Acute infection and ischemic stroke:
A tale of unrecognized biases

Let me try to help you decide whether you want to keep reading beyond the second paragraph.

To follow this commentary you need to know enough about causal diagrams, which means knowing the three possible sources of an association between two variables (causal paths, confounding paths, and open induced paths). If you don't know what I am talking about, first read the principles of causal diagrams in an article of mine[1] (pages 58-59), or elsewhere. You also need to know when and where we draw a dashed line in a causal diagram. If you don't, read another commentary here titled "colliding bias (part 1): misnomers and the missing dashed line". You might also consider another option: keep reading – even if only to realize what an eye opener you have been missing in research methods.

*

It is highly unlikely that acute infection, with sepsis among its scary sequelae, has a *precisely* null effect on ischemic stroke – the null hypothesis which many set up to reject. By precisely null, I mean a probability ratio of 1 for the contrast between exposed (acute infection) and unexposed. Not 1.01, not 1.001, and not even 1.000000000000000000001. Exactly 1. Sound unreasonable? It is. Infection can cause septic shock, and shock can cause ischemic stroke because brain tissue might die when its blood supply is compromised. The cascade "infection→sepsis→shock →stroke" alone tells us that the probability ratio should not be exactly 1 (setting aside miraculous nullification by another path through which acute infection *prevents* stroke.) Next, ask yourself whether you truly care about rejecting any null effect hypothesis – if the true effect size happens to be small enough, say, a probability ratio of 1.01. (I mean the *true* effect size, not your *estimate* of the truth.)

So much for the null testing ritual, and for those who think that research should inform us whether acute infection is a cause of ischemic stroke. Yes, it is. (And no, a small *p*-value does not lend support to the estimate. The phrase "statistically significant estimate" is misinterpreted, widely.[2])

The task at hand is, therefore, estimation. What is the magnitude of the effect of acute infection on ischemic stroke? Is the effect size near 1 or near 2? Is it about 5 or about 10? How much does that effect vary as a function of the time difference between the two variables? That's what various research designs should estimate for us, for a start.

And they tell us different stories, many of which are not explained by differences in the "hazard period", the period during which the effect is assumed meaningful. Table 1 gives you an idea of what was reported by studies in the past 20 years.[3-11] At least two studies differ from others on the order of twofold or more: odds ratios of 6 to 8 versus odds ratios of 2 to 3. One study shows a twofold internal disagreement between two research designs (case-crossover vs. cohort).[a] Another reported odds ratios of 1.5 or smaller. Quite a puzzle.

Table 1. Estimated effects (mostly odds ratios) of acute infection on ischemic stroke.

First author	Hazard period	Estimate	Exposure	Design
Grau, 1995	7 days 14 days	~5 ~3	Infection	Case-control
Bova, 1996	1 week	~3	Infection	Case-control
Grau, 1998	1 week	~3	Infection	Case-control
Paganini-Hill, 2003	1 week	~1 ~1.5	Infection	Case-control Case-crossover
Smeeth, 2004	14 days	2-3*	Infection	Case-series (case-crossover)
Clayton, 2008	7 days 8-28 days	~2 ~2	Respiratory infection	Case-control
Elkind, 2011	14 days 30 days 14 days 30 days	~8 ~7 ~4** ~2.5**	Hospitalized infection	Case-crossover Cohort
Levine, 2014	14 days	~2	Infection	Case-crossover
Cowan, 2016	14 days 30 days	~8 ~6	Hospitalized infection	Case-crossover

*rate ratio
**hazard ratio

Any critical mind should look for explanations. And there are many, including the play of chance which does not spare most estimates. I am looking for biased estimators, however, and I will do so by displaying causal diagrams for different study designs. Every type of bias can be revealed and explained by a causal diagram – so long as the diagram correctly displays crucial variables and crucial causal

---

[a] When the disease is rare (stroke) the hazard rate ratio, the rate ratio, and the odds ratio should all be similar, unless censoring was substantial.

connections. A causal diagram does not have to be complete. None is complete.

