

A Young Person's Guide to Empirical Legal Research. With Illustrations from the Field of Medical Malpractice

Ben C.J. van Velthoven*

1. Introduction

The analysis of law and legal systems over the past decades has been characterized by a steadily growing interest in an external perspective as opposed to the traditional internal perspective of the legal profession. The movement has been set in motion by social scientists and economists who wanted to know more about the 'law in action', leaving the 'law in the books' to legal scholars.¹ A central element of this line of research is empirical work. Empirical legal research can first of all be helpful by simply observing how and how often specific legal rules are actually applied in real-world situations. These observations can also be used to reflect on whether and how law and society are related in a causal way. This can, for instance, entail a fresh and free exploration of the data to develop some hypotheses, a first piece of theory, as to how a legal rule affects the behaviour of the citizens and organizations that are subject to the rule. But once a more or less full-fledged theory is available, observations can also be used for a serious statistical test of the theory's validity.

In the meantime, the interest in empirical legal research has also spread widely among legal scholars.² But they are generally not very well educated in the dos and don'ts of that line of research. Now, of course, there are many textbooks, both introductory and advanced, on social science research methods, statistics and econometrics.³ The setting of the presentation in these textbooks, the issues that are being addressed and the level of the treatment do not nicely fit in, however, with the needs of the JD student or any other legal academic who is taking the first steps in this unfamiliar field.

This article attempts to bridge the gap between the expertise and skills of the legal novice and the more specialized literature on empirical research methods.

* Associate professor of Law and Economics at Leiden Law School. I wish to thank Nienke van der Linden, Ali Mohammad and Charlotte Vrendenbargh from Leiden Law School and two anonymous reviewers and the editors of this journal for helpful comments on earlier drafts.

1 Well-known publication outlets are the *Journal of Law and Economics*, founded in 1958, the *Law and Society Review*, launched in 1966, the *Journal of Legal Studies*, since 1972, and the *Journal of Empirical Legal Studies*, since 2004.

2 See, e.g., Siems 2014, part II.

3 See, e.g., Moore et al. 2015, Wooldridge 2013 and Field 2013, to mention just three introductory texts.

It lays out the general principles of empirical legal research and discusses a series of stumbling blocks on the way forward.⁴ As such, it should be helpful to the legal academic both when starting his own empirical research project and in judging the quality of the work of other scholars. Unfortunately, a presentation of methodological principles can easily degenerate into a dry catalogue of definitions. To try and get the reader involved in the consecutive steps of the discussion, medical malpractice law will be used as a central source of illustration. That field of law has already been extensively studied in the empirical literature.

The article proceeds as follows. Section 2 first sketches the main features of medical malpractice law. Section 3 introduces the main research themes within empirical legal research. Section 4 discusses the methodology, and the stumbling blocks, of descriptive research. Section 5 does the same for explanatory research. Section 6 concludes with some final comments.

2. Medical malpractice law⁵

Accident losses occur in many ways. People get hurt, for instance, by adverse events in the course of medical treatment that is supposed to cure them. In many instances, such an adverse event is just a medical complication that sometimes, alas but unavoidably, arises in the course of a treatment that is otherwise performed with all due professional care. Some injuries, however, result from medical error. When the physician fails to meet the profession's customary standard of adequate care, the patient can sue him for negligence. To obtain compensation under the law of medical malpractice, a particular application of tort law, the plaintiff must show that a duty of care existed; that the defendant failed to conform to the required standard of care, either by his acts or by failure to act; that the plaintiff suffered harm; and that the breach of duty was the proximate cause of the harm.

The tort system performs several important functions in society. First, it provides a forum for victims to be heard and to oblige injurers to make up for unduly risky behaviour (corrective justice). In close connection, it provides compensation for the harm as a result of the negligence of others, thereby acting as a source of insurance (distributive justice). Thirdly, by imposing sanctions on negligent behaviour, it provides incentives for potential injurers to take appropriate care and reduce the number of injuries (prevention or deterrence).

As such, on paper, medical malpractice law may be well devised. Since the 1970s, however, it has been the source of heated public debate in the US, initiated by three medical malpractice crisis periods characterized by significant increases in the premiums and contractions in the supply of medical liability insurance.

4 My exposition is very indebted to Epstein & King 2002. However, while their argument is mostly directed at law professors and takes over 130 pages, mine is less than 20 pages and, I hope, expressly attuned to JD students and other beginning legal researchers.

5 For more details, see Van Velthoven & Van Wijck 2012 and the references given there.

In response to these crises, US states have over the years enacted a variety of reforms in their tort systems. Apparently, the law in action does not always function the way it is designed in the books. Thus, it is very useful to start a separate, empirical line of legal study.

