APAMI2020 Poster Presentation Sessions | APAMI 2020 | Poster Presentation Sessions

Artificial Intelligence

Sun. Nov 22, 2020 3:00 PM - 4:00 PM Room E-2 (Congress center 5F - Conference Room 53)

[AP2-E2-4-05] An Electronic Phenotyping Algorithm to Identify Cases of Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis in the MID-NET Database

*Rieko Izukura[1], Takeshi Nakahara[3], Takamichi Ito[3], Chinatsu Nojiri[2], Tadashi Kandabashi[2], Takanori Yamashita[2], Atsushi Takada[2], Yoshiaki Uyama[4], Naoki Nakashima[2] （1. Social Medicine, Department of Basic Medicine, Faculty of Medical Sciences, Kyushu University, Japan, 2. Medical Information Center, Kyushu University Hospital, Japan, 3. Department of Dermatology, Graduate School of Medical Sciences, Kyushu University, Japan, 4. Pharmaceuticals and Medical Devices Agency, Japan）

Keywords: Phenotyping Algorithm, Adverse Effects, Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis

Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) is a well-known drug-induced adverse effect. We developed a phenotyping algorithm to identify SJS/TEN with high positive predictive value and sensitivity. The initial algorithm was modified by adding the principal variables based on the gradient boosting decision tree and clinical perspectives. Consequently, it was improved to the phenotyping algorithm to identify SJS/TEN with an increased positive predictive value from 8.5% to 76.5% and a sensitivity of 76.5%. It is necessary to further refine this algorithm by evaluating its robustness in other hospitals.

An Electronic Phenotyping Algorithm to Identify Cases of Stevens-Johnson Syndrome /Toxic Epidermal Necrolysis in the MID-NET Database

Rieko Izukura[a], Takeshi Nakahara[c], Takamichi Ito[c], Chinatsu Nojiri[b], Tadashi Kandabashi[b], Takanori Yamashita[b], Atsushi Takada[b], Yoshiaki Uyama[d] and **Naoki Nakashima[b]**

[a] *Social Medicine, Department of Basic Medicine, Faculty of Medical Sciences, Kyushu University, Japan*
[b] *Medical Information Center, Kyushu University Hospital, Japan*
[c] *Department of Dermatology, Graduate School of Medical Sciences, Kyushu University, Japan*
[d] *Pharmaceuticals and Medical Devices Agency, Japan*

Abstract

Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) is a well-known drug-induced adverse effect. We developed a phenotyping algorithm to identify SJS/TEN with high positive predictive value and sensitivity. The initial algorithm was modified by adding the principal variables based on the gradient boosting decision tree and clinical perspectives. Consequently, it was improved to the phenotyping algorithm to identify SJS/TEN with an increased positive predictive value from 8.5% to 76.5% and a sensitivity of 76.5%. It is necessary to further refine this algorithm via an evaluation of its robustness in other hospitals.

Keywords:
Phenotyping Algorithm, Adverse Effects, Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis

Introduction

The Medical Information Database Network (MID-NET) that is composed of data on administrative claims, Diagnosis Procedure Combination (DPC) and the structured electronic medical record (s-EMR) including laboratory data values, is newly available in Japan [1]. The MID-NET was constructed as a secondary resource for analyzing various health outcomes related to pharmaceutical products and contains the clinical information of approximately 5.05 million people that can be used for research purposes. In data-driven or retrospective studies, accurate identification of targeted cases from the database is crucial, wherein a phenotyping algorithm constitutes an essential tool.

Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) is a well-known drug-induced adverse effect with an incidence of 1–2 cases per million [2]. We developed a phenotyping algorithm to identify SJS/TEN cases in the MID-NET using machine learning (ML) techniques.

Materials and Methods

We retrospectively analyzed 304,191 cases at Kyushu University Hospital from April 1, 2009 to March 31, 2018.

