



Introduction to clinical research based on modern epidemiology

Junichi Hoshino¹

Received: 11 February 2020 / Accepted: 28 February 2020 / Published online: 24 March 2020
© The Author(s) 2020

Abstract

Over the past 20 years, recent advances in science technologies have dramatically changed the styles of clinical research. Currently, it has become more popular to use recent modern epidemiological techniques, such as propensity score, instrumental variable, competing risks, marginal structural modeling, mixed effects modeling, bootstrapping, and missing data analyses, than before. These advanced techniques, also known as modern epidemiology, may be strong tools for performing good clinical research, especially in large-scale observational studies, along with relevant research questions, good databases, and the passion of researchers. However, to use these methods effectively, we need to understand the basic assumptions behind them. Here, I will briefly introduce the concepts of these techniques and their implementation. In addition, I would like to emphasize that various types of clinical studies, not only large database studies but also small studies on rare and intractable diseases, are equally important because clinicians always do their best to take care of many kinds of patients who suffer from various kidney diseases and this is our most important mission.

Keywords Propensity score · Instrumental variables · Bootstrapping · Competing risk · Multiple imputation · Intractable disease · Modern epidemiology

Introduction

After the development of the concept of evidence-based medicine (EBM), which has been in widespread use since the 1990 [1], many kinds of evidence-based clinical guidelines have been published. Thanks to advances in computer technologies and internet access, clinicians can now easily view these guidelines and apply the updated knowledge, which largely contributes to the improvement in the quality of care and uniformity of medical services provided. On the other hand, it also reveals the difficulty of implementing EBM for intractable diseases, rare diseases, and chronic diseases with a long history of progression because these diseases are often difficult to show sufficient evidence with “hard” outcomes based on “high-quality interventional clinical trials”. In nephrology, it is often difficult to show

sufficient evidence for many diseases, and the US government recently started to approve of the use of an eGFR decline of 30% as a proxy outcome in clinical trials [2]. In addition, because of the very strict inclusion criteria of interventional trials, the implementation of the evidence written in the guidelines is often limited in nephrology clinical practice.

The mission of clinicians is to provide the best daily clinical practice based on updated clinical knowledge, which is improving daily. If evidence is not sufficient for bedside clinical practice, we have to solve clinical questions by ourselves using “real-world” clinical data. In general, evidence levels of observational studies are considered lower than those of interventional studies because of the existence of many biases: selection bias, information bias, missing data, publication bias, etc. However, recent advances in epidemiology (also known as modern epidemiology), information technology, and computer technology allow us to conduct high-quality observational studies. Advances in computer and internet technologies have been tremendous. Now, submitting a manuscript with analog photography by international mail has become a folk tale. It has been only 20 years since the first iMac by Apple Inc. and only 15 years since the release of Gmail by Google. Advances in these technologies

This article was presented as the Oshima Award memorial lecture at the 59th annual meeting of the Japanese Society of Nephrology, held at Yokohama, Japan in 2016.

✉ Junichi Hoshino
hoshino@toranomon.gr.jp

¹ Toranomon Hospital, Nephrology Center, 2-2-2, Toranomon, Minato-ku, Tokyo 105-8470, Japan

have dramatically changed not only human life but also clinical research. In basic science, many advanced technologies have been launched, such as mass spectrometry, whole-genome sequencing, omics analyses, and single-cell analyses. These advanced technologies have dramatically changed many fields of medical sciences not only in basic science but also in clinical science. Thanks to these advances, all clinical researchers can perform advanced epidemiological and statistical techniques with their own personal computer, and clinicians have become more familiar with observational studies. However, statistical techniques are only one element for good clinical research. The other three elements are relevant research questions based on clinical experience, high-quality databases, and sufficient passion to accomplish clinical studies. These elements are equally essential for good clinical research (Fig. 1).

Here, I would like to introduce some fundamentals of clinical research and recent advances in the field of epidemiology.

Fundamentals of clinical research

As shown in Fig. 2, many clinical studies begin with clinical questions at the bedside. These questions are structured through the PE(I)CO format. It is essential to clarify definitions of the target population (P), exposures (E) or interventions (I), comparisons (C), and outcomes (O) as a first step in clinical research. All clinical research is conducted in accordance with the ethics guidelines for medical research for human subjects. Therefore, almost all clinical research should be started only after the approval of the ethics review committee. A good database (DB) is one that anyone can understand and reanalyze. Therefore, a codebook—a list of codes for each variable—and the updated version history