3. The research question

Any serious empirical legal study starts from a research question. In essence, the research questions with regard to the law and the legal system can be classified into three main themes, which can be summarized in three simple statements:

- How does the law in action operate?
- What does law bring about?
- How does law come about?

3.1. *How does the law in action operate?*

The first main theme is *descriptive* in nature. Under this heading the researcher may want to know whether parties in a dispute actually invoke the legal rules that have been made available in the books. A further question may be to what degree the parties actually succeed in obtaining the outcomes they are legally entitled to. In the context of medical malpractice law, for instance, it is important to find out whether the tort system is adequately performing its functions of correction and compensation. Thus, the researcher may set out to assess how many victims of medical negligence in fact get full compensation for harm.⁶

3.2. *What does law bring about?*

The second theme groups research into the *effects* of legal rules on private individuals, on organizations and on society as a whole. Starting from the general idea that law provides incentives to citizens and organizations to refrain from certain activities and to initiate others, the researcher sets out to find the causal relationships between legal rules, human behaviour and social consequences. Hence, this line of research has an *explanatory* character. As to medical malpractice law, damage claims upon negligent behaviour might stimulate health care professionals to act in accordance with the standards of due care. Empirical research can show whether the incentives really work preventatively.⁷

6 In the US only 2.6 per cent of victims of medical negligence appear to file a claim, and between 73 and 91 per cent of those claims result in actual compensation. See Van Velthoven 2009, p. 461-468.

7 Consistent evidence of effects on physician behaviour and physician supply has not yet emerged in the US. See Eisenberg 2013.

3.3. *How does law come about?*

The third theme is about the *origin* of law and is also explanatory in nature. As to the processing of court cases, the researcher may want to know which considerations and incentives affect judges and juries when actually passing judgment in a trial, and to what degree. For society might put great value on courts that are truly independent and without any kind of bias. In the US, for instance, empirical research can try to sort out whether juries in medical malpractice cases tend to be pro-plaintiff, as is frequently claimed.⁸ Another line of research relates to statutory law where decision-making ultimately takes place in the political arena. Here, empirical research can analyse how the interests of the more or less silent majority are weighed against those of big business and other highly active pressure groups. The researcher can, for instance, investigate the role of the American Medical Association in informing policymakers on the impact of medical malpractice law in the US.⁹

4. Research methods I: descriptive methods

Once the research question has been formulated, the researcher must choose the proper set of methods to find the answer. A descriptive question (How does the law in action operate?) requires a different methodology from an explanatory question (What brings law about? How does law come about?). This section addresses descriptive research methods; the next section will be dedicated to explanatory research methods.

4.1. *Introduction*

Finding an answer to a descriptive question is, in essence, a matter of collecting and summarizing *data*. These data may originate from interviews or case studies, being more or less qualitative in nature. But more frequently, the data will be quantitative, counting the number of instances a specific outcome obtains, the amount involved and so on.

When it comes to medical malpractice cases, the national statistical offices generally do not collect and publish directly relevant figures. Still, useful figures may be found on a structural basis in the administrative files of the medical liability insurance companies (claims, settlements), the courts (trials, verdicts) and the health care system. Interesting data may also be taken from surveys, incidentally held or repeated at regular intervals. Various national Paths to Justice surveys, for instance, question citizens on the justiciable problems they encounter in

8 As a rule, juries in the US seem to cope quite well with the conflicting evidence they are asked to judge. See Diamond & Salerno 2013. For the incentives affecting judges and juries see also Kornhauser 2012.

9 To find that its information is seriously misleading. See Eisenberg 2013, p. 520-522. See on the role of the media also Haltom & McCann 2004.

everyday life and on the way they handle these problems from their emergence to the eventual outcome.¹⁰

Ideally, the researcher would have data that directly relate to the relevant *population*, the entire group of entities he is interested in.¹¹ Furthermore, his data would be measured in a valid and reliable manner. *Validity* is the extent to which a measure reflects the underlying concept; *reliability* is the extent to which it is possible to reproduce the value that has been measured. In practice, various stumbling blocks pop up.

4.2. Operationalizing and measuring

For one thing, the concept at hand needs to be clearly defined to be able to measure it. Frequently, however, that is more easily said than done. If the researcher is interested in rather abstract notions such as procedural justice, legitimacy or fairness, it is obviously not a simple task to *operationalize* these concepts, that is to isolate observable elements that may serve as proxy for the concept at hand. Difficulties may also arise in the *measurement* of more concrete items. Conceptually, it may be quite clear what is meant by medical errors. But the number of court verdicts in medical liability cases or the number of claims filed with medical liability insurers surely will not give a reliable estimate of the true scale of the problem. Many victims refrain from legal action because they are too sick, do not want to jeopardize the treatment by their practitioner or hang back from the costs and the emotional stress of court proceedings.