Development of the phenotyping algorithm of SJS/TEN

1. Creation of an initial phenotyping algorithm

- Data source: ICD10 code[*1] of e-EMR and DPC
- Index date: EMR: Diagnozed date, DPC: admission date

[*1]ICD10: International Classification of Diseases, 10th revision codes

(SJS [L511]) or (TEN [L512])
or (Similar disease[*2] [L00, L080, L270, L271, L510, L519])
[*2] e.g.) Drug-induced eruption

2. Data extraction in the MID-NET and random sampling

Total 200 cases were randomly selected from all possible cases that were extracted by the initial phenotyping algorithm.

3. Review by two physicians per case

Two dermatologists individually reviewed the 200 cases; thereafter, they re-evaluated the cases regarding which there was disagreement.

4. Calculation of weighted kappa coefficient (κ-coefficient) and positive predictive value (PPV)

We calculated the κ-coefficient as a summary of the coincidence degree between each pair of physicians and the PPV.

5. Modification of [SJS/TEN algorithm A] using the gradient boosting decision tree (GBDT)

A predictive model for SJS/TEN was estimated referring to the largest area under the curve (AUC) [3], the PPV, or the sensitivity within the 200 cases (the deemed sensitivity).

6. Estimation of all possible cases and recalculation of sensitivity

We recalculated the sensitivity as per the following assumption: The number of true cases that are identified using the initial algorithm is a real number of SJS/TEN cases.

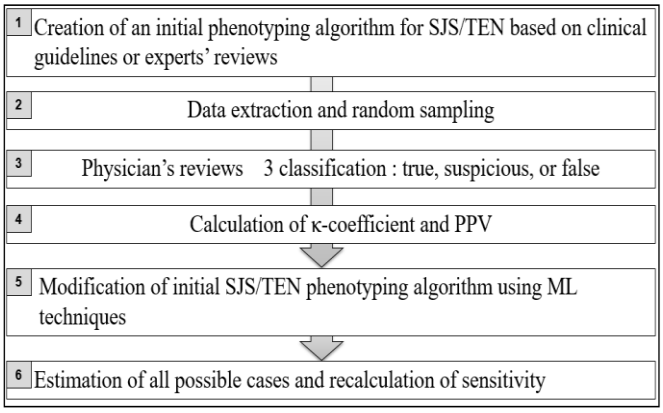

Figure 1- Workflow of phenotyping algorithm development

Results

Data extraction and random sampling

Total 357 cases were detected using the initial SJS/TEN algorithm. Of these, 200 cases were randomly selected for the review of experts.

Expert review and calculation

The κ-coefficient was 0.77 (95% confidence interval: 0.65–0.89). Total 17 cases were identified as true, with a PPV of 8.5%. SJS/TEN did not exist among patient who were diagnosed with similar diseases. In the next step, the diagnostic code of similar diseases was deleted from the algorithm.

Modification of the initial SJS/TEN phenotyping algorithm

Total 1880 explanatory variables for the GBDT were extracted from the 200 cases. These consisted of the following structured data codes or values in the MID-NET: disease diagnosis, pharmaceutical products, laboratory test values, surgery implementation, medical intervention in Japanese administrative and claims, and patient's age and gender (Figure 2).

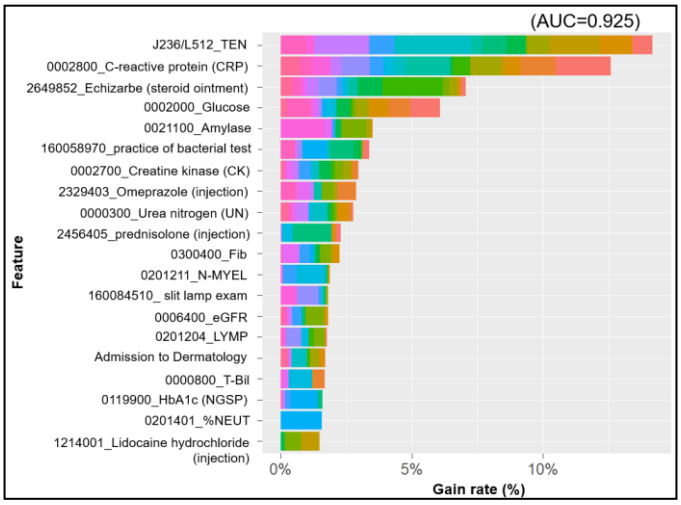

Figure 2- Results of the GBDT
1880 explanatory variables, true: n = 17, false: n = 183

