Characteristics of Hospital Differences in Missing of Clinical Laboratory Test Results in a Multi-hospital Observational Database Contributing to MID-NET® in Japan

Maki Komamine (komamine-maki@pmda.go.jp)

Department of Biostatistics, Kyoto University School of Public Health, Kyoto, Japan bOffice of Medical Informatics and Epidemiology, Pharmaceuticals and Medical Devices Agency, Tokyo, Japan

https://orcid.org/0000-0002-8390-0922

Yoshiaki Fujimura

Tokushukai Information System Incorporated

Yasuharu Nitta

Kishiwada Tokushukai Hospital

Masatomo Omiya

Kyoto University School of Public Health

Masaaki Doi

Kyoto University School of Public Health

Tosiya Sato

Kyoto University School of Public Health

Research article

Keywords: Drug safety, clinical laboratory test, database, missing data, observational study, pharmacoepidemiology

Posted Date: September 21st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-74875/v1

Version of Record: A version of this preprint was published at BMC Medical Informatics and Decision Making on June 6th, 2021. See the published version at https://doi.org/10.1186/s12911-021-01543-5.

Title page

Title:

Characteristics of hospital differences in missing of clinical laboratory test results in a multi-hospital observational database contributing to MID-NET® in Japan

Authors:

Maki Komamine[a,b]*, Yoshiaki Fujimura[c], Yasuharu Nitta[d], Masatomo Omiya[a], Masaaki Doi[a], Tosiya Sato[a]

Affiliations:

[a]Department of Biostatistics, Kyoto University School of Public Health, Kyoto, Japan

[b]Office of Medical Informatics and Epidemiology, Pharmaceuticals and Medical Devices Agency, Tokyo, Japan

[c]Head Office, Tokushukai Information System Incorporated, Osaka, Japan

[d]Kishiwada Tokushukai Hospital, Osaka, Japan

***Corresponding author:**

Maki Komamine

Department of Biostatistics, Kyoto University School of Public Health

Yoshida-konoecho, Sakyo-ku, Kyoto 606-8501, Japan

Telephone: +81-75-753-4475

Email: komamine-maki@pmda.go.jp

Prior posting and presentations: The contents of this paper have not been previously posted or presented.

Abstract

Background: In Japan, a multiple-hospital observational database system, the Medical Information Database Network (MID-NET®), was launched for post-marketing drug safety assessments. These assessments will be based on datasets with missing laboratory results. The characteristics of missing data considering hospital differences have not been evaluated. We assessed the missing proportion and the association between missingness and a factor through case studies using a database system, a part of MID-NET®.

Methods: Seven scenarios using laboratory results before the prescription of the assessed drug as baseline covariates and data from 10 hospitals of Tokushukai Medical

Group were used. The missing proportion and the association between missingness and patient background were investigated per hospital. The associations were assessed using the log of adjusted odds ratio (log-aOR). Additionally, an ad hoc survey was conducted to explore other factors affecting the missingness.

Results: For some laboratory tests, missing proportions varied among hospitals, such as 7.4%–44.4% of alkaline phosphatase (ALP) and 8.1%–31.2% of triglyceride (TG) among statin users. The association between missingness and affecting factors also differed among hospitals for some factors; example, the log-aOR of hospitalization associated with missingness of TG was -0.41 (95% CI, -1.06 to 0.24) in hospital 3 and 1.84 (95% CI, 1.34 to 2.34) in hospital 4. In the ad hoc survey focusing on ALP, hospital-dependent differences in the ordering system settings were observed.

Conclusions: Hospital differences in missing data appeared in some laboratory tests in our multi-hospital observational database, which could be attributed to the affecting factors, including the patient background.

Keywords:

Drug safety, clinical laboratory test, database, missing data, observational study, pharmacoepidemiology

Background

Observational databases, including health insurance claims and electronic medical records (EMRs), are crucial data sources for regulatory decision-making, providing clinical evidence on the usage and potential benefits or risks of a medical product.[1-5] Particularly, laboratory test results are useful sources of covariates or outcome measures in pharmacoepidemiological studies, including post-marketing drug safety assessments.[6,7]

The appropriate use of these data is difficult because some data obtained during routine medical care may be missing in datasets for analysis.[8,9] Missing covariate data is a critical issue for observational studies requiring confounding adjustments. Various methods have been proposed to overcome improper handling of missing data that can result in bias.[8,10] The features of missing data (e.g., missing proportion and factors associated with missingness) and sources of missing data are crucial for choosing appropriate missing data methods.[9,10]

Missing proportions and factors associated with missingness can differ across data partners in databases covering multiple sites or hospitals. The variability in missing data among data partners is a critical issue for applying the missing data method. For three sites contributing to the US Food and Drug Administration Mini-Sentinel Distributed

Database (MSDD), Raebel *et al.*[9] reported that the missing proportion of baseline laboratory results and factors associated with missingness varied by site. Differential missingness across sites was attributed to multiple factors, such as the type of data partner (e.g., only with laboratory results of outpatients) and patient background. The authors recommended applying a missing data method in a site-specific manner.

In Japan, the Medical Information Database Network (MID-NET®) was launched as a national project in April 2018 for post-marketing drug safety assessments.[6,7,11] This multi-hospital observational database system comprises 23 mid-sized and large hospitals from 10 collaborative organizations.[12] Unlike those of the MSDD, all collaborative hospitals of the MID-NET® are the same type of data partners and have EMRs as data sources of laboratory results. Hospital differences in missing laboratory results may still exist because of hospital-dependent potential factors (e.g., laboratory test measurement policies) and patient-dependent factors. Although laboratory results covered by the MID-NET® project are quality-checked and standardized extensively,[12] the features of missing data considering hospital differences have not been thoroughly evaluated.

We used data from 10 MID-NET®-collaborative hospitals and seven exposure-outcome scenarios using laboratory results as baseline covariates to investigate the

characteristics of hospital differences in missing data as follows: (i) we investigated the frequency of laboratory result records and quantified the missing proportion; (ii) we assessed the association between the missingness and a factor affecting missingness; and (iii) we conducted an ad hoc survey to explore other factors affecting hospital differences in missing data. In some scenarios using laboratory results as outcome measures, we performed a supplementary investigation of the frequency of laboratory test records after the prescription date.

Methods

Target hospitals and database

The MID-NET® is a distributed and closed network system in which each collaborative organization has a database system containing claims data, diagnosis procedure combination data, and EMRs.[12] The collaborative organizations consist of seven individual and three group hospitals. Each group hospital database collectively stores data from their MID-NET®-contributing hospitals. The largest group hospital, Tokushukai Medical Group comprising 10 hospitals, was selected for investigating hospital differences with one database system.

The selected hospitals differ in size and serve as regional core hospitals with an emergency department. Hospital names are provided in Supplementary Table S1. We assigned hospital identification numbers 1–10 to ensure privacy in the results. EMRs in the database system for MID-NET®-collaborative organizations of Tokushukai Medical Group contain laboratory results, including those from the emergency department. The database does not capture hospital-specific data (e.g., laboratory test measurement policies and number of patients or beds).

Definition of missing data

The observational database has two basic sources of missing laboratory results: a laboratory test was not conducted, and a laboratory test was conducted but not recorded.[8,9] Because the two sources were difficult to distinguish, we defined missing data as follows: "data that would be meaningful for analysis but not available during a specific period."

Missingness should be confirmed during a patient's continuous consecutive observation. Therefore, we recreated the observation period for each patient by connecting hospital visits data. We then adopted five periods to confirm the missingness of laboratory results (the "target period") as baseline covariates or outcome measures:

for baseline covariates, 1) 90 days before the first prescription date (including the date) or 2) 180 days before the first prescription date (including the date); for outcomes, 3) period from prescription date to observation period end, 4) period from 365 days after first prescription date, or 5) 84 days after the first prescription date. The first and second periods were adopted by referring to previous cohort studies using a laboratory test as baseline covariate[13-15] and a previous study assessing missing data in the MSDD for 183 days.[9] The third and fourth periods were adopted for cases where all outcomes were included and for cases where the study interest was the only outcome after a certain period from the prescription date, respectively. The 365 days in the fourth period was created by referring to the mean follow-up period in a previous study of our scenario.[16] The last period was adopted for scenario 3, considering the follow-up period used in clinical trials (8 weeks) and different treatment intervals for each patient.[17]

Frequency of laboratory result records and missing proportion

Frequencies of records in patients with laboratory result records of interest during a target period in each scenario were considered to assess the missing proportions. For a laboratory result used as baseline covariate, we counted the number of records per target period for each patient. Multiple records from the same day were outside the study

144 objective and counted as one record. We then calculated the percentage of patients for
145 each number of records in the overall cohort. The percentage of patients without a
146 record, namely missing proportion, was also calculated for each hospital cohort.
147 For a laboratory result used as an outcome measure, we counted the number of
148 records per target period and calculated the percentage in the overall cohort. In the
149 analysis using the third target period, we calculated quartiles, along with the maximum
150 and minimum values of the period, because of patient-dependent target period
151 variations.
152
153 *Association between missingness and a potential factor*
154 We assessed hospital differences in the association between the missingness of
155 laboratory result records before the prescription date and a potential factor affecting the
156 missingness by fitting a logistic regression model in each hospital cohort of an
157 individual scenario. Potential factors included sex, age, year of cohort entry,
158 hospitalization, complications, concomitant medication, and class number of
159 concomitant medications (Table 1). Complications or concomitant medications not
160 observed in each hospital cohort were excluded from the covariates of hospital-specific
161 logistic regression models. Each factor's association was evaluated by the log of

adjusted odds ratio (log-aOR) and 95% confidence interval (95% CI). In the model for scenario l ($l = 1, \cdots, L$), we used the following notation: Y_{ijl}, a missing data indicator (1 when missing or 0 otherwise); X_{ijl}, covariates; i, individuals of each hospital; j, number of laboratory tests; and K_l, number of covariates of each hospital. We fitted logistic models as

$$\text{logit}\big(Pr(Y_{ijl} = 1|\boldsymbol{X}_{ijl})\big) = \alpha + \boldsymbol{X}_{ijl}{}'\boldsymbol{\beta}_{jl},$$

where $\boldsymbol{X}_{ijl} = (X_{ij1l}, \ldots., X_{ijK_ll})'$.