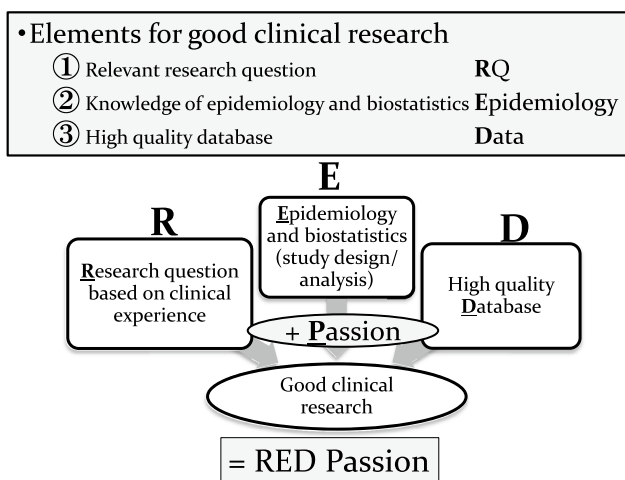


Fig. 1 Elements for good clinical research

Seven steps of clinical research

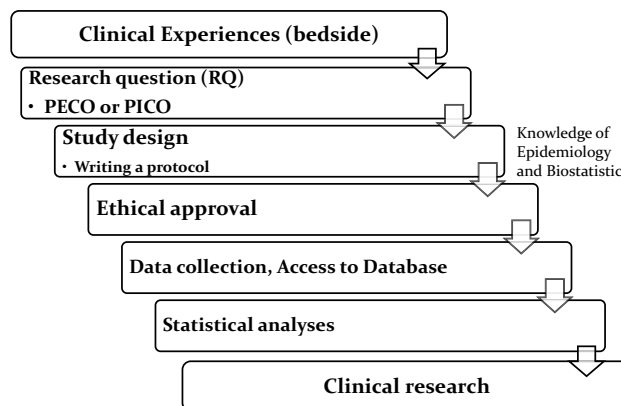


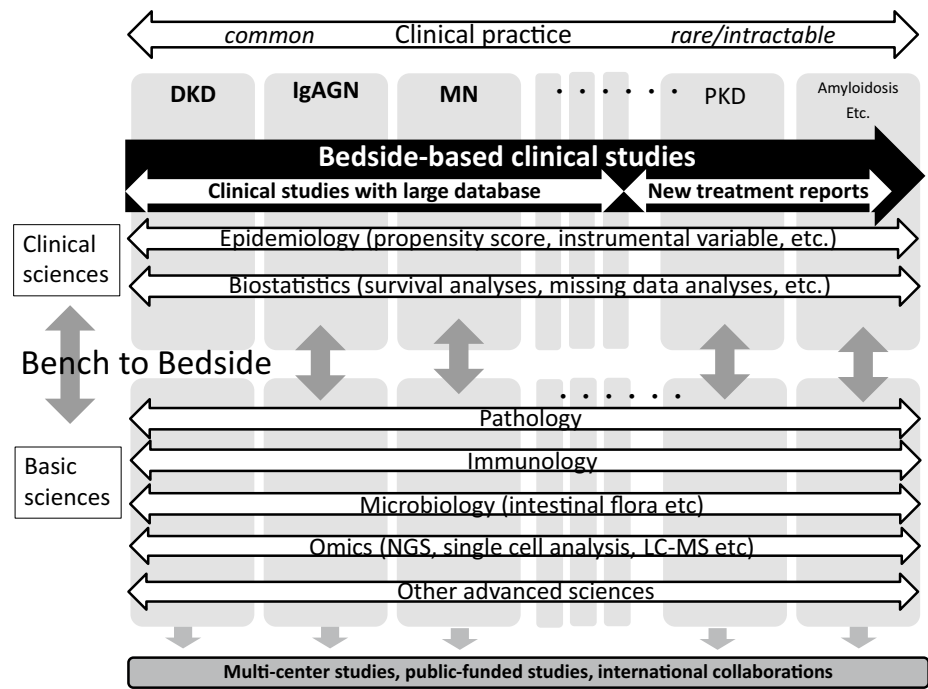
Fig. 2 Seven steps of clinical research. PECO population, exposure, comparison, and outcome; PICO population, intervention, comparison, and outcome

should be stored together with the DB. In addition, it is also important to create an “alive” DB that is formatted in consideration of later statistical processing, rather than a “dead” DB that contains a mix of text and numerical characters. These “alive” DBs make it easy to use analytic software in the analysis process. DB creation and DB cleaning are time-consuming and labor-intensive tasks, but researchers should keep in mind that high-quality, traceable DBs are necessary for good clinical research.

We need to understand the strengths and weaknesses of large DBs. It is generally considered that a large, multicenter DB is the best because of its statistical power and low selection bias. However, large DBs often has difficulty obtaining additional data that researchers are interested in, and data access is often limited. Moreover, these larger DBs may contain more missing data and outliers than in-house data in general. Therefore, researchers should understand the mechanism of missing data and how to handle missing data and outliers before analyzing the data.

“Bench to bedside” has been an important concept for our medical researchers, since all advanced basic sciences are developed with the aim of treating patients. However, this concept may lead researchers to focus on one specific disease, since it may be the shortest route for researchers’ success to focus on one specific disease and to deepen their knowledge. If we consider the association between areas of kidney diseases and research technologies, advanced research techniques such as genome sequencing, omics analysis, single-cell analysis, and epigenomics can be applicable to many kinds of diseases and thus could be considered the horizontal axis (Fig. 3). Modern epidemiology could also be on the horizontal axis in the clinical research field. The combination of these horizontal technologies are essential to deepen knowledge. On the other hand, it is often difficult

Fig. 3 Associations between areas of kidney diseases and research technologies in clinical studies. *DKD* diabetic kidney disease, *IgAGN* IgA nephropathy, *MN* membranous nephropathy, *PKD* polycystic kidney disease



for clinical researchers to limit their areas of interest on the vertical axis because doctors cannot select patients to care for in clinical practice. In addition to these elements, the types of clinical studies may vary depending on the number of patients available. For common diseases, a large DB and knowledge of epidemiology and biostatistics may be essential for good clinical research. However, for rare diseases, papers on new treatments or even case reports are equally important for patients. These studies often do not require sophisticated statistical techniques. Therefore, we need to understand that the necessity of knowledge of epidemiology and biostatistics may depend on the commonality of a disease, and it may be natural for clinical researchers to have an interest in various kidney diseases and publish “bedside-based” clinical studies.