**Here is the notation** I will be using**:**

E: exposure status (truth)
E_{dx}: a diagnosis of infection
E^*: the study version (measurement) of E
D: ischemic stroke (truth)
D^* the study version (measurement) of D
H_1: hospitalization status at the hazard period
H_2: hospitalization status shortly after D
C: any confounder (a shared cause of E and D).
Z: any cause of hospitalization
R: any shared cause of exposure and hospitalization
S: selection status for a case-control study

A cohort study: acute infection and stroke

Figure 1 shows several elements of a cohort study, some of which are shared by other designs. The effect of interest is identified by a question mark above the arrow $E{\rightarrow}D$, and the "hazard period" between E and D is shown below the arrow (usually counted in days in the case of infection and stroke). In a classic measurement, the true variable is a cause of the measured variable ($E{\rightarrow}E^*$, $D{\rightarrow}D^*$).

Figure 1.
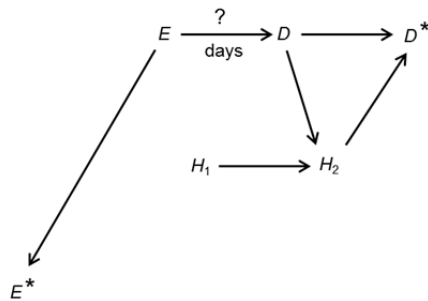
If you wonder about the path $H_1{\rightarrow}H_2$, here are two arguments: First, in accord with an axiom of causality, a variable at one time is a cause of that variable at any later time.[12] Second, it is easy to propose causal paths by which hospitalization (or not) at one time affects the decision to admit a patient later, especially when the interval is short (and here it is short). In fact, the shorter the interval between a variable at two time points, such as H_1 and H_2, the stronger the effect of the former on the latter.

Figure 1 is a good illustration of basic Information bias. The bias exists whenever the association between E^* and D^* (the association we estimate) differs from the association between E and D (the

association we wish to estimate). In the simplest case, all sources of information bias are attributed to a single path between E^* and D^*: $E^*{\leftarrow}E{\rightarrow}D{\rightarrow}D^*$.

This is not the case here, however, nor in many other examples. Stroke is a severe disease for which people are usually hospitalized ($D{\rightarrow}H_2$), and there is little doubt that location (hospital, elsewhere) affects the extent and type of diagnostic modalities. Therefore, hospitalization status affects the diagnosis of ischemic stroke ($H_2{\rightarrow}D^*$). As a result, we observe a second path of information bias: $E^*{\leftarrow}E{\rightarrow}D{\rightarrow}H_2{\rightarrow}D^*$ (Figure 1).

Note that ascertaining D^* from hospital records does not block the path $H_2{\rightarrow}D^*$, which reflects the reality of medical care. The path as depicted can be blocked only by conditioning on H_2 – for example, by restricting the *entire* sample to hospitalized patients after the hazard period. (The consequences will be examined in another design.) Alternatively, albeit unrealistically, the effect $H_2{\rightarrow}D^*$ may be null if researchers set up a measurement protocol that ignores information from external medical sources.[b]

As far as information bias is concerned, Figure 1 is good enough for a cohort study of many exposures, but not for an exposure that may lead to hospitalization, such as acute infection. A key arrow should be added, $E{\rightarrow}H_1$ (Figure 2).

Figure 2.
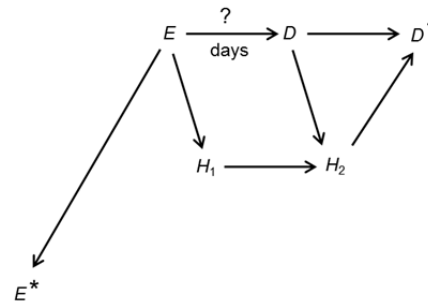
This new arrow is a troublemaker. We now observe a *third* path of information bias. Here are all three:

$$E^*{\leftarrow}E{\rightarrow}D{\rightarrow}D^*$$
$$E^*{\leftarrow}E{\rightarrow}D{\rightarrow}H_2{\rightarrow}D^*$$
$$E^*{\leftarrow}E{\rightarrow}H_1{\rightarrow}H_2{\rightarrow}D^*$$

---

[b] We rely on effect modification between D, H_2, and a research protocol for measuring D^*. If interested, read a commentary here titled "On effect modification and its applications".

