Do Case-Control Studies Always Estimate Odds Ratios?

Jeremy A. Labrecque, Myriam M.G. Hunink, M. Arfan Ikram and M. Kamran Ikram

Correspondence address: Dr. Jeremy Labrecque, Department of Epidemiology, Erasmus MC, PO Box 2040, 3000 CA, Rotterdam, The Netherlands (email: j.labrecque@erasmusmc.nl)

Affiliations: Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands

Funding: No funding declared.

Conflicts of interest: Dr. Hunink receives Royalties from Cambridge University Press for a textbook on Medical Decision Making, reimbursement of expenses from the European Society of Radiology (ESR) for work on the ESR guidelines for imaging referrals, reimbursement of expenses from the European Institute for Biomedical Imaging Research (EIBIR) for membership of the Scientific Advisory Board, and research funding from the American Diabetes Association, the Netherlands Organization for Health Research and Development, the German Innovation Fund, and the Society for Medical Decision Making.

Running head: What do Case-Control Studies Estimate?

Abstract

Case-control studies are an important part of the epidemiologic literature, yet there remains a lot of confusion about how to interpret estimates from different case-control study designs. We demonstrate that not all case-control study designs estimate odds ratios. In contrast, case-control studies in the literature often report odds ratios as their main parameter even when using designs that do not estimate odds ratios. Therefore, only studies using specific case-control designs should report odds ratios, whereas the case-cohort and incidence density sampled case-control studies must report risk ratio and incidence rate ratios, respectively. This also applies to case-control studies conducted in open cohorts which often estimate incidence rate ratios. We also demonstrate the misinterpretation of case-control study estimates in a small sample of highly-cited case-control studies in general epidemiologic and medical journals. We therefore suggest greater care be taken when considering which parameter is to be reported from a case-control study.

Keywords: case-control studies; control sampling; odds ratio; risk ratio; incidence rate ratio

The case-control study is an important but often misunderstood study design. In our current understanding, a case-control study is nothing but an efficiently conducted cohort study achieved by sampling a subset of potential controls to get a measure of exposure distribution among them. Much has been written trying to clear the confusion in conducting and reporting on case-control studies (1–4) but there remain misunderstandings about how to interpret results from case-control studies. In 2008, Knol et al (1) reviewed 150 case-control studies in 20 journals to survey which parameter they reported—odds ratios, risk ratios or incidence rate ratios—and which parameter their design may have allowed them to estimate. They found that 90% of the studies only reported an odds ratio (OR) despite the fact that the majority used designs that estimate risk ratios or incidence rate ratios. It appears as though the attitude toward case-control studies is that it is always correct to present an OR but that, with some designs, the estimate also has a second interpretation either as a risk ratio or rate ratio. Here, we aim to prevent further such confusion about the parameter estimated in case-control studies by explaining why many commonly-used case-control designs do not, in fact, estimate ORs.

What does a case-control study estimate?

The findings of Knol et al (1) should make epidemiologists pause. How is it that the parameter estimated in a case-control design—a design which is required learning to obtain a degree in epidemiology at any level—is misinterpreted in the majority of studies using it in the leading medical and epidemiologic journals? This OR-centric view of case-control studies can also be seen in epidemiologic textbooks: "[I]n a case-control study the relative risk cannot be calculated directly,"(5, p. 208) or, "the primary measure of effect in a case-control study is the OR."(6, p. 45)

Much of this confusion is due to a mismatch between the statistical interpretation of a two-by-two table from a case-control study (Table 1) and its epidemiologic reality. Let us begin with the most often taught example, a case-control study conducted within a closed cohort where controls are sampled at the end of follow-up. To simplify things, let us imagine we have all the cases meaning we know the true values in column $Y = 1$ of our 2×2 table. In this design, we sample a fraction, f, of the participants from column $Y = 0$ Therefore, we know the value of a and c as well as the values $b * f$ and $d * f$. If we know the sampling fraction (f), we can calculate b and d allowing us to calculate any measure of association. (7)