Scenarios

Seven cohort study scenarios using laboratory results as baseline covariates were created (Supplementary Figure S1). Scenarios 1–5 were original scenarios; scenarios 6 and 7 were incorporated to compare our results with those of Raebel *et al.*[9] Scenario setting details are provided in Table 1. The backgrounds of the original scenarios 1–5 were as follows.

Scenario 1: Risk of diabetes associated with antipsychotic drug use

Glucose metabolism disorder is considered a risk of second-generation antipsychotics (SGAs).[18,19] We created a scenario with a cohort with new antipsychotic

users to compare the diabetes risk of SGAs with that of first-generation antipsychotics (FGAs), considering blood glucose level and HbA1c as baseline covariates and outcome measures. Depending on the study's interests, the target population may have been formed of patients without diabetes based on the baseline laboratory results. Thus, we also created a sub-cohort that only included patients confirmed to be diabetes-free using baseline blood glucose or HbA1c (National Glycohemoglobin Standardization Program; NGSP) as follows: excluding patients with blood glucose of ≥200 mg/dL or HbA1c(NGSP) of ≥6.5%, or without a record of blood glucose and HbA1c.

Scenario 2: Risk of hepatic injury associated with statin use

Hepatic injury is considered as a risk common to all statins and mentioned in package inserts as a severe adverse effect. The attention level differs among statins (atorvastatin and rosuvastatin are contraindicated for patients with decreased liver function). Observational studies demonstrated that the hepatic injury risk of atorvastatin use, particularly that of high-dose use, is higher than that of other statins,[16] and only a few studies indicated a similar risk in rosuvastatin and atorvastatin users.[20] We then created a scenario comparing the hepatic injury risk of atorvastatin with that of other statins, including rosuvastatin, considering low-density lipoprotein cholesterol (LDL-

198 chol), triglyceride (TG), alanine aminotransferase (ALT), aspartate transaminase (AST),
199 and alkaline phosphatase (ALP) as baseline covariates, and ALT, AST, and ALP as
200 outcome measures.
201
202 Scenario 3: Effect of uric acid synthesis inhibitor use on uric acid level
203 The uric acid-lowering effect of febuxostat was non-inferior to that of allopurinol in
204 a Japanese phase III clinical trial.[17] Because patients with renal impairments were
205 excluded from the trial's target population, the effect on an overall population is
206 unclear. We created a scenario comparing the uric acid-lowering effect of febuxostat
207 with that of allopurinol, considering serum uric acid and serum creatinine as baseline
208 covariates, and serum uric acid as outcome measure.
209
210 Scenario 4: Risk of hyponatremia associated with proton pump inhibitor use
211 Hyponatremia, a risk of lansoprazole use, is listed as a serious adverse effect in the
212 lansoprazole package insert in Japan, but not in those of other proton pump inhibitors
213 (PPIs). A case–control study indicated that other PPIs are associated with an increased
214 hyponatremia risk.[21] We created a scenario comparing the hyponatremia risk of

215 lansoprazole with that of other PPIs using serum sodium and serum creatinine as
216 baseline covariates, and serum sodium as outcome measures.
217
218 Scenario 5: Risk of acute pancreatitis associated with oral antidiabetic drug use
219 Acute pancreatitis is considered as a risk of dipeptidyl peptidase-4 inhibitor (DPP-
220 4I) use and listed in the DPP-4I package insert as a severe adverse effect in Japan. Some
221 observational studies demonstrated that the acute pancreatitis risk associated with DPP-
222 4Is may not be higher than that associated with other oral antidiabetic agents.[22-24] We
223 created a scenario comparing the acute pancreatitis risk of DPP-4I with that of other oral
224 antidiabetic agents, including biguanide, sulfonylurea, or α-glucosidase inhibitor, using
225 blood glucose level, HbA1c, and serum amylase as baseline covariates.
226
227 *Protocol approval and statistical analysis*
228 Our study protocol was approved by the Kyoto University Graduate School and
229 Faculty of Medicine Kyoto University Hospital Ethics Committee in November 2018
230 (R1793). Statistical analyses were performed using SAS version 9.4 (SAS Institute,
231 Cary, NC, USA).
232

Results

Study cohorts

The overall cohorts were identified as follows: scenario 1: 3430 new antipsychotics users; scenario 2: 6195 new statin users; scenario 3: 3481 new users of uric acid synthesis inhibitors; scenario 4: 10,372 new PPI users; scenario 5: 2994 new users of oral antidiabetics; scenario 6: 965 new users of combinations of antimicrobials with warfarin; and scenario 7: 1007 new SGA users (Supplementary Figures S1–8). Patient characteristics and their numbers in each hospital cohort are provided in Supplementary Tables S2–S6. The background of some patients differed among hospitals.

Frequency of laboratory result records and missing proportion

In the overall cohort, the frequency of laboratory result records within 90 days before prescription differed among laboratory tests except for ALT and AST (Figure 1). In most laboratory tests, patients with one record were the most frequent, although some had multiple records. In scenario 1, the percentage of patients with multiple records was higher for blood glucose than for HbA1c. The missing proportions (shaded bars, Figure 1) were <30%, except for HbA1c and serum amylase in scenarios 1 and 5; example, 29.2% of ALP in scenario 2, 22.4% of serum creatinine in scenario 4, 13.8% of blood

glucose in scenario 7, 12.8% of blood glucose in scenario 1, 9.7% of international normalized ratio (INR) in scenario 6, and 4.0% of blood glucose in scenario 5. Extending the target period to 180 days did not substantially change these missing proportions (Supplementary Figure S9).

In each hospital cohort, missing proportions within 90 days before prescription differed among hospitals for some laboratory tests; example, 5.2%–41.3% of blood glucose in scenario 1, 7.4%–44.4% of ALP in scenario 2, 8.1%–31.2% of TG in scenario 2, 4.7%–21.9% of INR in scenario 6, 1.4%–39.1% of blood glucose in scenario 7 (Figure 4). In scenario 1, the blood glucose missing proportion was higher in hospital 10 than in the other hospitals. In scenario 2, the missing proportion variations of ALT/AST and ALP differed among hospitals. Specifically, hospital 3 showed a large difference among these tests, whereas hospital 6 did not. Similar to the overall cohort results, extending the target period to 180 days did not substantially change the hospital differences (Supplementary Figure S12).

The frequency of laboratory result records after prescription differed from that before the prescription (Figure 2), for example, the percentage of patients with only one record decreased, and the missing proportion slightly increased (shaded bars, Figure 2). Limiting the target period to 365 days improved the missing proportion (Supplementary

Figure S10) despite a decrease in patient numbers: scenario 1: 881 patients; scenario 2: 2901 patients; and scenario 4: 2333 patients. In scenario 3, the missing proportion was 33.2% within 84 days after prescription (shaded bars, Figure 3). In scenario 1, the blood glucose missing proportion among the sub-cohort was 19.9%, which was 3.1% lower than that among the cohort (Supplementary Figure S11).

Association between missingness and a potential factor

Scenarios 6 and 7 were excluded from analysis because of the low patient numbers in the hospital cohorts. The degree of association between missingness and a factor differed among hospitals for some factors (Figure 5). For example, in scenario 2, the log-aOR of associating hospitalization with missingness of TG was <0 in hospital 3 (log-aOR, -0.41 [95% CI, -1.06 to 0.24]) but >0 in hospital 4 (log-aOR, 1.84 [95% CI, 1.34 to 2.34]).

Because hospital differences in the missing proportions within 180 days before prescription did not substantially vary from that within 90 days, this analysis was limited to the latter target period.

Ad hoc survey

The missing proportion of ALT/AST and ALP suggested an influence from hospital-dependent mechanical factors. The missing proportions may vary among these liver function tests because they measure different parameters. However, the degree of variation differed widely between hospitals 3 and 6.

Laboratory tests are ordered individually or in a group. Grouping can differ for each hospital because it can be customized. We assumed the effect of groupings on a chance of performing laboratory tests, namely missingness, and assessed the inclusion of ALT, AST, and ALP in groupings in hospitals 3 and 6 by confirming some groupings. We could not perform quantitative assessment and instead used the electronic laboratory ordering system because our database did not contain grouping data. We identified differences in some grouping settings; specifically, ALP was often grouped along with ALT or AST in hospital 6 but not in hospital 3.

Discussion

We evaluated seven scenarios in a multi-hospital observational database system, a part of the MID-NET®, to investigate hospital differences in missing laboratory results

303 for baseline covariates. In addition to these differences, we examined factors affecting
304 the frequency of laboratory result records and missing data sources.
305 Variations in purpose for performing laboratory tests might have caused
306 differences in the frequency of laboratory result records among laboratory tests or
307 scenarios. In routine medical care, laboratory tests are performed to diagnose diseases
308 and assess or monitor physiological functions.[25] For example, assessing and monitoring
309 physiological functions could have contributed to regular laboratory testing and
310 multiple records, such as serum creatinine in scenario 4. Variations in test intervals
311 allowed by the health insurance in Japan (e.g., blood glucose, maximum of 60 per
312 month for type 2 diabetes, and HbA1c, once per month) may have also affected the
313 frequency. The period for confirming the missingness should be created considering
314 these factors and the study objective.
315 Several factors contributed to missing laboratory results in our database. Few
316 studies have systematically referred to missing data sources, except the MSDD-based
317 study by Raebel *et al.*[9] Here, the missing data sources included type of data partner,
318 patient location where tests were conducted (e.g., emergency department), collectability
319 from outside of contracted laboratories, and patient backgrounds. Our database had
320 some common and different sources compared to this previous study. Patient

backgrounds were considered to affect the missing data in our database, similar to observations in the previous study. However, the contribution of the other three factors to the missing data may be limited, although this was not quantitatively assessed. All 10 hospitals in our study are the same type of data partner and had EMR-based laboratory results, including those of the emergency department. Laboratory tests assessed were mainly performed in the hospital and not outsourced. A new potential source was the grouping of laboratory tests. Other remaining potential factors included the policy for performing laboratory tests, which was considered at the planning stage but not assessed because of a lack of data.