How the process of operationalization and measurement may evolve can be illustrated by the problem of finding out how many victims of medical malpractice in fact get full compensation.¹² To set the results in the proper perspective, the researcher should start at the base of the so-called dispute pyramid, consisting of all adverse medical events that result from negligent behaviour. This set of cases is not readily available, however, from administrative files or national statistics. It can only be constructed from searching through the medical files of the health care system. A further complication is that these files can only be judged on their merits by independent medical experts. To do so, these medical experts have to apply the judicial criteria of due care, which (making things somewhat simpler) generally coincide with the medical profession's own customary standard of adequate care. Clearly, building a database like this is difficult to organize, time-consuming and expensive. As a consequence, databases like this are constructed only on an irregular footing,¹³ and when they are constructed it is generally done by sampling.¹⁴ Once the set of medical errors is identified, the next step is to try and match these medical errors to the actual claims filed with medical liability

10 Pleasence et al. 2013 give a recent overview.

11 The population may be a group of persons, but the term may also refer to a set of trials, a set of countries, and so on, depending on the specific entities under study.

12 For more details, see Van Velthoven 2009.

13 See Studdert et al. 2000 for a large-scale study of medical records in Utah and Colorado in 1992.

14 See the next subsection.

insurers. This matching process is an awkward job of its own, given phonetic name variations and other administrative inconsistencies. The matching process, however, is essential to single out *false negatives* and *false positives*. False negatives are the patients that were the victim of negligent behaviour, but did not file a claim. False positives are the claimants that, at least through the eyes of independent medical experts, did not suffer from medical malpractice after all. In the final step, the researcher can then assess the indemnities paid by liability insurers to the victims of medical errors, as a result of either an out-of-court settlement (the great majority) or a court verdict (a small percentage of cases). By then the researcher is ready to report how many victims of medical malpractice are actually compensated.

4.3. Sampling

An issue deserving separate treatment here is that much available data are obtained from a *sample*, a (small) part of the population. Often it is just too costly or impractical to collect information on all the entities in the population. Extrapolating sample findings to the entire population, however, implies *descriptive inference*: “the process of using the facts we know to learn about facts we do not know”.¹⁵ This generalizing about the world on the basis of observing just a part of it goes along with inherent uncertainty. It will yield valid information only if the observations are a representative, unbiased subset. The absence of *selection bias* is best guaranteed if the observations are a *random selection*.¹⁶

An important caveat is in order, though. Random sampling cannot avoid selection bias if the set of observable cases from which the sample is drawn is itself not representative of what the researcher is interested in. The threat of *selection effects* is especially acute in empirical legal research into law enforcement and civil litigation processes. As for civil litigation, the researcher in general does not directly observe the original disputes as and when they arise between plaintiffs and defendants. The disputes become visible to him only when the plaintiffs decide to file a claim at a court or an insurance company. From there on, the administrative files will provide sufficient information to follow the main steps of the dispute until the verdict or the payment of damages. The empirical legal researcher might then feel tempted to work with the cases that are available. For he could argue that it is surely better to have some knowledge about the world of disputes than no knowledge at all. However, this maxim may turn out to be wrong, depending on the research question. The central problem is that at each step from the origin

15 Epstein & King 2002, p. 29.

16 Under equal probability sampling all entities in the population have an equal chance of being included in the sample. Sometimes, however, stratified random sampling may be preferable. Take settlements, for instance. The frequency distribution of settlement amounts is generally very skewed, with quite a lot of rather modest values and only a few extreme amounts. It is quite well possible that an equal probability sample, by chance, would end up without any of those extreme values. Distinguishing different classes of settlements and drawing separate equal probability samples within each class may yield a more useful and reliable sample. See further Epstein & King 2002, 108-114.

of a civil dispute until the final conclusion, cases will be dropping out. At each step, plaintiffs may decide to take no further action and to withdraw from the dispute, or to settle the dispute by an informal agreement with the defendant. These cases that are dropping out, and the reason why, will not generally be registered in the administrative files. That in itself would still not be a problem if it might safely be assumed that the cases dropping out are a random, unbiased subset of the original set of civil disputes. But it is highly unlikely that this assumption holds, for the plaintiffs who take no further action where other plaintiffs carry on will have a reason to do so. Either these plaintiffs are different (financial means? attitude towards risk? social skills?) or their disputes differ (type of dispute?, amount at stake?). As a consequence, the set of cases at each subsequent step of the dispute process is a non-random selection of the set of cases at the previous step, and thus also a non-random selection of the original set of disputes between plaintiffs and defendants. The researcher who does not take these selection effects into account is most susceptible to arrive at biased conclusions. In the case of medical malpractice, to derive conclusions about the scale of medical error or the average size of damage payments from court files only surely leads to a misperception of what is actually going on.