That several, open paths connect two variables does not necessarily imply a stronger association. (Negative confounding is a good example). But open paths can add up, and they do add up here. Each of the extra paths of information bias is expected to generate a positive association between E^* and D^*. The bias is compounded.

We might consider blocking the third path by conditioning on H_1; for example, restricting the sample of exposed and unexposed to people who were (or were *not*) hospitalized during the hazard period. That cohort design will indeed block paths through H_1, but will create many more bypassing paths through any other cause (Z) of H-variables and through shared causes (R) of H_2 and E (Figure 3).

Figure 3.
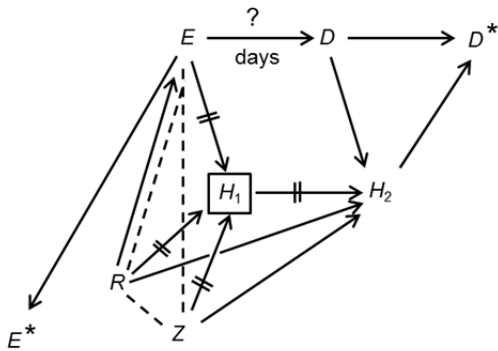
Which brings up a generic lesson: Conditioning on an effect of the exposure opens the door to new problems.

A cohort study: "hospitalized infection" and stroke

Figure 4 depicts the causal structure of a cohort study in which the exposure is "hospitalized infection", namely, infection followed by hospitalization. I added an interim E_{dx} (say, a physician's diagnosis of infection) between E and E^* and drew arrows from E_{dx} and H_1 to E^*. To simplify, I depicted Z and omitted R.

In this design E^* is derived from two variables, $E^* = f(E_{dx}, H_1)$, as follows:

If $E_{dx}=1$ and H_1=hospitalized, then $E^*=1$. Otherwise, $E^*=0$.

Figure 4.
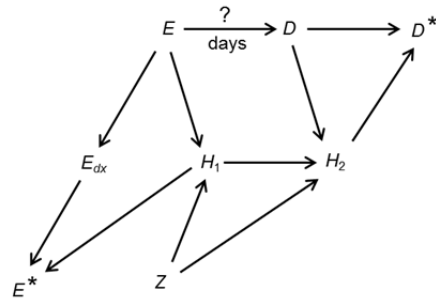
Although not previously recognized, this method of measuring E is subject to two (alternative) major flaws. You may choose between the flaw of thought bias and the flaw of information bias, as explained next.

The variable "hospitalized infection status" itself is not a natural variable in causal reality. It is a derived variable, and like any derived variable (the output of a function), it affects nothing. To assume that "hospitalized infection" is a cause of ischemic stroke is one of many examples of thought bias,[13,14] the bias that arises by attributing causal properties to made-up variables.[c]

The usual escape route from thought bias is well known: You do not claim that "hospitalized infection" causes stroke. Rather, you claim that the variable serves to impute the value "severe infection" of E. You argue that you are estimating the effect of severe infection, acknowledging that the unexposed group contains milder forms of infection. It is not the case of thought bias; it is negligible information bias (which you would probably call "measurement error").

Unfortunately, the argument fails. Not every imputation of the values of one variable from other variables is permissible in science.[13] In particular, E should not be imputed by a process that is known to add bias *by itself*, a process with built-in extra bias above and beyond difficult-to-avoid measurement error. And not a negligible amount of bias – a lot of bias. The derivation of E^* from E_{dx} and H_1 creates four new paths of information bias through H_1 (Figure 4, above):

$$E^* \leftarrow H_1 \rightarrow H_2 \rightarrow D^*$$
$$E^* \leftarrow H_1 \leftarrow Z \rightarrow H_2 \rightarrow D^*$$
$$E^* \leftarrow H_1 \leftarrow E \rightarrow D \rightarrow D^*$$
$$E^* \leftarrow H_1 \leftarrow E \rightarrow D \rightarrow H2 \rightarrow D^*$$

---

[c] The disturbing idea was proposed in 2010 and has been criticized, or endorsed, precisely zero times since then. Dormant scholars? Cognitive dissonance? Silent agreement? Make a guess.