Our database had hospital differences in the missing proportion and association between the missingness and a factor affecting missingness. As described above, there were few missing data sources in our database. Patient backgrounds were a substantial source, and the grouping of laboratory tests to order remains a potential source. In some patient backgrounds, the association with the missingness differed among hospitals. Additionally, hospital differences in missing blood glucose in scenario 1 were diminished by limiting the study subjects to patients over 21 years of age in the additional analysis (Supplementary Figure S13). In the ad hoc survey focusing on ALP with substantial hospital differences in missing proportion, hospital-dependent

differences in the setting of some groupings of laboratory tests were observed. In our database, hospital-dependent potential missing data sources exist, but the corresponding data are not available for analysis. Therefore, missing data methods should consider the effect of the hospital (i.e., such as using a hospital-specific approach).

Variations in the type of missing data sources among databases accounted for the difference in the missing proportion. In scenarios 6 and 7, differences among hospitals were lower than those among sites in a previous study[9] (INR from scenario 6: 2.8%–21.9% vs. approximately 8.0%–80.0%; blood glucose from scenario 7: 1.4%–30.6% vs. 41.1%–72.3%). Although study population differences caused these variations, differences in missing data sources among databases may also contribute.

This study had several strengths. First, we investigated the characteristics of hospital differences in missing laboratory results using a part of the MID-NET®. As these characteristics also exist in the entire MID-NET®, our findings will provide guidance for using MID-NET®, which is a national project. Second, we observed hospital differences in the missing data and discussed the missing data source affecting these differences: patient background and grouping of laboratory tests to order. Finally, we observed various missing proportions by including multiple laboratory tests. The

356 variety contributed to characterizing the hospital differences in missing data, although
357 the laboratory tests used were limited.
358 Nonetheless, there were some limitations. First, laboratory tests not covered by our
359 study may have other missing data characteristics. Second our results may not be
360 generalizable to the entire MID-NET®. There were differences among the 10
361 Tokushukai Medical Group hospitals that exist in the entire MID-NET®. However, a
362 non-difference observed among the 10 hospitals does not assure it is a non-difference in
363 the entire MID-NET®. Other hospitals may have different factors affecting the missing
364 proportion or their hospital differences. For example, the 10 Tokushukai Medical Group
365 hospitals are mainly general hospitals, whereas the other hospitals are mostly
366 specialized hospitals. As the latter provides medical care to patients referred from other
367 hospitals and clinics, referral rates may be a factor.
368
369 **Conclusions**
370 We concluded that hospital differences in the missing data appeared in some
371 laboratory tests in a multiple-hospital observational database system contributing to the
372 MID-NET® because of factors such as patient background, although all hospitals are the
373 same type of data partner. Importantly, these differences were found in the entire MID-

NET®. Since data of hospital-dependent factors affecting missingness are not available in MID-NET®, missing data methods should be applied while considering the effect of each hospital (e.g., use a hospital-specific approach). Further studies should investigate the influence of these hospital differences on outcome parameter estimations.

Declarations

Ethics approval and consent to participate

Our study protocol was approved by the Kyoto University Graduate School and Faculty of Medicine Kyoto University Hospital Ethics Committee in November 2018 (R1793).

Consent for publication

Not applicable.

Availability of data and materials

Due to the terms of use for MID-NET®, the dataset used for analysis cannot be made openly available; the terms limit the use to approved analysts and do not allow analysts to share individual datasets from the predetermined secure environment. This study used the database system for MID-NET®-collaborative organizations of the Tokushukai

Medical Group, a part of MID-NET®, and not the entire MID-NET®. However, we followed the terms of use for MID-NET®, because the datasets used for this analysis were included in the entire MID-NET®.

Competing interests

Maki Komamine is employed by the Pharmaceuticals and Medical Devices Agency and has no financial or personal relationships with other people or organizations that could inappropriately influence or bias the contents of this paper. Other authors have no financial or personal relationships with other people or organizations that could inappropriately influence or bias the contents of this paper.

List of abbreviations

ALP: alkaline phosphatase

ALT: alanine aminotransferase

AST: aspartate transaminase

CI: confidence interval

DPP-4I: dipeptidyl peptidase-4 inhibitor

EMRs: electronic medical records

FGA: first-generation antipsychotics

INR: international normalized ratio

LDL-chol: low-density lipoprotein cholesterol

Log-aOR: log of adjusted odds ratio

MID-NET: Medical Information Database Network

MSDD: Mini-Sentinel Distributed Database

NGSP: National Glycohemoglobin Standardization Program

PPI: proton pump inhibitors

SGA: second-generation antipsychotics

TG: triglyceride

Funding

Not applicable.

Authors' contributions

MK, MO, MD, and TS conceptualized the study, MK analyzed the data. MK wrote the initial draft of the manuscript. MK, YF, YN, MO, MD, and TS contributed to the

interpretation of findings and manuscript revisions. All authors have read and approved the final version of the manuscript.

Acknowledgements

We thank Dr. Yoshiaki Uyama and the Tokushukai Medical Group for their assistance. Views expressed here are those of the authors and do not necessarily represent the official views and findings of the Pharmaceuticals and Medical Devices Agency.

References

1. U.S. Food and Drug Administration. Framework for FDA's Real-World Evidence Program. 2018. https://www.fda.gov/media/120060/download. Accessed 31 July 2020.
2. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, LaVange L, Marinac-Dabic D, Marks PW, Robb MA, Shuren J, Temple R, Woodcock J, Yue LQ, Califf RM. Real-world evidence—what is it and what can it tell us? N Engl J Med. 2016;375:2293–7.
3. Robb MA, Racoosin JA, Sherman RE, Gross TP, Ball R, Reichman ME, Midthun K, Woodcock J. The US Food and Drug Administration's Sentinel Initiative:

445 expanding the horizons of medical product safety. Pharmacoepidemiol Drug Saf.
446 2012;21:9–11.
447 4. Blake KV, Prilla S, Accadebled S, Guimier M, Biscaro M, Persson I, Arlett P,
448 Blackburn S, Fitt H. European Medicines Agency review of post-authorisation
449 studies with implications for the European Network of Centres for
450 Pharmacoepidemiology and Pharmacovigilance. Pharmacoepidemiol Drug Saf.
451 2011;20:1021–9.
452 5. Ishiguro C, Takeuchi Y, Uyama Y, Tawaragi T. The MIHARI project: establishing
453 a new framework for pharmacoepidemiological drug safety assessments by the
454 Pharmaceuticals and Medical Devices Agency of Japan. Pharmacoepidemiol Drug
455 Saf. 2016;25:854–9.
456 6. Pharmaceuticals and Medical Devices Agency. Summary of MID-NET® Study:
457 No.2018-001. https://www.pmda.go.jp/files/000233987.pdf; 2020 Accessed 31 July
458 2020.
459 7. Pharmaceuticals and Medical Devices Agency. Summary of MID-NET® study: No.
460 2018-002. https://www.pmda.go.jp/files/000234446.pdf; 2020 Accessed 31 July
461 2020.
462 8. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for handling missing

463 data in electronic health record derived data. EGEMS (Wash DC). 2013;1:1035.
464 9. Raebel MA, Shetterly S, Lu CY, Flory J, Gagne JJ, Harrell FE, Haynes K,
465 Herrinton LJ, Patorno E, Popovic J, Selvan M, Shoaibi A, Wang X, Roy J. Methods
466 for using clinical laboratory test results as baseline confounders in multi-site
467 observational database studies when missing data are expected. Pharmacoepidemiol
468 Drug Saf. 2016;25:798–814.
469 10. Eekhout I, de Boer RM, Twisk JW, de Vet HC, Heymans MW. Missing data: a
470 systematic review of how they are reported and handled. Epidemiology.
471 2012;23:729–32.
472 11. Yamada K, Itoh M, Fujimura Y, et al. The utilization and challenges of Japan's MID-
473 NET ® Medical Information Database Network in postmarketing drug safety
474 assessments: a summary of pilot pharmacoepidemiological studies.
475 Pharmacoepidemiol Drug Saf. 2019;28:601–8.
476 12. Yamada K, Itoh M, Fujimura Y, Kimura M, Murata K, Nakashima N, Nakayama
477 M, Ohe K, Orii T, Sueoka E, Suzuki T, Yokoi H, Ishiguro C, Uyama Y; MID-
478 NET® project group. Establishment of the MID-NET® medical information
479 database network as a reliable and valuable database for drug safety assessments in
480 Japan. Pharmacoepidemiol Drug Saf. 2019;28:1395–1404.

13. Masato T, Masahito O, Takaaki M, Inagaki N, Kawakami K. Comparative effectiveness of sodium-glucose cotransporter-2 inhibitors versus other Classes of glucose-lowering medications on renal outcome in type 2 diabetes. Mayo Clin Proc. 2020;95:265–73.

14. Chamberlain AM, Cohen SS, Weston SA, Fox KM, Xiang P, Killian JM, Qian Y. Relation of cardiovascular events and deaths to low-density lipoprotein cholesterol level among statin-treated patients with atherosclerotic cardiovascular disease. Am J Cardiol. 2019123:1739–44.

