

Research Article

Assess the Application of the E-Value in the Unmeasured Confounder Evaluation of Observational Pharmaceutical Studies

Lihong Huang,^{1,2} Jianbing Ma ,³ Xiaochun Qiu,⁴ and Tao Suo ⁵

¹Department of Biostatistics, Zhongshan Hospital, Fudan University, Shanghai 200032, China

²Evidence-based Medicine Center, Fudan University, Shanghai 200032, China

³School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China

⁴Library of Shanghai Jiao Tong University School of Medicine, Shanghai 200240, China

⁵General Surgery Department of Zhongshan Hospital, The General Surgery Institute, Fudan University, Shanghai 200032, China

Correspondence should be addressed to Jianbing Ma; 1938193637@qq.com and Tao Suo; suo.tao@zs-hospital.sh.cn

Received 10 September 2021; Accepted 27 September 2021; Published 12 October 2021

Academic Editor: Zhu Xiao

Copyright © 2021 Lihong Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Public health is very important in big cities, and data analysis on public health studies is always a demanding issue that determines the study effectiveness. E-value was proposed as a standard sensitivity analysis tool to assess unmeasured confounders in observational studies, but its value is doubted. To evaluate the usefulness of E-value, in this paper, we collected 368 observational studies on drug effectiveness evaluation published from 1998 to September 2019 (out of 3426 searched studies) and evaluated the features of E-value. We selected the effects of primary outcomes or the largest effects in terms of hazard ratio, risk ratio, or odds ratio. Effects were transformed into estimated effect sizes following a standard E-value computation. In all 368 studies, the disease with the highest percentage was infections and infestations, at 21.7% (80/368). Our results showed that the median relative effect size was 1.89 (Q1-Q3: 1.41–2.95), and the corresponding median E-value was 3.19 with 95% confidence interval lower bound 1.77. Smaller studies yielded larger E-values for the effect size estimate and the relationship was considerably attenuated when considering the E-value for the lower bound of 95% confidence interval on the effect size. Notably, E-values have a monotonic, almost linear relationship with effect estimates. We found that E-value may cause misimpressions on the unmeasured confounder, and the same E-value does not reflect the varying nature of the unmeasured confounders in different studies, and there lacks a guidance on how E-value can be deemed as small or large, all of which limits the capability of E-value as a standard sensitivity analysis tool in real applications.

1. Introduction

Public health issues are drawing more and more attentions since they may cause harm to a large proportion of the population, especially in big cities where the density of population is high. One main problem of dealing with public health events is that these events often involve large amount of data and complex factors that may affect the results. In practice, it is quite important, but also in the meantime, very difficult, to assess if the results are reliable given that some of the factors are inevitably not addressed, due to ignorance or missing, called “confounding analysis.”

A formal definition of “confounding” is in terms of dependence of counterfactual outcomes and exposure, possibly conditional on covariates, while a “confounder” is defined as a preexposure covariate C for which there exists a set of other covariates X such that effect of the exposure on the outcome is unconfounded conditional on (X, C) and for no proper subset of (X, C) is the effect of the exposure on the outcome unconfounded, given the subset [1]. A study with a good design can control the confounders from the beginning, such as randomized controlled trials. However, when a randomized clinical trial (RCT) is not available, or the analysis is required to be based on real world data, an

increased attention has to be paid for the application of observational studies [2–4].

When treatment is not randomized, confounder is the common bias which is not easily controllable, especially unmeasured confounders [5, 6]. An important approach to evaluating evidence for causation in the face of unmeasured confounding is “sensitivity analysis,” that considers how strong unmeasured confounding would have to be to explain away the association [7, 8]. Such methods often require additional untestable assumptions. Some assume a single binary confounder [9, 10], and others assume there is no interaction between the effects of the exposure and the confounder on the outcome [11, 12]. Recently, VanderWeele and Ding introduced the “E-value,” a new measure which quantifies the minimum strength of association between the unmeasured confounder(s) and exposure and outcome to fully explain away a specific treatment outcome association, conditional on the measured covariates [13, 14]. The E-value is a general tool for sensitivity analysis that does not require assumptions about the nature of the unmeasured confounder. The authors recommended to report E-values in all epidemiologic investigations, as unmeasured confounding is often the central challenge in assessing evidence for causality in observational research. E-values can assess the robustness of the unmeasured confounding, thereby supplementing *P* values. Given the novelty of E-value with many appealing features, its applications are increasing rapidly.

The U.S. Food and Drug Administration published “Framework for FDA’s real-world evidence program” at the end of 2018, in which retrospective observational study was covered that such a study should identify the population and determine the exposure/treatment from historical data [15]. In light of this regulation, in this paper, we explore the use of E-value to evaluate the potential impact of unmeasured confounder(s) in retrospective observational pharmaceutical studies.

2. Methods

2.1. Selection of Studies. We selected observational pharmaceutical studies that reported results on drug effects. We always chose the most prominent associations stated in the abstracts of the articles in terms of hazard ratio (HR), risk ratio (RR), or odds ratio (OR), among which some were from the subgroup analysis. The exclusion criteria are shown in Figure 1.

Figure 1 describes the process how we selected the final studies from the whole search results.

