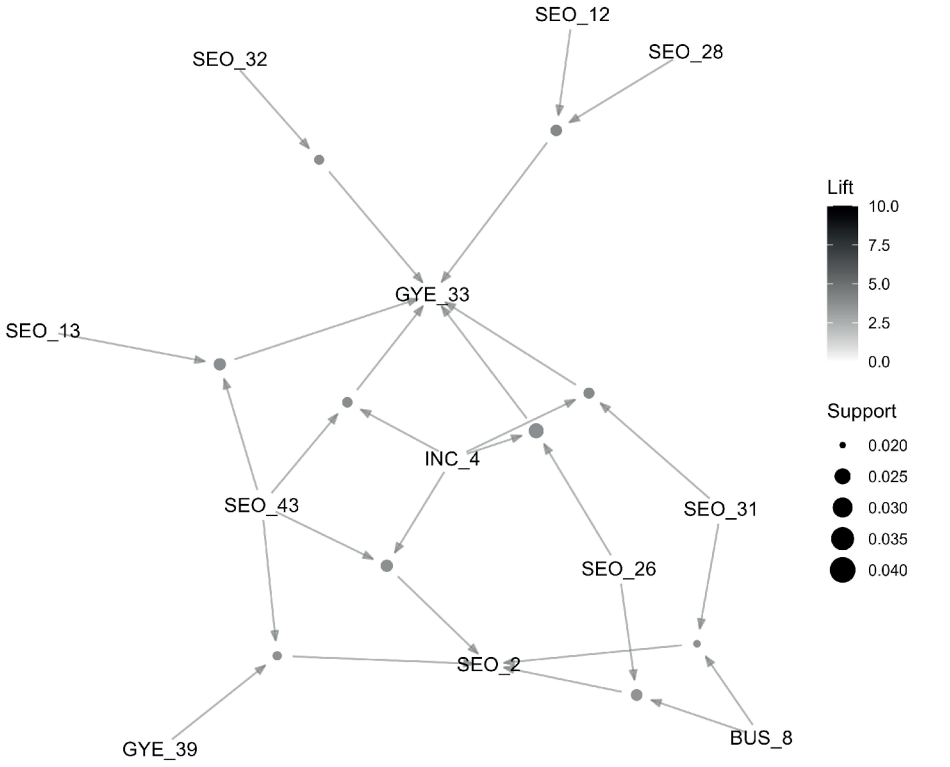

Figure 1. Association rule graph for total clinical trials. The top 10 association rules ranked by lift are presented. Vertices denote trial site codes, while the size and color of the circles linking the vertices denote support and lift of the association rule, respectively. Trial sites were coded using a three-character province code (i.e., Busan, BUS; Chungcheongbuk-do, CHB; Chungcheongnam-do, CHN; Daegu, DAG; Daejeon, DAJ; Gangwon-do, GAN; Gwangju, GWA; Gyeonggi-do, GYE; Gyeongsangbuk-do, GYB; Gyeongsangnam-do, GYN; Incheon, INC; Jeju-do, JEJ; Jeollabuk-do, JEB; Jeollanam-do, JEN; Sejong, SEJ; Seoul, SEO; Ulsan, ULS) and a unique number, generated in alphabetical order and grouped by province.

The diabetes domain had significantly different trial collaboration compared to the other domains, as indicated by the significantly different lift metrics. (**Table 1**, **Fig. 4**).

DISCUSSION

The results of our study showed that there were differences in trial collaboration in Korea according to study phase and clinical trial domain. Phase 1 clinical trials are exclusively conducted among a group of trial sites. Trials in other phases are more interconnected; however, several clusters of trial sites were identified. Collaboration in trials is also significantly different across trial domains.

