**Supplemental Material**

Appendix 1. Detailed description of the EMR database from the University of Tokyo Hospital

Appendix 2. Detailed description of the definition of observation periods for each patient

Appendix 3. Definition of the occurrence of liver injury

Appendix 4. Detailed description of the candidate covariates for the high-dimensional propensity score models

Appendix 5. Detailed algorithm for the construction of the high-dimensional propensity score

Appendix 6. Stabilized inverse probability of treatment weight

Appendix 7. Stabilized inverse probability of censoring weight

Supplementary Figure 1. Schematic example of the definitions of observation periods in a hypothetical patient

Supplementary Figure 2. The distribution of the estimated high-dimensional propensity scores for the treated and untreated episodes in the pooled cohorts for macrolides, penicillin-based antibiotics, and fluoroquinolones

Supplementary Figure 3. The cumulative incidence of liver injury per 1,000 treated and untreated episodes during the 30-day follow-up period after antibiotic prescription

Supplementary Table 1. Number of users for each target antibiotic

Supplementary Table 2. Sensitivity analysis results using unweighted models that adjusted for the deciles of hdPS

Supplementary Table 3. Sensitivity analysis results of gradual decreases of the covariates’ coefficients toward zero in the hdPS model

# **Appendix 1. Detailed description of the EMR database from the University of Tokyo Hospital**

The database included information on patient characteristics (anonymized personal identifiers, age, and sex); dates of hospital visits, admission, and discharge; prescribed medications; diagnoses; and laboratory test results. Medications were coded according to National Health Insurance Drug Codes provided by Japan’s Ministry of Health, Labour and Welfare. Diagnoses were recorded and identified using International Classification of Diseases, 10th revision (ICD-10) codes.

# **Appendix 2. Detailed description of the definition of observation periods for each patient**

Multiple hospitalizations or outpatient visits within 100 days were considered a single observation period for each patient. If the interval between two hospitalizations or outpatient visits exceeded 100 days, we treated the intervals as unobserved periods. It should be noted that each patient could contribute multiple observation periods. A schematic example of these observation periods is presented in Supplementary Figure 1. If the duration of a patient’s hospitalization extended beyond the study period, the duration was truncated before January 1, 2011 or after December 31, 2015, where appropriate.

# **Appendix 3. Definition of the occurrence of liver injury**

The occurrence of liver injury was defined as either (1) an increase of 200% or more of the upper limit of the normal range of alanine aminotransferase or conjugated bilirubin; or (2) a same-day combined increase in aspartate aminotransferase, alkaline phosphatase, and total bilirubin provided that one of these values showed an increase of 200% or more of the upper limit of the normal range according to the criteria specified in an international consensus meeting [1].

# **Appendix 4. Detailed description of the candidate covariates for the high-dimensional propensity score models**

We examined two dimensions of candidate covariates: concomitant medications and medical history. Concomitant medications were identified as drugs with a prescription period that overlapped the index date of a target episode. These drugs were categorized according to the first four digits of their National Health Insurance Drug Codes (692 variables), which designate a specific therapeutic class. As the prescription periods for ointments, transdermal patches, eye drops, and “pill-in-the-pocket” medications could not be identified, we assigned a single day as the prescription duration for these medications. Moreover, we excluded target antibiotics from the candidate concomitant medications if their class effects were being estimated. The other covariate dimension of medical history was defined as diseases or medical conditions that had been diagnosed within 90 days before the index date of each episode. These were identified using the first three digits of the International Classification of Diseases, 10th revision codes (2,039 variables).

# **Appendix 5. Detailed algorithm for the construction of the high-dimensional propensity score**

**Ranking of candidate covariates**

In this study, the 200 candidate covariates were ranked based on their relative hazard for liver injury (LI) occurrence and the prevalence of each variable in the exposed and unexposed episodes using the method describe in Schneeweiss et al. [2]. The procedure was conducted as follows:

***Step 1:*** For each candidate covariate, the hazard ratio (HR) for 30-day LI occurrence was calculated using a univariate Cox model. Patients were censored after the passage of 29 days from the index date or the observation end date, whichever was earlier. Similar to the estimation of the intention-to-treat effect, changes in treatment status within 30 days were ignored.

***Step 2:***Using the formula proposed by Bross [3], the score of each candidate covariate was calculated based on its potential to bias the relationship between exposure (treatment) and event. The scores were obtained from the following equations:





where *P*C1 and *P*C0 represent the prevalence of each candidate covariate in the exposed and unexposed episodes, respectively. Variables with a higher score have a higher potential of introducing bias into the exposure-event relationship.