5. Research methods II: Explanatory methods

5.1. Introduction

The other two main types of empirical legal research discussed in section 3 – studying What brings law about? and How does law come about? – are explanatory in nature. To find an answer to those kinds of research questions calls for *causal inference*. That is, the researcher is looking for a particular factor (or set of factors) that led to a particular outcome. The relationship between factor and outcome is said to be causal if without the particular factor the particular outcome would not have surfaced. The outcome in this relationship is frequently referred to as the *dependent variable*, the causal factor is the *explanatory* or *independent variable*. Causal inference is tightly related to theorizing, where a *theory* is understood to be a reasoned and precise speculation about how a specific legal rule or legal system actually works, a coherent story about the real world. Such a theory is of interest only if it is subject to proof. Thus, to avoid any misunderstanding with legal academics who tend to use the term in a rather different way from other disciplines, a theory here is not in any way normative; it has a purely positivist connotation. A proper theory that is subject to proof can be used to generate testable implications. These *hypotheses* tell us what we would expect to observe in the real world if the theory is right.

Medical malpractice law provides a useful illustration. Economic theory, firmly resting on rational-choice principles, predicts that more patients will file claims upon a removal of procedural barriers to litigation or an increase in the average adjudicated indemnity payment. This, in turn, will raise liability pressure on medical practitioners, because more cases have to be defended and liability premiums will go up. Medical practitioners will take steps to try and avoid these liability

problems. One option is to raise the level of care (more diagnostic tests, less risky procedures, extra office visits), and the other option is to cut back on services (more referrals of risky patients, early retirement, moving to another state). These two reaction patterns go by the name of *positive* and *negative defensive medicine*, respectively. The chain of causation thus results in a testable hypothesis: a change in the tort system that increases (decreases) liability pressure on medical practitioners will result in more (less) positive and negative defensive medicine.

It should be noted that the relationship between causal inference and theorizing is two-sided. If there is already a useful theory at hand to develop a hypothesis about how the causal factor affects the outcome, then causal inference can be used to *test* the hypothesis. The result of the test tells us whether the theory is right or wrong. If the hypothesis proves wrong, we have a clear indication that the theory should be rejected, or at least revised. If the hypothesis passes the test, the theory has got empirical support. But alas, no one test shall ever be able to prove the theory 100% right. Even if all swans that were observed up till now were white, we cannot be sure that all swans in the entire world are white. For the possibility remains that the next swan to appear is a black one.

For many research questions, however, there is as yet no useful theory. Empirical legal science is still in its infancy, to be honest. In such a case there is as yet no well-developed hypothesis to be tested by causal inference. All the same, the available data can still be very helpful to get a grip on how legal rules and legal systems actually work. The researcher can start an *explorative search* of the data to single out the factor (or factors) that might have a causal effect on the outcome. This explorative search may yield enough insight to develop a first hypothesis, maybe even the sketchy outline of a theory. Notice, however, that the same data that were used in the explorative search cannot be used later for an independent test of the new hypothesis. After all, that hypothesis has been developed so as to fit the data, and by construction cannot be proven wrong. For a serious test the researcher has to find a new set of data.

5.2. *Variation in the causal factor*

The method of causal inference has several aspects that deserve to be singled out, for each of them may give rise to stumbling blocks of its own.

First, central to the research method is *variation in the causal factor*. The causal factor must be a variable that takes on different values: yes or no, more or less, the new rule or the old rule, and so on. If there is no variation in the causal factor, causal inference is impossible. It is then just not possible to confront two states of the world that differ in the causal factor (yes/no, more/less, and so on) to observe the differential impact on the outcome.

That simple fact is itself a serious impediment to many potentially interesting research projects. Suppose the researcher is interested in the effect of tort law on the scale of medical malpractice in Dutch hospitals. Unfortunately for him, the rules for medical liability have remained essentially unaltered in the Netherlands in recent times. So causal inference will not give him any clue. As an alternative he

might turn to the US, where several waves of tort reform over the past three decades have led to substantial changes in medical liability rules. Moreover, medical liability rules not only vary over time in the US, they also differ geographically between the states. Thus, a panel¹⁷ set of US data contains enough variation to start a promising quest for the impact of medical liability law.¹⁸ Afterwards, the researcher will have to face the difficult question of whether the findings for the US can be transplanted to the Netherlands, given the institutional and economic differences between the two countries.

5.3. Establishing the causal effect

When the researcher has assured himself that the causal factor has sufficient variation, the next step is to establish the value of the dependent variable for different values of the causal factor. The difference in outcome gives the *causal effect*, if any. In order to determine the causal effect, the researcher might set up a real experiment where he can vary the causal factor himself and directly observe any differences in the outcome. Or he might carry out a thought experiment to compare the outcome as it has actually been observed in the real world with the outcome in a hypothetical world where the causal factor would have had another value.