As demonstrated in our previous study, there have been concerns regarding the geographic accessibility of clinical trials worldwide. Disparities in geographic accessibility have been reported in various studies in the United States [21]. In particular, early phase trials are vulnerable to potential selection biases due to limited accessibility to trial sites [22]. The phenomenon may be attributed to the complex features of the study operation, and trial sites are limited in large centers located in metropolitan areas [23]. A similar phenomenon was noted in pediatric clinical trials, where trial sites were clustered and fragmented [8].

https://doi.org/10.12793/tcp.2024.32.e17

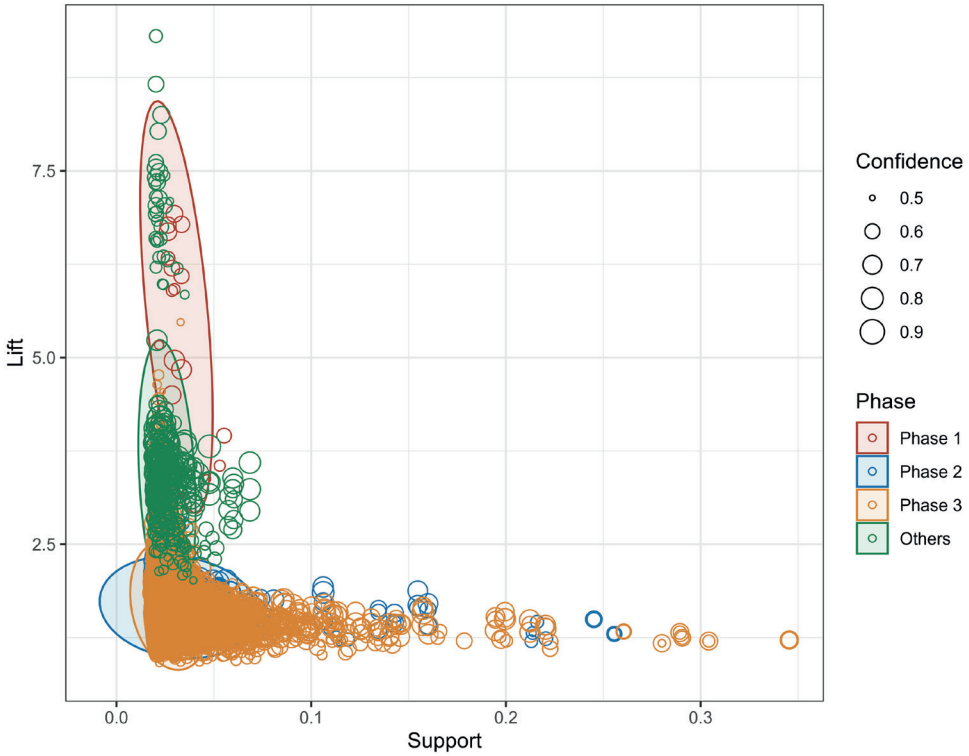

Figure 2. Distribution of association rule quality metrics by study phase. The size of the circles denotes confidence, and the colors indicate the phase. Normalized distributions are represented as ellipses.

To facilitate collaboration among trial sites, it is essential to understand the current status of collaboration. Social network analysis has been used to analyze collaboration networks [24]. This analysis can evaluate the overall structure of networks using quantitative measures and easily visualize potential collaborations. When combined with social network analysis, association rule mining can effectively identify collaboration groups and provide a straightforward association between specific parties.

However, the drawbacks of association rule mining should be carefully considered. It is highly sensitive to initial thresholds (i.e., minimum support or confidence), and important associations could be omitted. There have been attempts to find these omitted but important associations using unsupervised learning methods, such as k-means clustering [25]. Additionally, identified association rules cannot be easily interpreted without expertise in the relevant domain [26]. Nonetheless, association rule mining can be valuable for exploring potential associations in large datasets.

We assume that fragmented collaboration in clinical trials may be a potential source of bias in multi-center trials. Multi-center trials are typically considered to reduce the biases present in single-center trials and to enhance external validity [27,28]. However, our study reveals that preferences in selecting trial sites could introduce new biases. Additionally, this could necessitate a 'trial cluster-level' analysis of treatment effects to ensure external validity.