We identified studies by searching MEDLINE and Embase databases (last search on September 29th, 2019, using keywords (“*drug screening*”/exp OR “*drug screening*”) OR (“*drug effect*”/exp OR “*drug effectiveness*”) OR “*drug evaluation*”) AND (“*observational study*”/exp OR “*observational study*”) OR “*real world study*”). We only considered articles published in English for inclusion in the analysis, and we only kept studies that measured direct efficacy on HR, RR, or OR outcomes. We excluded systematic reviews, articles on animal tests, Chinese Medicine, case reports, computer simulation, methodology, policy demonstration,

and machine learning algorithms, as well as other research studies not focusing on drug effectiveness evaluation. Two reviewers screened all titles/abstracts to apply the inclusion/exclusion criteria independently and in duplicate. The two reviewers resolved any discrepancy at each stage through a consensus process, and all studies were screened by two reviewers in parallel; the resulting list of the included articles/trials was discussed by all researchers to ensure the accuracy of the final decision.

2.2. Data Extraction. We collected data regarding the drug efficacy outcome on the sample size, the effect measure (hazard ratio, relative risk, or odds ratio), the adjusted effect size estimate, the event proportion of the effect, and the associated 95% confidence interval. Note that when multiple treatments were measured, we only focus on the most prominent contrast in terms of HR, RR, or OR, whichever deviates from the null effect the most, assuming it would yield the largest effect, thus the most robustness to unmeasured confounding [16]. The commonality of an outcome (common or uncommon) is defined as the result if the total number of events (at the end of follow-up) is great than or equal to 15% of the total number of participants.

We prespecified the following rules to select a single effect size per study. We gave priority to the primary outcome; when no primary outcome was specified or multiple treatments were measured, we selected the one giving the largest association with the selected outcome. The corresponding extracted information was retrieved from the full-text whenever necessary. Reviewer discrepancies were resolved following a consensus procedure.

2.3. Data Analyses. All effect sizes were transformed into an estimated relative effect (ERE) by the following procedure [17]:

- (a) If the outcome was uncommon and the ratio was greater than 1, we set the ERE to the original HR, RR, or OR.
- (b) If the ratio was less than 1, regardless if the outcome was common or uncommon, it was inverted for the subsequent consistent operations. So, it came to that all EREs were more than or equal to 1 afterwards. Here, we should note that once a relative effect is inverted, we also invert the corresponding 95% confidence interval (CI) bounds such that (low, high) becomes (1/high, 1/low).
- (c) After step (b), if the ratio was HR and the outcome was common, we set ERE to $(1 - 0.5^{\text{sqrt}(\text{HR})}) / (1 - 0.5^{\text{sqrt}(1/\text{HR})})$.
- (d) After step (b), if the ratio was OR and the outcome was common, we set ERE to $\text{sqrt}(\text{OR})$.

Using the ERE, we calculated its E-value using the following equation:

$$E - \text{Value} = \text{ERE} + \text{sqrt}(\text{ERE} * (\text{ERE} - 1)). \quad (1)$$

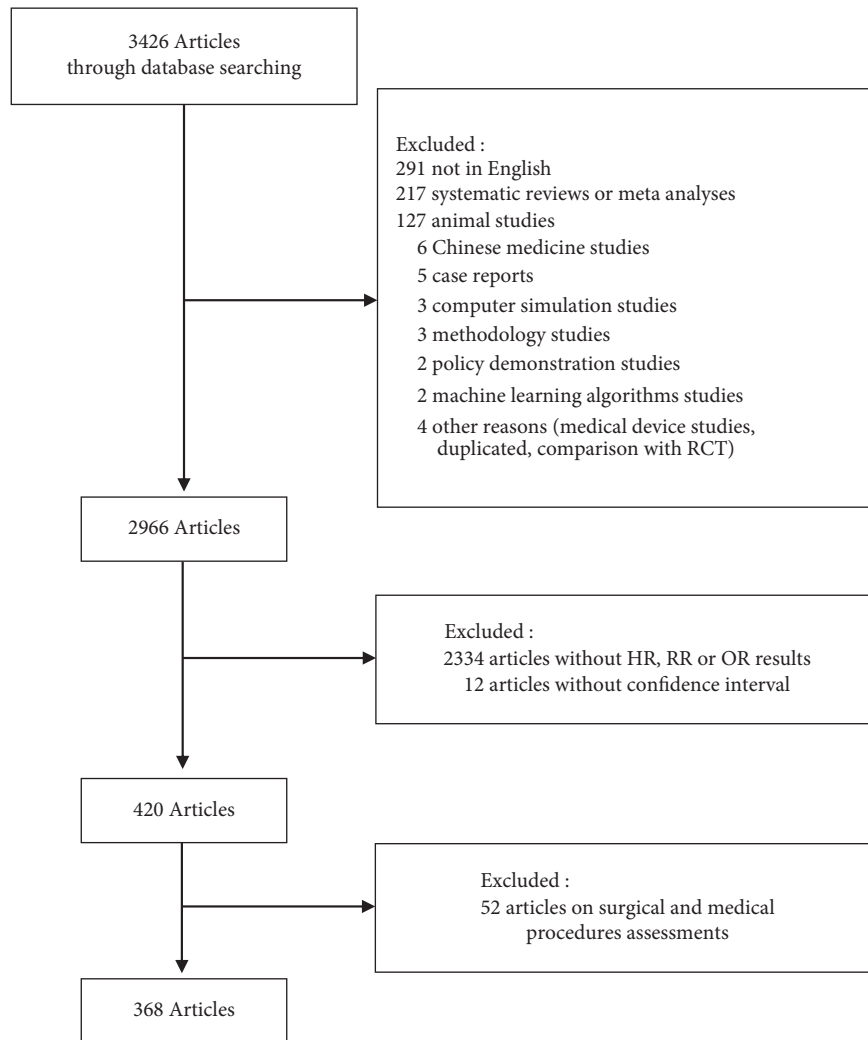


FIGURE 1: Selection of observational pharmaceutical studies.