Missing data and missing data analysis

Missing data not only leads to biased results but also potentially loses many observations, which reduces the power of the analysis. There are four famous approaches to handle the problem of missing data: (1) complete case analysis, (2) imputation, (3) weighting methods, and (4) model-based approaches. A complete case analysis is often the default of statistical software, but it may cause biased results and the power of the analysis will be reduced, as mentioned above. Commonly used imputation procedures include hot-deck imputation, mean imputation, and regression imputation, where the missing values are estimated by known variables. Multiple imputation may be the most popular method for

imputation. It refers to the procedures of replacing each missing value by a vector of $D \geq 2$ imputed values. The D values are ordered in the sense that D complete datasets can be created from the vectors of imputations; replacing each missing value by the first component in its vector of imputations creates the first completed dataset, replacing each missing value by the second component in its vector creates the second completed dataset, and so on. The model-based procedure defines a model for the observed data and bases inferences on the likelihood or posterior distribution under that model, with the parameters estimated by procedures such as maximum likelihood [3]. In addition, a mixed effects model may be an option to handle missing data. See Little’s textbook or other papers for more information [3, 4].

Modern epidemiology and clinical research

Biases have always been a major problem for clinical researchers. Randomization is an excellent research design that can control all biases, including unknown biases. However, ethical issues, high costs, lack of external validity, and “uncontrolled” patients have been major limitations of this study design. On the other hand, because observational studies lack randomized assignment of participants into treatment conditions, researchers must employ statistical procedures to balance the data before assessing the treatment effects. In recent years, many excellent methods to control these biases have been developed, such as propensity score analysis, inverse probability weighting, marginal structural modeling, bootstrapping, instrumental variable, etc. [5–8].

Given the low cost and easy accessibility of DBs and the necessity of bias control, observational studies are considered an excellent tool for new researchers.

As shown in Fig. 4, the number of clinical papers using these modern epidemiological techniques has increased dramatically over the past couple of decades. It may be possible that these modern techniques will become familiar to all clinical researchers—such as Cox proportional hazards modeling in survival analysis—in the near future. Before implementation, we must keep in mind that there are some basic assumptions in a regression analysis: normally distributed errors (with a mean of zero), the independence of covariates (no multicollinearity), no correlation between the

residual terms (autocorrelation), and homoscedasticity of the errors (equal variance around the line).

Propensity score analyses

In observational studies, treated and untreated subjects often differ systematically on prognostic factors leading to treatment selection bias or confounding in estimating the effect of a treatment on an outcome. Over the past 40 years, researchers have recognized the need to develop more efficient approaches for assessing treatment effects from observational studies, and statisticians (e.g., Rosenbaum & Rubin) and econometricians (e.g., Heckman) have developed a new approach called propensity score analysis [9–11]. A propensity score (PS) is the probability being assigned a treatment (or exposure) to accomplish data balancing when treatment assignment is unignorable, to evaluate treatment effects using non-experimental approaches and to reduce multidimensional covariates into a one-dimensional score (Fig. 5a).

There are five steps in a propensity score analysis: (1) selecting the variables for the PS model, (2) estimating the PSs, (3) applying the PS methods, (4) assessing the balance, and (5) estimating the treatment effect [12]. The PSs are calculated by logistic or probit regression models consisting of all the confounders in the given analysis set ranging from 0 to 1. In addition to all confounders, outcome predictors irrespective of the exposure can improve the precision of an estimated treatment effect without increasing the bias [12, 13]. Note that variables that are strongly related to the treatment but not to the outcome (instrumental variables) or

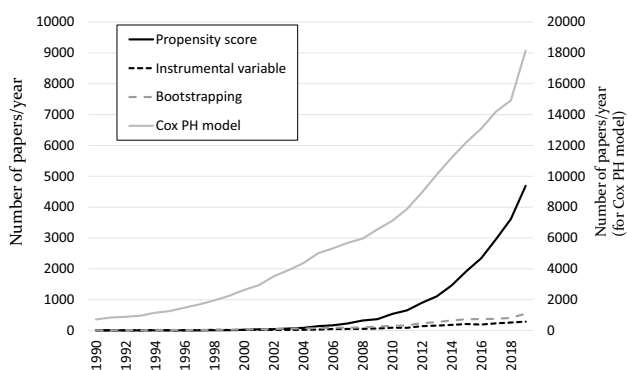
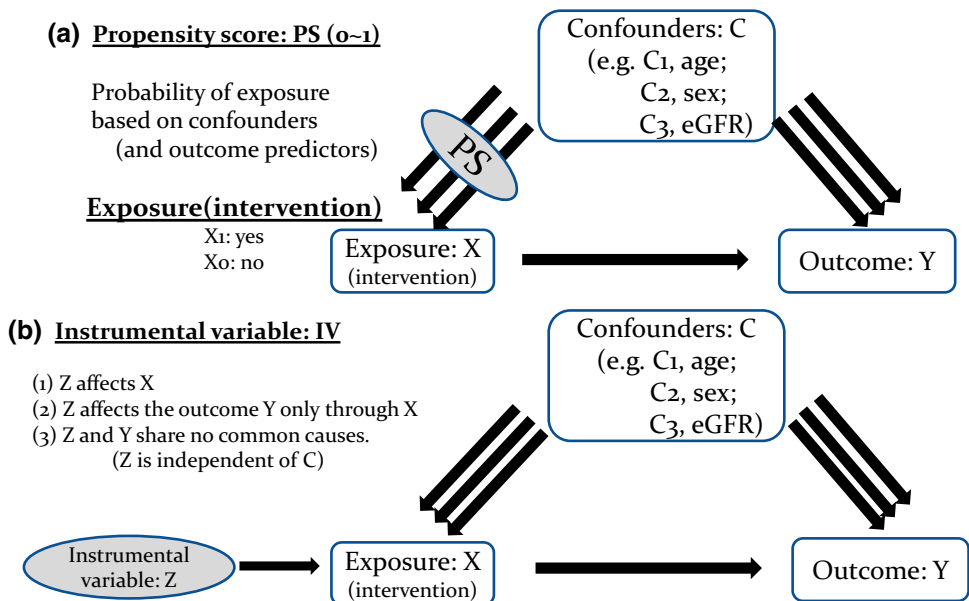