If we do not know f, what can we estimate knowing only a, c, $b * f$ and $d * f$? Most will answer, correctly, that we can estimate an OR. This is where the mismatch between statistical interpretation and epidemiologic reality begins. When an estimate is described as an OR, we think immediately of the *disease* OR which is the ratio of disease odds in the exposed group, a/b, divided by the disease odds in the unexposed group, c/d. Case-controls studies cannot directly estimate the disease OR, however, because we have not sampled everyone in column $Y = 0$ and therefore do not know b and d, only $b * f$ and $d * f$. What they can calculate is an *exposure* OR: the odds of exposure in the cases, a/c, divided by the odds of exposure in the non-cases, $b * f/d * f = b/d$. In this design, the exposure OR, which we refer to as a case/non-case exposure OR, is equal to the disease OR:

$$\text{Case/non-case exposure OR} = \frac{a/c}{bf/df} = \frac{ad}{bc} \times \frac{f}{f} = \frac{ad}{bc} = \frac{a/b}{c/d} = \text{Disease OR}$$

We can repeat this procedure with the case-cohort design where we sample from the 'Total' column in Table 1, again, with a sampling fraction *f*. We can calculate an exposure OR but this time with a different denominator. We divide the odds of exposure in the cases (a/c) by the odds

of exposure in the total population $((c+d)*f/(a+b)*f)$. Note that it is not possible for the case/non-case exposure OR from the previous paragraph, $\frac{a/c}{b/d}$, to be equal to this exposure OR $\frac{a/c}{(a+b)/(c+d)}$, which we will call the case/cohort exposure OR. Therefore, even though we have followed the same statistical procedure as in the previous paragraph, the epidemiologic reality of this estimate is different. The case/cohort exposure OR is *not equal* to the disease OR but the risk ratio:

$$\text{Case/cohort exposure OR} = \frac{a/c}{(a+b)f/(c+d)f} = \frac{a(c+d)}{c(a+b)} \times \frac{f}{f} = \frac{a(c+d)}{c(a+b)} = \frac{a/(a+b)}{c/(c+d)}$$

$$= \text{Risk Ratio}$$

Note that the exposure OR from a case-cohort design is *not an approximation* of the risk ratio. It is, in fact, a mathematically equivalent way of expressing the risk ratio. Here we see the mismatch between statistical interpretation and epidemiologic reality. Although we have only used two-by-two tables to this point, we could also use logistic regression to analyze our case-cohort study. We are taught that the exponentiated coefficient from a logistic regression must be interpreted as an OR. When data from a case/cohort design are analyzed with logistic regression, the exponentiated coefficient *cannot* be interpreted as an OR but *only* as a risk ratio.

The same is true for designs that use sample person-time rather than participants. In these designs the denominator of the exposure OR is the odds of exposure across a sample f of all person-time: $PT_1 * f/PT_0 * f = PT_1/PT_0$. This case/person-time exposure OR is equal to the incidence rate ratio:

$$\text{Case/person-time exposure OR} = \frac{a/c}{fPT_1/fPT_0} = \frac{aPT_0}{cPT_1} \times \frac{f}{f} = \frac{aPT_0}{cPT_1} = \frac{a/PT_1}{c/PT_0} = \text{Rate Ratio}$$

Again, the case/person-time exposure OR is not equal to the disease OR and yet most studies with designs that estimate the incidence rate ratio report ORs.(1) Either they believe these designs can be interpreted as disease ORs or are choosing to report exposure ORs. Furthermore, calling it an OR not only defeats the purpose of using these designs, but can be misleading to the reader who assumes, reasonably, that when the term OR is used without qualifier, it refers to the disease OR.

The same logic as for case-control studies with incidence density sampling can be applied to open cohorts that either match on time or are conducted in populations where the prevalence of exposure is constant and are therefore sampling person-time.(2) Estimates from these designs can only be interpreted as incidence rate ratios and not ORs. This is important because, according to Knol et al 2008 (1), these designs are the most common case-control study design and also appear to be the most often misinterpreted.

Why is this important?

If the estimate from case-cohort or incidence density sampling designs *is not equal to* the disease odds ratio, why do studies employing these designs continue to refer to their estimates as ORs? An even simpler way of thinking about this is that a risk ratio or rate ratio cannot be equal to a disease OR (unless all are equal to one). Therefore, it is only possible for an estimate to have, at most, one of these interpretations. Using the case/cohort design and referring to the estimate as an OR is equivalent to using a model that estimates risk ratios (*e.g.* log-binomial regression) and calling the parameter an OR.