To organize a comparative discussion of research designs, the *Maryland Scientific Methods Scale* is a useful instrument. This scale was developed in the context of evaluating the multitude of studies on the effects of crime prevention programmes.¹⁹ It is a simple 5-point scale for methodological quality, ranging from 1 (low) to 5 (high).

Assume we want to test the hypothesis that an increase in medical liability pressure on obstetricians and gynaecologists stimulates the adoption of caesarean sections as opposed to vaginal deliveries. The argument behind the hypothesis is that upon an adverse outcome, practitioners will have more trouble in court defending their choice of a vaginal delivery, which may be less invasive for the mother, but is more risky for the child. The hypothesis might be tested by simply observing whether caesarean section rates are lower in US states with lower medical liability premiums. Finding this to be true, however, would not prove anything. States with a relatively low caesarean section rate could differ in numerous other ways from other states (for instance, in the average age and health condition of the mothers), and the differences could be attributable to any of these other factors. There is also no causal order, as the observed medical liability premiums and caesarean section rates refer to the same time frame. So the research design is of a very low methodological quality and would get the rating 1.

17 A panel contains data on various entities over time. It combines the elements of a cross section, which consists of observations of various entities at the same moment in time, and a time series, which refers to one entity at different points in time.

18 For reviews of the corresponding literature see Van Velthoven & Van Wijck 2012, Eisenberg 2013 and Zeiler & Hardcastle 2013.

19 Sherman et al. 2006, chapter 2. Throughout their book Sherman et al. are looking for evidence-based policies, trying to find out what works, what does not work and what is promising.

The hypothesis might also be tested by observing whether the caesarean section rate goes down after a tort reform measure that lowers medical liability premiums. This research design has a causal order because the tort reform measure precedes the fall in the caesarean section rate. But it is again rather weak because the fall in the caesarean section rate may be attributable to numerous alternative explanations. Perhaps the average age of the mothers went down at about the same time, perhaps the decrease was merely a continuation of a trend that already set in before, or perhaps the decrease was just a coincidence because the caesarean section rate for one reason or other was unusually high and then returned to its normal level. This design would get the rating 2.

The problem of the former two research designs is to disentangle the effects of medical liability pressure from alternative explanations. That could be achieved by confronting two sets of states, an *experimental group* that is subject to the intervention (a change in medical malpractice law), and a *control group* that does not have the intervention but is in all other aspects equal to the experimental group. The *randomized experiment* is usually considered to be the strongest research design we can arrive at (rating 5). By randomly assigning entities to both the experimental and the control groups, those in the experimental condition will be equivalent to those in the control condition on all other aspects that may affect the outcome. Notice that equivalent does not mean equal, because of the natural variability among entities. Equivalent implies that any differences between the two groups apart from the intervention are a matter of chance. The effects of chance will average out if sufficiently large numbers of entities are assigned to each group.

Randomized experiments, however, are as yet in short supply in empirical legal research. For instance, to investigate the impact of medical liability law by randomly assigning hospitals and practitioners, or areas, to an experimental and a control group would probably run into serious implementation problems. Cold feet and ethical objections also take their toll.²⁰ If so, the researcher must find a less demanding way to control for other aspects that can influence the outcome. One way to avoid setting up a randomized experiment is to look for a *natural experiment*. In a natural experiment the variation in the causal factor is not planned by the researcher, it happens because of some external event. The external event may nevertheless produce very useful data, if it creates an experimental situation (after the event) that is comparable in all other respects to the control situation (before the event).²¹

20 This, of course, raises the question of whether it is (more) ethical to continue current practices without any real, evidence-based knowledge about the personal and social effects.

21 I am not aware of any empirical study on the effects of medical malpractice law that takes advantage of a natural experiment. Several natural experiments, however, have contributed to our empirical knowledge on the effects of crime and punishment. In Washington DC, for example, the number of police in the streets was suddenly reinforced on a terrorist alert. Research has exploited this change to study the effect of police on the extent of crime. Cf. Klick & Tabarok 2010, p. 129-132.

Another way forward is to turn to some *quasi-experimental research design*, that is try to find an alternative situation (the control setting) that is sufficiently comparable to the experimental setting and compare the outcome before and after the intervention in both settings. In terms of our example, examine what happened with the caesarean section rate in a state where medical liability pressure went down as a result of some tort reform measure, find some comparable control state without tort reform and observe the development in the caesarean section rate there too. The differential development in the caesarean section rates between the two states gives an estimate of the effect of the intervention. This *difference-in-difference estimate* at the same time takes account of other factors that may have changed over time and of all other aspects that make the states differ. To give a simple example, suppose the caesarean rate in the experimental state was .45 before and fell to .30 after the change in medical malpractice law, while in the control state the caesarean rate was .34 before and .25 after the intervention. In the experimental state the caesarean section rate crime fell by .15, but this apparently cannot be fully attributed to the change in medical malpractice law, as in the control state the caesarean section rate fell too, by .09. Clearly, there were also other factors changing over time. If the two states are really comparable, in the sense that the other factors changing over time were equally active in both states, then an equal part of the drop in the caesarean section rate in the experimental state should be attributed to these other factors. The remainder ($.15 - .09 = .06$) is the causal effect of the change in medical malpractice law.