Fig. 4 The number of published papers using modern epidemiological techniques. Cox PH model: Cox’s proportional hazard model. The number of published papers was calculated by PubMed on Feb 1, 2020

Fig. 5 Schematic diagrams of propensity scores and instrumental variables. *eGFR* estimated glomerular filtration rate

Schematic diagrams of propensity scores and instrumental variables



weakly related to the outcome should not be included in the PS model because such variables could amplify the bias in the presence of unmeasured confounding [12–15]. Assuming that there are no unmeasured confounders, treated and untreated subjects with the same PS tend to have similar distributions of measured confounders [16]. For PS estimation, the *c*-statistics is often cited as a measure of “fit” of a PS model; that is, the ability of the model to predict treatment assignment using observed covariates [17–19]. The *c*-statistics takes on values between 0.5 (classification no better than a coin flip) to 1.0 (perfect classification). A very high or very low *c*-statistics implies the reduced utility of the PS approach. However, we have to keep in mind that the purpose of PS estimation is to balance risk factors for the outcome between treatment groups to eliminate confounding. Reliance on the *c*-statistics in selecting a PS model may provide false confidence that all confounders have been balanced between treatment groups [17]. For balance assessments, the standardized difference (SDif) is often used to assess the balance between the selected treated and untreated subjects. It is preferable that the SDif be less than 0.1 [20]. If not, changing the caliper of the PSs (usually 20% of the standard deviation of the PSs), selecting other matching methods, changing the replacement status of the matching selection, or rechecking the choice of variables for the PS model (e.g., checking interactions and linearity) are often performed. Recently, post-matching the *c*-statistics of the PS model has been suggested as an overall measure of the balance across covariates [21].

There are four major analytic approaches that use PSs: PS matching, PS adjustment, PS stratification, and PS weighting (Fig. 6). PS matching is the most widely used and probably most understandable PS analytical method [12]. On the other hand, because this method selects only matched subjects, it may result in an undesirable loss of study participants. Note that the exclusion of unmatched subjects from the analysis not only affects the precision of the effect estimate but also has consequences on the generalizability of the results. The other three methods—PS adjustment, PS stratification, and PS weighting—do not need to consider the loss of study participants. PS weighting aims to reweight the treated and untreated subjects to make them more representative of the population of interest without the loss of participants. Inverse probability weighting (IPW) using PSs has recently become popular. In IPW, the use of stabilizing weights could help “normalize” the range of the inverse probabilities and increase the efficiency of the analysis [12]. However, we should keep in mind that IPW tends to overweight participants with extremely small (or large) PSs. Freedman and Berk reported that PS weighting was optimal only under three circumstances: (1) when the study subjects are independently and identically distributed, (2) when selection is exogenous, and (3) when the selection

equation is properly specified [22]. For treatment effect estimation, careful interpretation of the treatment effect estimate is needed. PS matching typically focuses on the effect of the treatment in either the treated or the untreated subjects, not on the average treatment effect on the whole population [23, 24]. PS adjustment and PS stratification give conditional treatment effect estimates [25]. However, marginal structural modeling methods using IPW estimate a marginal treatment effect in randomized studies; thus, the estimate can be directly interpreted as the average causal treatment effect between treated and untreated patients [26, 27]. See the textbook and papers on PS for more information [9, 12, 28, 29].

Longitudinal data analyses

Survival analysis is one of the most familiar analyses for medical scientists. On the other hand, survival data have two common features that are difficult to handle with conventional statistical methods: censoring and time-dependent covariates. Kaplan–Meier analysis and Cox proportional hazard modeling are probably the most widely used techniques in our field (Fig. 4), although there are many different methods, including exponential regression, log-normal regression, Weibull AFT modeling, competing risks models, and discrete-time methods. Cox proportional hazards models have a basic assumption called the proportional hazards assumption: the hazard ratio comparing two groups is constant. This assumption is often tested by log–log plots and Schoenfeld residuals, which should be tested before estimation.

