RESEARCH

Exact-matching algorithms using administrative health claims database equivalence factors for real-world data analysis based on the target trial emulation framework

Yuichi Nishioka[1,2] · Emiri Morita[1] · Saki Takeshita[1] · Sakura Tamamoto[1] ·
Tomoya Myojin[1] · Tatsuya Noda[1] · Tomoaki Imamura[1]

Abstract

Real-world data have become increasingly important in medical science and healthcare. A new, effective, and practically feasible statistical design is needed to unlock the potential of real-world data that decision-makers and practitioners can use to meet people's healthcare needs. In the first half of the study, we validated our proposed new method by simulation, and in the second half, we conducted a clinical study on actual real-world data. We proposed the "Exact Matching Algorithm Using Administrative Health Claims Database Equivalence Factors (AHCDEFs)" using a target trial emulation framework. The simulation trials were conducted 500 times independently, considering the misclassification and chance errors of all variables and competing events of outcome. Two conventional methods, multivariate and propensity score analyses, were compared. Next, we estimated the effect of specific health guidance provided in Japan on the prevention of diabetes onset and medical expenditures. Our proposed novel method for real-world data returns improved estimates and fewer type I errors (the probability of erroneously determining that there is a difference when, in fact, there is no difference) than conventional methods. We quantitatively demonstrated the effectiveness of specific health guidance in Japan in preventing the onset of diabetes and reducing medical expenditures during five years. We proposed a new method for analyzing real-world data and an exact-matching algorithm using AHCDEFs. The larger the number of patients available for analysis, the more the AHCDEFs that can be matched, thereby removing the influence of confounding factors. This method will generate significant evidence when applied to real-world data.

✉ Yuichi Nishioka
  y_n@naramed-u.ac.jp

1 Department of Public Health, Health Management and Policy, Nara Medical University, 840 Shijyo-cho, Kashihara City, Nara 634-8521, Japan

2 Department of Diabetes and Endocrinology, Nara Medical University, Kashihara, Nara, Japan

 Springer

Abbreviations

AHCDEFs	Administrative Health Claims Database Equivalence Factors
BMI	Body Mass Index
EMA	Exact-Matching Algorithm using AHCDEFs
HbA1c	Hemoglobin A1c
HSLRG	Health Science and Labour Research Grants
RCT	Randomized Controlled Trial
SD	Standard Deviation
TTE	Target Trial Emulation

1 Background

Real-world data have become increasingly important in medical science and healthcare (Mellinghoff et al. 2022; Sun et al. 2018). Administrative claims databases contain larger data in many countries, and have great potential for new scientific discovery, solving problems, and making decisions that are otherwise unfeasible (Hernán and Robins 2016). Meanwhile, new, effective, and practically feasible statistical designs are needed to unlock the potential of real-world data; decision makers and practitioners can apply the results and conclusions to better meet the medical and healthcare needs of our society (Frieden 2017; Baumfeld Andre et al. 2020).

With the target trial emulation (TTE) framework, an increasing amount of real-world data is being utilized (Matthews et al. 2022; Xie et al. 2020; Madenci et al. 2021; Takeuchi et al. 2021). However, TTE is not always properly understood. The idea of TTE is not to bring observational studies closer to randomized controlled trials (RCTs), but to bring clinical studies, including observational studies and RCTs, closer to the "ideal study" that can settle a research question (Hernán and Robins 2016). For example, in an ideal comparison, the same people are divided and different interventions are used to compare outcomes. This is clearly an ideal study, but an impossible one. Therefore, in RCTs, randomization is used, which is mimicked in propensity score (PS) analysis (Liang et al. 2014). Randomization is a method of approaching the ideal, and not the goal. Based on the TTE approach, a study has greater comparability if all factors are matched, in all pairs, between the comparison and control groups.

One study suggests that the correctness of the estimates requires designing the study and analyzing the data based on principle (Hoffman et al. 2022). The purpose of this study is to verify a new method for real-world data analysis, based on the TTE framework.

2 Methods

In the first half of the study, we validated the new method by simulation, and in the second half, using this method, we conducted a clinical study on actual real-world data. As the subject of the latter half of the study, we estimated the effect of the specific health guidance provided in Japan, on the prevention of diabetes onset and medical expenditures. This subject was chosen because the effect of lifestyle guidance in preventing the onset of diabetes is almost self-evident (Muramoto et al. 2014), making it difficult to conduct an RCT from an ethical perspective. Therefore, real-world evidence is required to estimate this effect.

This observational cohort study used administrative claims data and was approved by the Ethics Committee of Nara Medical University (approval no. 1123–7, October 8th, 2015).