15. Chang CH, Kusama M, Ono S, Sugiyama Y, Orii T, Akazawa M. Assessment of statin-associated muscle toxicity in Japan: a cohort study conducted using claims database and laboratory information. BMJ Open. 2013;3: e002040

16. Clarke AT, Johnson PC, Hall GC, Ford I, Mills PR. High dose atorvastatin associated with increased risk of significant hepatotoxicity in comparison to simvastatin in UK GPRD cohort. PLoS One. 2016;11:e0151587.

17. Pharmaceuticals and Medical Devices Agency: Review report of febuxostat. http://www.pmda.go.jp/files/000223354.pdf; 2010 Accessed 31 July 2020.

497 18. Newcomer JW, Haupt DW, Fucetola R, Melson AK, Schweiger JA, Cooper BP,
498 Selke G. Abnormalities in glucose regulation during antipsychotic treatment of
499 schizophrenia. Arch Gen Psychiatry. 2002;59:337–45.
500 19. Lindenmayer J, Nathan A, Smith R. Hyperglycemia associated with the use of
501 atypical antipsychotics. J Clin Psychiatry. 2001; 62: 30–8.
502 20. Chang CH, Chang YC, Lee YC, Liu YC, Chuang LM, Lin JW. Severe hepatic
503 injury associated with different statins in patients with chronic liver disease: a
504 nationwide population-based cohort study. J Gastroenterol Hepatol. 2015;30:155–
505 62.
506 21. Falhammar H, Lindh JD, Calissendorff J, Skov J, Nathanson D, Mannheimer B.
507 Associations of proton pump inhibitors and hospitalization due to hyponatremia: A
508 population-based case-control study. Eur J Intern Med. 2019;59:65–9.
509 22. Chang CH, Lin JW, Chen ST, Lai MS, Chuang LM, Chang YC. Dipeptidyl
510 peptidase-4 inhibitor use is not associated with acute pancreatitis in high-risk type 2
511 diabetic patients: A nationwide cohort study. Medicine (Baltimore). 2016;95:e2603.
512 23. Hong JL, Buse JB, Jonsson Funk M, Pate V, Stürmer T. The risk of acute pancreatitis
513 after initiation of dipeptidyl peptidase 4 inhibitors: Testing a hypothesis of subgroup
514 differences in older US adults. Diabetes Care. 2018;41:1196–203.

515 24. Kim YG, Kim S, Han SJ, Kim DJ, Lee KW, Kim HJ. Dipeptidyl peptidase-4
516 inhibitors and the risk of pancreatitis in patients with type 2 diabetes mellitus: A
517 population-based cohort study. J Diabetes Res. 2018;2018:5246976.
518 25. Beaulieu-Jones BK, Lavage DR, Snyder JW, Moore JH, Pendergrass SA, Bauer
519 CR. Characterizing and managing missing structured data in electronic health
520 records: Data analysis. JMIR Med Inform. 2018;6:e11.

Table 1. Description of study scenarios.

#	Scenario question	Study cohort	Laboratory test of interest	Factors used for assessing the association with missingness
1	Whether SGA users have a higher risk of diabetes than FGA users?	Inclusion criteria: ・ Patients initiated with any SGA or FGA during the study period (1 January 2015 to 31 December 2017). ・ New users of monotherapy for SGA or FGA: Patients with no prescription of the drugs for >180 days before the first prescription of SGA or FGA. Exclusion criteria: ・ Patients with a diagnosis of dementia (ICD10: F01, F03, G30, G310, etc.) within the 180 days before the first prescription of any SGA or FGA. ・ Patients with a diagnosis of diabetes (ICD10: E10, E11, E12, E13, E14, O24) within 180 days before the first prescription of any SGA or FGA. Sub-cohort exclusion criteria: ・ Patients with the following criteria within the 90 days before the first prescription date: blood glucose was ≥200 mg/dL or HbA1c(NGSP) was ≥6.5%. ・ Blood glucose or HbA1c was not recorded within the 90 days before the first prescription date.	Baseline covariate: ・ Blood glucose, HbA1c Outcome: ・ Blood glucose, HbA1c Note: ・ We both count HbA1c (JDS) and HbA1c (NGSP) since HbA1c (JDS) can convert to HbA1c (NGSP)	Sex, age, year of cohort entry, hospitalization, first visit, emergency care (at the date of the first prescription), class number of concomitant medications, complications, concomitant medication (180 days before the date of the first prescription of any antidiabetic drug) ・ Complications: Hepatitis, liver cirrhosis, chronic pancreatitis, hypertension, hyperlipidemia, hyperthyroidism, Cushing's syndrome, primary aldosteronism, pancreatic cancer, liver cancer, pheochromocytoma, hemochromatosis, schizophrenia, mood disorder, neurotic disorder, cancer other than liver and pancreatic ・ Concomitant medication: Beta-blockers, thiazide diuretics, antidepressants, corticosteroids, interferon prepared, high-calorie transfusion agents, immunosuppressants
2	Whether users of other statins have a different risk of hepatic injury than atorvastatin users?	Inclusion criteria: ・ Patients initiated with any statin (rosuvastatin, pitavastatin, pravastatin, simvastatin, fluvastatin, or atorvastatin) during the study period (1 January 2015 to 31 December 2017). ・ New users of monotherapy for statin as in scenario 1.	Baseline covariate: ・ ALT, AST, ALP, Bilirubin ・ LDL-chol, TG Outcome: ・ ALT, AST, ALP	Sex, age, year of cohort entry, hospitalization, first visit, emergency care (at the date of the first prescription), class number of concomitant medication, complications, concomitant medication (180 days before the date of the first prescription of any antidiabetic drug)

		Exclusion criteria: ・ Patients with a diagnosis of hepatic injury. (ICD10: B18, K70-K76, K770, K778) within 180 days before the first prescription of any statin.		・ Complications: Chronic kidney disease, heart failure, acute myocardial infarction, hypertension, cerebrovascular diseases, diabetes mellitus, peripheral vascular disease ・ Concomitant medication: Antiepileptic drugs, fibrates, ezetimibe, anti-gout preparations, antithyroid agent, NSAIDs, antifungal drugs, anti-tuberculosis agents, therapeutic agents for chronic hepatitis B or C
3	Does the uric acid-lowering effect differs between febuxostat and allopurinol in a population, including patients with renal dysfunction?	Inclusion criteria: ・ Patients initiated with uric acid synthesis inhibitor (febuxostat or allopurinol) during the study period (1 January 2015 to 31 December 2017). ・ New users of monotherapy for uric acid synthesis inhibitor as in scenario 1. Exclusion criteria: Patients with a period from prescription to end of observational period fewer than 84 days.	Baseline covariate: ・ Serum uric acid ・ Serum creatinine Outcome: ・ Serum uric acid	Sex, age, year of cohort entry, hospitalization, first visit, emergency care (at the date of the first prescription), class number of concomitant medications, complications, concomitant medication (180 days before the date of the first prescription of any antidiabetic drug) ・ Complications: Leukemia, heart failure, acute myocardial infarction, hypertension, cerebrovascular diseases, hyperlipidemia, diabetes mellitus, renal failure, other liver diseases, malignant tumors ・ Concomitant medication: Other anti-gout preparations, NSAIDs, ARB, ACE inhibitors, beta-blockers, calcium channel blockers, diuretics, new quinolone antibiotic, aminoglycoside antibiotic
4	Whether lansoprazole users have a higher risk of hyponatremia than users of other PPIs?	Inclusion criteria: ・ Patients initiated with any PPI (lansoprazole, omeprazole, rabeprazole, esomeprazole, or vonoprazan) during the study period (1 January 2016 to 31 December 2017).	Baseline covariate: ・ Serum sodium ・ Serum creatinine Outcome: ・ Serum sodium	Sex, age, year of cohort entry, hospitalization, first visit, emergency care (at the date of the first prescription), class number of prescriptions of concomitant medication, complications, concomitant medication (180 days before the date of the first prescription of any antidiabetic drug) ・ Complications:

		・ New users of monotherapy for PPI: Patients with no prescription of the drugs for >180 days before the first prescription of PPI. Exclusion criteria: ・ Patients with a diagnosis of hyponatremia (diseases that contain “sodium” in ICD10: E871) within the 180 days before the first prescription of any PPI.		renal failure, liver cirrhosis, pancreatitis, heart failure, inflammatory bowel disease, adrenal insufficiency, hypothyroidism, helicobacter pylori infection ・ Concomitant medication: antiepileptic drugs, ARB, ACE inhibitors, diuretics, antidepressants, antipsychotics, corticosteroids, NSAIDs, anti-gout preparations, beta-blockers, calcium antagonist, aminoglycoside antibiotics, new quinolone antibiotic
5	Whether DPP-4I users have a higher risk of acute pancreatitis than users of other oral antidiabetic agents?	Inclusion criteria: ・ Patients initiated with oral antidiabetic agents (DPP-4I, BG, SU, or α-GI) during the study period (1 January 2015 to 31 December 2017). ・ New users of monotherapy for oral antidiabetic agents, as in scenario 1. Exclusion criteria: ・ Patients with a diagnosis of acute pancreatitis (ICD10: K850, K851, K853, K858, K859) within the 180 days before the first prescription of any other oral antidiabetic agent.	Baseline covariate: ・ Blood glucose ・ HbA1c ・ Serum amylase Note: ・ We both count HbA1c (JDS) and HbA1c (NGSP) since HbA1c (JDS) can convert to HbA1c (NGSP)	Sex, age, year of cohort entry, hospitalization, first visit, emergency care (at the date of the first prescription), class number of concomitant medications, complications, concomitant medication, examination/operation (180 days before the date of the first prescription of any antidiabetic drug) ・ Complications: Alcoholic pancreatitis, inflammatory bowel disease, peptic ulcer disease, gallstone disease, acute appendicitis, diverticulitis, pancreatic cancer, cholangiocarcinoma, duodenal cancer, alcoholic pancreatitis, chronic kidney disease, hepatitis, liver cirrhosis, fatty liver, alcoholic liver disease, heart failure, acute myocardial infarction, hypertension, cerebrovascular diseases, hyperlipidemia, peripheral vascular disease, diabetic complication ・ Concomitant medication: Antiepileptic drugs, ARB, ACE inhibitors, diuretics, antiarrhythmics class I and III, thiazolidinediones, glinides, SGLT2 inhibitors, insulin, GLP-1 receptor agonists, corticosteroids, estrogen, NSAIDs, codeine, PPIs, H2 antagonists, 5-aminosalicylic acid agents