The above workflow of E-value calculations is also shown in Figure 2.

Figure 2 describes the workflow of E-value calculations.

We also computed the E-value for the lower bound (LB) of the 95% CI using the following equation (note that some of the bounds had been inverted along with their corresponding effects):

$$E - \text{Value} = \text{LB} + \text{sqrt}(\text{LB} * (\text{LB} - 1)). \quad (2)$$

As standard errors were also used for illustration, for figure clarity, we used logged standard errors (LSE) as

$$\text{LSE} = \frac{(\text{Ln}(\text{UB}) - \text{Ln}(\text{LB}))}{3.92}, \quad (3)$$

where UB and LB are upper and lower bounds of 95% CI. For abbreviation, in all the figures in this paper, instead of logged standard error, we simply use the phrase standard error when there is no confusion.

The number of publications were summarized annually and by disease categories in system organ class (SOC) of MedDRA (Medical Dictionary for Regulatory Activities).

We summarized the distribution of the EREs, E-values, and standard errors across all studies by the median, 25% and 75% percentiles in original values and logged values. In order to assess the association between the effect size estimate and study size, we produced a funnel plot showing the standard error against the effect size. We also produced scatter plots depicting the logged E-value and logged effect size against the standard errors. Finally, we produced scatter plots showing the EREs against the E-values.

We used *Python 3.7* for all analyses. The codes on ERE and E-value calculation are listed in Appendix. All data and statistical codes are available on request.

3. Results

We selected all the 3426 articles from 1998 to the last searching day of September 2019. We identified 368 articles that were eligible for the final analysis (Figure 2), and the details of the study characteristics are shown in Table 1.

During the study period, the number of publications was increased. The studies after 2014 accounted for more than 65% (65.2%, 240/368) of all publications (Figure 3(a)). In all

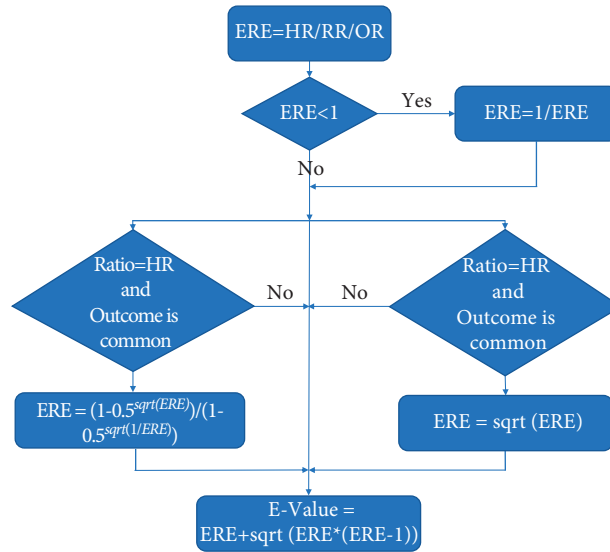


FIGURE 2: E-value calculation workflow.

TABLE 1: Study characteristics of the observational pharmaceutical studies.

Characteristics	Total	Infections and infestations	Cardiac disorders	Neoplasms benign, malignant, and unspecified
Study Numbers (N)	368	80	51	26
Sample size, median (Q1, Q3)	634 (205, 4721)	314 (157, 1064)	3005 (610, 14033)	506 (105, 4036)
Publication year, median (Q1, Q3)	2016 (2012, 2018)	2016 (2012, 2017)	2014 (2010, 2017)	2016 (2011, 2018)
Effect measure:				
Hazard ratio, <i>n</i> (%)	156 (42.4)	25 (31.3)	35 (68.6)	12 (46.2)
Risk ratio, <i>n</i> (%)	31 (8.4)	7 (8.8)	3 (5.9)	3 (11.5)
Odds ratio, <i>n</i> (%)	181 (49.2)	48 (60.0)	13 (25.5)	11 (42.3)
Uncommon (<15%) outcome, <i>n</i> (%)	203 (55.2)	37 (46.3)	34 (66.7)	15 (57.7)
Protective effect (effect size <1), <i>n</i> (%)	148 (40.2)	24 (30.0)	28 (54.9)	17 (65.4)
Negative result, <i>n</i> (%)	47 (12.8)	11 (13.8)	9 (17.6)	3 (11.5)

368 studies, the disease with the highest percentage of the observational pharmaceutical studies was infections and infestations, at 21.7% (80/368), followed with cardiac disorders (13.9%, 51/368) and neoplasms benign, malignant, and unspecified (including cysts and polyps) (7.1%, 26/368). The distribution of the studies by disease categories in SOC of MedDRA is shown in Figure 3(b).

Figure 3(a) presents the numbers of publications in years from 1998 to 2019. Note that our search day was Sep. 29th, 2019, so the number of publications in the whole year 2019 would be higher.

Figure 3(b) depicts the proportions of outcomes by the Meddra disease categories. It can be seen that the odds ratio was used the most in observational pharmaceutical studies, as expected. *Note.* Blocks 1–16 represent “infections and infestations,” “cardiac disorders,” “neoplasms benign, malignant, and unspecified (incl cysts and polyps),” “nervous system disorders,” “musculoskeletal and connective tissue disorders,” “not a single disease,” “general disorders and administration site conditions,” “endocrine disorders,”

“respiratory, thoracic, and mediastinal disorders,” “injury, poisoning, and procedural complications,” “renal and urinary disorders,” “gastrointestinal disorders,” “psychiatric disorders,” “vascular disorders,” “metabolism and nutrition disorders,” and “all others.”