Of note, analyzing the current status of trial collaboration is essential when planning a clinical trial with decentralized elements. Clinical trials with decentralized components involve various stakeholders and service providers outside traditional trial sites [29]. The trial

https://doi.org/10.12793/tcp.2024.32.e17

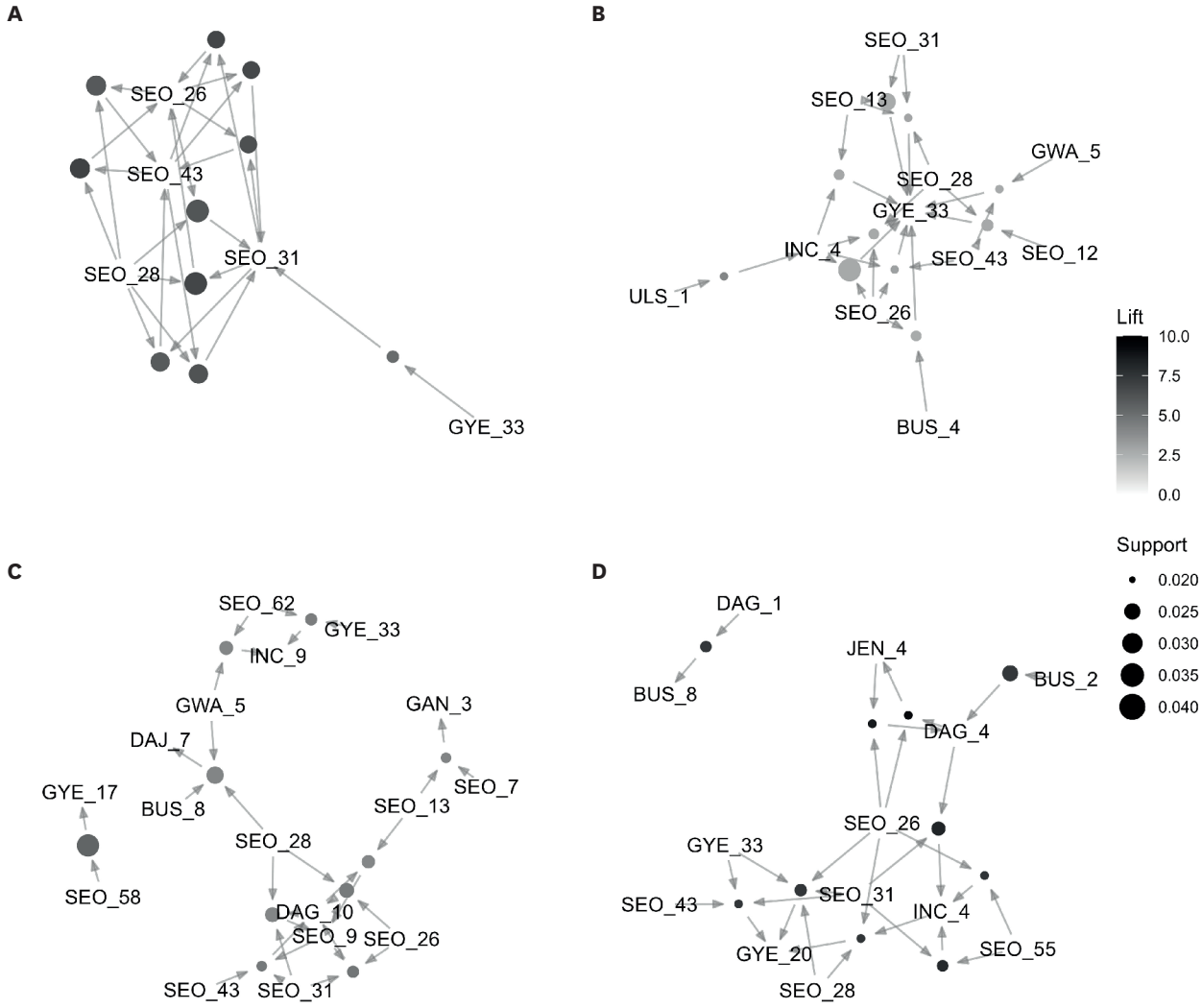

Figure 3. Association rule graphs by study phase: phase 1 (A), phase 2 (B), phase 3 (C), and others (D). The top 10 association rules ranked by lift are presented. Vertices denote trial site codes, while the size and color of the circles linking the vertices denote support and lift of the association rule, respectively. Trial sites were coded using a three-character province code (i.e., Busan, BUS; Chungcheongbuk-do, CHB; Chungcheongnam-do, CHN; Daegu, DAG; Daejeon, DAJ; Gangwon-do, GAN; Gwangju, GWA; Gyeonggi-do, GYE; Gyeongsangbuk-do, GYB; Gyeongsangnam-do, GYN; Incheon, INC; Jeju-do, JEJ; Jeollabuk-do, JEB; Jeollanam-do, JEN; Sejong, SEJ; Seoul, SEO; Ulsan, ULS) and a unique number, generated in alphabetical order and grouped by province.

design must facilitate coordination with central coordinating centers and network hospitals. In particular, decentralized elements in oncology are relatively under-implemented, despite the increasing demand for better accessibility [30]. Therefore, understanding the current status of trial collaboration could help in designing and adopting decentralized elements in future trials.