6	Does the combination of users of interacting antimicrobials and warfarin have a higher risk of bleeding than users of non-interacting antimicrobials and warfarin?	Inclusion criteria: ・ Patients initiated with antimicrobials with (interacting antimicrobials) or without bleeding risk (non-interacting antimicrobials) during the study period (1 January 2015 to 31 December 2017). ・ New users of monotherapy for antimicrobials as in scenario 1. Exclusion criteria: ・ Patients prescribed warfarin less than twice before the first prescription of any antimicrobials ・ Patients aged less than 21 at the first prescription of any antimicrobials Note: ・ The settings were created referring to the previous study.[9]	Baseline covariate: ・ INR	Non-conducted
7	Whether users of olanzapine, quetiapine, or risperidone have a higher risk of diabetes than aripiprazole users?	Inclusion criteria: ・ Patients initiated with any SGA (olanzapine, quetiapine, risperidone, or aripiprazole) during the study period (1 January 2015 to 31 December 2017). ・ New users of monotherapy for SGA. Exclusion criteria: ・ Patients with a diagnosis of diabetes (ICD10: E10, E11, E12, E13, E14, O24) before the first prescription of any SGA. ・ Patients with a diagnosis of polycystic ovary syndrome (ICD10: E282) before the first prescription of any SGA. ・ Patients aged less than 21 at the first prescription of any antimicrobials Note: ・ The setting was created referring to a previous study.[9]	Baseline covariate: ・ Blood glucose	Non-conducted

Abbreviations: PPI, proton pump inhibitor; ICD, international classification of diseases; ARB, angiotensin II receptor blocker; ACE, angiotensin-converting-enzyme; NSAID, nonsteroidal anti-inflammatory drug; SGA, second-generation antipsychotic; FGA, first-generation antipsychotic; HbA1c, hemoglobin A1c; JDS; Japan diabetes society, NGSP, national glycohemoglobin standardization program; ALT, alanine transaminase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; LDL-chol, low-density lipoprotein cholesterol; TG, triglyceride; DPP-4I, dipeptidyl peptidase-4 inhibitor; BG, biguanide; SU, sulfonylurea; α-GI, α-glucosidase inhibitor; INR, international normalized ratio.

Location in the text file: If possible, Table 1 should be placed between the subsections “Scenarios” and “Protocol approval and statistical analysis.”

Figures legends

Figure 1. Frequency of laboratory results recorded within 90 days before prescription in overall cohort.

The assessment includes scenarios 1–7. The vertical axis shows the percentage of patients at each number of records in the overall cohort. The missing proportions (shaded bars) differed among laboratory tests or scenarios but the differences were mostly less than 30%.

Abbreviations: HbA1c, hemoglobin A1c; LDL-chol, low-density lipoprotein cholesterol; TG, triglyceride; ALT, alanine transaminase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; INR, international normalized ratio.

Figure 2. Frequency of laboratory results recorded from prescription to end of observation period in overall cohort.

The assessment includes scenarios 1, 2, and 4. (a) shows the frequency, in which the vertical axis shows the percentage of patients at each number of records in the overall cohort. (b) is box plot of the target period. The missing proportions after prescription (shaded bars) differed from that before prescription.

Abbreviations: HbA1c, hemoglobin A1c; ALT, alanine transaminase; AST, aspartate aminotransferase; ALP, alkaline phosphatase.

Figure 3. Frequency of laboratory results recorded within 84 days after prescription in overall cohort.

The assessment was for scenarios 3. The vertical axis shows the percentage of patients at each number of records in the overall cohort. The missing proportion within 84 days after prescription was approximately 30% of patients.

Figure 4. Missing proportion within 90 days before prescription in each hospital.

The assessment includes scenarios 1–7. The vertical axis shows the missing proportion in each hospital cohort. There were some laboratory tests with hospital differences in the missing proportions.

Abbreviations: HbA1c, hemoglobin A1c; LDL-chol, low-density lipoprotein cholesterol; TG, triglyceride; ALT, alanine transaminase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; INR, international normalized ratio.

Figure 5. The varying association between missingness and affecting factors among hospitals. The assessment includes scenarios 1–5. The presented results are examples of the factors that have been suggested to affect hospital differences in association with missingness.

Abbreviations: OR; odd ratio, HbA1c, hemoglobin A1c; LDL-chol, low-density lipoprotein cholesterol; TG, triglyceride; ALT, alanine transaminase; ALP, alkaline phosphatase; AMI, acute myocardial infarction; HF, heart failure; UA, uric acid; NSAIDs, non-steroidal anti-inflammatory drugs.

Figures

Figure 1.

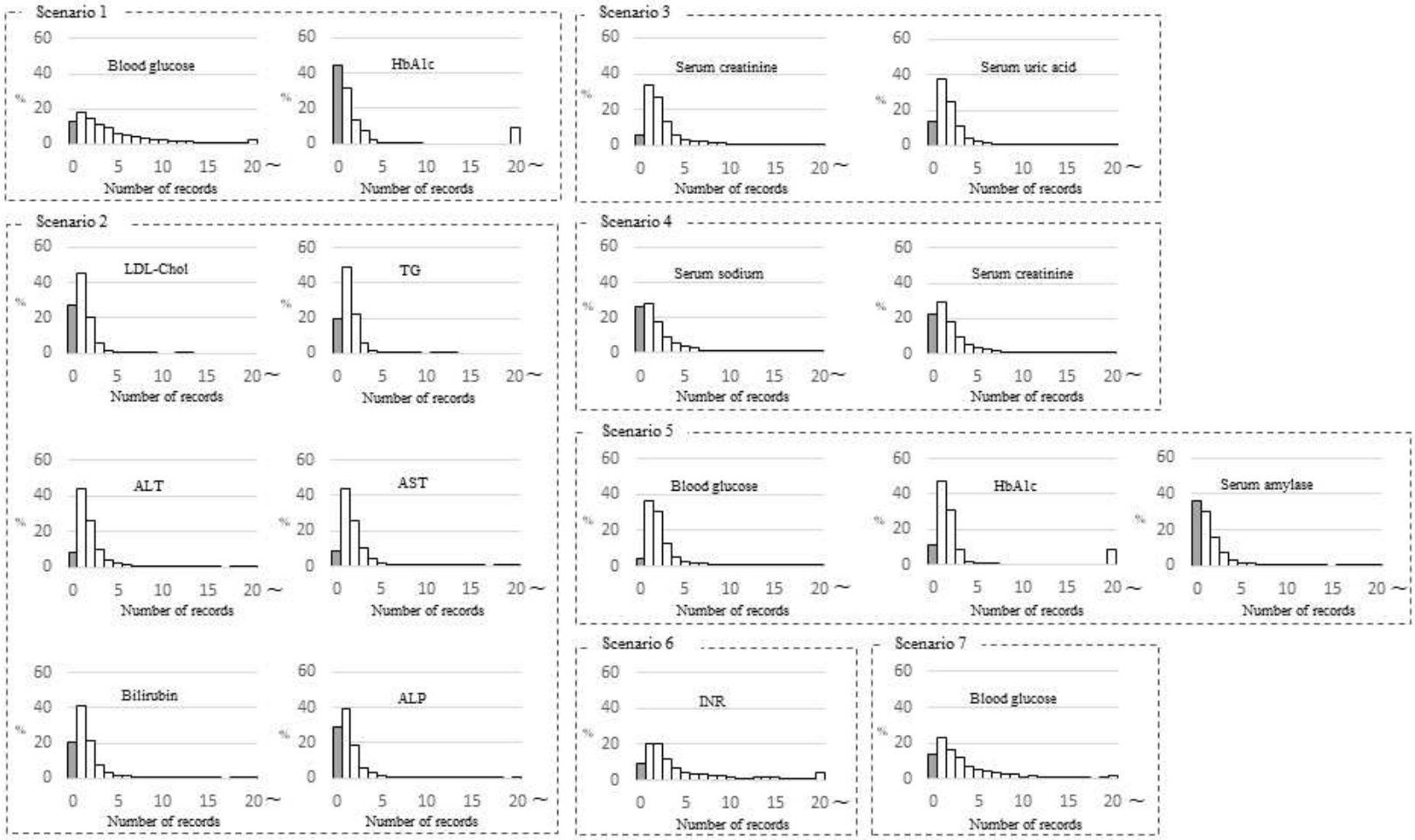


Figure 2.