***Step 3:*** Each candidate covariate was ranked in descending order of the score values.

**Logistic regression models for estimating the high-dimensional propensity score**

To address the issues of separability in a model that includes a large number of covariates, we used Firth’s penalized likelihood logistic regression model to estimate the high-dimensional propensity score.

# **Appendix 6. Stabilized inverse probability of treatment weight**

Let *A* and *L* denote an exposure (*A* = 1 if exposed, 0 if unexposed) and the set of baseline covariates, respectively; we observe (*Ai*, *Li*) for each patient *i*. The stabilized inverse probability of treatment weight *SWiA* for a patient *i* was defined as

,

where *fA*(*a*) is probability of *A* taking a value *a* and *fA|L*(*a|l*) is conditional probability of *A* = *a* given *L* = *l* [4]. The denominator of this expression calculates the probability that the patients (episodes) received their own treatment (exposed or unexposed) given their baseline covariates. Because the propensity score (PS) was defined in this study as the estimated probability of treatment that is conditional on baseline covariates, the denominators correspond with PS or 1−PS for the exposed or unexposed episodes, respectively. The numerator of the weight, which is the stabilizing term, is the proportion of exposed or unexposed episodes to all episodes in each pooled cohort.

# **Appendix 7. Stabilized inverse probability of censoring weight**

The stabilized inverse probability of censoring weight *SWtC* of follow-up day *t* (ranging from 1 to 30, where *t* = 1 is the index date) for each episode was defined as:

 ,

where *Ct* indicates if there is a treatment change on day *t* (*Ct*= 1 if treatment status changed on day *t*; *Ct*= 0 if otherwise), *Yt* indicates if there is an event occurrence on day *t* (*Yt*= 1 if LI occurred on day *t*; *Yt*= 0 if otherwise), *Et* indicates if the observation period ends on day *t* (*Et*= 1 if the observation period ended on day *t*; *Et*= 0 if otherwise), *A* indicates if an episode is exposed (*A* = 1) or unexposed (*A* = 0), and *L* represents the baseline covariates for each episode [4]. *C*0 had a value of zero for all episodes. Since the mechanism of treatment change may differ between the exposed (change to unexposed) and unexposed (change to exposed) episodes, separate models were used to calculate the weights for each comparison group. Therefore, the condition stipulated by *A* was not needed in all models. We assumed that the temporal ordering of each measurement was *Yt*, *Et*, and *Ct* within each day *t* of observation, and that these indicators never returned to 0 after attaining a value of 1. The weight *SWtC* was constructed as follows:

***Step 1:*** For each day *t* rangingfrom days 2 to 30, we extracted episodes that were under observation and had not yet experienced an event before day *t*−1 (i.e., *Yt*−1 = 0) and treatment change before day *t*−2 (i.e., *Ct*−2 = 0). The episodes where the observation end date was day *t*−1 (i.e., *Et*−1 = 1) were then deleted from the datasets of day *t* because we had defined the change of treatment to occur after event onset or end of observation. Extracted episodes were examined using the day variable *T* = *t* and the indicator of treatment changes. We aggregated these 29 datasets (one for each day from days 2–30) as a pooled dataset for the calculation of inverse probability of censoring weight. Note that the dataset for follow-up day 1 (index date) was not included in the pooled dataset because the probability of treatment changes for all episodes on the index date is zero. These pooled datasets were separately constructed for each group of episodes (exposed or unexposed).

***Step 2:*** For each pooled dataset, the denominator of *SWtC* for each follow-up day of each episode was calculated using a Firth’s penalized likelihood logistic regression model to estimate the probability that there was no change in treatment status. This model included 100 covariates selected for the calculation of PS, patient age, sex, year, hospitalization status on the index date, and day variable *T*. The numerator of *SWtC* (stabilized term) for each follow-up day of each episode was also calculated using a Firth’s penalized likelihood logistic regression model for estimating the probability that there was no change in treatment status; this model only included the day variable *T* as a covariate.

***Step 3*:** The weight *SWtC* for each follow-up day *t* of each episode was calculated based on the terms described in *Step 2*.