The literature on case-control studies sometimes uses confusing language which may lead to some of these misconceptions. One textbook, referring to a case-cohort design, states that, “the expected

EOR [exposure odds ratio] from this case-control study would closely approximate the risk ratio from a corresponding follow-up study, even if the follow-up study was never done!" (8, p. 84) In fact, such a design does more than approximate the risk ratio, it is an estimator of the risk ratio. Another popular textbook states, "relative risks cannot be calculated directly from a case-control study," because case-control studies only obtain an "estimate of relative risks based on the odds ratios that are obtained in the case-control studies." (5, p. 208) Again, this is not correct. Relative risks can be directly calculated from case/cohort designs and this does not rely on any special relationship between the risk ratio and the disease OR. The risk ratio and the case/cohort exposure OR are mathematically equivalent. Even statements such as, "using a case-cohort design, one can estimate the risk ratio," (7, p. 124) are potentially ambiguous as the reader may think an alternative parameter can also be presented.

We wish to point out that we are not advocating for the use of any of the terms we are using here except for teaching. These terms are simply to point out that there are three different types of exposure OR which do not share the same properties and it is therefore incorrect to assume that all exposure ORs can be interpreted as a disease OR.

Have things improved?

We ran a modified version of the review by Knol et al (1) selecting the two most-cited case-control studies in the past five years from each of the following journals: *Lancet*, *New England Journal of Medicine*, *JAMA*, *BMJ*, *Annals of Internal Medicine*, *American Journal of Epidemiology*, *International Journal of Epidemiology*, *Epidemiology*, *European Journal of Epidemiology* and the

Journal of Epidemiology and Community Health. The search strategy and related code to run the search can be found in Web Appendix 1 as well as a table of the parameters reported.

Of the 20 studies we reviewed, 19 reported an OR and one reported a hazard ratio (Web Table 1). The latter was a nested case-control study with a known sampling fraction allowing the authors to analyze their case-control data as though it were a cohort. Furthermore, 13 studies used a design that realistically estimated the incidence rate ratio and only four used designs that estimated ORs. In some studies, ambiguity in the description of control selection made it difficult to determine which parameter was being estimated.

Two studies (9,10) explicitly mentioned that their sampling design allows them to interpret their estimates as incidence rate ratios yet report ORs as their main parameters. For example, Friis et al (10, p. 349) state, "With the nested case–control design and risk set sampling of control participants, the OR provides unbiased estimates of the corresponding incidence rate ratios in the underlying source population, without distortion by competing risks." Despite this awareness, the authors present ORs as their main parameter. As we have argued, this reflects a long-standing misconception about case-control studies: rather than providing authors an option of whether to report an OR or incidence rate ratio (or risk ratio as the case may be), the study design and in particular the sampling strategy for the controls directly determines what parameter is being estimated.

Conclusion

Many epidemiologists before us have laid out proofs and explanations for why some case-control study designs can be interpreted as risk ratios and incidence rate ratios.(1–3) What has been missing from the literature and textbooks, is the clarification that these study designs not only *can* be interpreted as risk ratios or incidence rate ratios but *must* be interpreted as such. Lack of understanding of this point can be clearly seen in the literature where ORs are reported as the parameter of interest regardless of the design. Furthermore, a clear definition of the different types of exposure odds has been lacking. The term exposure OR should not be used without being clear about who is in the denominator. Finally, it is important to know that in a case-control study the *sampling strategy determines* which measure of association you are estimating and should be reported clearly.

References

1. Knol MJ, Vandenbroucke JP, Scott P, et al. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *American Journal of Epidemiology*. 2008;168(9):1073–1081.

2. Vandenbroucke JP, Pearce N. Case-control studies: Basic concepts. *International Journal of Epidemiology*. 2012;41(5):1480–1489.

3. Miettinen OS. Estimability and Estimation in Case-Referent Studies. *American Journal of Epidemiology*. 1976;103(2):226–235.

4. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *American Journal of Epidemiology*. 1982;116(3):547–553.