Notice that when the approach above is applied to just one experimental and one control unit, the estimate of the causal effect may be valid (unbiased). But the estimate is almost certainly not reliable because of the element of chance in choosing the two units. Thus, it is preferable (the rating for methodological quality passes from 3 to 4) to apply the difference-in-difference technique to a larger number of experimental and control units, taking care they are sufficiently comparable in all other relevant aspects.²²

5.4. Statistical significance and effect size

When it comes to the causal effect, empirical researchers want to establish whether the effect on the outcome is *statistically significant*. It is almost sure, then, that the outcome would not have obtained its actual value without the intermediation of the causal factor.

In statistical testing the researcher is opposing two hypotheses. Under the *null hypothesis* the causal factor has no effect on the outcome. The alternative hypothesis holds that there is a causal effect. Statistical analysis cannot directly prove the causal effect to be true. But it provides a method to evaluate whether the data give enough confidence to reject the null hypothesis and accept the causal effect indirectly. For this, the researcher must choose a *significance level*, the maximum amount of error he is willing to make in rejecting the null hypothesis and

22 One method to ensure the comparability of experimental and control units that is coming into fashion in the field of crime studies is *matching*. See, e.g., Wermink et al. 2013.

accepting the causal effect. Generally, the significance level is set at 5%.²³ This means that the researcher accepts a chance of at most 5% that he holds the causal effect true while in actual practice it is not. Thus, a statistically significant effect is always surrounded by (a small amount of) uncertainty.

Uncertainty also holds the other way round. When the statistical test does not result in a rejection of the null hypothesis, that does not mean for sure that there is no causal effect. All it says is that the researcher did not find enough evidence of a causal effect. But that can be an outcome of chance. Maybe his number of observations was just not large enough to be sufficiently sure about the effect. Thus, results of statistical analysis are always surrounded by (the stumbling block of) uncertainty. To get rid of that uncertainty, we should have a time capsule to rerun history with and without the intervention.

For policy purposes it is generally not enough to just know that the effect of the causal factor is statistically significant, that is non-zero. The *size* of the effect may be of interest too. Consider tort reform through a cap on non-economic damages. Such a cap, if effective, will reduce the medical liability pressure on practitioners. It may also reduce the costs of the health care system, to the benefit of all those citizens that pay a contribution somehow or other. But it directly erodes the compensation for the victims of negligent adverse medical events. And the reduction in medical care as a result of fewer diagnostic tests being performed and more low-quality professionals staying in business may be harmful to public health in general. How should the policymakers (courts, parliament) weigh these pros and cons, if they only know that the effects are non-zero but have no quantitative information on the size of the effects? By how much do medical liability premiums go down? By how much are positive and negative medicine pushed back? By how much is the average health condition of the patients affected?

A statistical technique frequently used to get insight in the quantitative size of a causal effect is *regression analysis*. In its most simple form, regression is about a linear relationship between the outcome, the dependent variable Y and the causal effect, the explanatory variable X_1 . Recognizing that the outcome may also be influenced to a greater or lesser extent by other factors, these are added to the regression equation as control variables, X_2 to X_k . The regression equation reads:

$$Y = \alpha + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_k \cdot X_k + \varepsilon, \quad (1)$$

where α is a constant, β_1 till β_k are the parameters or coefficients that inform us about the effect of the various independent variables on outcome Y , and ε is the so-called error term, representing all other factors that are not explicitly taken along in the equation, including chance. In a medical malpractice study, for instance, Y might represent the caesarean section rate in an area, while X_1 might be an indicator for medical liability pressure, and X_2 till X_k might be variables

23 But sometimes researchers set it, more stringently, at 1%, or, less demanding, at 10%.

that control for the average age and health condition of the mothers and other socio-economic, demographic and institutional factors of relevance. The causal factor X_1 (as well as all other explanatory variables) may be a continuous variable that can take any value on the measurement scale (e.g., the average level of medical liability premiums). But it can also be a so-called dummy variable that takes only the values 1 or 0, say, indicating whether the cap on non-economic damages is on or off.