# **References**

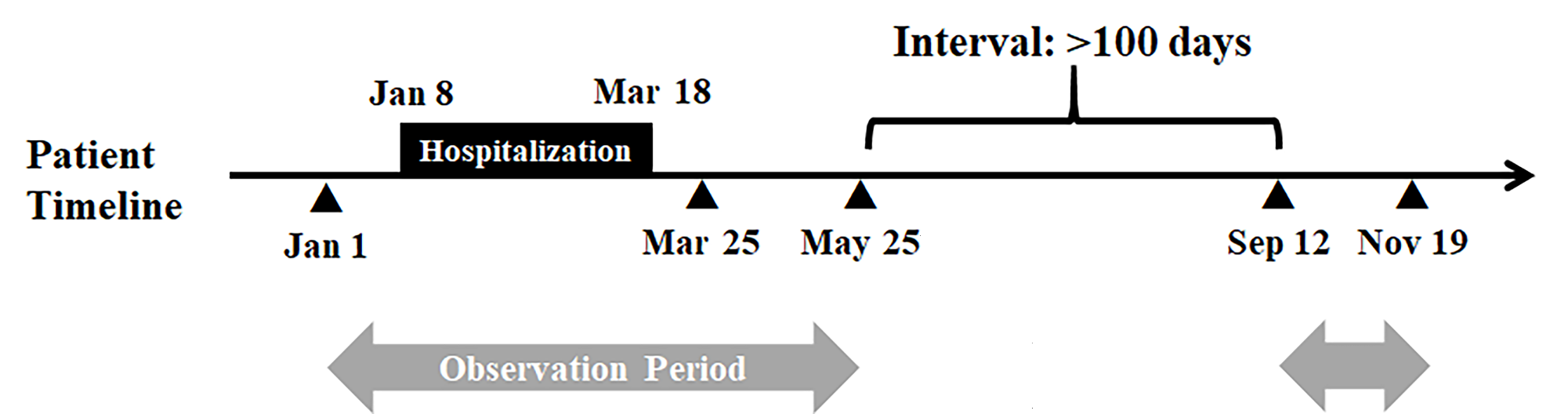
1. Benichou C. Criteria of drug-induced liver disorders. Report of an international consensus meeting. J Hepatol. 1990 Sep;11(2):272-6. PubMed PMID: 2254635.

2. Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology. 2009 Jul;20(4):512-22. doi: 10.1097/EDE.0b013e3181a663cc. PubMed PMID: 19487948; PubMed Central PMCID: PMC3077219.

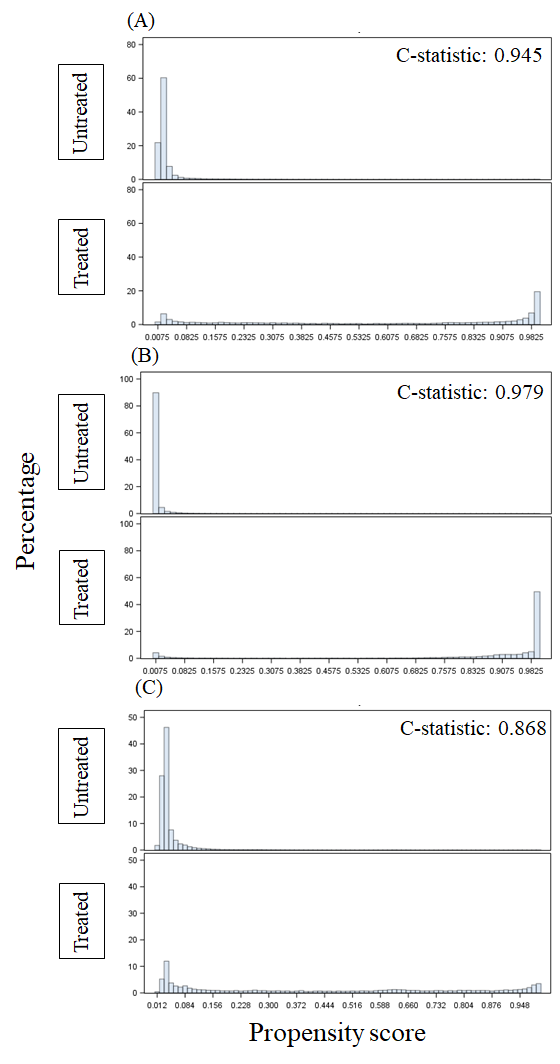
3. Bross ID. Spurious effects from an extraneous variable. J Chronic Dis. 1966 Jun;19(6):637-47. PubMed PMID: 5966011.

4. Toh S, Hernan MA. Causal inference from longitudinal studies with baseline randomization. Int J Biostat. 2008 Oct 19;4(1):Article 22. doi: 10.2202/1557-4679.1117. PubMed PMID: 20231914; PubMed Central PMCID: PMC2835458.