5. Gordis L. Epidemiology. 4th ed. Philadelphia: Elsevier/Saunders; 2009.

6. Kestenbaum B, Adeney KL, Weiss NS, et al. Epidemiology and biostatistics: An introduction to clinical research. Dordrecht ; New York: Springer; 2009.

7. Rothman KJ, Greenland S, Lash T. Modern Epidemiology. Third edit. Philadelphia: Lippincott Williams & Williams; 2008.

8. Kleinbaum DG, Sullivan KM, Barker ND. A Pocket Guide to Epidemiology. New York, NY: Springer New York; 2007.

9. Wiese AD, Griffin MR, Schaffner W, et al. Opioid Analgesic Use and Risk for Invasive Pneumococcal Diseases. *Annals of Internal Medicine*. 2018;168(6):396-404.

10. Friis S, Riis AH, Erichsen R, et al. Low-dose aspirin or nonsteroidal anti-inflammatory drug use and colorectal cancer risk: A population-based, case-control study. *Annals of Internal Medicine*. 2015;163(5):347–355.

Table 1. Two by two table from a fully enumerated cohort.

Exposure	***Y*=1**	***Y*=0**	**Total**	**Person-time**
E=1	*a*	*b*	N_1	PT_1
E=0	*c*	*d*	N_0	PT_0

Y outcome, *E* exposure, N_1 total unexposed, N_0 total exposed, PT_1 person-time exposed, PT_0 person-time unexposed

Web Appendix 1

Article Title: **Do Case-Control Studies Always Estimate Odds Ratios?**

Table of contents:

Search strategy

Search terms: ("International journal of epidemiology"[Journal] OR "European journal of epidemiology"[Journal] OR "American journal of epidemiology"[Journal] OR "Epidemiology (Cambridge, Mass.)"[Journal] OR "Journal of epidemiology and community health"[Journal] OR "Lancet (London, England)"[Journal] OR "Annals of internal medicine"[Journal] OR "BMJ (Clinical research ed.)"[Journal] OR "JAMA"[Journal] OR "The New England journal of medicine"[Journal]) AND "case-control studies"[MeSH] AND ("2014/07/30"[PDAT] : "2019/07/30"[PDAT])

We did not find any case-control studies in the *Journal of Clinical Epidemiology* in the past five years and therefore replaced that journal with the *European Journal of Epidemiology*. We chose the most cited articles in this time period to focus on the articles with the most impact. Only two articles from each journal were selected for a total of 20 instead of the 150 from the original Knol et al article because our goal was not to replicate their finding but to see whether this practice continues today.

Using the search terms above, a list of references is obtained from the PubMed webiste. This list can be saved as a text file which can then be used in the R code below to retrieve the number of citations for each article:

```
library(readr)
library(magrittr)
library(rentrez)

# Import txt saved from the pubmed search
tab <-  read_file("pubmed_result_5yrs.txt")

# Get PMIDs from file
PMIDs <- strsplit(tab,"PMID: ") %>% sapply(.,FUN = function(x) {
    x <- sub("\r\n","",x)
    substring(x,1,8) %>% return
})

# Take the references and break them into groups of maximum 200 so the
# entrez_summary does not time out
list_records <- split(PMIDs, ceiling(seq_along(PMIDs)/200))

# Get information on references and recombine into one data frame
refs <- lapply(list_records, function(x) {
  rec <- entrez_summary(db="pubmed", id=x)
  q <- extract_from_esummary(rec,
                             elements = c("uid","source","sorttitle",
                                          "pmcrefcount","sortfirstauthor"),
                             simplify = F)
  q <- lapply(q,function(x) {
    as.vector((unlist(x)))})
  q <- do.call(rbind,q)
})

refs <- do.call(rbind,refs) %>% as.data.frame
names(refs) <- c("pmid", "joural", "title","citations")

# Write retrieved data to csv
write.csv(refs, file = "review_5yrs.csv")
```

We then selected the top two cited articles that applied the case-control design and reported a measured of association.

Web Table 1: A table containing information on the 20 references we in our literature review including parameter reported, control selection and parameter based on control selection. References can be found at the end of the table.