2.1 Outline for a new method: exact-matching algorithm using administrative health claims database equivalence factors

This novel method is based on a very simple idea: according to the TTE concept, an ideal study in terms of comparability (= Target Trial) should have all covariates, except exposure, matched in both groups. However, the key to this simple approach is to match each of the covariates, not just the representative values of the covariates, as in an RCT or the PS in an observational study. This differs from traditional statistical methods, in that all measurable factors, including interactions of any order, are controlled. It is also characterized by the fact that the differences in the covariates are always zero and not distributed, regardless of the stratification of the analysis. Administrative health claims database equivalence factors (AHCDEFs) are collected from administrative claims databases and can confound exposure and outcome. The weighting factor for the group with a smaller number of people therein without exposure, and the group with exposure, among those with perfectly matched AHCDEFs, was set to 1, so that the sum of the weighting factors for the group without exposure and the group with exposure, was equal. For example, consider a simple case in which the AHCDEFs are age, gender, and systolic and diastolic blood pressure: consider a 42-year-old male with a systolic/diastolic blood pressure of 130/80 mmHg. The control for that case is a 42-year-old male with a blood pressure of 130/80 mmHg. Here, all four items are matched in the case and the subject; along with the four items, the 11 interaction terms of any combination of the four items are also controlled. The difference in this approach, compared to the conventional method that assumes agreement of representative values in the population, is that we are able to control for interaction terms of any order in the model.

2.2 Simulation procedure

The trials were conducted 500 times, independently (t = 1, 2, $\cdots$, 500) considering the misclassification and chance errors of all variables and competing events of outcome. The simulation process followed a sufficient causal model (Liang et al. 2014). Thus, it is only when there are equal or more than one sufficient causes that an event occurs. For each patient (n = 1, 2, $\cdots$, 50000), we constructed a confounding model of exposure (X_{t1n}) and outcome (Y_{tn}) of interest. That is, there was no difference in the true proportions of outcome occurrence between the groups with and without exposure, and the true odds ratio of outcome occurrence for the group exposed to the group without exposure was 1. X_{tin} is the low-to-moderate correlated variable. X_{t1n} is observable, $X_{t2n}, \cdots, X_{t100n}$ contain both observable and unobservable variables. Any of the latter 99 variables causes Y. to estimate the effect of X_{t1n} on Y_{tn}, when only some component causes are observable and known and there may be some noncausal factors being mistaken as causal factors. Thus, we have

$$P\left(Y_{tn} = 1 | X_{tin}, K_{ti}\right) = \frac{exp\left(\beta_{t1}X_{t1n} + \sum_{i=2}^{100} \beta_{ti}X_{tin}K_{ti}\right)}{1 + exp\left(\beta_{t1}X_{t1n} + \sum_{i=2}^{100} \beta_{ti}X_{tin}K_{ti}\right)}$$

If X_{tin} is an observable and known risk factor, $K_{ti} = 1$, otherwise $K_{ti} = 0$ (i = 2, $\cdots$, 100). For some i, $K_{ti} = 1$ and X_{tin} is a sufficient cause of Y. β_{ti} is the estimated effect of X_{tin} on Y_{tn} if $K_{ti} = 1$. In summary, X_{t1n} is a factor that is observed and is not a sufficient cause of Y. Among the remaining 99 X_{tin}, there is at least one that is a sufficient cause of Y with $K_{ti} = 1$. When $K_{ti} = 0$, X_{tin} is either unobservable or observable (as not a risk factor).

2.3 Details of the simulation

To set low-to-moderate correlated variables $X_{t1n}, X_{t2n}, \cdots, X_{t100n}$, independent random sampling of V_{tin} and U_{tn} was performed from uniform [0,1) distributions. All V_{tin} are $500 \times 100 \times 50{,}000$ independent samples, and all U_{tn} are $500 \times 50{,}000$ independent samples. Independent random sampling of P_{ti} was performed from uniform [0.3,0.6) distributions. All P_{ti} are 500×100 independent samples. For $j = 1, \cdots, 9$, A_{tji} takes on a random value, drawn from a uniform [0.5, 0.9) distribution. All A_{tji} are $500 \times 9 \times 100$ independent samples. G_{tjin} is an unknown value based on an exact threshold value, with no misclassifications. $G_{tjin} = 1$, if a linear combination of variables $V_{tin}P_{ti} + U_{tn}(1 - P_{ti})$ > A_{tji}. $G_{tjin} = 0$, if a linear combination of variables $V_{tin}P_{ti} + U_{tn}(1 - P_{ti}) \leq A_{tji}$. X_{tin} is the known value, based on the expected threshold value of a uniform [0.5, 0.9) distribution, that is, 0.7 with some misclassifications. $X_{tin} = 1$ if the linear combination of variables $V_{tin}P_{ti} + U_{tn}(1 - P_{ti})$ > 0.7. $X_{tin} = 0$ if a linear combination of the variables $V_{tin}P_{ti} + U_{tn}(1 - P_{ti}) \leq 0.7$.