After inverting, effect sizes were necessary so that all relative effects were greater than 1, the selected studies had median relative effect of 1.89 (Q₁ = 1.41 and Q₃ = 2.95), and the corresponding median E-value for all selected studies was 3.19, which means that, in an observational pharmaceutical research, it would typically assume an uncontrolled confounder that is associated with both the exposure and the outcome by a relative effect of 3.19 each to turn the estimate into a null estimate (Table 2).

From the funnel plot in Figure 4, although most of the selected studies showed statistically significant associations, there were still a few negative results published, and smaller studies (i.e., with larger standard errors) necessarily showed larger effect size. Figure 5(c) shows that smaller studies yielded larger E-values for the effect size estimate.

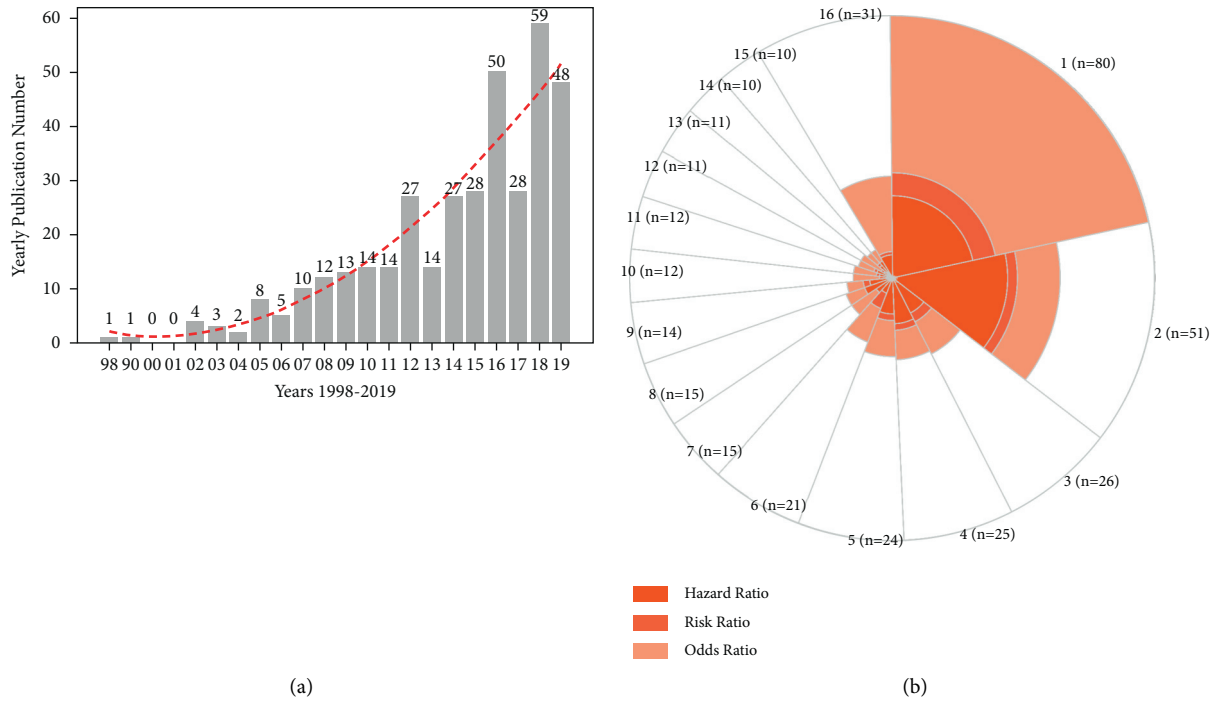


FIGURE 3: Description of the observational pharmaceutical publications: (a) yearly publication number and (b) proportion of outcomes by disease categories in SOC of MedDRA.

TABLE 2: Effect size estimates, precision, and E-values in observational pharmaceutical studies.

Disease	Statistics	Median (Q1, Q3)	Logged: median (Q1, Q3)
All (N = 368)	Relative effect size	1.89 (1.41, 2.95)	0.64 (0.35, 1.08)
	Standard error	0.63 (0.22, 2.28)	0.27 (0.13, 0.49)
	E-value	3.19 (2.18, 5.35)	1.16 (0.78, 1.68)
	E-value for the 95% CI lower bound	1.77 (1.31, 2.71)	0.57 (0.27, 1.00)
Infections and infestations (N = 80)	Relative effect size	2.18 (1.64, 2.88)	0.78 (0.50, 1.06)
	Standard error	1.31 (0.44, 3.86)	0.38 (0.21, 0.61)
	E-value	3.78 (2.67, 5.20)	1.33 (0.98, 1.65)
	E-value for the 95% CI lower bound	1.69 (1.28, 2.68)	0.52 (0.25, 0.98)
Cardiac disorders (N = 51)	Relative effect size	1.54 (1.30, 1.93)	0.43 (0.27, 0.66)
	Standard error	0.31 (0.14, 0.59)	0.15 (0.10, 0.26)
	E-value	2.45 (1.93, 3.27)	0.90 (0.66, 1.18)
	E-value for the 95% CI lower bound	1.54 (1.22, 2.08)	0.43 (0.20, 0.73)
Neoplasms benign, malignant and unspecified (incl cysts and polypos) (N = 26)	Relative effect size	1.85 (1.58, 2.59)	0.61 (0.46, 0.95)
	Standard error	0.59 (0.30, 2.19)	0.29 (0.15, 0.49)
	E-value	3.10 (2.54, 4.63)	1.13 (0.93, 1.53)
	E-value for the 95% CI lower bound	2.06 (1.56, 3.25)	0.72 (0.45, 1.17)

CI, confidence interval on the effect size. Note that, for consistency, we inverted effect size was necessary so that all effects were greater than 1. Here, for logged standard error (LSE), we use the equation $LSE = (\ln(UB) - \ln(LB)) / 3.92$, where UB and LB are upper and lower bounds of 95% CI.