Additionally, competing risks are an important concept in clinical research, since survival analysis is often applied to study death or other events of interest. However, an

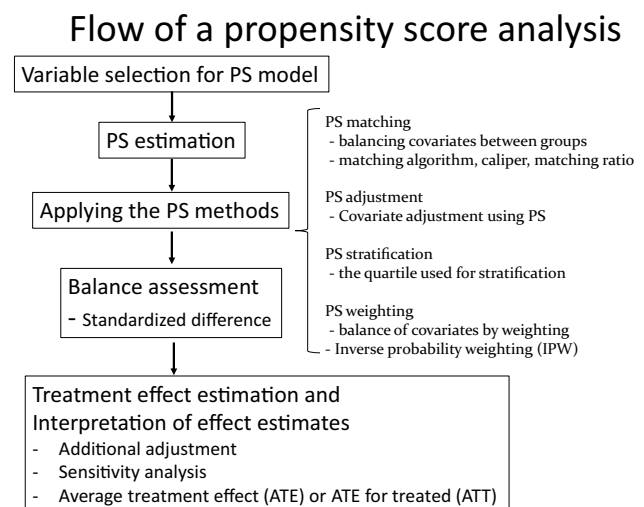


Fig. 6 Flow of a propensity score analysis. *PS* propensity score

important assumption of standard survival analyses such as the Kaplan–Meier method is that censoring is “independent”, which implies that the censored patients at a certain time point should be representative of those still at risk at same time. This condition is usually the case when a patient is lost to follow-up. In oncology and cardiovascular medicine, the analytical problem of competing risks has been acknowledged for many years. In nephrology, death and ESRD are competing risks in ESRD risk studies because death before ESRD prevents ESRD from occurring. If death before ESRD was censored, it may result in biased estimates because death is strongly related to future ESRD events [30, 31]. The cause-specific Cox hazard model is often suitable for studying associations between covariates and the instantaneous risk of a clinical event. However, the cause-specific hazard model cannot produce predicted probabilities of the event of interest without additional models for the competing risk event, because the predicted probability of ESRD must take into account both the incidence of ESRD and death. In this case, the sub-distribution hazard (SHR) approach proposed by Fine and Gray [32] is considered a preferable approach for prognostic studies, because this method can be used to predict the future risk of ESRD, taking into account the attrition due to death [5, 30]. Note that the SHR model directly provides individual probabilities of an event, given a patient’s characteristics, but it cannot be interpreted as a hazard ratio (HR) [6].

Qualifying the effect of treatment duration in survival analysis is a major problem for researchers, since only people who survive for a long time can receive a treatment for a long time. A direct comparison of subjects with longer and shorter follow-ups would be biased. For analyses of these longitudinal data with time-varying confounders, Robins reported a g formula to measure the healthy worker survivor effect [7]. Although this formula is complicated, Hernan recently reported a simple three-step approach to estimate the effect of treatment duration on survival outcomes using observational data [8]. The first step is duplicating people to assign them to treatment duration strategies at time zero, eliminating immortal time bias [33–35]. The second step is censoring the duplicates when they deviate from their assigned treatment strategies through follow-up. The introduced selection bias can be eliminated by the third step with inverse probability weighting to adjust for the potential selection bias introduced by censoring. Applications of these new approaches can be found in the areas of nephrology [36–38], infectious diseases [39, 40], gastroenterology [34], and urology [41].

Instrumental variable method

Instrumental variable (IV) analysis is one of the methods used to control for confounding and measurement error in observational studies so that causal inferences can be made.

This method was invented in the 1920s in economics and has appeared in the health sciences [42]. Suppose X and Y are the exposure and outcome of interest, respectively, and we can observe their relation to a third variable Z (Fig. 5b). Let Z be associated with X but not associated with Y except through its association with X . Here, Z is called an IV [33]. That is, an IV is a factor associated with the exposure but not with the outcome. For example, the price of alcohol can affect the likelihood of expectant mothers drinking alcohol, but there is no reason to believe that it directly affects the child’s birth-weight [43]. There are three assumptions of IV: (1) Z affects X , (2) Z affects the outcome Y only through X , and (3) Z and Y share no common cause [42, 44]. The obvious example of an IV is in randomized controlled trials, since the random treatment assignment Z is independent of confounders and affects Y only through X . IV analyses are promising for the estimation of therapeutic effects from observational data as they can circumvent unmeasured confounding [45]. However, even if the IV assumption holds, Boef et al. reported that IV analysis will not necessarily provide an estimate closer to the true effect than conventional analyses, as this result depends on the estimates’ bias and variance [46]. They also reported that IV methods have the most value in large studies if considerable unmeasured confounding is likely and a strong and plausible instrument is available.

Bootstrapping

In clinical research, we always assume that results from a sample population (e.g., CKD patients in our hospital) imply the same results in the source population (e.g., CKD patients in Japan or around the world). Given the central limit theorem and law of large numbers, the mean of all the samples from the population will be approximately equal to the mean of the population if the samples are large enough. Furthermore, all the samples will follow an approximate normal distribution, with all variances being approximately equal to the variance in the population, divided by each sample’s size. The original concept of bootstrapping was proposed by Bradley Efron [47, 48]. The basic idea of bootstrapping is that inference about a population from sample data can be modeled by resampling the sample data with replacement and performing inference about a sample from the resampled data. For example, after generating sufficiently large randomly selected pseudo-sample sets (e.g., 500 bootstrapping sample sets) from the original sample, the distribution of these statistics across the bootstrapping sample sets (e.g., the mean of the distribution of the means of bootstrapping sample sets) will be approximately equal to the distribution of the original sample statistics (e.g., the mean of the distribution of the original sample). This method is very

useful when inferring the unknown distribution of the original sample statistics. This kind of simulation method has been developed and implemented in medicine [47, 49–51].