Author	Year	Journal	Parameter reported	Parameter based on design	Assumption for inter-pretation	Underlying cohort type	Control selection	Notes
Steinmaus et al	2014	American Journal of Epidemiology	OR	OR	Exposure prevalence constant	Open	"Controls without lung, bladder, or kidney cancer were randomly selected from the Chilean Electoral Registry and were frequency-matched by sex and 5-year age group. This registry contained more than 95% of people older than 40 years of age who were recorded in the national census."	Had the authors not excluded people with cancers from the referent group they would have estimated an incidence rate ratio.
Risch et al	2015	American Journal of Epidemiology	OR	IRR	Exposure constant	Open	"Over the same time period, we used random digit dialing (preceded by letters sent to eligible households) to identify potential control subjects."	
Friis S et al	2015	Annals of Internal Medicine	OR	IRR	None required	Closed	"Using risk set sampling and applying the same selection criteria as for case patients, we matched 10 population control participants for each case according to birth year, sex, and area (county/region)."	"With the nested case-control design and risk set sampling of control participants, the ORs provide unbiased estimates of the corresponding incidence rate ratios in the underlying source population, without distortion by competing risks (41)." The authors recognize they are estiamting IRRs but report ORs nonetheless.
Wiese et al	2018	Annals of Internal Medicine	OR	IRR	None required	Closed	"Controls were matched to cases on the index date, as well as age (individual years) and county of residence on that date. A subject could serve as a control for multiple cases and could later become an IPD case."	"Importantly, the nested case-control odds ratio provides an unbiased estimator of incidence rate ratios with negligible or no loss of precision" The authors recognize they are estiamting IRRs but report ORs nonetheless.

Author	Year	Journal	Parameter reported	Parameter based on design	Assumption for inter-pretation	Underlying cohort type	Control selection	Notes
Park et al	2015	BMJ	Hazard ratio	HR	None required	Closed	"Our control group (n=420 386) comprised 5% annual random samples of the source population during the study period."	The authors multiplied by the sampling ratio to directly estimate the HR. However, given how they sampled their controls, they could have estimated the IRR without using the sampling fraction.
Billioti de Gage et al	2014	BMJ	OR	IRR	None required	Closed	"Each person with dementia (case) was matched on sex, age group (70-74, 75-79, 80-84, or $\geq$ 85), and duration of follow-up (6, 7, 8, 9, or 10 years) at the index date with four controls by using an incidence density sampling strategy."	
Kalkbrenner et al	2015	Epidemiology	OR	IRR/OR	None required	Open	"For this study, birth certificate rosters were sampled for a comparison group that represented the source population, as follows: In North Carolina, we selected a 15% random sample of births in the study counties and birth years without regard to autism status, stratified by birth year. After removing infant deaths, adoptions, and multiple births, the North Carolina dataset included 2,645 children (1994), 2,729 children (1996), 3,088 children (1998), and 4,806 children (2000). In California, we selected a 3% random sample of 1996 births in the study area, and, after removing multiple births, infant deaths, known autism cases, and 16 children without a street address (which included 14 adopted children), the California control sample included 2,311 children."	The autism cases were included with the controls in the NC data which means that IRRs were estimated. In the California data, autism cases were excluded with the controls meaning that they estimated an OR.

Author	Year	Journal	Parameter reported	Parameter based on design	Assumption for inter-pretation	Underlying cohort type	Control selection	Notes
Metayer et al	2014	Epidemiology	OR	IRR	Exposure constant	Open	See this publication for descripition of control selection: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3652629/	Many of the studies included used some type of matching. If the matching was done on time then no additional assumption is needed. If not matched on time then the assumption about constant exposure is needed.
Papantoniou et al	2016	European Journal of Epidemiology	OR	IRR	Exposure constant	Open	"Control subjects are women with no history of breast cancer living in the same catchment area as cases. Controls were selected randomly from the rosters of General Practitioners at the Primary Health Centers (PHC) participating in the study that cover nearly all the population living in the corresponding area."	
Brattström et al	2015	European Journal of Epidemiology	OR	OR	Exposure constant	Open	"Eligible controls were Swedish residents not found in the Trauma Registry. A random sample from the general population of 36,910 age, gender and municipality matched controls were extracted from the register of total population."	Education level would not have changed for most people in this sample therefore, if those who had experienced trauma had not been eliminated from the controls, this study would have estimated an IRR rather than an OR. The same applies to co-morbidity as an exposure. Co-morbidity was evaulated 8 years before the trauma event but cases could not serve as controls for cases that occur earlier meaning that this study estimated ORs but could have estimated IRRs by allowing cases to serve as controls for cases that occurred earlier.