In fact, smaller studies show larger effect size estimates and large effect size estimates give larger E-values (Figures 5(a) and 5(c)). Figures 5(b) and 5(d) show that the relationship

is considerably attenuated when considering the E-value for the lower bound of 95% confidence interval on the effect size.

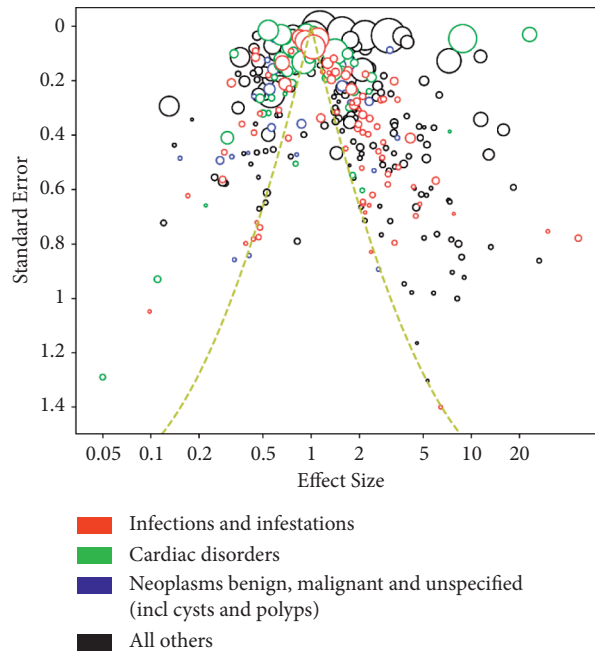


FIGURE 4: Funnel plot.

Figures 5(a) and 5(c) plot the effect size and E-value against the standard error, respectively. Figures 5(b) and 5(d) show the 95% CI lower bounds of the effect size and E-value against the standard error, respectively. The red curves indicate the least square simulation.

Figure 6 shows the association between the estimated effect size and E-value. For all investigated effect sizes, including both common and uncommon HRs (Figure 6(a)), RRs (Figure 6(b)); note that RRs do not change by the commonality), and both common and uncommon ORs (Figure 6(c)), E-values have a monotonic, almost linear relationship with effect estimates. Monotonic rise or monotonic decline depends on whether the estimated effect size is greater than or less than 1.

Figure 6 plots the association between estimated effect size and E-value. An outcome is common if the total number of events (at the end of follow-up) is great than or equal to 15% of the total number of participants. HR and OR are transformed into an estimated effect size based on the outcome commonality, so in Figures 6(a) and 6(c), two curves indicate the common and uncommon least square fit curves. RR does not change on the outcome commonality, so there is only one curve in Figure 6(b).

4. Discussion

Observational studies are noninterventional clinical study designs that are covered by FDA real world evidence program [16]. However, observational studies may be much less convincing due to the lack of randomization and the presence of confounding bias [17]. While traditional regression-based and propensity score analyses provide some control of confounding, they can only take into account factors that are measured. To assess how much unmeasured

confounding factors may pose to a study, researchers may conduct a sensitivity analysis, and here, the E-value analysis answers the question: how strong would the unmeasured confoundings have to be to negate the observed results? In this paper, we calculated the E-values for the published observational pharmaceutical studies, aiming to evaluate the features of this sensitivity indicator.

With the development of the electronic medical data, observational studies were applied more and more in drug evaluation, as indicated by the increasing trend of publication numbers. Although most of the published studies showed statistically significant associations, there were still few negative results published, which indicated the importance of the drug evaluation in the real clinical practice. Similar with the previous study [18], our results also showed the E-values based on the lower bound of the confidence interval are less influenced by study size, suggesting it is better to report the E-value based on the confidence interval, rather than the E-value based on the estimate. However, our results showed that the overall E-value for the observational pharmaceutical studies was 3.19 with 95% CI lower bound 1.77; it is hard for us to determine whether this estimated E-value is bigger enough to eliminate the concern about the unmeasured confounder(s).

As observed by Localio et al. [19], it is worth noting that our publication data-based results also showed that the E-value is almost linearly related to the absolute value of the effect estimate and the relationship is monotonic; the more the effect estimate deviates from the null, the larger the E-value is. In our results, the studies with small sample sizes will have larger standard errors on the effect size and will yield larger effect size estimates as compared to larger studies. As a monotonic consequence, smaller studies with larger effect size are more likely to give larger E-values, thus giving the misimpression of being more robust to unmeasured confounder(s).