C_{tji} takes a random value, drawn from the Bernoulli distribution, with a probability of success of 0.05. All C_{tji} are $500 \times 9 \times 50{,}000$ independent samples, which indicate whether a linear combination of variables $V_{tin}P_{ti} + U_{tn}(1 - P_{ti})$ is a component of the j th possible sufficient cause. Let $O_{tjn} = 1$, when all components for the jth possible sufficient cause are active in the n th observation of t th trial, that is $\sum_{i=2}^{100} G_{tjin} C_{tji} = \sum_{i=2}^{100} C_{tji}$. If $\exists t, j$ $\sum_{i=2}^{100} G_{tjin} C_{tji} < 2$, all variables are redetermined through the same j th random process in the t th trial. F_{tj} takes on a random value, drawn from the Bernoulli distribution, with a probability of success of 0.5. All F_{tj} are $500 \times 9 \times 50{,}000$ independent samples, which show whether each of the nine possible sufficient causes is a real sufficient cause. If there is no real sufficient cause, that is, $\exists t \sum_{j=1}^{9} F_{tj} = 0$, then all variables are redetermined through the same random process in the t th trial.

E_{tn}, and Q_{tn} take a random value, drawn from the Bernoulli distribution, with a probability of success of 0.001. All E_{tn} are $500 \times 50{,}000$ independent samples, which indicate competing events for the outcome. All Q_{tn} are $500 \times 50{,}000$ independent samples, indicating a small error in the outcome. $Y_{tn} = 1$ where $Q_{tn} = 1$, and where $\sum_{j=1}^{9} O_{tjn} F_{tj} \geq 1$ and $E_{tn} = 0$; otherwise $Y_{tn} = 0$. K_{ti} is a random value, drawn from the Bernoulli distribution, with a probability of success of $0.1 + \sum_{j=1}^{9} C_{tji} F_{tj}$, which indicates whether X_{ti} is an observable and known risk factor for the outcome.

2.4 Adjustment by new methods

In the above model, there was no difference in the true proportion of outcome occurrence between the groups with and without exposure. However, the bias in the

model resulted in an apparent difference in the proportion of outcomes. To visualize the extent to which the bias was adjusted by our proposed method, we simulated $P(Y_{tn} = 1|X_{t1n} = 1) - P(Y_{tn} = 1|X_{t1n} = 0)$ with and without adjustment.

2.5 Comparison with conventional methods

Two conventional methods, multivariate analysis and PS, were compared using this method (Seeger et al. 2005; Schneeweiss et al. 2009). In the model, the true odds ratio of outcome occurrence for the group exposed to the group without exposure was 1. Specifically, the odds ratios β_{ti} of the three methods were compared. Type I error is the probability of erroneously determining that there is a difference when, in fact, there is no difference. Similarly, the probabilities of Type-I errors for the three methods were determined. In the PS analysis, PSs themselves were used as covariates in the regression model, because the variance in both groups was not very different (Rubin 1979).

2.6 Application to real-world data: data source

In the demonstration of the methodology, the study cohort comprised anonymized data of individuals in the Kokuho database (KDB) of Nara prefecture in Japan. This data provided information on personal identifiers, date, age group, sex, description of the procedures performed, World Health Organization International Classification of Diseases (ICD-10) diagnosis codes, medical care received, medical examinations conducted without the results, prescribed drugs, and specific health check-ups, including results from 2013 to 2021.

2.7 Study population

We included data on individuals who underwent specific health check-ups between April 2014 and March 2021. Those who were not followed up in the past year and those who were prescribed diabetic medications in the past year were excluded.

2.8 Definition of diabetes

A validated algorithm was used to define diabetes based on claims data from Japan. This algorithm (74.6% sensitivity and 88.4% positive predictive value), for detecting people with diabetes, had three elements: the diagnosis-related codes for diabetes without the "suspected" flag, the medication codes for diabetes, and these two codes on the same record (Nishioka et al. 2022). This algorithm cannot detect people with diabetes who have not consulted a doctor and those on diet and exercise therapy only, but it can identify most of them on medication.

2.9 Effect of the specific health guidance

Among those who received the specified health check-up during the six-month period, those who received health guidance were identified and designated as the health guidance group. Those whose age, sex, body mass index (BMI), abdominal circumference,