As per the GBDT, C-reactive protein (CRP) was the most contributive variable (AUC = 0.925). However, it cannot be denied that the elevated CRP level was affected by other comorbidities. Moreover, two (glucose and amylase) of the top five principal variables could not clinically explain SJS/TEN. Eventually, based on the expert's perspective, we excluded CRP and included medical intervention and pharmaceutical products as shown in Table 1. The PPV and sensitivity of the modified algorithm were 76.5% and 76.5%, respectively (F value = 0.76).

Table 1- Final variables for the SJS-TEN phenotyping algorithm and their detection rate

Data resource	*Disease diagnosis*	*Pharmaceutical products*			*Medical intervention in administrative claims*
Variables	SJS: L511 TEN: L512	Steroid products	Immuno-globulin	Antibiotic ophthalmic agents	Plasmapheresis
Target (n)	n (%)	n (%)	n (%)	n (%)	n (%)
True-positive n = 17	17 (100)	16 (94.1)	4 (23.5)	16 (94.1)	3 (17.6)
Negative n = 183	35 (19.1)	66 (36.1)	0 (0.0)	26 (14.2)	0 (0.0)

Detection rate: (case applied to variable/targeted case) ×100

The modified phenotyping algorithm is shown as follows:

(SJS [L511]) or (TEN [L512]) and (≥2 variables in below)

Diagnostic code		SJS: L511, TEN: L512
Steroid	injection:	First 3 digits of YJ7*: 245
	Eye drop:	First 4 digits of YJ7:1315
	Ointment:	YJ7: 2649852
Immunoglobulin		First 4 digits of YJ7:1319
Antibiotic ophthalmic agents		First 4 digits of YJ7:6343
Plasmapheresis		administrative claims: 140008210

PPV = 76.5%, sensitivity = 76.5%. F value = 0.76

*YJ7: Japanese drug code consists of 7 digits

Estimation of all possible cases and recalculation of the sensitivity

Twenty-three SJS/TEN cases were estimated in the hospital. The sensitivity was also recalculated as 76.5%.

Discussion

Compared to TEN with severe skin conditions, the cases extracted based on the SJS diagnostic code could not be evaluated accurately owing to the lack of clinical information. This could be one of the reasons that the features of SJS/TEN were not enough described in the GBDT. SJS/TEN diagnosis mainly depends on the pathological results and clinical presentations; therefore, it is difficult to filter out all the SJS/TEN cases using information regarding medication used for SJS/TEN. If the database contained the clinical details, it would enable further improvement in the quality of the SJS/TEN algorithm. Owing to the relatively small sample size, it was challenging to ultimately validate the reproducibility of the sensitivity for SJS/TEN algorithm. We recommend that it is verified using different data sets at other hospitals.

Conclusions

We developed the SJS/TEN phenotyping algorithm using the ML technique. This algorithm should be used depending on the study purpose as it was created with small sample sizes. It is expected that the robustness of this algorithm is examined using data sets in other hospitals.

Acknowledgments and Ethics

This study was supported by the Japan Agency for Medical Research and Development Grant Number 17mk0101088h0001. This study was approved by the ethics review committee of Kyushu University (No. 30-423).

References

[1] Pharmaceuticals and Medical Devices Agency; Aavailable from: *https://www.pmda.go.jp/safety/mid-net/0001.html*

[2] Anthony P, *et al.* Treatment of Toxic Epidermal Necrolysis by a Multidisciplinary Team. A Review of Literature and Treatment Results. *Burns.* 2018; 44(4):807-15.

[3] Tianqi C, Carlos G. Xgboost: A scalable tree boosting system. Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016; 785-94.

Address for correspondence

Rieko Izukura, RN, PhD.
Medical Information Center, Kyushu University Hospital
3-1-1, Maidashi, Higashi-ku, Fukuoka, Japan
E-mail: ochamame0830@med.kyushu-u.ac.jp