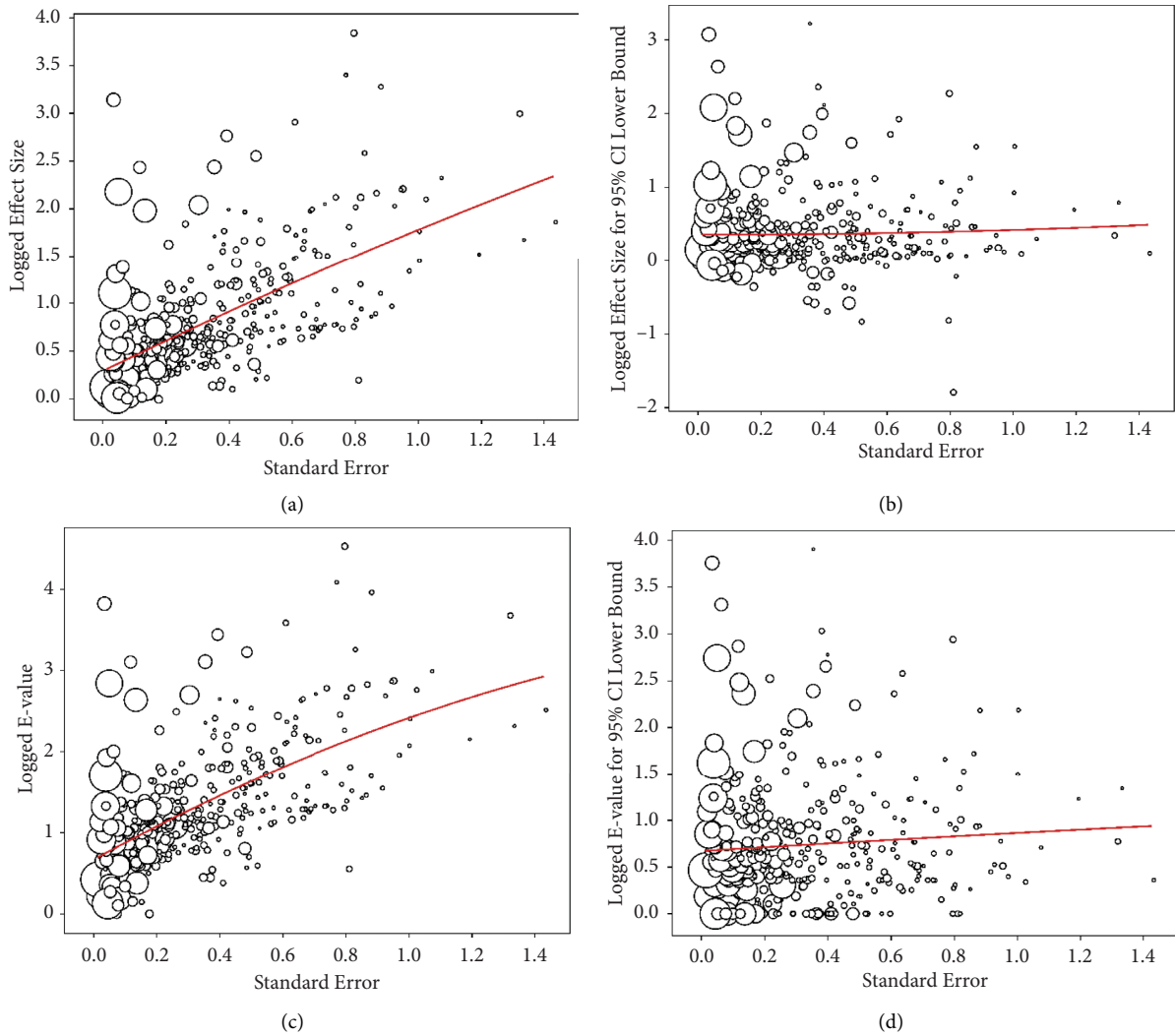


FIGURE 5: Scatter plots depicting the estimated effect size or E-value against standard error. (a) Logged effect size. (b) Logged effect size for the lower limit of 95% CI. (c) Logged E-value. (d) Logged E-value for the lower limit of 95% CI.

Therefore, E-value is an alternative approach to sensitivity analyses for unmeasured confounding in observational studies, with some advantages and disadvantages. It has two major appealing features. First, in contrast to standard methods for sensitivity, it requires much less assumptions from investigators, such as assuming that exposure prevalence and confounder prevalence are at the point that maximizes confounding. Second, it is intuitive because the lowest limit is 1, and the higher the E-value is, the stronger the unmeasured confounder must be to explain the observed association [20]. Notably, E-value does have some limitations. First, E-values have a monotonic, almost linear relationship with effect estimates, and, therefore, provide little additional information, thus have the risk to provide the misimpression of the unmeasured confounder. Second, from the formula of the E-value, it is a transformation of effect estimates, so a given effect estimate always produces the same E-value, but the reality of the unmeasured confounder varies from study to study. Third, there is no specific guidance on the range within which an E-value should be

deemed as small and thus the residual confounder is still a serious threat [21].

It is worth mentioning that causal inference from the observational study is a vital problem, but it comes with strong assumptions. Most method assume that we observe all confounders. Therefore, it is crucial to diagnose the unmeasured confounder in the observational study. Ideally, if we can control the measured confounders very well and determine that the unmeasured confounders do not exist, then there is no need to worry about the impact of the confounders. Several methods have been reported for detecting confounders, Wang and Blei developed the algorithm “deconfounder” which uses the multiplicity of causes to infer unmeasured confounders [22] and “negative controls” which is to repeat the experiment under conditions in which it is expected to produce a null result and verify that it does indeed produce a null result and also is a tool for evaluating confounders in observational studies [23]. As a sensitivity tool of detecting unmeasured confounders, E-value had been applied in some observational studies, and