medical expenses in the past year, the number of days of outpatient visits in the past year, and the number of days of hospitalization in the past year matched those of the health guidance group were identified and designated as the control group. In the health guidance and control groups, odds ratios for new onset diabetes were calculated. Generalized estimating equations were used in the analysis, and a binomial distribution was assumed for the outcome of diabetes. The link function was assumed to be a logit. In an additional analysis, average medical expenditures per person per month, for both groups, were observed over time. All statistical tests were two-tailed, and P-values < 0.05 were considered statistically significant. All statistical analyses were performed using the Microsoft SQL Server 2016 Standard (Microsoft Corp., Redmond, WA, USA) and IBM SPSS Statistics for Windows, version 27.0 (IBM, Armonk, NY, USA).
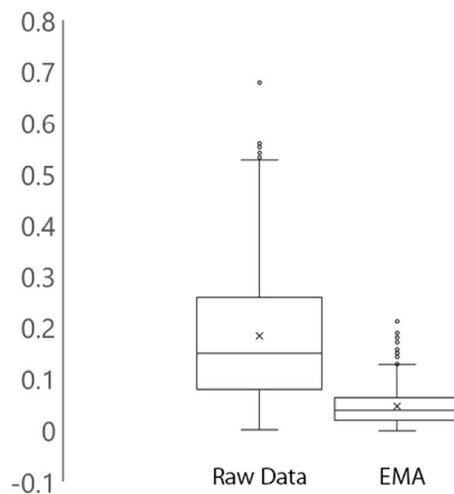
Fig. 1 Box plot of estimated difference in incidence proportions of outcome with and without exposure EMA; Exact Matching Algorithm using AHCDEFs AHCDEFs; Administrative Health Claims Database Equivalence Factors Trials conducted 500 times independently $(t = 1, 2, \cdots, 500)$. The simulation process followed a sufficient causal model. Box plots depict descriptive statistics of the estimated values. The mean is represented by x, the median by the line across, and the first (Q1) and third (Q3) quartiles by the bottom and top of the box, respectively. The box is drawn from Q1 to Q3, with a horizontal line drawn in the middle to denote the median. The interquartile range (IQR) is the distance between the upper and lower quartiles. The boundaries are based on range 1.5 IQR value. From above Q3, a distance of 1.5 times the IQR is measured, and a whisker is drawn up to the largest observed data point from the dataset that falls within this distance. Similarly, a distance of 1.5 × the IQR is measured below Q1, and a whisker is drawn down to the lowest observed data point from the dataset that falls within this distance. All other observed data points outside the whisker boundaries are plotted as outliers on a box plot, as small circles
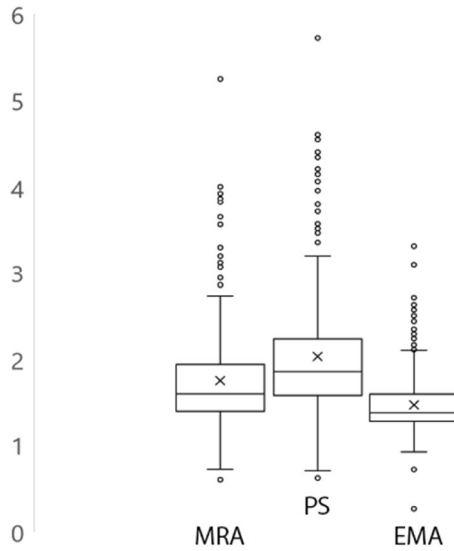
Fig. 2 Box plot of estimated odds ratios of outcome with and without exposure by the methods of adjustment Trials were conducted 500 times independently (t = 1, 2, ⋯ , 500). The simulation process followed a sufficient causal model. Box plots depict the descriptive statistics of the estimated odds ratios. The mean is represented by x, the median by the line across, and the first (Q1) and third (Q3) quartiles by the bottom and top of the box, respectively. The box is drawn from Q1 to Q3, with a horizontal line drawn in the middle to denote the median. The interquartile range (IQR) is the distance between the upper and lower quartiles. The boundaries are based on range 1.5 IQR value. From above Q3, a distance of 1.5 times the IQR is measured, and a whisker is drawn up to the largest observed data point from the dataset that falls within this distance. Similarly, a distance of 1.5×the IQR is measured below Q1, and a whisker is drawn down to the lowest observed data point from the dataset that falls within this distance. All other observed data points outside the whisker boundaries are plotted as outliers on a box plot, as small circles. MRA; Multivariate Regression Analysis PS; Propensity Score EMA; Exact Matching Algorithm using AHCDEFs AHCDEFs; Administrative Health Claims Database Equivalence Factors

3 Results

3.1 Adjustment by new methods

We estimated $P\left(Y_{tn} = 1 | X_{t1n} = 1\right) - P\left(Y_{tn} = 1 | X_{t1n} = 0\right)$ in Fig. 1. The new method tended to make the estimated value closer to the true value of 0 and reduced the variation in the estimated values.

3.2 Comparison with conventional methods

Figure 2 shows the estimated odds ratios of outcome occurrence for the group exposed to the group without exposure, among the three methods. Compared to the multivariate model and PS, the estimates for the new method approach the true odds ratio of 1, with a smaller scatter. Table 1 shows the probability of a type I error when the odds ratios were estimated using the three methods and the univariate model. The probability of a type I error for the new method was 6.6%, while the univariate, multivariate, and PS models all had a probability of 97% or higher.