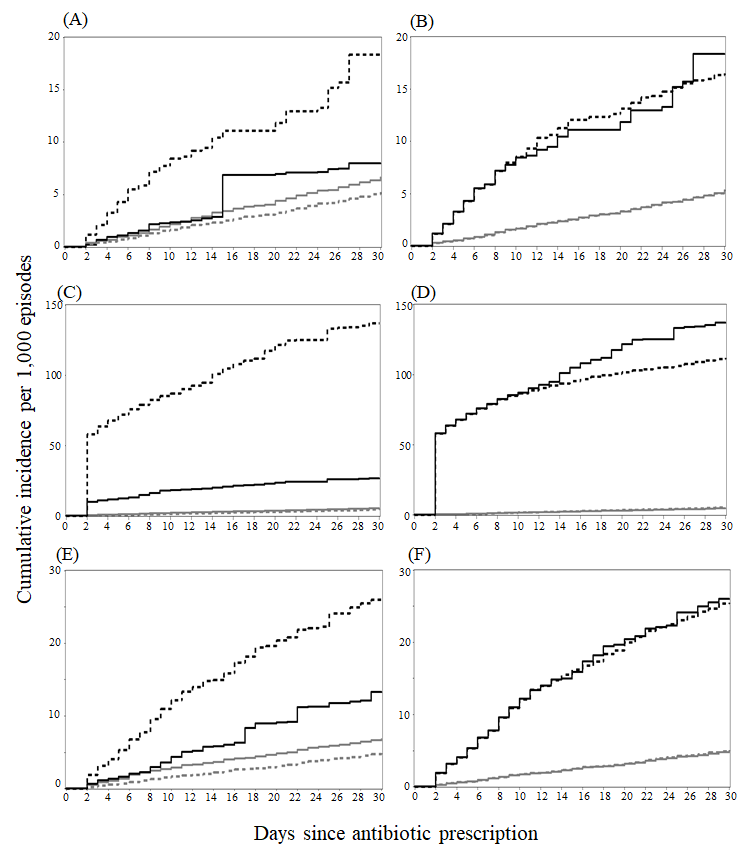
**Supplementary Figure 1.** Schematic example of the definitions of observation periods in a hypothetical patient. The triangles, bar, and double-headed arrows represent outpatient visits, duration of hospitalization, and observation periods, respectively. The patient visited the hospital as an outpatient on Jan 1, Mar 25, May 25, Sep 12 and Nov 19; and was hospitalized from Jan 8 to Mar 18. In this case, the first observation period was from Jan 1 to May 25 because there were no intervals between outpatient visits and hospitalization that exceeded 100 days during this period. However, the interval between May 25 and the next outpatient visit (Sep 12) was more than 100 days, and the patient was considered to be unobserved during this period. The second observation period, which covered two outpatient visits, began from Sep 12 and ended on Nov 19.



**Supplementary Figure 2.** The distribution of the estimated high-dimensional propensity scores for the treated and untreated episodes in the pooled cohorts for (A) macrolides, (B) penicillin-based antibiotics, and (C) fluoroquinolones.



**Supplementary Figure 3.** The cumulative incidence of liver injury per 1,000 treated (black lines) and untreated episodes (gray lines) during the 30-day follow-up period after antibiotic prescription. The cumulative incidence curves in (A) macrolides, (C) penicillin-based antibiotics, and (E) fluoroquinolones are represented by dashed lines (inverse probability of censoring [*SWtC*] weighted) and solid lines (inverse probability of treatment and censoring [*SWA* × *SWtC*] weighted). The cumulative incidence curves in (B) macrolides, (D) penicillin-based antibiotics, and (F) fluoroquinolones are represented by dashed lines (unweighted) and solid lines (inverse probability of treatment and censoring [*SWtC*] weighted).



