医薬品副作用データベースを用いる際のデータクリーニングの必要性の初期検討

西島 信明, 東 華奈子, 荒屋 さやか, 伊澤 和俊, 井上 和美, 草刈 貴子,
地下 春佳, 須賀 沙弥香, 鈴木 有未, 髙木 毅, 高杉 安彦, 村上 育美,
渡邊 舞, 川上 淳二, 瀬戸 極, 北島 行雄*

Preliminary Study on the Need for Data Cleaning in PMDA's Drug Adverse Event Report Database

Nobuaki NISHIJIMA, Kanako AZUMA, Sayaka ARAYA, Kazutoshi IZAWA,
Kazumi INOUE, Takako KUSAKARI, Haruka JIGE, Sayaka SUGA, Yumi SUZUKI,
Takeshi TAKAGI, Yasuhiko TAKASUGI, Ikumi MURAKAMI, Mai WATANABE,
Junji KAWAKAMI, Kiwamu SETO and Yukio KITAJIMA*

Summary

Data from the Japanese Adverse Drug Event Report database (JADER) became available for download through the website of the Pharmaceuticals and Medical Devices Agency (PMDA), Japan, in April 2012. Analysis of this database, e.g., by comparing a target drug with other similar marketed drugs, should be useful to investigate and assess the relatedness of detected adverse events (AEs).

However, it would be possible to miss a signal if duplicate reports are present in the database or if data notation is not standardized, and therefore data cleaning might be required to minimize such issues. We therefore considered several factors that may affect signal detection, using 3 tables (Case List Table, Drug Information Table, and Adverse Drug Reaction Table) in JADER, and investigated whether data cleaning is required in each case.

In the Case List Table, we focused on duplicate reports from 2 separate data sources: medical institutions and pharmaceutical companies. In the Drug Information Table, we focused on differences in drug notation: one-byte and two-byte characters, lower- and upper-case letters, with and without hyphenation, and with and without spacing, as well as drug names in Japanese and English. In the Adverse Drug Reaction Table, we focused on recurrence of AEs within a short time period, using an anticancer drug as an example.

Our study showed that the result of signal detection could change when one-byte character and two-byte character drug names were unified. Also, unification of drug names to Japanese or English may be required to improve signal detection, at least for some drugs.

Our findings indicate that care is necessary while using JADER. We suggest that the utility of the database could be increased by data cleaning and unification.