Table 1 Number of type I errors of each adjustment method

	Univariate regression model	Multivariate regression model	Propensity score (PS) model	Exact-matching algorithms using AHCDEFs
Trial (times)	500	500	500	500
Type I error (times)	497	485	492	33
Proportion of type I error	99.4%	97.0%	98.4%	6.6%

AHCDEFs; Administrative Health Claims Database Equivalence Factors

The trials were conducted 500 times independently ($t = 1, 2, \cdots, 500$) considering the misclassification and chance errors of all variables and competing events of outcome. The simulation process followed a sufficient causal model. Thus, for an event to occur, only when equal to or more than 1 sufficient cause $(X_{t2n}, \cdots, X_{t100n})$ to occur. For each patient ($n = 1, 2, \cdots, 50,000$), we constructed a confounding model of exposure (X_{t1n}) and outcome (Y_{tn}) of interest. That is, the true odds ratio of outcome occurrence for the group exposed to the group without exposure was 1. Type I error is an erroneous statistical determination of the odds ratio, which is not equal to 1 when there is no difference. Although the probability of type I error is usually controlled within 5%, it can actually be much higher than 5% if the confounding factors are poorly adjusted

Univariate regression model

$$P\left(Y_{tn} = 1|X_{tin}\right) = \frac{exp\left(\beta_{t1}X_{t1n}\right)}{1+exp\left(\beta_{t1}X_{t1n}\right)}$$

Multivariate regression model:

$$P\left(Y_{tn} = 1|X_{tin}, K_{ti}\right) = \frac{exp\left(\beta_{t1}X_{t1n}+\sum_{i=2}^{100}\beta_{ti}X_{tin}K_{ti}\right)}{1+exp\left(\beta_{t1}X_{t1n}+\sum_{i=2}^{100}\beta_{ti}X_{tin}K_{ti}\right)}$$

Propensity score model:

$$P\left(Y_{tn} = 1|X_{tin}, K_{ti}\right) = \frac{exp\left(\beta_{t1}X_{t1n}+PS\left(X_{t1n}=1|X_{tin}, K_{ti}\right)\right)}{1+exp\left(\beta_{t1}X_{t1n}+PS\left(X_{t1n}=1|X_{tin}, K_{ti}\right)\right)}$$

$$PS\left(X_{t1n} = 1|X_{tin}, K_{ti}\right) = \frac{exp\left(\sum_{i=2}^{100}\beta_{ti}X_{tin}K_{ti}\right)}{1+exp\left(\sum_{i=2}^{100}\beta_{ti}X_{tin}K_{ti}\right)}$$

Exact Matching Algorithms using AHCDEFs:

$$P\left(Y_{tn} = 1|X_{tin}, K_{ti}\right) = \frac{exp\left(\beta_{t1}X'_{t1n}+\sum_{i=2}^{100}\beta_{ti}X'_{tin}K_{ti}\right)}{1+exp\left(\beta_{t1}X'_{t1n}+\sum_{i=2}^{100}\beta_{ti}X'_{tin}K_{ti}\right)}$$

If $\frac{NumberofPatients\left(\forall i \geq 2, X_{tin}=X_{tiN}|X_{t1n}=1\right)}{NumberofPatients\left(\forall i \geq 2, X_{tin}=X_{tiN}|X_{t1n}=0\right)} < 1,$

$$X'_{tin} = \begin{cases} X_{tin} * \frac{NumberofPatients\left(\forall i \geq 2, X_{tin}=X_{tiN}|X_{t1n}=1\right)}{NumberofPatients\left(\forall i \geq 2, X_{tin}=X_{tiN}|X_{t1n}=0\right)} & ifX_{t1n}=0) \\ X_{tin} & ifX_{t1n}=1) \end{cases}$$

If $\frac{NumberofPatients\left(\forall i \geq 2, X_{tin}=X_{tiN}|X_{t1n}=1\right)}{NumberofPatients\left(\forall i \geq 2, X_{tin}=X_{tiN}|X_{t1n}=0\right)} > 1,$

$$X'_{tin} = \begin{cases} X_{tin} * \frac{NumberofPatients\left(\forall i \geq 2, X_{tin}=X_{tiN}|X_{t1n}=0\right)}{NumberofPatients\left(\forall i \geq 2, X_{tin}=X_{tiN}|X_{t1n}=1\right)} & ifX_{t1n}=1) \\ X_{tin} & ifX_{t1n}=0) \end{cases}$$

AHCDEFs are X_{tin} if $K_{ti} = 1$ and $i \geq 2$, including at least one real sufficient cause

3.3 Effects of specific health guidance in preventing the onset of diabetes mellitus

In total, 6332 individuals (4964 duplicate pairs) were enrolled. Of these, 240 were diagnosed with diabetes. Table 2 shows the characteristics of the study cohort. The odds ratio for type 2 diabetes in a participant was 0.75 (95%CI 0.58–0.97), if the participant was provided specific health guidance, compared to those who did not.