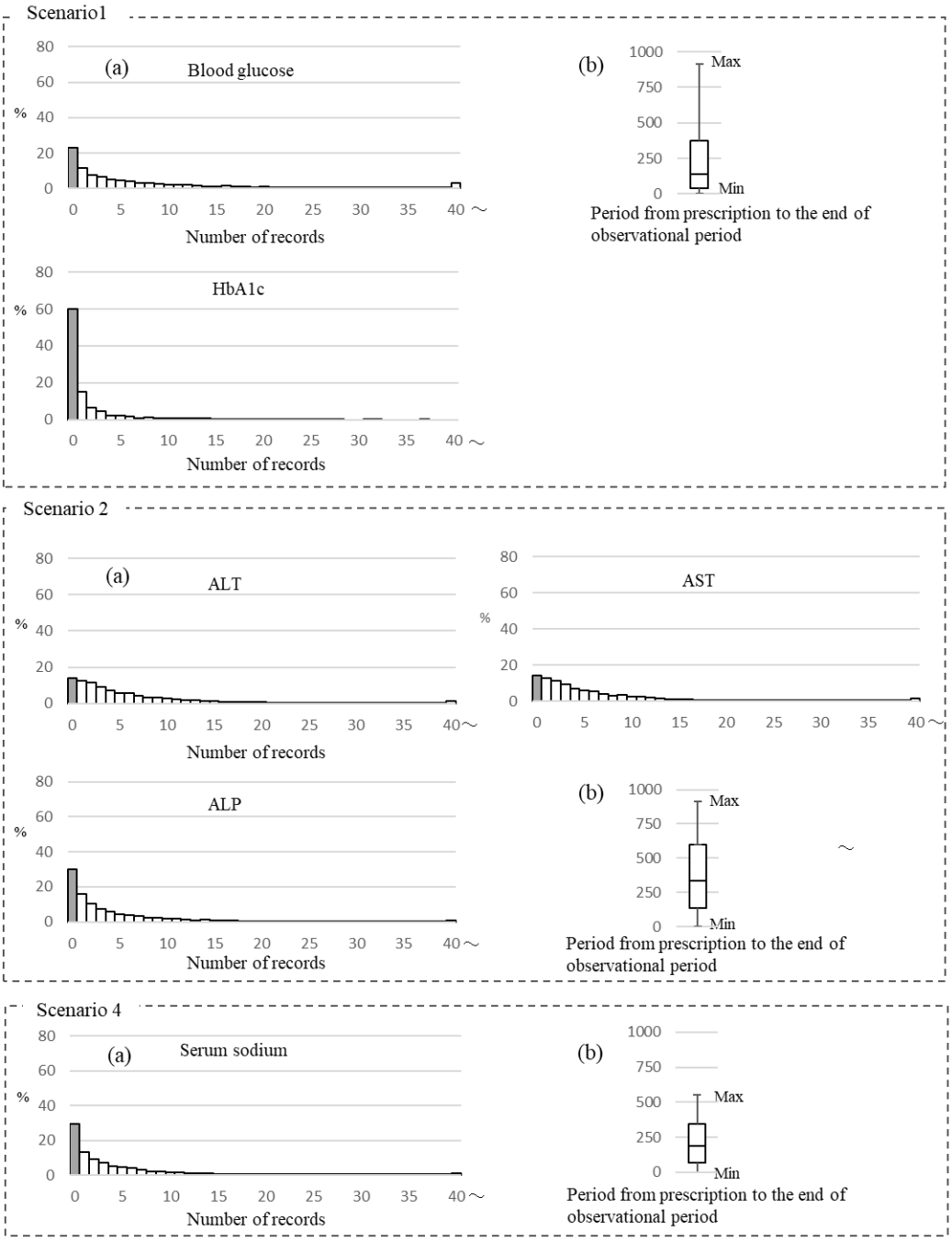

Scenario1
(a)
Blood glucose
%
Number of records
HbA1c
(b)
Max
Min
Period from prescription to the end of observational period
Scenario 2
ALT
AST
ALP
Scenario 4
Serum sodium

Figure 3.

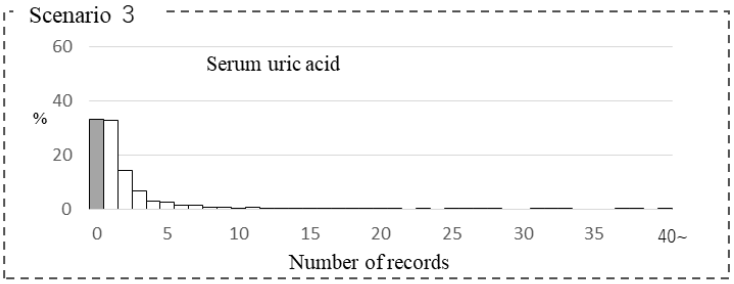

Scenario 3
Serum uric acid
%
60
40
20
0
0
5
10
15
20
25
30
35
40~
Number of records

Figure 4.

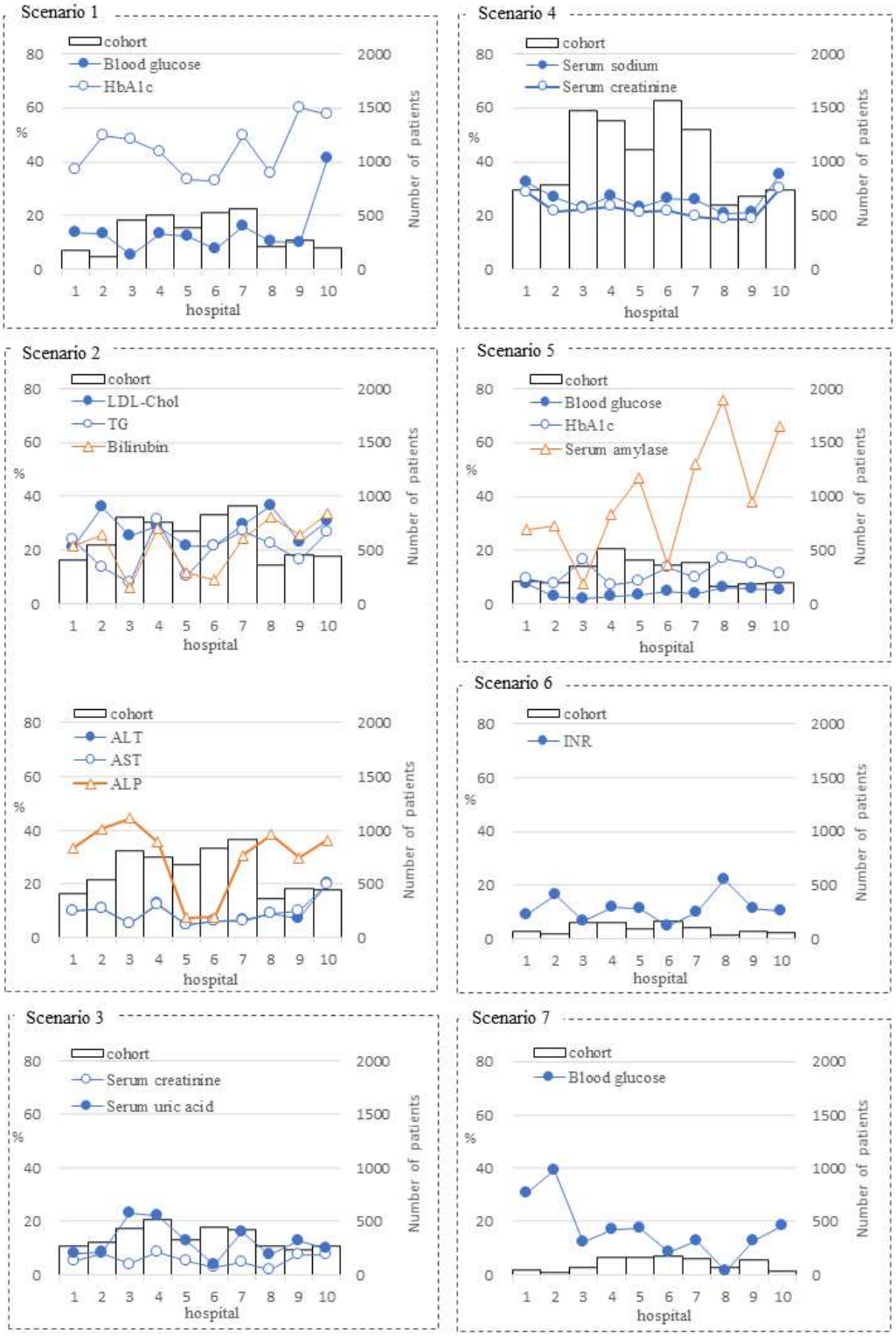
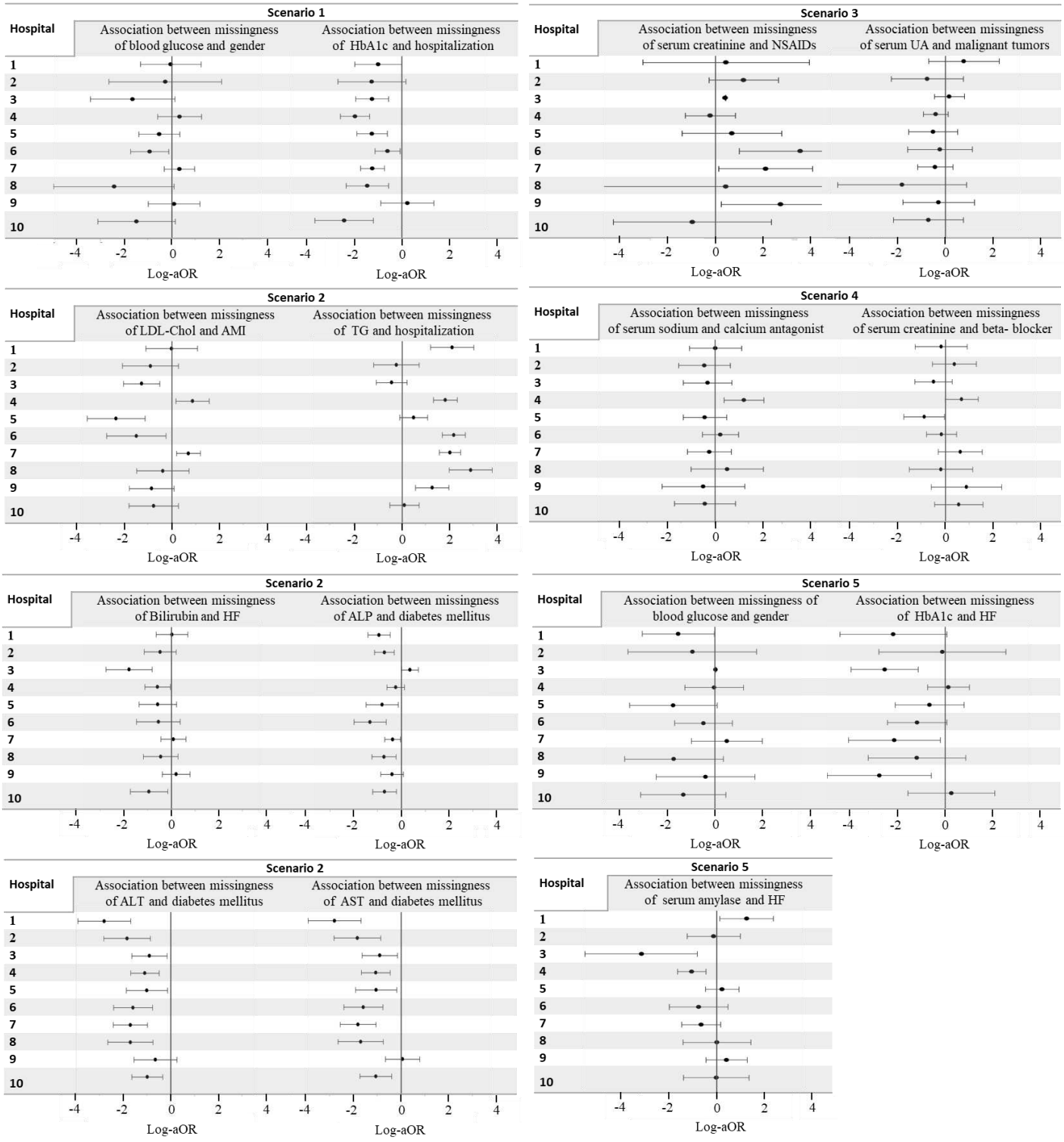