Author	Year	Journal	Parameter reported	Parameter based on design	Assumption for inter-pretation	Underlying cohort type	Control selection	Notes
O'Neill et al	2015	International Journal of Epidemiology	OR	IRR	None required	Open	US:"Cases were identified from population-based cancer registries, and controls were randomly selected from birth records, with ratios ranging from 1:1 to 10:1."; UK: "One (or two since 2000) control birth records were also routinely selected, individually matched on sex, district and subdistrict of registration and being born within 6 months of the case. No controls were diagnosed with childhood cancer at time of selection."	
Toporcov et al	2015	International Journal of Epidemiology	OR	IRR	None required	Open	"Most of these were hospital-based casecontrol studies, and in the majority of these studies, the control subjects were matched to cases with regard to age, sex and additional characteristics (such as study centre, hospital and race/ethnicity)."	
Oksuzyan S et al	2015	Journal of Epidemiol Community Health	OR	IRR	Exposure constant	Open	"Controls were selected randomly from CBR and matched to cases (1 to 1) on date of birth (±6 months) and sex."	It was not necessary to link the birth and cancer registry. Random sampling from the birth registry would have provided the exposure distribution in the source population to estimate a risk ratio.
Dawson AL et al	2016	Journal of Epidemiol Community Health	OR	OR	Controls have no potential to become cases	Open	"Controls were live-born infants with no major birth defects born in the same catchment areas as cases, and were selected at random from birth hospital logs or vital records."	

Author	Year	Journal	Parameter reported	Parameter based on design	Assumption for inter-pretation	Underlying cohort type	Control selection	Notes
Lewis JD et al	2015	JAMA	OR	IRR	None required	Closed	"For each bladder cancer case, one cohort member who was alive, under follow-up, and without bladder cancer at the case patient's diagnosis was randomly selected as a control after matching on sex, age (±2.5 years), and time from entry into the diabetes registry to index date (±6 months)."	
Chung WH et al	2014	JAMA	OR	OR	None required	Open	"Drug-tolerant patients who had received phenytoin for more than 3 months without evidence of adverse reactionswere enrolled as controls from the departments of neurology orneurosurgery of theCGMHhealth systeminTaiwanin 2002-2014."	
Cao-Lormeau VM et al	2016	Lancet	OR	IRR	Exposure constant	Open	"a first control group (CTR1, n=98) was recruited among patients hospitalised or consulting for non-febrile illness at the CHPF" //"a second control group (CTR2, n=70) was recruited among age-matched (± 10 years) patients with RT-PCR-confirmed ZIKV infection, but who did not develop any neurological complication."	The exposure is not constant in the long term but likely is on short time scale on which Zika causes Guillain-Barré.

Author	Year	Journal	Parameter reported	Parameter based on design	Assumption for inter-pretation	Underlying cohort type	Control selection	Notes
O'Donnell MJ et al	2016	Lancet	OR	IRR	Exposure constant	Open	"Controls were either community-based or hospital-based. Hospital-based controls were patients admitted to hospital or those attending an outpatient clinic for disorders or procedures not related to stroke or transient ischaemic attack, or visitors or relatives of other inpatients.10 Specific approaches to identifying sources of community-based controls were not prespecifi ed, because standardised approaches might not be feasible in all settings. However, site guidance on preferred and acceptable sources for hospital-based controls were provided at each site"	
Jaiswal S et al	2017	New England Journal of Medicine	OR	IRR	Exposure constant	Open	"In these studies, case participants consisted of those with early-onset myocardial infarction who were selected at the time of the index presentation to hospitals, and controls were persons from the same medical centers who did not have cardiovascular disease."	
Steri M et al	2017	New England Journal of Medicine	OR	OR	None required	Open	"Coincident associations were assessed in case-control sets of 2934 patients with multiple sclerosis, 411 patients with SLE, and 3392 controls from across Sardinia, as well as in a population cohort (SardiNIA study) of 6921 volunteers from the Lanusei valley in Sardinia."	