Challenges for intractable kidney disease: importance, relevance and novelty

The treatment of intractable diseases requires a great deal of effort and passion, but from the viewpoint of clinical research, these diseases are often rare, and small studies, including case series papers, are valuable for bedside treatment. For example, clinical studies of pediatric nephritic syndrome and amyloidosis have been published in major journals despite the limited number of participants [52, 53]. These examples suggest that not only methodology and DB issues but also clinical relevance and novelty are very important in clinical research. For example, in 1996, we started using arterial embolization therapy for intractable PKD patients on dialysis [54]. This new treatment method can dramatically reduce the size of enlarged kidneys and improve their symptoms as well as nutritional conditions, and this method has spread across the world. Additionally, although diabetic nephropathy is a common disease, its pathophysiology is largely unclear. Therefore, we examined the association between pathological changes in diabetic nephropathy and clinical outcomes [55–62]. Other examples can be found in many areas, such as in the areas of treatment of polycystic liver disease [63, 64], amyloidosis [65–68], and severe ischemic limb treatment [69–72]. Clinical relevance and novelty sometimes overcome the limitation of the number of patients, especially in the areas of rare and intractable diseases.

Closing remarks

Here, I briefly introduce modern epidemiology techniques, as well as the importance of novelty and relevance in all research areas of interest. PS analyses, IV methods, competing risk analysis, and other recently developed methods are becoming more familiar to clinical researchers. Thanks to recent advances in the field of computer science and internet technologies, it is now very easy to learn and perform these techniques by ourselves. On the other hand, there are many unsolved research questions, especially in the areas of rare diseases and intractable diseases. In these areas, it is not always necessary to use these modern techniques; rather, it is important to treat patients with sincerity and with passion. A case report or case series should be valuable in these areas. I hope that many doctors will become interested in clinical

research, publish many excellent clinical papers, and as a result, contribute to improving patients' quality of life.

Acknowledgements The author is grateful to all members of Toranomon Hospital, especially to Dr. Akira Yamada, Dr. Shigeo Hara, Dr. Kenmei Takaichi, and Dr. Yoshifumi Ubara, who always taught clinical practice and clinical research kindly and with passion, and grateful to Prof. Kunihiro Yamagata, Prof. Takashi Wada, and Dr. Ikuto Masakane for the opportunity to perform nationwide clinical studies.

Funding This work was supported by JH's research grants from the Okinaka Memorial Institute for Medical Research (Grant no. 2019).

Compliance with ethical standards

Conflict of interest The author received competitive research funds from Otsuka Pharm. and lecture fee from Ono Pharm.

Human and animal rights This article does not contain any original studies with human participants or animals.

Informed consent Informed consent is not required, because this article is not involve human participants.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what isn't. *BMJ*. 1996;312:71–2.
2. Levey AS, Inker LA, Matsushita K, Greene T, Willis K, Lewis E, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *Am J Kidney Dis*. 2014;64:821–35.
3. Little AJ, Rubin B. *Statistical analysis with missing data*. 2nd. Edition ed. Hoboken: Wiley; 2002.
4. Lunt M. *A guide to imputing missing data with stata*. 2011.
5. Li L, Yang W, Astor BC, Greene T. Competing risk modeling: time to put it in our standard analytical toolbox. *J Am Soc Nephrol*. 2019;30:2284.
6. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170:244–56.
7. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods—application to control of the healthy worker survivor effect. *Math Modell*. 1986;7:1393–512.
8. Hernán MA. How to estimate the effect of treatment duration on survival outcomes using observational data. *BMJ*. 2018;360:k182.