Scenario 1
cohort
Blood glucose
HbA1c
Scenario 4
cohort
Serum sodium
Serum creatinine
Scenario 2
cohort
LDL-Chol
TG
Bilirubin
cohort
ALT
AST
ALP
Scenario 5
cohort
Blood glucose
HbA1c
Serum amylase
Scenario 6
cohort
INR
Scenario 3
cohort
Serum creatinine
Serum uric acid
Scenario 7
cohort
Blood glucose
%
Number of patients
hospital

Figure 5.


Scenario 1
Hospital
Association between missingness of blood glucose and gender
Association between missingness of HbA1c and hospitalization
Scenario 3
Association between missingness of serum creatinine and NSAIDs
Association between missingness of serum UA and malignant tumors
Scenario 2
Association between missingness of LDL-Chol and AMI
Association between missingness of TG and hospitalization
Scenario 4
Association between missingness of serum sodium and calcium antagonist
Association between missingness of serum creatinine and beta- blocker
Scenario 2
Association between missingness of Bilirubin and HF
Association between missingness of ALP and diabetes mellitus
Scenario 5
Association between missingness of blood glucose and gender
Association between missingness of HbA1c and HF
Scenario 2
Association between missingness of ALT and diabetes mellitus
Association between missingness of AST and diabetes mellitus
Scenario 5
Association between missingness of serum amylase and HF
1
2
3
4
5
6
7
8
9
10
-4
-2
0
2
4
Log-aOR

Figures

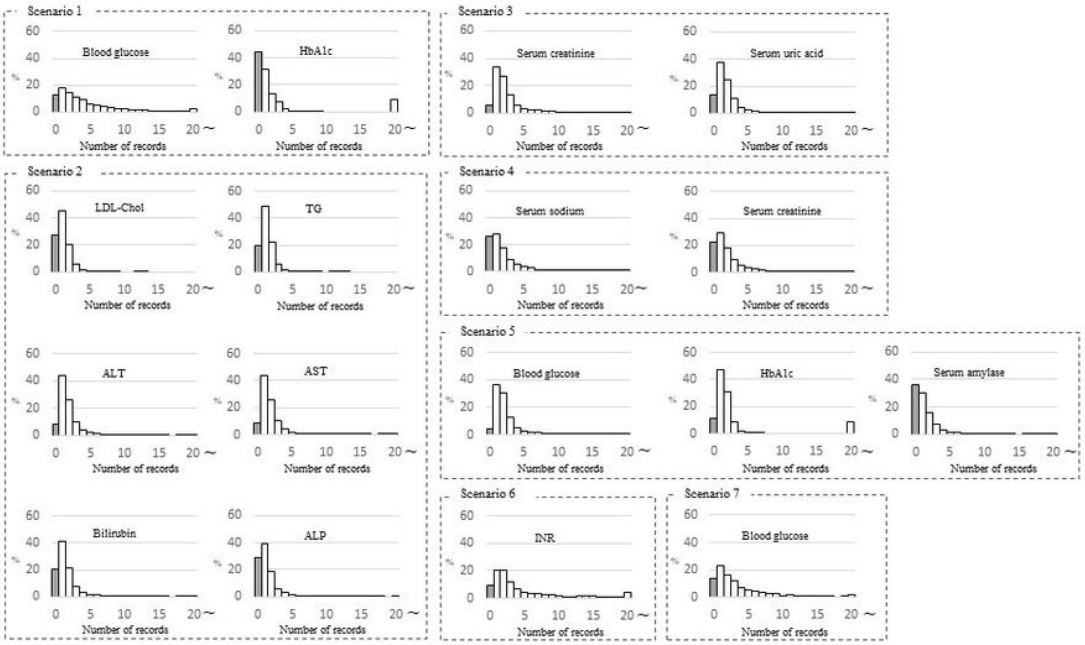


Figure 1

Frequency of laboratory results recorded within 90 days before prescription in overall cohort. The assessment includes scenarios 1–7. The vertical axis shows the percentage of patients at each number of records in the overall cohort. The missing proportions (shaded bars) differed among laboratory tests or scenarios but the differences were mostly less than 30%. Abbreviations: HbA1c, hemoglobin A1c; LDL-chol, low-density lipoprotein cholesterol; TG, triglyceride; ALT, alanine transaminase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; INR, international normalized ratio.

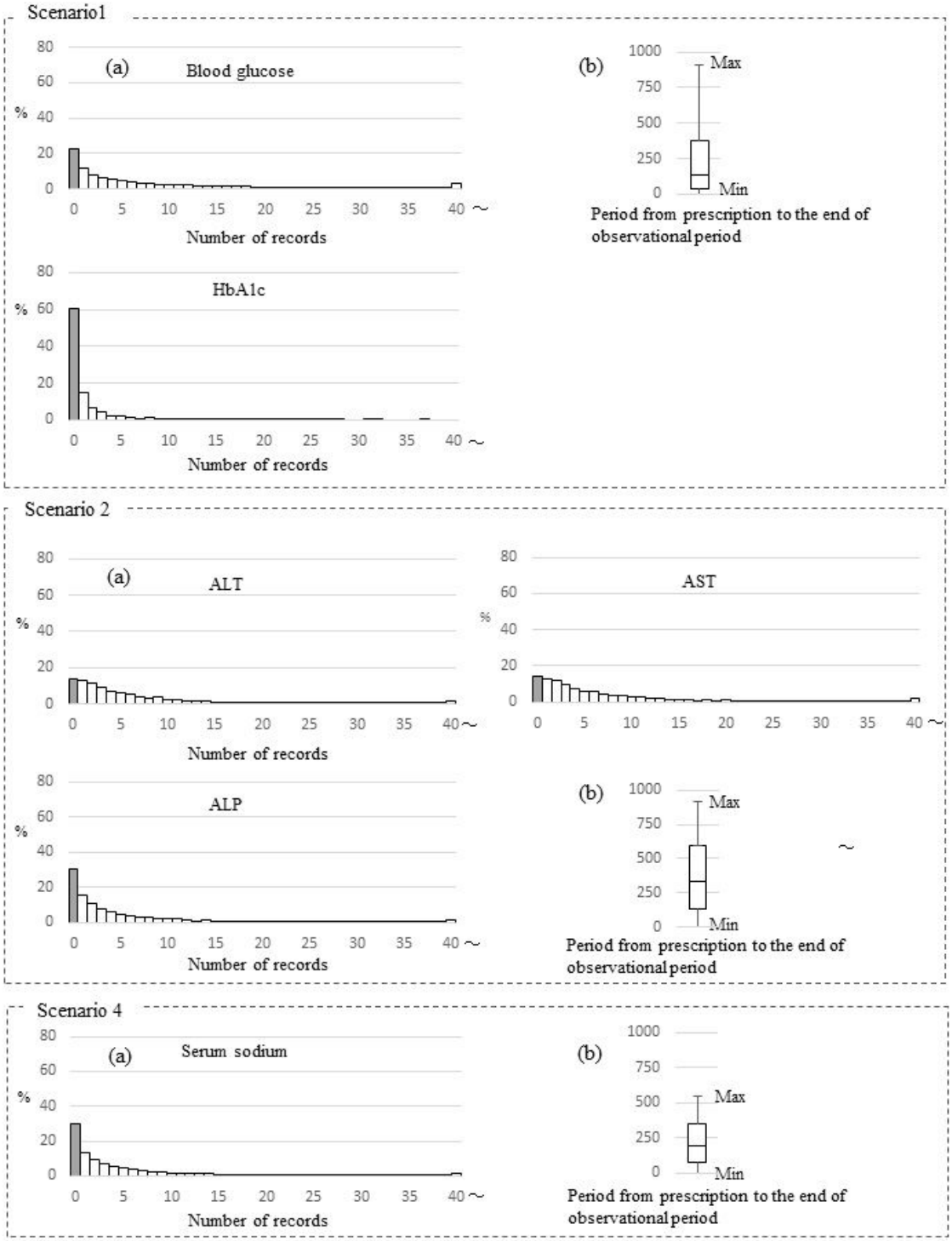


Figure 2

Frequency of laboratory results recorded from prescription to end of observation period in overall cohort. The assessment includes scenarios 1, 2, and 4. (a) shows the frequency, in which the vertical axis shows the percentage of patients at each number of records in the overall cohort. (b) is box plot of the target period. The missing proportions after prescription (shaded bars) differed from that before prescription.