References

1. Steinmaus C, Ferreccio C, Yuan Y, et al. Elevated lung cancer in younger adults and low concentrations of arsenic in water. *American Journal of Epidemiology.* 2014;180(11):1082–1087.

2. Risch HA, Yu H, Lu L, et al. Detectable Symptomatology Preceding the Diagnosis of Pancreatic Cancer and Absolute Risk of Pancreatic Cancer Diagnosis. *American Journal of Epidemiology.* 2015;182(1):26–34.

3. Friis S, Riis AH, Erichsen R, et al. Low-dose aspirin or nonsteroidal anti-inflammatory drug use and colorectal cancer risk: A population-based, case-control study. *Annals of Internal Medicine.* 2015;163(5):347–355.

4. Wiese AD, Griffin MR, Schaffner W, et al. Opioid Analgesic Use and Risk for Invasive Pneumococcal Diseases. *Annals of Internal Medicine.* 2018;168(6):396.

5. Park TW, Saitz R, Ganoczy D, et al. Benzodiazepine prescribing patterns and deaths from drug overdose among US veterans receiving opioid analgesics: Case-cohort study. *BMJ.* 2015;350:1–8.

6. De Gage SB, Moride Y, Ducruet T, et al. Benzodiazepine use and risk of Alzheimer's disease: Case-control study. *BMJ.* 2014;349:1–10.

7. Kalkbrenner AE, Windham GC, Serre ML, et al. Particulate matter exposure, prenatal and postnatal windows of susceptibility, and autism spectrum disorders. *Epidemiology.* 2015;26(1):30–42.

8. Metayer C, Milne E, Dockerty JD, et al. Maternal Supplementation with Folic Acid and Other Vitamins and Risk of Leukemia in Offspring: A Childhood Leukemia International Consortium Study. *Epidemiology.* 2014;25(6):811–822.

9. Papantoniou K, Castaño-Vinyals G, Espinosa A, et al. Breast cancer risk and night shift work in a case-control study in a Spanish population. *European Journal of Epidemiology.* 2016;31(9):867–878.

10. Brattström O, Eriksson M, Larsson E, et al. Socio-economic status and co-morbidity as risk factors for trauma. *European Journal of Epidemiology.* 2015;30(2):151–157.

11. O'Neill KA, Murphy MFG, Bunch KJ, et al. Infant birthweight and risk of childhood cancer: International population-based case control studies of 40 000 cases. *International Journal of Epidemiology.* 2015;44(1):153–168.

12. Toporcov TN, Znaor A, Zhang ZF, et al. Risk factors for head and neck cancer in young adults: A pooled analysis in the INHANCE consortium. *International Journal of Epidemiology.* 2015;44(1):169–185.

13. Oksuzyan S, Crespi CM, Cockburn M, et al. Race/ethnicity and the risk of childhood leukaemia: A case-control study in California. *Journal of Epidemiology and Community Health.* 2015;69(8):795–802.

14. Dawson AL, Tinker SC, Jamieson DJ, et al. Twinning and major birth defects, National Birth Defects Prevention Study, 1997-2007. *Journal of Epidemiology and Community Health.* 2016;70(11):1114–1121.

15. Lewis JD, Habel LA, Quesenberry CP, et al. Pioglitazone use and risk of bladder cancer and other common cancers in persons with diabetes. *JAMA - Journal of the American Medical Association.* 2015;314(3):265–277.

16. Chung WH, Chang WC, Lee YS, et al. Genetic variants associated with phenytoin-related severe cutaneous adverse reactions. *JAMA - Journal of the American Medical Association.* 2014;312(5):525–535.

17. Cao-Lormeau V-M, Blake A, Mons S, et al. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: A case-control study. *The Lancet.* 2016;387(10027):1531–1539.

18. O'Donnell MJ, Chin SL, Rangarajan S, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): A case-control study. *The Lancet.* 2016;388(10046):761–775.

19. Jaiswal S, Natarajan P, Silver AJ, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *New England Journal of Medicine.* 2017;377(2):111–121.

20. Steri M, Orrù V, Idda ML, et al. Overexpression of the Cytokine BAFF and Autoimmunity Risk. *New England Journal of Medicine*. 2017;376(17):1615–1626.