The first path delivers a strong association between E^* and D^* because every segment is a strong effect. The second path is not a single path: Every cause (Z) of hospitalization status creates such a path between E^* and D^*, so replicate this path an unknown number of times. The strength of the last two paths largely depends on the strength of $E \rightarrow D$ (possibly weak.) Collectively, however, these paths are a rich source of information bias, which can easily explain a stronger association of ischemic stroke with "hospitalized infection" than with "infection".

Look back at Table 1 and find the single cohort design. The exposure was "hospitalized infection" and the rate ratio in a hazard period of 14 days was about 4. How much information bias was built into the estimator (Figure 4)?

A case-control study: acute infection and stroke

A case-control study calls for preferential selection of people with the disease. The sample contains a much higher ratio of diseased people (cases) to disease-free (controls) than that ratio in the source population. This feature of the case-control design is captured by drawing an arrow between D^* and S (selection status). And since only the selected people are studied, conditioning on S is inherent in any research design (Figure 5).

If D^* were the only cause of S, no problem would arise; the odds ratio is immune to uni-path colliding bias.[15] But that's not always the case. Location of the person is often another cause of S, thereby turning S into a collider. For example, in a case-control study of hospitalized cases and *non-hospitalized* controls, hospitalization status affects S ($H_2 \rightarrow S$). Furthermore, the effect is modified by D^*: When $D^*=1$ (stroke), the probability of being selected is much higher if the person is hospitalized than not hospitalized. And when $D^*=0$ (stroke free), the probability of being selected is much higher if the person is *not* hospitalized than hospitalized. Effect modification between D^* and H_2 on S=selected implies a new associational component (a dashed line) besides the arrow $H_2 \rightarrow D^*$ (Figure 5).[16]
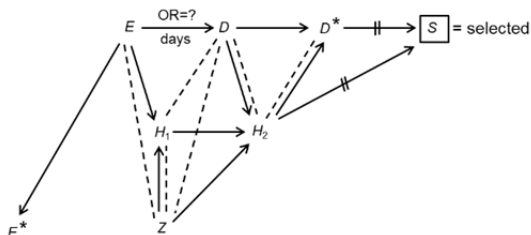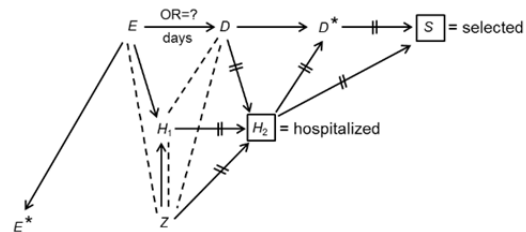
Figure 5.

Figure 5 is loaded with other dashed lines. They all arise because conditioning on an effect of a collider is equivalent to conditioning on a collider. For example, E and Z are associated after conditioning on S, because E and Z collide at H_1, and S is an effect of H_1 ($H_1 \rightarrow H_2 \rightarrow S$). I did not depict R, any shared cause of infection and hospitalization. You can imagine how much clutter we would have seen in Figure 5 if R was added (more dashed lines and more paths of bias).

Some induced paths in Figure 5 are sources of information bias. They are paths between E^* and D^* that contribute to the difference between the estimated association (E^* with D^*) and the association of interest (E with D). Other induced paths account for colliding bias. They travel between E and D themselves: $E \rightarrow H_1 -- D$ and $E -- Z -- D$, for instance. They contribute to the association between E and D even if information bias is absent; even if E^* and D^* are exact copies of E and D, respectively.

Is there a remedy? Is there a way to eliminate all these induced paths? Let's try to condition on H_2 (Figure 6) – that is, to sample hospitalized cases and hospitalized controls.
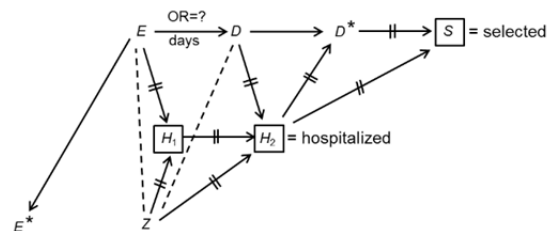
Figure 6.

Paths of information bias have been eliminated, but paths of colliding bias remain, including paths through R, shared causes of infection and hospitalization (not shown).