Our study has several limitations. We only analyzed trial sites and did not include principal investigators, which may limit the interpretability of the results. Moreover, association rules are not causative; thus, the identified collaborations should be interpreted alongside other evidence of collaboration. Classifying the clinical trial domain of a trial using keywords may not fully reflect the characteristics of the trials. Further investigation in global registries and comparisons with other countries, such as *ClinicalTrials.gov*, would be required.

https://doi.org/10.12793/tcp.2024.32.e17

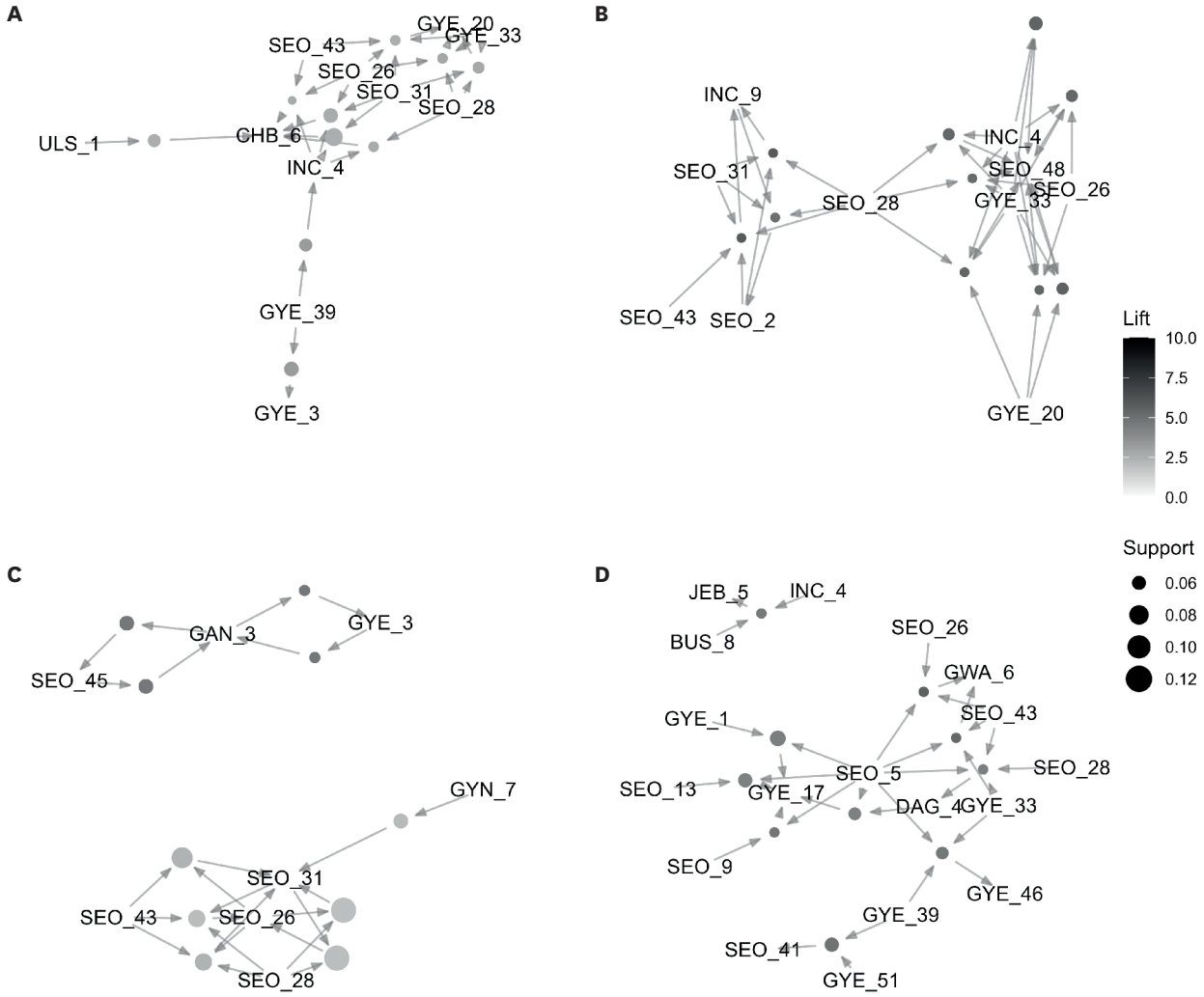

Figure 4. Association rule graphs by clinical trial domain: lung cancer (A), breast cancer (B), pediatrics (C), and diabetes (D). The top 10 association rules ranked by lift are presented. Vertices denote trial site codes, while the size and color of the circles linking the vertices denote support and lift of the association rule, respectively. Trial sites were coded using a three-character province code (i.e., Busan, BUS; Chungcheongbuk-do, CHB; Chungcheongnam-do, CHN; Daegu, DAG; Daejeon, DAJ; Gangwon-do, GAN; Gwangju, GWA; Gyeonggi-do, GYE; Gyeongsangbuk-do, GYB; Gyeongsangnam-do, GYN; Incheon, INC; Jeju-do, JEJ; Jeollabuk-do, JEB; Jeollanam-do, JEN; Sejong, SEJ; Seoul, SEO; Ulsan, ULS) and a unique number, generated in alphabetical order and grouped by province.

In conclusion, association rule mining can identify collaborations among trial sites. Collaboration in phase 1 trials is relatively more exclusive than in other phases, and aspects of collaboration differ among clinical trial domains.

REFERENCES

1. Gresham G, Meinert JL, Gresham AG, Meinert CL. Assessment of trends in the design, accrual, and completion of trials registered in ClinicalTrials.gov by sponsor type, 2000-2019. JAMA Netw Open 2020;3:e2014682. PUBMED | CROSSREF
2. Unverzagt S, Prondzinsky R, Peinemann F. Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review. J Clin Epidemiol 2013;66:1271-1280. PUBMED | CROSSREF
3. Chung KC, Song JW; WRIST Study Group. A guide to organizing a multicenter clinical trial. Plast Reconstr Surg 2010;126:515-523. PUBMED | CROSSREF