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Supplementary Table 1.** Number of users for each target antibiotic | | | | | | |
| Antibiotic Class | Antibiotic | | | Number of users\* | | |
| Macrolides | Erythromycin ethylsuccinate | | | 127 | | |
|  | Erythromycin stearate | | | 295 | | |
|  | Erythromycin lactobionate | | | 30 | | |
|  | Roxithromycin | | | 501 | | |
|  | Clarithromycin | | | 6,147 | | |
|  | Azithromycin hydrate | | | 3,628 | | |
|  | Spiramycin acetate | | | <10† | | |
|  | Josamycin | | | <10† | | |
|  |  | | |  | | |
| Penicillin-based antibiotics | Benzylpenicillin potassium | | | 57 | | |
| Ampicillin hydrate | | | 155 | | |
|  | Ampicillin sodium | | | 1,844 | | |
|  | Ampicillin sodium/sulbactam sodium | | | 6,899 | | |
|  | Ampicillin hydrate/cloxacillin sodium | | | 33 | | |
|  | Ampicillin sodium/cloxacillin sodium hydrate | | | 77 | | |
|  | Amoxicillin hydrate | | | 7,556 | | |
|  | Amoxicillin hydrate/potassium clavulanate | | | 1,948 | | |
|  | Piperacillin sodium | | | 422 | | |
|  | Tazobactam/piperacillin | | | 5,606 | | |
|  | Sultamicillin tosylate hydrate | | | 646 | | |
|  |  | | |  | | |
| Fluoroquinolones | Levofloxacin hydrate | | | 17,036 | | |
|  | Tosufloxacin tosylate hydrate | | | 464 | | |
|  | Ciprofloxacin hydrochloride | | | 700 | | |
|  | Ciprofloxacin | | | 346 | | |
|  | Pazufloxacin mesylate | | | 45 | | |
|  | Moxifloxacin hydrochloride | | | 116 | | |
|  | Garenoxacin mesylate Hydrate | | | 1627 | | |
|  | Sitafloxacin hydrate | | | 23 | | |
| \*Number of patients who were prescribed each antibiotic from 2011 to 2015 at the University of Tokyo Hospital. | | | | | | |
|
| **Supplementary Table 2.** Sensitivity analysis results using unweighted models that adjusted for the deciles of hdPS | | | | | |
|
| Antibiotic Class | | Estimator | Point Estimate | | 95% CI |
| Macrolides | | Intent-to-treat HR | 1.11 | | 0.82–1.52 |
|  | |  |  | |  |
|  | | Per-protocol HR | 1.34 | | 0.94–1.90 |
|  | |  |  | |  |
| Penicillin-based antibiotics | | Intent-to-treat HR | 3.89 | | 3.26–4.65 |
|  | |  |  | |  |
|  | | Per-protocol HR | 6.35 | | 4.72–8.56 |
|  | |  |  | |  |
| Fluoroquinolones | | Intent-to-treat HR | 1.67 | | 1.41–1.98 |
|  | |  |  | |  |
|  | | Per-protocol HR | 2.11 | | 1.73–2.58 |
| Abbreviations: CI, confidence interval; hdPS, high-dimensional propensity score; HR, hazard ratio. | | | | | |
|

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Supplementary Table 3.** Sensitivity analysis results of gradual decreases of the covariates’ coefficients toward zero in the hdPS model | | | | | |
|
| Antibiotic Class | Covariate | Coefficient | Estimator | Point Estimate | 95% CI |
| Macrolides | Concomitant use of other otological drugs\* | 4.0 | Intent-to-treat HR | 1.19 | 0.74–1.93 |
|  |  | Per-protocol HR | 1.18 | 0.65–2.13 |
|  |  | 2.0 | Intent-to-treat HR | 1.23 | 0.76–1.99 |
|  |  |  | Per-protocol HR | 1.23 | 0.68–2.22 |
|  |  | 0.0 | Intent-to-treat HR | 1.22 | 0.76–1.98 |
|  |  |  | Per-protocol HR | 1.23 | 0.68–2.22 |
|  |  |  |  |  |  |
| Penicillin-based antibiotics | Concomitant use of other antibacterial drugs against gram-positive or -negative bacteria† | 6.0 | Intent-to-treat HR | 3.35 | 2.65–4.23 |
|  |  | Per-protocol HR | 5.50 | 4.27–7.09 |
|  | 4.0 | Intent-to-treat HR | 3.85 | 3.11-4.75 |
|  |  | Per-protocol HR | 6.90 | 5.36–8.88 |
|  |  | 2.0 | Intent-to-treat HR | 4.03 | 3.21–5.06 |
|  |  |  | Per-protocol HR | 7.37 | 5.67–9.59 |
|  |  | 0.0 | Intent-to-treat HR | 4.00 | 3.16–5.05 |
|  |  |  | Per-protocol HR | 7.29 | 5.61–9.48 |
| \*Original coefficient: 4.71, † Original coefficient: 7.53. Abbreviations: CI, confidence interval; hdPS, high-dimensional propensity score; HR, hazard ratio. | | | | | |
|
|