9. Guo S, Fraser M. Propensity score analysis. California: SAGE Publications Inc.; 2010.
10. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
11. Heckman JJ. Dummy endogenous variables in a simultaneous equation system. *Econometrica*. 1978;46:931–59.
12. Ali MS, Groenwold RH, Belitser SV, Pestman WR, Hoes AW, Roes KC, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *J Clin Epidemiol*. 2015;68:112–21.
13. Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol*. 2011;174:1213–22.
14. Pearl J. Invited commentary: understanding bias amplification. *Am J Epidemiol*. 2011;174:1223–7 **discussion pg 8–9**.
15. Ali MS, Groenwold RH, Pestman WR, Belitser SV, Roes KC, Hoes AW, et al. Propensity score balance measures in pharmacoepidemiology: a simulation study. *Pharmacoepidemiol Drug Saf*. 2014;23:802–11.
16. Rosenbaum PR, Rubin DB. Difficulties with regression analyses of age-adjusted rates. *Biometrics*. 1984;40:437–43.
17. Westreich D, Cole SR, Funk MJ, Brookhart MA, Stürmer T. The role of the c-statistic in variable selection for propensity score models. *Pharmacoepidemiol Drug Saf*. 2011;20:317–20.
18. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*. 2006;59:437–47.
19. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiol Drug Saf*. 2004;13:841–53.
20. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Statist*. 1985;39:33–8.
21. Franklin JM, Rassen JA, Ackermann D, Bartels DB, Schneeweiss S. Metrics for covariate balance in cohort studies of causal effects. *Stat Med*. 2014;33:1685–99.
22. Freedman DA, Berk RA. Weighting regressions by propensity scores. *Eval Rev*. 2008;32:392–409.
23. Hill J. Discussion of research using propensity-score matching: comments on 'A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003' by Peter Austin, *Statistics in Medicine*. *Stat Med*. 2008;27:2055–61 **discussion 66–9**.
24. Lunt M. Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. *Am J Epidemiol*. 2014;179:226–35.
25. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Systematic differences in treatment effect estimates between propensity score methods and logistic regression. *Int J Epidemiol*. 2008;37:1142–7.
26. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11:561–70.
27. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–60.
28. Wijeyesundera DN, Austin PC, Beattie WS, Hux JE, Laupacis A. Outcomes and processes of care related to preoperative medical consultation. *Arch Intern Med*. 2010;170:1365–74.
29. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*. 2006;59:437–47.
30. Noordzij M, Leffondre K, van Stralen KJ, Zoccali C, Dekker FW, Jager KJ. When do we need competing risks methods for survival analysis in nephrology? *Nephrol Dial Transplant*. 2013;28:2670–7.
31. Grams ME, Coresh J, Segev DL, Kucirka LM, Tighiouart H, Sarnak MJ. Vascular disease, ESRD, and death: interpreting competing risk analyses. *Clin J Am Soc Nephrol*. 2012;7:1606–14.
32. Fine JP, Gray RJ. A Proportional hazards model for the subdistribution of a competing risk. *J Am Statist Assoc*. 1999;94:496–509.
33. Zheng W, Luo Z, van der Laan MJ. Marginal structural models with counterfactual effect modifiers. *Int J Biostat*. 2018;2018:14.
34. Garcia-Albeniz X, Hsu J, Hernan MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol*. 2017;32:495–500.
35. Hernan MA. Counterpoint: epidemiology to guide decision-making: moving away from practice-free research. *Am J Epidemiol*. 2015;182:834–9.
36. Zhang Y, Thamer M, Kaufman J, Cotter D, Hernan MA. Comparative effectiveness of two anemia management strategies for complex elderly dialysis patients. *Med Care*. 2014;52(Suppl 3):S132–S139139.
37. Akizawa T, Kurita N, Mizobuchi M, Fukagawa M, Onishi Y, Yamaguchi T, et al. PTH-dependence of the effectiveness of cinacalcet in hemodialysis patients with secondary hyperparathyroidism. *Sci Rep*. 2016;6:19612.
38. Tentori F, Albert JM, Young EW, Blayney MJ, Robinson BM, Pisoni RL, et al. The survival advantage for haemodialysis patients taking vitamin D is questioned: findings from the Dialysis Outcomes and Practice Patterns Study. *Nephrol Dial Transplant*. 2009;24:963–72.
39. Cain LE, Logan R, Robins JM, Sterne JA, Sabin C, Bansi L, et al. When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: an observational study. *Ann Intern Med*. 2011;154:509–15.
40. Cain LE, Saag MS, Petersen M, May MT, Ingle SM, Logan R, et al. Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therapy. *Int J Epidemiol*. 2016;45:2038–49.
41. Garcia-Albeniz X, Chan JM, Paciorek A, Logan RW, Kenfield SA, Cooperberg MR, et al. Immediate versus deferred initiation of androgen deprivation therapy in prostate cancer patients with PSA-only relapse. An observational follow-up study. *Eur J Cancer*. 2015;51:817–24.
42. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000;29:722–9.
43. Ahn C. Chapter 19—biostatistics used for clinical investigation of coronary artery disease. In: Aronow WS, McClung JA, editors. *Translational research in coronary artery disease*. Boston: Academic Press; 2016. p. 215–221.
44. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
45. Almond D, Doyle JJ, Kowalski AE, Williams H. Estimating marginal returns to medical care: evidence from at-risk newborns. *Q J Econ*. 2010;125:591–634.
46. Boef AG, Dekkers OM, Vandenbroucke JP, le Cessie S. Sample size importantly limits the usefulness of instrumental variable methods, depending on instrument strength and level of confounding. *J Clin Epidemiol*. 2014;67:1258–64.
47. Efron B, Tibshirani R. *An introduction to the bootstrap*. New York: Chapman and Hall; 1994.