Owing to the error term it will generally not be possible to calculate the true values of the parameters in the regression equation from a data set with observations on the dependent and independent variables. However, there are specialized software packages (such as SPSS, SAS, STATA and Eviews) that can be of help in providing an approximation, or *estimation*, of the parameter values and in establishing the statistical significance of each individual parameter.

In essence, regression analysis belongs to the class of quasi-experimental research designs. Once the parameters α and β_1 till β_k have been estimated, the outcome Y can be calculated from the regression equation both for a situation with the intervention and for a situation without the intervention while holding all other relevant factors constant. The difference between the two outcomes, the causal effect, is dictated by the parameter β_1 . The sign of that parameter (+ or -) tells us whether the effect of the causal factor X_1 is positive or negative, its size is the quantity by which the outcome changes per unit increase in the causal factor.

This is not the place to discuss regression techniques in greater detail.²⁴ Three important stumbling blocks for successful regression analysis, however, deserve to be mentioned here. First, when the number of observations is (too) small it may not be possible to filter out the effect of the causal factor, even when it is there. Secondly, when the explanatory variables are not independent enough or have an insufficient amount of variation, it may not be possible to determine the impact of each individual factor in a reliable manner. Thirdly, when an explanatory factor is left out that is correlated with any of the other explanatory variables, the estimation result may be biased. The effect of the omitted variable may be included in the parameters of the other explanatory variables, but we do not know by how much. The implication is not, however, to avoid misunderstanding, that a regression equation can be meaningful only if it includes *all* factors that may potentially influence the outcome. Leaving out explanatory factors for which no data are available will not bias the results, as long as they are not correlated with the independent variables in the regression equation.

5.5. About causality

Apart from randomized experiments, empirical research designs cannot provide compelling evidence of a *cause-and-effect link*. They establish a correlation or association between (variations in) the causal factor and the outcome. Two stumbling blocks remain: (a) a third factor may be responsible for both the change in the

24 See, e.g., Moore et al. 2015, Field 2013 or Wooldridge 2013.

causal factor and the outcome, and (b) the direction of causation may be the other way round. In general, it will be impossible to prove that any such alternative explanation for the observed link between the causal factor and the outcome is totally out of the question.

Even if a relationship cannot be proved to be causal, for practical purposes it may be considered as such. The following set of criteria gives some hold in deciding whether the evidence of causality is *good enough*, provided the quasi-experimental research design had sufficient quality:

- the association is strong;
- the association is consistent, as it is found not just in one study, but in several independent studies;
- higher doses of the causal factor are associated with stronger responses in the outcome;
- the alleged cause precedes the effect in time; and
- the alleged cause is plausible, because it is in accordance with theoretical reasoning.

The *direction* of causation is an issue in empirical legal research that deserves separate attention. Frequently, where the argument runs that X is causing Y, it is also very likely that Y is causing X, at least to some degree. Take, for instance, medical liability law. When claims go up in size or number as a result of more patient-friendly tort rules, pressure on the medical profession to act with more care will increase. But at the same time, the mutual insurance companies may get into financial difficulties, unleashing political pressure from the well-organized medical profession to advocate and obtain tort reform. If so, the researcher has to cope with two-sided causation: changes in tort law may be both the cause and the effect of changes in liability pressure on medical practitioners. If that *simultaneity* problem is not addressed, the regression results may be severely biased.²⁵

6. Some final comments

The path of empirical research is strewn with obstacles and stumbling stocks, as the discussion in this article has made clear. Between the lines it has also become clear that the stumbling blocks can be tackled in various ways. The researcher should carefully check the definition and measurement of the data he is going to employ; he should consider whether his conclusions are affected by selection effects; he should not use the same data set for explorative research and testing; when an opportunity exists to employ more data, he should do so; in testing he should use only an experimental or quasi-experimental research design. In addition, statistical methods exist to diagnose (and sometimes solve) problems of correlation between explanatory variables, omitted variables, simultaneity and selection effects. Alas, practical difficulties generally do not allow all stumbling

25 Zeiler & Hardcastle 2013.

blocks to be removed to full satisfaction at the same time. By implication, the researcher should present his conclusions with proper modesty, as his findings are surely not the final answer to the research question at hand. As long as we cannot rerun history, all empirical knowledge is weighed down by some degree of uncertainty.

But that should not dissuade us from taking new steps along that same path of empirical research, if only because it is the only way to test our theoretical ideas about how the world is working. If the researcher is working in accordance with the standards of empirical science, he will be transparent about the origin and quality of his data,²⁶ present his research design in such detail that it is open to replication and discuss both the strengths and the weaknesses of his conclusions. As a result, the reader has all the information he needs to assess the findings, and the degree of uncertainty surrounding them. There is probably more risk and danger in sticking to untested theoretical ideas whose truth is no more than a pure guess, than in engaging in empirical research and coming up with results whose degree of uncertainty can be reasonably assessed.