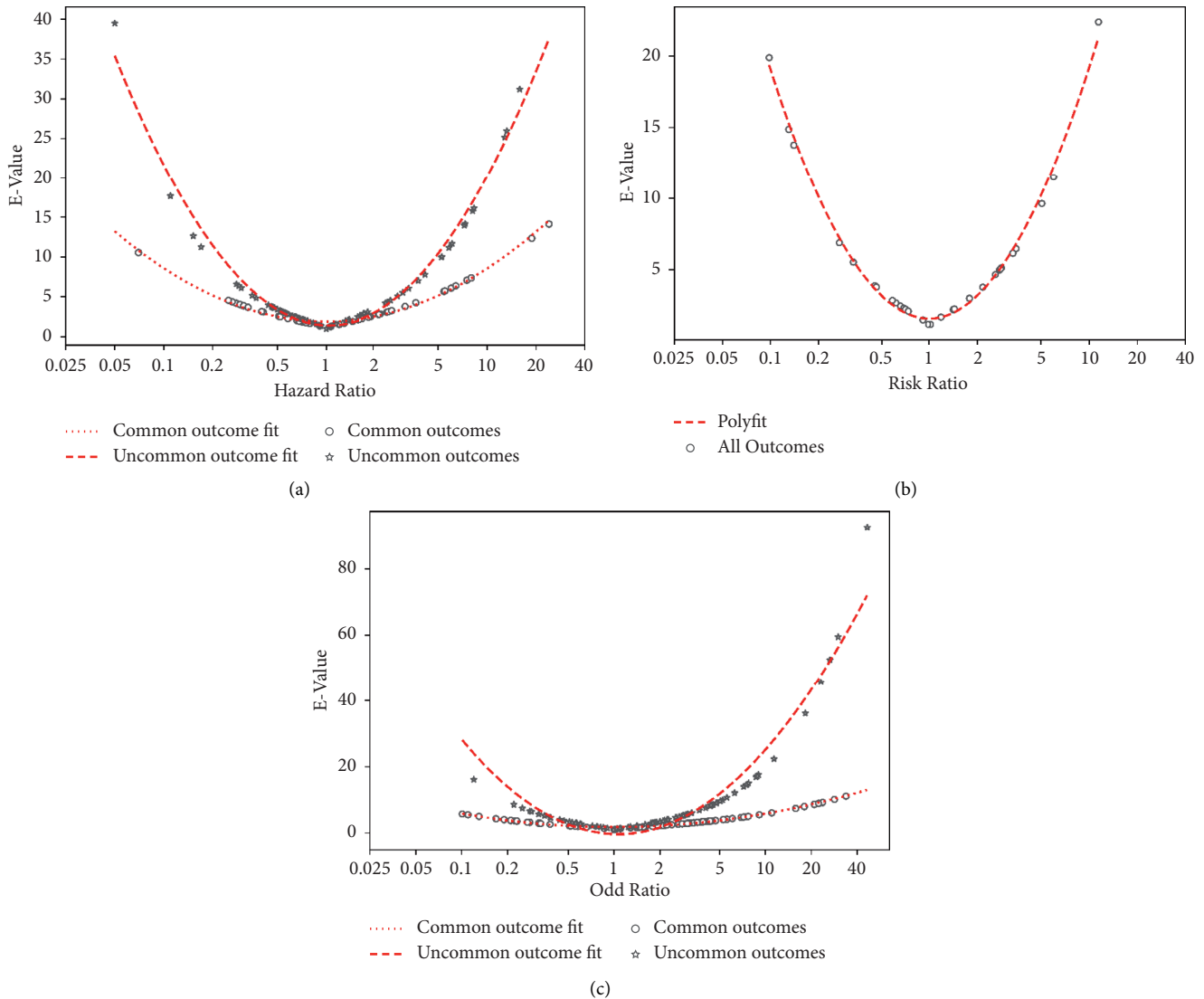


FIGURE 6: Association between estimated effect size and E-value. (a) Hazard ratio. (b) Risk ratio. (c) Odds ratio.

it was reported that 87 papers with 516 E-values were identified till the end of 2018 in an empirical assessment of the observational studies [24]. Because of the above several limitations of E-value, we have summarized E-value should not be a substitute for careful consideration of potential sources of unmeasured confounders; if used, the evaluation criteria should be interpreted in the context of expected confounding in specific fields.

5. Conclusion

With lots of data and factors being involved in public health events in big cities, confounding analysis has to play a key role in determining the quality of the study results. In particular, confounding control is a big challenge in observational pharmaceutical studies, especially the unmeasured confounder(s). E-value analysis could provide information on how strong the unmeasured confounders would have to be to negate the observed results, but

researchers may still ask for other analysis methods to remedy the limitations of E-value. Researchers need to consider confounder in a systematic, thorough, and balanced way. Sensitivity analysis is a common applied method to evaluate the unmeasured confounder. More and better methods are expected for the observational pharmaceutical studies.

Appendix

The Python code on ERE and E-value calculation

```
import math
def standardev(r):
    return r + math.sqrt(r*(r-1))
def relativer(r,whichr,common = False):
    myr = r
    if(common):
```



```

if(whichr == 'hr'):
    myr = (1-pow(0.5,math.sqrt(r)))/(1-pow(0.5,-
math.sqrt(1/r)))
elif(whichr == 'or'):
    myr = math.sqrt(r)
return myr
def rtoev(r,whichr,common = False):
    return standardev(relativer(r,whichr,common))
def rrcitoev(rrci):
    l = rrci[0]
    if(l<=1): return 1
    else: return l+math.sqrt(l*(l-1))
def geteffectsize(hr,rr,anor):
    es = hr
    index = 'hr'
    isrev = False
    if (hr>0 and hr<1):
        es = 1/hr
        isrev = True
    if(rr>0 and max(rr,1/rr)>es):
        es = max(rr,1/rr)
        index = 'rr'
    if(rr<1):
        isrev = True
    if(anor>0 and max(anor,1/anor)>es):
        es = max(anor,1/anor)
        index = 'or'
    if(anor<1):
        isrev = True
    return es, index, isrev

```

Data Availability

The data used to support the findings of this study are available online at MEDLINE and Embase databases (last search on September 29th, 2019, using keywords (“drug screening”/exp OR “drug screening”) OR (“drug effect”/exp OR “drug effectiveness”) OR “drug evaluation”) AND (“observational study”/exp OR “observational study”) OR “real world study”).