Abbreviations: HbA1c, hemoglobin A1c; ALT, alanine transaminase; AST, aspartate aminotransferase; ALP, alkaline phosphatase.

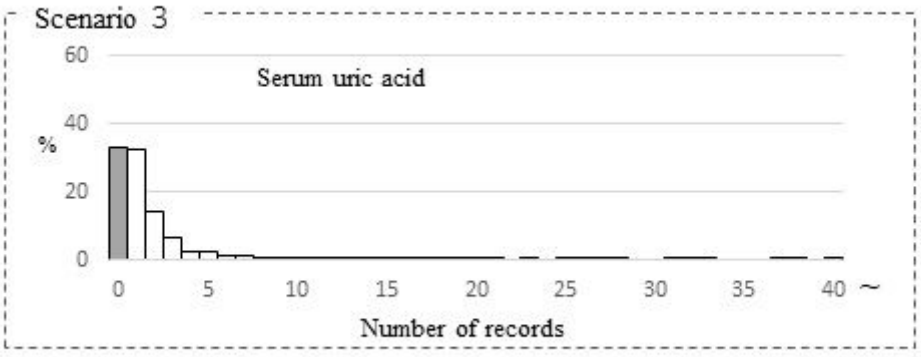


Figure 3

Frequency of laboratory results recorded within 84 days after prescription in overall cohort. The assessment was for scenarios 3. The vertical axis shows the percentage of patients at each number of records in the overall cohort. The missing proportion within 84 days after prescription was approximately 30% of patients.

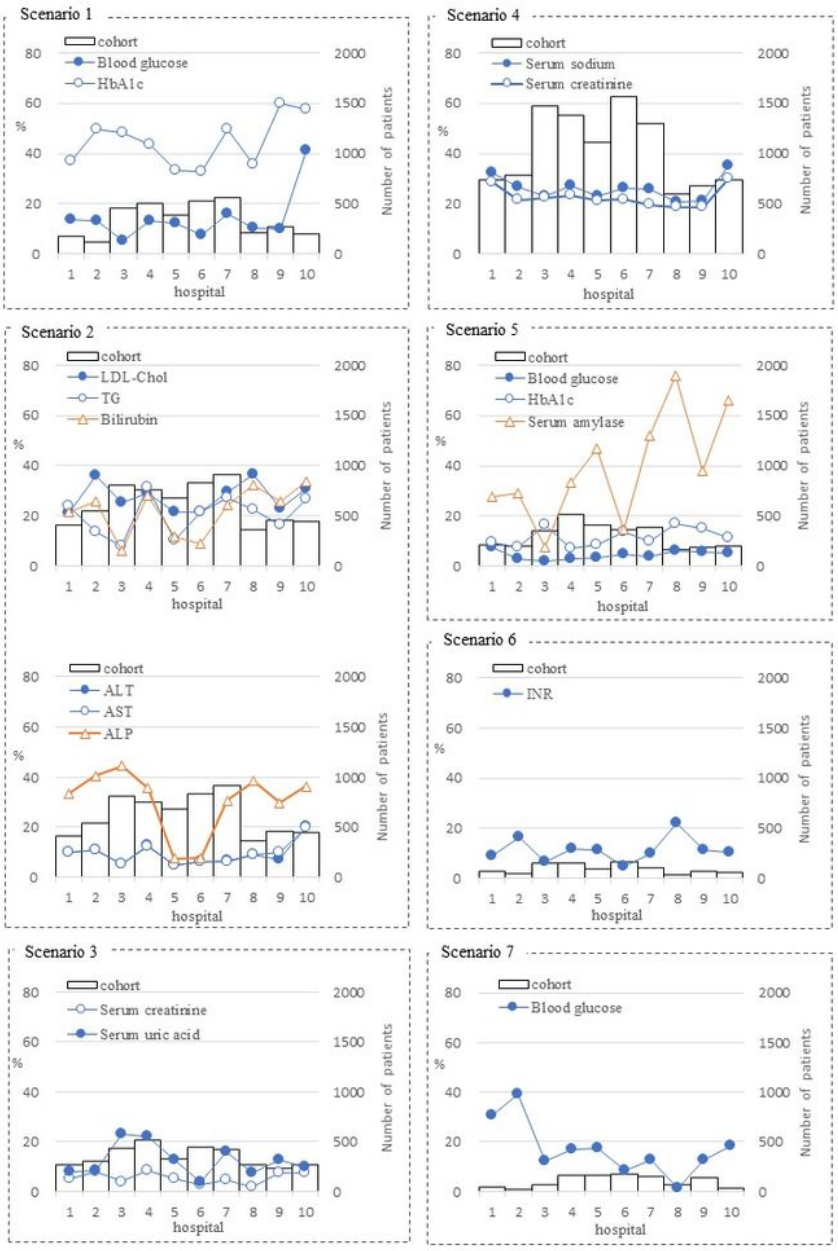


Figure 4

Missing proportion within 90 days before prescription in each hospital. The assessment includes scenarios 1–7. The vertical axis shows the missing proportion in each hospital cohort. There were some laboratory tests with hospital differences in the missing proportions. Abbreviations: HbA1c, hemoglobin A1c; LDL-chol, low-density lipoprotein cholesterol; TG, triglyceride; ALT, alanine transaminase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; INR, international normalized ratio.

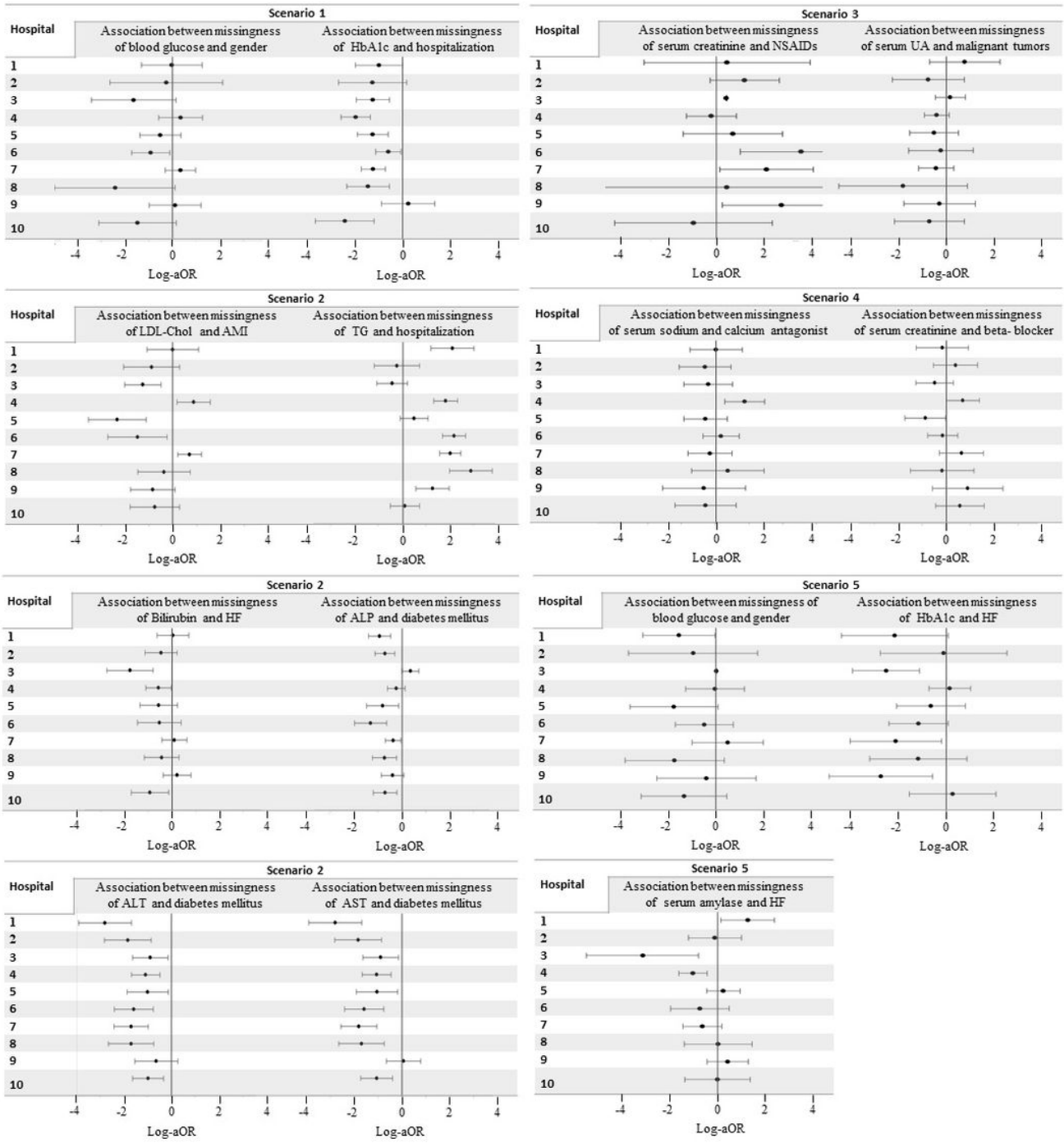


Figure 5

The varying association between missingness and affecting factors among hospitals. The assessment includes scenarios 1–5. The presented results are examples of the factors that have been suggested to affect hospital differences in association with missingness. Abbreviations: OR; odd ratio, HbA1c, hemoglobin A1c; LDL-chol, low-density lipoprotein cholesterol; TG, triglyceride; ALT, alanine transaminase; ALP, alkaline phosphatase; AMI, acute myocardial infarction; HF, heart failure; UA, uric acid; NSAIDs, non-steroidal anti-inflammatory drugs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplements.pdf