Can the problem be solved by conditioning on H_1 as well (Figure 7)?

Figure 7.

Commentary

No, it is impossible to eliminate all paths of colliding bias through *Z* (Figure 7) and through *R* (not shown). To block these paths, we would need to condition on both types of variables (which are not confounders). At any rate, none of the studies in Table 1 correspond to Figure 7. It is difficult to carry out a study in which cases and controls must have been in the hospital during the entire (preceding) hazard period. All cases must have suffered stroke while already hospitalized.

How large is the bias?

It is difficult to quantify, but it is easy to argue in qualitative terms. The magnitude of the bias depends on the strength of open induced paths, which in turn depends, in part, on the strength of dashed line segments. The magnitude of the association behind a dashed line depends, in part, on the magnitude of effect modification between the colliding variables:[d] the weaker the modification on the probability ratio scale, the "weaker" the dashed line and the smaller the bias.

Consider Figure 6, for example. The strength of *D*--*Z* depends on the magnitude of effect modification between *D* and *Z* with respect to H_2=hospitalized. We may argue that effect modification is weak in this case because stroke is a severe disease, and the decision to hospitalize for a severe disease should not differ much according to the value of any other cause of hospitalization. The same argument might hold for the dashed line *D*--H_1. Nonetheless, that's not necessarily the case for *Z*--H_1 and *E*--*Z*. Effect modification may be strong so those dashed lines might indicate strong associations (and large bias). Recall that *Z* (and *R*) could stand for many variables.

A matched case-control study: acute infection and stroke

Figure 8 shows the causal diagram for a classic matched case-control study (matching on *C*).

Figure 8.
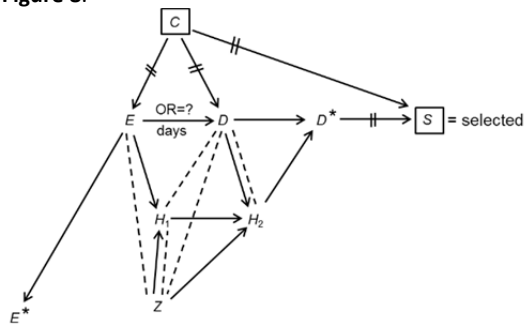
---
d A partially proven conjecture

The matched confounder (*C*) affects selection status (*S*), and that's the only difference from a diagram for an unmatched case-control study in which we condition on *C*. In both designs, we must condition on the confounder because matching controls to cases does not remove confounding bias by the matching variable(s).[17]

As before, I depicted *Z* to show how causes of hospitalization add more paths of colliding bias. I did not depict *R*, any shared cause of infection and hospitalization. Again, you can imagine how much clutter we would have seen in Figure 8 if *R* was added (more dashed lines and more paths of bias). Some of the case-control studies in Table 1 match Figure 8, reporting odds ratios in the region of 2 to 3. How much bias was built into those estimators?

I hope the lesson is clear. When the exposure affects hospitalization, and hospitalization affects the diagnosis, and the diagnosis affects selection – many new roads are paved: roads of bias.

A case-crossover study: snow shoveling and stroke

Suppose we are interested in estimating the effect of shoveling snow (not infection) on ischemic stroke using a 1:1 matched case-control study. We recruit James who has had a stroke and find Michael, a matched control, say, matched on weight (and on sex, of course). Next, we ask James and Michael whether they shoveled snow any time during a hazard period of 12 hours (Figure 9). Lastly, we use the set of such case-control pairs to compute the odds ratio.

Figure 9.
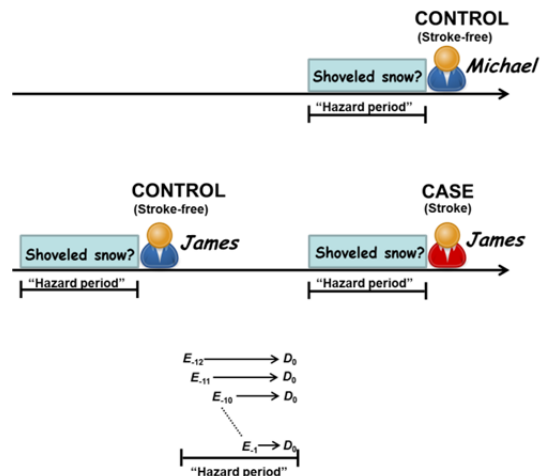
Alternatively, we may use a case-crossover design.[18]