Conflicts of Interest

The authors have no conflicts of interest to declare.

Authors' Contributions

H. H. and J.M. contributed to the study design. H. H. and X. Q. contributed to data collection. H. H. and J. M. performed statistical analysis and interpretation and drafted the manuscript. H. H., J. M., and T. S. revised the manuscript.

All authors contributed to critical revision of the manuscript and approved its final version.

Acknowledgments

The present study was supported by the Sichuan Provincial Key Research and Development Program (no. 2021YFG0345 to Jianbing Ma) and National Natural Science Foundation of China (no. 81903407 to Lihong Huang).

References

- [1] T. J. VanderWeele and I. Shpitser, “On the definition of a confounder,” *Annals of Statistics*, vol. 41, no. 1, pp. 196–220, 2013.
- [2] J. P. Vandenbroucke, “When are observational studies as credible as randomised trials?” *The Lancet*, vol. 363, no. 9422, pp. 1728–1731, 2004.
- [3] A. R. Feinstein, “Current problems and future challenges in randomized clinical trials,” *Circulation*, vol. 70, no. 5, pp. 767–774, 1984.
- [4] N. Black, “Why we need observational studies to evaluate the effectiveness of health care,” *BMJ*, vol. 312, no. 7040, pp. 1215–1218, 1996.
- [5] S. Greenland and H. Morgenstern, “Confounding in health research,” *Annual Review of Public Health*, vol. 22, no. 1, pp. 189–212, 2001.
- [6] F. Shaya and A. Gu, “Confounding in health services research: issues and solutions,” *Research in Human Capital and Development*, vol. 16, pp. 67–78, 2007.
- [7] P. R. Rosenbaum, *Sensitivity to Hidden Bias*, Springer, New York, NY, USA, 2002.
- [8] P. R. Rosenbaum, “Design sensitivity and efficiency in observational studies,” *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 692–702, 2010.
- [9] P. R. Rosenbaum and D. B. Rubin, “Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome,” *Journal of the Royal Statistical Society: Series B*, vol. 45, no. 2, pp. 212–218, 1983.
- [10] G. W. Imbens, “Sensitivity to exogeneity assumptions in program evaluation,” *The American Economic Review*, vol. 93, no. 2, pp. 126–132, 2003.
- [11] J. J. Schlesselman, “Assessing effects of confounding variables,” *American Journal of Epidemiology*, vol. 108, no. 1, pp. 3–8, 1978.
- [12] D. Y. Lin, B. M. Psaty, and R. A. Kronmal, “Assessing the sensitivity of regression results to unmeasured confounders in observational studies,” *Biometrics*, vol. 54, pp. 948–963, 1998.
- [13] T. J. Vanderweele and P. Ding, “Sensitivity analysis in observational research: introducing the E-value,” *Annals of Internal Medicine*, vol. 167, no. 4, pp. 268–274, 2017.
- [14] P. Ding and T. J. VanderWeele, “Sensitivity analysis without assumptions,” *Epidemiology*, vol. 27, no. 3, pp. 368–377, 2016.
- [15] USFDA, *Framework for FDA’s Real-World Evidence Program*, Food and Drug Administration, Silver Spring, MA, USA, 2018.
- [16] D. C. Klonoff, “The new FDA real-world evidence program to support development of drugs and biologics,” *Journal of Diabetes Science and Technology*, vol. 14, no. 2, pp. 345–349, 2020.
- [17] M. A. Hernán and J. M. Robins, *Causal Inference*, CRC, Boca Raton, FL, USA, 2010.
- [18] L. Trinquart, A. L. Erlinger, J. M. Petersen, M. Fox, and S. Galea, “Applying the E value to assess the robustness of

- epidemiologic fields of inquiry to unmeasured confounding,” *American Journal of Epidemiology*, vol. 188, no. 6, pp. 1174–1180, 2019.
- [19] A. R. Localio, C. B. Stack, and M. E. Griswold, “Sensitivity analysis for unmeasured confounding: E-values for observational studies,” *Annals of Internal Medicine*, vol. 167, no. 4, pp. 285–286, 2017.
- [20] S. Haneuse, T. J. VanderWeele, and D. Arterburn, “Using the E-value to assess the potential effect of unmeasured confounding in observational studies,” *Jama*, vol. 321, no. 6, pp. 602–603, 2019.
- [21] J. P. A. Ioannidis, Y. J. Tan, and M. R. Blum, “Limitations and misinterpretations of E-values for sensitivity analyses of observational studies,” *Annals of Internal Medicine*, vol. 170, no. 2, p. 108, 2019.
- [22] Y. Wang and D. M. Blei, “The blessings of multiple causes,” *Journal of the American Statistical Association*, vol. 114, no. 528, pp. 1574–1596, 2019.
- [23] M. Lipsitch, E. Tchetgen Tchetgen, and T. Cohen, “Negative controls,” *Epidemiology*, vol. 21, no. 3, pp. 383–388, 2010.
- [24] M. R. Blum, Y. J. Tan, and J. P. A. Ioannidis, “Use of E-values for addressing confounding in observational studies—an empirical assessment of the literature,” *International Journal of Epidemiology*, vol. 49, no. 5, pp. 1482–1494, 2020.