4. Harris PA, Dunsmore SE, Atkinson JC, Benjamin DK Jr, Bernard GR, Dean JM, et al. Leveraging the expertise of the CTSA program to increase the impact and efficiency of clinical trials. JAMA Netw Open 2023;6:e2336470. PUBMED | CROSSREF
5. Tang M, Joensuu H, Simes RJ, Price TJ, Yip S, Hague W, et al. Challenges of international oncology trial collaboration-a call to action. Br J Cancer 2019;121:515-521. PUBMED | CROSSREF
6. Marina N, Bielack S, Whelan J, Smeland S, Krailo M, Sydes MR, et al. International collaboration is feasible in trials for rare conditions: the EURAMOS experience. Cancer Treat Res 2009;152:339-353. PUBMED | CROSSREF
7. Kataoka K, Nakamura K, Caballero C, Evrard S, Negrouk A, Shiozawa M, et al. Collaboration between EORTC and JCOG-how to accelerate global clinical research partnership. Jpn J Clin Oncol 2017;47:164-169. PUBMED | CROSSREF
8. Duffett M, Brouwers M, Meade MO, Xu GM, Cook DJ. Research collaboration in pediatric critical care randomized controlled trials: a social network analysis of coauthorship. Pediatr Crit Care Med 2020;21:12-20. PUBMED | CROSSREF
9. Villalona-Calero MA, Malhotra J, Chung V, Xing Y, Gray SW, Hampel H, et al. Integrating early-stage drug development with clinical networks; challenges and opportunities: the city of hope developing experience. J Clin Med 2023;12:4061. PUBMED | CROSSREF
10. Huh KY, Hwang JG, Lee S. Trends of clinical trials from 2014 to 2016 in South Korea. Transl Clin Pharmacol 2018;26:172-176. PUBMED | CROSSREF
11. Huh KY, Yu KS, Lim HS, Kim H. Trends of clinical trials from 2017 to 2019 in Korea: an integrated analysis based on the Ministry of Food and Drug Safety (MFDS) and the Clinical Research Information Service (CRIS) registries. Transl Clin Pharmacol 2021;29:186-196. PUBMED | CROSSREF
12. Huh KY, Yu KS, Song I. Analysis of the distribution of trial sites in South Korea using social network analysis. Transl Clin Pharmacol 2023;31:1-12. PUBMED | CROSSREF
13. Vu K, Clark RA, Bellinger C, Erickson G, Osornio-Vargas A, Zaïane OR, et al. The index lift in data mining has a close relationship with the association measure relative risk in epidemiological studies. BMC Med Inform Decis Mak 2019;19:112. PUBMED | CROSSREF
14. Altaf W, Shahbaz M, Guergachi A. Applications of association rule mining in health informatics: a survey. Artif Intell Rev 2017;47:313-340. CROSSREF
15. Ogasawara M, Sugimori H, Iida Y, Yoshida K. Analysis between lifestyle, family medical history and medical abnormalities using data mining method–association rule analysis. In: Khosla R, Howlett RJ, Jain LC (eds). Knowledge-based intelligent information and engineering systems. KES 2005. Berlin: Springer; 2005.