Table 2 Background after exact-matching algorithm using AHCDEFs

AHCDEFs	Specific health guidance provided	Specific health guidance not provided
Number of patients	4964	4964
Number of females (proportion)	1049 (21.1%)	1049 (21.1%)
Age Mean $\pm$ SD	67.4 $\pm$ 5.4	67.6 $\pm$ 5.6
BMI Mean $\pm$ SD	24.4 $\pm$ 1.5	24.4 $\pm$ 1.5
Abdominal girth Mean $\pm$ SD	88.7 $\pm$ 3.4	88.7 $\pm$ 3.4
HbA1c Mean $\pm$ SD	5.6 $\pm$ 0.2	5.6 $\pm$ 0.2
Medical expenditures in the past year (USD) Mean $\pm$ SD	545 $\pm$ 489	575 $\pm$ 489
Number of days in hospital in the past year Mean $\pm$ SD	0.0 $\pm$ 0.2	0.0 $\pm$ 0.2
Visiting hospital days in the past year Mean $\pm$ SD	5.8 $\pm$ 5.9	6.2 $\pm$ 5.9

AHCDEFs; Administrative Health Claims Database Equivalence Factors

SD; standard deviation

BMI; body mass index [kilograms/meter^2]

Abdominal girth [cm]

HbA1c; hemoglobin A1c [%]

USD/JPY was 150 (October 21, 2022)

Those who received health guidance during the six-month period among those who received the specified health check-up were identified and designated as the health guidance group. Those whose age, sex, BMI, abdominal circumference, medical expenses in the past year, number of days of outpatient visits in the past year, and number of days of hospitalization in the past year matched those of the health guidance group were identified and designated as the control group. Odds ratios for new onset of diabetes were determined for the health guidance group and its control group that were able to select a control group

3.4 Effects of specific health guidance in reducing medical expenditures

Figure 3 presents data on medical expenditures per person per month, with and without specific health guidance. After adjusting for background factors, medical expenditures were lower in the group that received health guidance throughout the study period. The results of simple aggregation and PS matching are shown in the Supplementary Figure as a comparison with conventional methods. These are different from the results of the new method employed in this study, and baseline adjustment was considered inadequate.

4 Discussion

In this study, we showed that our proposed novel method for real-world data returned improved estimates and fewer type I errors than conventional methods. Using this new method, we also quantitatively demonstrated the effectiveness of specific health guidance in Japan, in preventing the onset of diabetes and reducing medical expenditures during five years.

In contrast, most previous studies have not shown the effectiveness of specific health guidance in Japan. Creating reliable evidence from complex longitudinal data is not easy, and many studies may have flaws in their designs (Hoffman et al. 2022; Groenwold 2021). The most important reason we were able to demonstrate the effectiveness of specific health
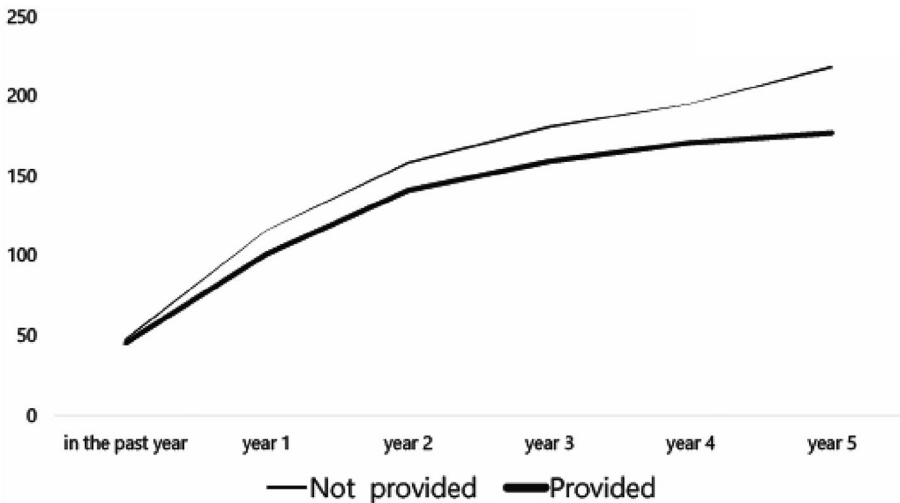
Fig. 3 Medical expenditures with and without specific health guidance Vertical axis; medical expenses per person per month (USD) Horizontal axis; time USD/JPY was 150 (October 21, 2022). Those who received health guidance during the six-month period among those who received the specified health check-up were identified and designated as the health guidance group. Those whose age, sex, BMI, abdominal circumference, medical expenses in the past year, number of days of outpatient visits in the past year, and number of days of hospitalization in the past year matched those of the health guidance group were identified and designated as the control group. The medical expenditure per person per month in both groups was observed over time

guidance for the first time in this study is that we did not adhere to a feasible RCT when setting up the target trial. It is impossible to perfectly match the background factors in an RCT. However, it is only by matching the background factors that we can make the best use of information from the measured factors. Specifically, even arbitrary order interactions between background factors can be incorporated into the model and adjusted accordingly.