48. Efron B. Bootstrap methods: another look at the jackknife. *Ann Statist.* 1979;7:1–26.
49. Hoshino J, Molnar MZ, Yamagata K, Ubara Y, Takaichi K, Kovesdy CP, et al. Developing an A1c based equation to estimate blood glucose in maintenance hemodialysis patients. *Diabetes Care.* 2013;36:922–7.
50. Royston P, Sauerbrei W. Bootstrap assessment of the stability of multivariable models. *Stata J.* 2009;9:547–70.
51. Geldof T, Popovic D, Van Damme N, Huys I, Van Dyck W. Nearest neighbour propensity score matching and bootstrapping for estimating binary patient response in oncology: a Monte Carlo simulation. *Sci Rep.* 2020;10:964.
52. Iijima K, Sako M, Nozu K, Mori R, Tuchida N, Kamei K, et al. Rituximab for childhood-onset, complicated, frequently relapsing nephrotic syndrome or steroid-dependent nephrotic syndrome: a multicentre, double-blind, randomised, placebo-controlled trial. *Lancet.* 2014;384:1273–81.
53. Jaccard A, Moreau P, Leblond V, Leleu X, Benboubker L, Hermine O, et al. High-dose melphalan versus melphalan plus dexamethasone for AL amyloidosis. *N Engl J Med.* 2007;357:1083–93.
54. Ubara Y, Takei R, Hoshino J, Tagami T, Sawa N, Yokota M, et al. Intravascular embolization therapy in a patient with an enlarged polycystic liver. *Am J Kidney Dis.* 2004;43:733–8.
55. Hoshino J, Furuichi K, Yamanouchi M, Mise K, Sekine A, Kawada M, et al. A new pathological scoring system by the Japanese classification to predict renal outcome in diabetic nephropathy. *PLoS ONE.* 2018;13:e0190923.
56. Hoshino J, Mise K, Ueno T, Imafuku A, Kawada M, Sumida K, et al. A pathological scoring system to predict renal outcome in diabetic nephropathy. *Am J Nephrol.* 2015;41:337–44.
57. Mise K, Hoshino J, Ubara Y, Sumida K, Hiramatsu R, Hasegawa E, et al. Renal prognosis a long time after renal biopsy on patients with diabetic nephropathy. *Nephrol Dial Transplant.* 2014;29:109–18.
58. Mise K, Hoshino J, Ueno T, Hazue R, Hasegawa J, Sekine A, et al. Prognostic value of tubulointerstitial lesions, urinary N-acetyl-beta-d-glucosaminidase, and urinary beta2-microglobulin in patients with type 2 diabetes and biopsy-proven diabetic nephropathy. *Clin J Am Soc Nephrol.* 2016;11:593–601.
59. Mise K, Hoshino J, Ueno T, Imafuku A, Kawada M, Sumida K, et al. Impact of tubulointerstitial lesions on anaemia in patients with biopsy-proven diabetic nephropathy. *Diabet Med.* 2015;32:546–55.
60. Mise K, Hoshino J, Ueno T, Sumida K, Hiramatsu R, Hasegawa E, et al. Clinical implications of linear immunofluorescent staining for immunoglobulin G in patients with diabetic nephropathy. *Diabetes Res Clin Pract.* 2014;106:522–30.
61. Mise K, Hoshino J, Ueno T, Sumida K, Hiramatsu R, Hasegawa E, et al. Clinical and pathological predictors of estimated GFR decline in patients with type 2 diabetes and overt proteinuric diabetic nephropathy. *Diabetes Metab Res Rev.* 2014;31:572–81.
62. Yamanouchi M, Furuichi K, Hoshino J, Toyama T, Hara A, Shimizu M, et al. Nonproteinuric versus proteinuric phenotypes in diabetic kidney disease: a propensity score-matched analysis of a nationwide, biopsy-based cohort study. *Diabetes Care.* 2019;42:891–902.
63. Hoshino J, Suwabe T, Hayami N, Sumida K, Mise K, Kawada M, et al. Survival after arterial embolization therapy in patients with polycystic kidney and liver disease. *J Nephrol.* 2015;28:369–77.
64. Hoshino J, Ubara Y, Suwabe T, Sumida K, Hayami N, Mise K, et al. Intravascular embolization therapy in patients with enlarged polycystic liver. *Am J Kidney Dis.* 2014;63:937–44.
65. Hoshino J, Kawada M, Imafuku A, Mise K, Sumida K, Hiramatsu R, et al. A clinical staging score to measure the severity of dialysis-related amyloidosis. *Clin Exp Nephrol.* 2016;21:300–6.
66. Hoshino J, Ubara Y, Sawa N, Sumida K, Hiramatsu R, Hasegawa E, et al. How to treat patients with systemic amyloid light chain amyloidosis? Comparison of high-dose melphalan, low-dose chemotherapy and no chemotherapy in patients with or without cardiac amyloidosis. *Clin Exp Nephrol.* 2011;15:486–92.
67. Hoshino J, Yamagata K, Nishi S, Nakai S, Masakane I, Iseki K, et al. Carpal tunnel surgery as proxy for dialysis-related amyloidosis: results from the Japanese society for dialysis therapy. *Am J Nephrol.* 2014;39:449–58.
68. Hoshino J, Yamagata K, Nishi S, Nakai S, Masakane I, Iseki K, et al. Significance of the decreased risk of dialysis-related amyloidosis now proven by results from Japanese nationwide surveys in 1998 and 2010. *Nephrol Dial Transplant.* 2016;31:595–602.
69. Horie T, Yamazaki S, Hanada S, Kobayashi S, Tsukamoto T, Haruna T, et al. Outcome from a randomized controlled clinical trial-improvement of peripheral arterial disease by granulocyte colony-stimulating factor-mobilized autologous peripheral-blood-mononuclear cell transplantation (IMPACT). *Circ J.* 2018;82:2165–74.
70. Hoshino J, Fujimoto Y, Naruse Y, Hasegawa E, Suwabe T, Sawa N, et al. Characteristics of revascularization treatment for arteriosclerosis obliterans in patients with and without hemodialysis. *Circ J.* 2010;74:2426–33.
71. Hoshino J, Ubara Y, Hara S, Sogawa Y, Suwabe T, Higa Y, et al. Quality of life improvement and long-term effects of peripheral blood mononuclear cell transplantation for severe arteriosclerosis obliterans in diabetic patients on dialysis. *Circ J.* 2007;71:1193–8.
72. Hoshino J, Ubara Y, Ohara K, Ohta E, Suwabe T, Higa Y, et al. Changes in the activities of daily living (ADL) in relation to the level of amputation of patients undergoing lower extremity amputation for arteriosclerosis obliterans (ASO). *Circ J.* 2008;72:1495–8.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.