Like many clever ideas, the underlying idea is remarkably simple at its core. Instead of matching

5

James with Michael – a stranger – we match James with James himself at some earlier time (say, a week earlier) when James was stroke-free (Figure 9). Self-matching implies matching not only on sex and weight (if it hasn't changed), but also on an unknown number of variables whose values have not changed between the early and late hazard periods. It does not imply matching on variables that have changed. Perhaps an irritating email arrived an hour before James suffered a stroke, but not a week earlier.[e]

Notice that the "hazard period" in either stranger-matching or self-matching allows for different intervals between E and D (bottom of Figure 9). One person might have shoveled snow at the beginning of the hazard period, and another – an hour before the end. The two effects may be different, and therefore, the odds ratio is estimating some average of different effect sizes. Keep this point in mind when you compare an odds ratio for a 28-day hazard period with an odds ratio for a 14-day hazard period. You are comparing one average with a *nested* average.

Figure 10 shows the causal diagram for a case-crossover design of snow shoveling and stroke. Three points should be noted:

1) There is no conditioning on H_2. James was not required to be hospitalized at his control time (when he was stroke free).

2) Z-type variables and R-type variables are also matched on (and conditioned on), so long as the value did not change.

3) The bias is attributed only to paths of information bias (Figure 10).

Figure 10.

---
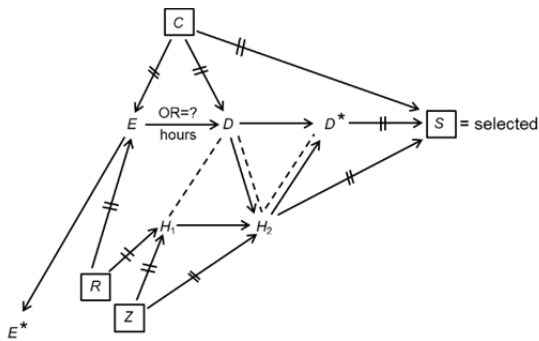
[e] Such causes (snow shoveling, an acute infection, an email message) are often called "triggers" – a superfluous term. A "trigger" is nothing more than a causal variable in some arbitrary proximity to the effect. Therefore, "trigger" vis-à-vis "cause" is a simple-minded distinction. There are no two kinds of causes.

Overall, a reasonable design, with difficult-to-avoid information bias. (Condition on H_2: require James to have been hospitalized right after *both* hazard periods).

A case-crossover study: acute infection and stroke

Let E be acute infection rather than snow shoveling, which means adding an arrow between E and H_1. The picture gets much worse (Figure 11). More paths of information bias and two paths of colliding bias ($E{\rightarrow}H_1{-}{-}D$; $E{\rightarrow}H_1{\rightarrow}H_2{-}{-}D$).

Figure 11.
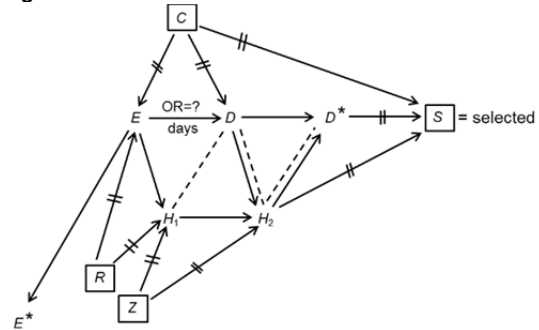
Two studies in Table 1 match Figure 11. They reported odds ratios on the order of 2 to 3.

A case-crossover study: "hospitalized infection" and stroke

Change "acute infection" to "hospitalized infection", and the bias machine is turned on even more (Figure 12): two paths of colliding bias and more than a dozen paths of information bias. (To simplify the diagram, I omitted Z and R which are conditioned on, if the value hasn't changed.)