Various types of evidence can be generated by refining the design by constantly pushing for an ideal study within the framework of TTE, in order to take advantage of the strengths of observational studies.

All the studies had unmeasured confounding factors. Observational studies have reported a high rate of type I errors (Liang et al. 2014), and the same result was obtained in this study. However, with our proposed exact-matching algorithm using AHCDEFs, the type I error probability is 6.6%, which is much lower than that of conventional methods. Although the type I error is still slightly higher than the acceptable range, we confirmed that it can be sufficiently reduced by refining the design using our method.

This study is the first to demonstrate the effectiveness of specific health guidance in reducing the incidence of diabetes and medical expenditures in Japan. In the United States, evidence suggests that multicomponent behavioral interventions in adults with obesity can lead to clinically significant improvements in reducing the incidence of type 2 diabetes among such adults and those with elevated plasma glucose levels (US Preventive Services Task Force et al. 2018). In 2008, Japan introduced a screening program to identify individuals with obesity and metabolic syndrome (Tsushita et al. 2018). All adults aged 40–74 years were required by law to participate every year (Fukuma et al. 2020). Therefore, this study presents the impact of this national health guidance intervention with an appropriate design, such that an RCT cannot be assembled.

Although RCTs are designed to answer a single question, they are expensive, time-consuming, and resource-intensive. It is not possible, in principle, to randomize a large sample population by including the various comorbidities and other confounding factors, and the patients included are generally younger with fewer comorbidities due to resource constraints. Therefore, the results are not immediately generalizable, leading to a large selection bias that compromises representativeness in this regard. Despite these major limitations, there is currently a worldwide mass production of "evidence" from sub-analyses and stratified analyses of past RCTs, which should be one per RCT, and which is used in daily practice. However, observational studies, when properly designed, have multiple strengths that make them a suitable complement to RCTs for decision-making.

Despite the notable findings that have emerged from this study, it had several limitations that must be acknowledged. First, the control for time-dependent confounding factors was not considered. However, this method has great potential in this respect, that is, based on the TTE concept, A*B time comparisons are performed and weighted 1/(A*B) in the groups with (A) and without exposure (B), which are perfectly matched in AHCDEFs. For each comparison, the termination of the observation for comparison patients is added to the end of the observation requirement (usually the occurrence of an outcome, the end of the study period, or withdrawal from the study), allowing control for time-dependent confounding. Second, this study did not model the administrative claims databases. Therefore, compatibility with high-dimensional propensity score adjustments cannot be examined (Schneeweiss et al. 2009). In principle, the two methods are expected to be highly compatible with each other, and future studies should be conducted to apply this method, to perfectly match the variables selected by the algorithm used in high-dimensional propensity score adjustment. Finally, we discuss the generalizability of the results. Generalizability is broadly compromised in order to increase comparability. However, this method is considered less susceptible to random errors than conventional methods, even when stratified analysis is performed. Thus, it is possible to make comparisons based on the circumstances of individual patients. Observational studies have significant limitations in dealing with real-measurement confounding and cannot replace RCTs. Although this method is effective for research questions for which an RCT would be difficult to conduct, it goes without saying that an RCT should be conducted if it is feasible to do so.

5 Conclusions

We propose a new method for analyzing real-world data and an exact-matching algorithm using AHCDEFs. The larger the number of patients available for analysis, the more AHCDEFs that can be matched, thereby removing the influence of confounding factors. It is expected that this method will generate significant evidence when applied to real-world data. In this process, it is desirable to clarify in detail the problems that may arise when applying this method.

Data availability The large datasets used and analyzed during the simulation study are available from the corresponding author on reasonable request. The KDB datasets generated and analyzed during the Japanese health guidance study are available from Nara prefecture, Japan, but restrictions apply to the availability of these data, which were used under license for the current study, and thus are not publicly available. The data will only be provided to those who have applied for and have received permission in advance, in Japan.

Declarations

Conflict of interest YN received consultation fees from Novo Nordisk. The other authors declare that they have no conflicts of interest.

Ethical approval All methods were conducted in accordance with relevant guidelines and regulations. All experimental protocols were approved by the Ethics Committee of Nara Medical University (approval no. 1123–7, October 8th, 2015). The need for informed consent was waived in view of the study design; all patient data were anonymized before analysis.

Consent to participate Not applicable.