16. Berka P, Rauch J. Mining and post-processing of association rules in the atherosclerosis risk domain. In: Khuri S, Lhotská L, Pisanti N (eds). Information technology in bio- and medical informatics, ITBAM 2010. Berlin: Springer, 2010.
17. Ministry of Food and Drug Safety. Public database for drug clinical trials, https://data.mfds.go.kr/OPCAC01P01?opdataSeq=44&startYear=&endYear=&callBackFn=fnPopupCallback. Accessed October 14, 2024.
18. Park HC. Association rule ranking function by decreased lift influence. J Korean Data Inf Sci Soc 2010;21:397-405.
19. Agrawal R, Srikant R. Fast algorithms for mining association rules. https://www.cse.msu.edu/~cse960/Papers/MiningAssoc-AgrawalAS-VLDB94.pdf. Accessed October 14, 2024.
20. Friendly M, Monette G, Fox J. Elliptical insights: understanding statistical methods through elliptical geometry. Stat Sci 2013;28:1-39. CROSSREF
21. Galsky MD, Stensland KD, McBride RB, Latif A, Moshier E, Oh WK, et al. Geographic accessibility to clinical trials for advanced cancer in the United States. JAMA Intern Med 2015;175:293-295. PUBMED | CROSSREF
22. Perni S, Moy B, Nipp RD. Disparities in phase 1 cancer clinical trial enrollment. Cancer 2021;127:4464-4469. PUBMED | CROSSREF
23. Baranda JC, Diaz FJ, Rubinstein L, Shields AF, Dayyani F, Mehta A, et al. Expanding access to early phase trials: the CATCH-UP.2020 experience. JNCI Cancer Spectr 2023;7:pkac087. PUBMED | CROSSREF
24. Lin G, Siddiqui S, Bernstein J, Martinez DA, Gardner L, Albright T, et al. Examining association between cohesion and diversity in collaboration networks of pharmaceutical clinical trials with drug approvals. J Am Med Inform Assoc 2021;28:62-70. PUBMED | CROSSREF
25. Lee KW, Park HC. Application of k-means clustering for association rule using measure of association. J Korean Data Inf Sci Soc 2008;19:925-936.

26. Mansingh G, Osei-Bryson KM, Reichgelt H. Using ontologies to facilitate post-processing of association rules by domain experts. Inf Sci 2011;181:419-434. CROSSREF

27. Bellomo R, Warrillow SJ, Reade MC. Why we should be wary of single-center trials. Crit Care Med 2009;37:3114-3119. PUBMED | CROSSREF

28. Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. Ann Intern Med 2011;155:39-51. PUBMED | CROSSREF

29. de Jong AJ, van Rijssel TI, Zuidgeest MGP, van Thiel GJMW, Askin S, Fons-Martínez J, et al. Opportunities and challenges for decentralized clinical trials: European regulators' perspective. Clin Pharmacol Ther 2022;112:344-352. PUBMED | CROSSREF

30. Fu S, Gerber DE, Beg MS. Decentralized clinical trials in oncology: are we ready for a virtual-first paradigm? J Clin Oncol 2023;41:181-185. PUBMED | CROSSREF