Figure 12.
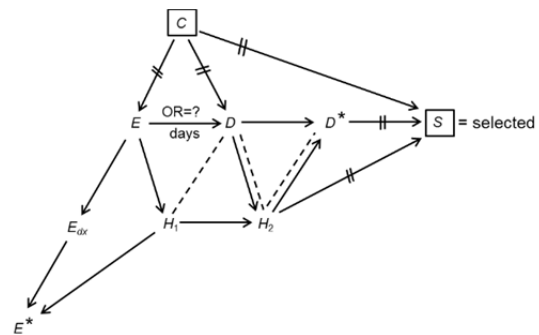
Here is the list of paths of bias in a case-crossover study of "hospitalized infection" and ischemic stroke:

Commentary

Colliding bias

$$E \rightarrow H_1 \text{--} D$$
$$E \rightarrow H_1 \rightarrow H_2 \text{--} D$$

Information bias

$$E^* \leftarrow E_{dx} \leftarrow E \rightarrow D \rightarrow D^*$$
$$E^* \leftarrow E_{dx} \leftarrow E \rightarrow D \rightarrow H_2 \rightarrow D^*$$
$$E^* \leftarrow E_{dx} \leftarrow E \rightarrow D \rightarrow H_2 \text{--} D^*$$
$$E^* \leftarrow E_{dx} \leftarrow E \rightarrow D \text{--} H_2 \rightarrow D^*$$
$$E^* \leftarrow E_{dx} \leftarrow E \rightarrow D \text{--} H_2 \text{--} D^*$$
$$E^* \leftarrow E_{dx} \leftarrow E \rightarrow H_1 \rightarrow H_2 \rightarrow D^*$$
$$E^* \leftarrow E_{dx} \leftarrow E \rightarrow H_1 \rightarrow H_2 \text{--} D^*$$
$$E^* \leftarrow H_1 \leftarrow E \rightarrow D \rightarrow D^*$$
$$E^* \leftarrow H_1 \leftarrow E \rightarrow D \rightarrow H_2 \rightarrow D^*$$
$$E^* \leftarrow H_1 \leftarrow E \rightarrow D \text{--} H_2 \rightarrow D^*$$
$$E^* \leftarrow H_1 \leftarrow E \rightarrow D \rightarrow H_2 \text{--} D^*$$
$$E^* \leftarrow H_1 \leftarrow E \rightarrow D \text{--} H_2 \text{--} D^*$$
$$E^* \leftarrow H_1 \rightarrow H_2 \rightarrow D^*$$
$$E^* \leftarrow H_1 \rightarrow H_2 \text{--} D^*$$

Look up two case-crossover studies of "hospitalized infection" in Table 1. They reported odds ratios on the order of 6 to 8. How much bias was built into the estimators? Probably a lot.[f]

If you don't have sufficient knowledge of causal diagrams, just consider the following paradox. Compared with other designs, a case-crossover study will control a larger number of "time-stable confounders", and such confounders – in the case of infection and stroke – are expected to create *positive* confounding. We expect shared causes of infection and stroke to be positively associated with both. In fact, this point was made by the authors of one study of hospitalized infection. Praising their case-crossover design over classic case-control studies, they wrote:

"Such studies are limited, however, by potential confounding due to underlying risk factors, such as smoking, that could lead to both infection and stroke".[9]

If so, case-crossover studies should have removed positive confounding, generating *smaller* estimates than estimates reported from other, confounding-prone, designs. Why, then, do we observe exactly the opposite? Why do two case-crossover studies of hospitalized infection report estimates that are twice as large as those reported in other designs? And why did the study quoted above find stronger effects of

"hospitalized infection" in a case-crossover analysis than in a cohort analysis?[9] Strong negative confounding (by what?) that overrode positive confounding is one explanation. Ample paths of bias is another (Figure 12). And no, it is not simply a stronger effect of "severe infection". That claim was already addressed and dismissed (Page 3). Nor does it explain, of course, *larger* estimates from a case-crossover analysis than a cohort analysis of "hospitalized infection" – in the very same study.[9]

One lesson is worth repeating, though: No one may claim to have a valid imputation of the value "severe infection" from H_1 when the imputation itself creates at least half a dozen paths of bias through H_1 (Figure 12). It is analogous to claiming that the defendant is guilty after planting incriminating evidence. The latter is bad practice of law; the former – bad practice of science.