References

Baumfeld Andre, E., Reynolds, R., Caubel, P., Azoulay, L., Dreyer, N.A.: Trial designs using real-world data: The changing landscape of the regulatory approval process. Pharmacoepidemiol. Drug Saf. **29**, 1201–1212 (2020)

Frieden, T.R.: Evidence for health decision making - Beyond randomized, controlled trials. N. Engl. J. Med. **377**, 465–475 (2017)

Fukuma, S., Iizuka, T., Ikenoue, T., Tsugawa, Y.: Association of the national health guidance intervention for obesity and cardiovascular risks with health outcomes among Japanese men. JAMA Intern. Med. **180**, 1630–1637 (2020)

Groenwold, R.H.H.: Trial emulation and real-world evidence. JAMA Netw. Open **4**, e213845 (2021)

Hernán, M.A., Robins, J.M.: Using big data to emulate a target trial when a randomized trial is not available. Am. J. Epidemiol. **183**, 758–764 (2016)

Hoffman, K.L., Schenck, E.J., Satlin, M.J., Whalen, W., Pan, D., Williams, N., Díaz, I.: Comparison of a target trial emulation framework vs cox regression to estimate the association of corticosteroids with COVID-19 mortality. JAMA Netw. Open **5**, e2234425 (2022)

Liang, W., Zhao, Y., Lee, A.H.: An investigation of the significance of residual confounding effect. BioMed Res. Int. **2014**, 658056 (2014)

Madenci, A.L., Wanis, K.N., Cooper, Z., Haneuse, S., Subramanian, S.V., Hofman, A., Hernán, M.A.: Strengthening health services research using target trial emulation: An application to volume-outcomes studies. Am. J. Epidemiol. **190**, 2453–2460 (2021)

Matthews, A.A., Danaei, G., Islam, N., Kurth, T.: Target trial emulation: applying principles of randomised trials to observational studies. BMJ **378**, e071108 (2022)

Mellinghoff, S.C., Bruns, C., Al-Monajjed, R., Cornely, F.B., Grosheva, M., Hampl, J.A., et al.: Harmonized procedure coding system for surgical procedures and analysis of surgical site infections (SSI) of five European countries. BMC Med. Res. Methodol. **22**, 225 (2022)

Muramoto, A., Matsushita, M., Kato, A., Yamamoto, N., Koike, G., Nakamura, M., et al.: Three percent weight reduction is the minimum requirement to improve health hazards in obese and overweight people in Japan. Obes. Res. Clin. Pract. **8**, e466–e475 (2014)

Nishioka, Y., Takeshita, S., Kubo, S., Myojin, T., Noda, T., Okada, S., et al.: Appropriate definition of diabetes using an administrative database: a cross-sectional cohort validation study. J. Diabetes Investig. **13**, 249–255 (2022)

Rubin, D.B.: Using multivariate matched sampling and regression adjustment to control bias in observational studies. J. Am. Stat. Assoc. **74**, 318–328 (1979)

Schneeweiss, S., Rassen, J.A., Glynn, R.J., Avorn, J., Mogun, H., Brookhart, M.A.: High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology **20**, 512–522 (2009)

Seeger, J.D., Williams, P.L., Walker, A.M.: An application of propensity score matching using claims data. Pharmacoepidemiol. Drug Saf. **14**, 465–476 (2005)

Sun, X., Tan, J., Tang, L., Guo, J.J., Li, X.: Real world evidence: Experience and lessons from China. BMJ **360**, j5262 (2018)

Takeuchi, Y., Kumamaru, H., Hagiwara, Y., Matsui, H., Yasunaga, H., Miyata, H., Matsuyama, Y.: Sodium-glucose cotransporter-2 inhibitors and the risk of urinary tract infection among diabetic patients in Japan: Target trial emulation using a nationwide administrative claims database. Diabetes Obes. Metab. **23**, 1379–1388 (2021)

Tsushita, K., Hosler, A.S., Miura, K., Ito, Y., Fukuda, T., Kitamura, A., et al.: Rationale and descriptive analysis of specific health guidance: the nationwide lifestyle intervention program targeting metabolic syndrome in Japan. J. Atheroscler. Thromb. **25**, 308–322 (2018)

US Preventive Services Task Force, Curry, S.J., Krist, A.H., Owens, D.K., Barry, M.J., Caughey, A.B., et al.: Behavioral weight loss interventions to prevent obesity-related morbidity and mortality in adults: US preventive services task force recommendation statement. JAMA **320**, 1163–1171 (2018)

Xie, Y., Bowe, B., Gibson, A.K., McGill, J.B., Maddukuri, G., Yan, Y., Al-Aly, Z.: Comparative effectiveness of SGLT2 inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and sulfonylureas on risk of kidney outcomes: Emulation of a target trial using health care databases. Diabetes Care **43**, 2859–2869 (2020)