It should further be acknowledged that empirical research is an iterative process. The results of the first empirical study in a new area should not be taken at face value. The findings may be unbiased when the research design has been methodologically adequate; but they are not necessarily reliable because of the element of chance. That element of chance can be reduced by consolidating the findings of several independent empirical studies in a systematic review or meta-evaluation. Moreover, research methods evolve through a process of trial and error.²⁷

To conclude, an empirical study (be it descriptive, explorative or testing hypotheses) does not have to be perfect to be of scientific and social value. So do not hesitate, and join the club.

References

Diamond & Salerno 2013

S.S. Diamond. & J.M. Salerno, 'Empirical analysis of juries in tort cases', in: J. Arlen (ed.), *Research Handbook on the Economics of Torts*, Cheltenham/Northampton MA: Edward Elgar 2013, p. 414-435.

Donohue 2015

J.J. Donohue III, 'Empirical evaluation of law: The dream and the nightmare', *American Law and Economics Review* 2015, 17(2), p. 313-360.

Eisenberg 2013

Th. Eisenberg, 'The empirical effects of tort reform', in: J. Arlen (ed.), *Research Handbook on the Economics of Torts*, Cheltenham/Northampton MA: Edward Elgar 2013, p. 513-550.

26 Better still, he will make his data publicly available as an addendum to his publication or on a website.

27 See Donohue 2015 for a thoughtful discussion of the Law & Economics literature on the deterrent impact of the death penalty.

Ben C.J. van Velthoven

Epstein & King 2002

L. Epstein & G. King, 'The rules of inference', *The University of Chicago Law Review* 2002, 69(1), p. 1-133.

Field 2013

A. Field, *Discovering Statistics Using SPSS*, Sage Publications 2013 (4e).

Haltom & McCann 2004

W. Haltom & M. McCann, *Distorting the Law: Politics, Media and the Litigation Crisis*, Chicago IL: University of Chicago Press 2004.

Klick & Tabarrok 2010

J. Klick & A. Tabarrok, 'Police, prisons, and punishment: the empirical evidence on crime deterrence', in: B.L. Benson & P.R. Zimmerman, *Handbook on the Economics of Crime*, Cheltenham/Northampton MA: Edward Elgar 2010, p. 127-144.

Kornhauser 2012

L.A. Kornhauser, 'Appeal and supreme courts', in: C.W. Sanchirico, *Procedural Law and Economics (Encyclopedia of Law and Economics, 2e, vol. 8)*, Cheltenham/Northampton MA: Edward Elgar 2012, p. 19-51.

Moore et al. 2015

D.S. Moore, G.P. McCabe & B. Craig, *Introduction to the Practice of Statistics*, New York: W.H. Freeman and Co 2015 (8e).

Pleasence et al. 2013

P. Pleasence, N.J. Balmer & R.L. Sandefur, *Paths to Justice: A Past, Present and Future Roadmap*, London: UCL Centre for Empirical Legal Studies 2013.

Sherman et al. 2006

L.W. Sherman, D.P. Farrington, B.C. Welsh & D.L. MacKenzie, *Evidence-Based Crime Prevention*, London/New York: Routledge 2006 (Revised edition).

Siems 2014

M. Siems, *Comparative Law*, Cambridge: Cambridge University Press 2014.

Studdert et al. 2000

D.M. Studdert et al., 'Negligent care and malpractice claiming behavior in Utah and Colorado', *Medical Care* 2000, 38(3), p. 250-260.

Van Velthoven 2009

B.C.J. van Velthoven, 'Empirics of tort', in: M. Faure (ed.), *Tort Law and Economics (Encyclopedia of Law and Economics, 2e, vol. 1)*, Cheltenham/ Northampton MA: Edward Elgar 2009, p. 453-498.

Van Velthoven & Van Wijck 2012

B.C.J. van Velthoven & P.W. van Wijck, 'Medical liability: do doctors care?', *Recht der Werkelijkheid* 2012, 33(2), p. 28-47.

Wermink et al. 2013

H.T. Wermink, R. Apel, P. Nieuwbeerta & A.J. Blokland, 'The incapacitation effect of first-time imprisonment: A matched samples comparison', *Journal of Quantitative Criminology* 2013, 29(4), p. 579-600.

Wooldridge 2013

J.M. Wooldridge, *Introductory Econometrics*, Cengage Learning 2013 (5e).

Zeiler & Hardcastle 2013

K. Zeiler & L. Hardcastle, 'Do damages caps reduce medical malpractice insurance premiums? A systematic review of estimates and the methods used to produce them', in: J. Arlen (ed.), *Research Handbook on the Economics of Torts*, Cheltenham/Northampton MA: Edward Elgar 2013, p. 551-587.