Epilogue

So how big is the effect of acute infection on ischemic stroke? What is the value of the causal parameter, setting modifiers aside?

Considering sources of bias, as exposed here, I am willing to place a bet on a range: an odds ratio (or equivalent) of 1.1 (min) to 3 (max) in a 14-day hazard period.

A tighter bet?

Somewhere between 1.1 and 1.9 for any infection which is not severe; between 2 and 3 for a severe infection; and nowhere near 6, or 7, or 8 – for either.

---

[f] I was asked to co-author the work by Cowan et al. and declined authorship after pointing out bias. Perhaps it was a small conflict of interests: I gave up the chance of adding another reader-reviewed article to my CV.

References

1. Shahar E, Shahar DJ. Causal diagrams and the cross-sectional study. *Clinical Epidemiology* 2013;5:57-65
2. A significant dialogue (on my home page): http://www.u.arizona.edu/~shahar/
3. Grau AJ, Buggle F, Heindl S, Steichen-Wiehn C, Banerjee T, Maiwald M, Rohlfs M, Suhr H, Fiehn W, Becher H, et al. Recent infection as a risk factor for cerebrovascular ischemia. *Stroke* 1995 Mar;26(3):373-9
4. Bova IY, Bornstein NM, Korczyn AD. Acute infection as a risk factor for ischemic stroke. *Stroke* 1996 Dec;27(12):2204-6

5. Grau AJ, Buggle F, Becher H, Zimmermann E, Spiel M, Fent T, Maiwald M, Werle E, Zorn M, Hengel H, Hacke W. Recent bacterial and viral infection is a risk factor for cerebrovascular ischemia: clinical and biochemical studies. *Neurology* 1998 Jan; 50(1) 196-203

6. Paganini-Hill A, Lozano E, Fischberg G, Perez Barreto M, Rajamani K, Ameriso SF, Heseltine PN, Fisher M. Infection and risk of ischemic stroke: differences among stroke subtypes. *Stroke* 2003 Feb;34(2):452-7.

7. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. *New England Journal of Med*. 2004 Dec 16;351(25):2611-8.

8. Clayton TC, Thompson M, Meade TW. Recent respiratory infection and risk of cardiovascular disease: case-control study through a general practice database. *European Heart Journal* 2008; 29:96-103

9. Elkind MS, Carty CL, O'Meara ES, Lumley T, Lefkowitz D, Kronmal RA, Longstreth WT Jr. Hospitalization for infection and risk of acute ischemic stroke: the Cardiovascular Health Study. *Stroke* 2011 Jul;42(7):1851-6

10. Levine DA, Langa KM, Rogers MAM. Acute infection contributes to racial disparities in stroke mortality. Neurology 2014;82:914-921

11. Cowan LT, Alonso A, Pankow JS, Folsom AR, Rosamond WD, Gottesman RF, Lakshminarayan K. Hospitalized infection as a trigger for acute ischemic stroke: The Atherosclerosis Risk in Communities Study. *Stroke* 2016 Jun;47:1612-7

12. Shahar E, Shahar DJ: Marginal structural models: much ado about (almost) nothing. *Journal of Evaluation in Clinical Practice* 2013 Feb;19(1):214-22

13. Shahar E, Shahar DJ: Causal diagrams, information bias, and thought bias. *Pragmatic and Observational Research* 2010;1:33–47

14. Shahar E, Shahar DJ: Causal diagrams and three pairs of biases. In: Epidemiology – Current Perspectives on Research and Practice (Lunet N, Editor). www.intechopen.com/books/epidemiology-current-perspectives-on-research-and-practice, 2012:pp. 31-62

15. The case-control study (a commentary on this website) http://www.u.arizona.edu/~shahar/commentaries.html

16. Shahar DJ, Shahar E. A theorem at the core of colliding bias. Submitted.

17. Shahar E, Shahar DJ. Causal diagrams and the logic of matched case-control studies. *Clinical Epidemiology* 2012;4:137-144 [Erratum: *Clinical Epidemiology* 2014;6:59]

18. Maclure M, Mittleman MA. Should we use a case-crossover design? *Annu Rev Public Health* 2000;21:193-221