Which type of lawyer wins most cases?
Abstract

For this study, information from 15,000 Dutch court cases from the period 2013–2016 was gathered using automated software crawlers. These crawlers were also used to gather information about 3,800 lawyers. The study is the first investigation into how automatic crawlers can be used to determine which party won a court case, as well as how this software can be used to generate information about the win-rate of lawyers. This method was used to find whether certain groups of lawyers are more likely to win a case. The groups are divided based on: years of experience, office size, home-field-advantage, gender and membership of a specialist organization. Using multivariate probit regression (n=1,269), it was determined that Dutch female lawyers who operate for plaintiff parties have a significant advantage over male plaintiff lawyers. In addition, based on the data that was extracted, it transpired that plaintiff lawyers enjoy a home field advantage.
Table of contents

Figures and tables ................................................................................................................................... 4
1. Introduction ......................................................................................................................................... 5
2. Literature review ................................................................................................................................. 7
   3.1. Bad markets, bad lawyers ......................................................................................................... 7
   3.2. Do lawyers make a difference? ................................................................................................. 9
   3.3. The litigation process ............................................................................................................... 9
   3.4. Defining the “winner” of a case ............................................................................................... 13
3. Academic and practical relevance .................................................................................................... 14
4. Performance predictors ...................................................................................................................... 16
   5.1. Years of experience .................................................................................................................. 16
   5.2. Size of the office ...................................................................................................................... 17
   5.3. Membership of a specialization association ............................................................................ 18
   5.4. Gender advantage .................................................................................................................... 20
   5.5. Home field advantage ............................................................................................................. 20
5. The model .......................................................................................................................................... 22
6. Data collection .................................................................................................................................... 24
   7.1. Extracting list of cases .............................................................................................................. 25
   7.2. Extracting data from individual court decisions ..................................................................... 26
   7.3. Generating data from lawyers ................................................................................................ 27
7. The data ............................................................................................................................................. 28
   8.1. Data on the court cases .......................................................................................................... 28
   8.2. Data on the lawyers ................................................................................................................ 30
8. The results .......................................................................................................................................... 32
9. Discussion ........................................................................................................................................... 36
   10.1. The number of observations .................................................................................................. 36
   10.2. The concept of winning ......................................................................................................... 36
   10.3. Females are more likely to win .............................................................................................. 37
   10.4. Home field advantage .......................................................................................................... 38
   10.5. Defending lawyer’s influence ............................................................................................... 38
   10.6. Low predictability ................................................................................................................. 39
10. Conclusion ....................................................................................................................................... 40
11. Literature .......................................................................................................................................... 41
Figures and tables

Table 1: Office size ................................................................................................................................. 17
Table 2: Specialization organizations .................................................................................................... 19
Table 3: Court cases samples ................................................................................................................ 29
Table 4: composition of court case samples ......................................................................................... 30
Table 5: Sample vs. population lawyers ................................................................................................. 31
Table 6: Data lawyers ............................................................................................................................. 32
Table 7: Random effects probit regression ............................................................................................. 34
Table 8 hypothesis ................................................................................................................................. 35

Figure 1: Galanter’s pyramid (Galanter 1983) ....................................................................................... 10
Figure 2: Model ..................................................................................................................................... 23
Figure 3: Data extraction process .......................................................................................................... 24
1. Introduction

Mr. X is a Dutch lawyer. He is a partner in a small law firm and has practiced law for more than thirty years. At first glance, Mr. X seems to be an excellent lawyer. However, few people know that Mr. X actually lost 57 of 61 Supreme Court cases in the past three years. Mr. X is an outlier, but his case shows the problem that clients face when looking for a lawyer. Clients lack the skill and knowledge to measure the quality of a lawyer so they try to estimate a lawyer’s quality based on parameters such as references from former clients, office size, experience and advertisements.

Mr. X’s noteworthy track record was found in our database, consisting of 15,000 Dutch court cases. The cases were analyzed automatically to determine the winner of each case. There could be a lot of underlying explanations for his performance, so it is unfair to conclude that Mr. X is a bad lawyer. However, as litigation is a lawyer’s core business, it stands to reason that the win-rate is at least an indicator of the lawyer’s quality. Until now, no research has been undertaken on the win-rate of lawyers in the Netherlands. This paper aims to fill this gap, or at least put a finger in the dam.

This research takes a big-data approach and uses some new techniques to obtain data. To the best of the author’s knowledge, it is the first time that this data collection approach has been used for an academic study in the Netherlands. The benefit of this approach is that large amounts of data can be collected in a small amount of time. For this study, 15,000 cases and 4,000 lawyers were analyzed. It is interesting to see to what extent this information can be collected and as such, this paper can be viewed as a reconnaissance. One of the major hurdles is to (automatically) determine the winner of a case. A disposition list was composed, consisting of words that indicate who won a case; for example, “orders to pay” indicates that the defending party has lost. In order to compose the disposition list, almost 300 court cases were analyzed manually by the author. The techniques used in this research can be beneficial for future researchers who wish to investigate the win-rate of lawyers and the causes of it. More information about the methodology is provided in chapters 6 and 7.

Using these techniques, this research aims to find whether certain groups of lawyers are more likely to win a case. To classify the groups of lawyers, the following parameters are used: years of experience, office size, home-field-advantage, gender and membership of a specialization organization. In general, there are two reasons to choose these parameters. First, these parameters are objectively determinable using data from the Dutch Bar Association. In addition, these parameters cannot be influenced easily by the lawyer (for example, the lawyer has to put a lot of effort into joining the specialization organization, which means that window dressing is unlikely).
The parameters are discussed more thoroughly in Chapter 4 and the results of the research can be found in chapters 7 and 8. The results are used to answer the following research question:

*Is there a correlation between years of experience, office size, home-field-advantage, gender or membership of a specialization organization and the likelihood of a Dutch lawyer winning a case in a Dutch court?*

A literature study is conducted in Chapter 2 in order to provide a more thorough understanding about conflicts and the role of the lawyer. In this study, the influence of the lawyer on the outcome of the case is discussed. Finally, Chapter 9 contains the discussion section in which the results are critically analyzed.

The data was gathered by Premonition Ltd., a United States-based company. Premonition provided the information for this research and also has provided the author with information about the datacrawling process, making possible verification of the reliability of the data. The datacrawling-process is described in Chapter 6. The information retrieved is strictly of a professional nature and is publicly available online. In order to protect their privacy, no individual lawyers are named. Furthermore, Dutch court documents are made anonymous before publication, so parties should not fear major privacy infringements. The Dutch Privacy Authority was informed about this research.
2. Literature review

3.1. Bad markets, bad lawyers

Finding a suitable lawyer is difficult. Like doctors, consultants and other service professionals, lawyers deliver a credence good.\(^1\) This means that clients generally do not possess the knowledge or skills to determine the quality of a lawyer. In the medical and accounting fields, procedures are fairly standardized, which means that the quality of a doctor or an accountant can be determined by comparing standards with actual performance. Such standardization does not exist in the legal world. Only the lawyer can, perhaps, determine whether they have done a good job.\(^2\) This means that the credence good problem is worse in the legal world, as even other lawyers cannot effectively determine the quality of legal services, apart from gross malpractice.\(^3\)

So, most clients cannot check the quality of legal services and have to trust their lawyer. This asymmetric information position may lead to a suboptimal level of service, sometimes even to the extent of malpractice or fraud.\(^4\) Two types of suboptimal services are associated with the credence good problem. A lawyer can underperform, which means that the client does not get enough knowledge or services to optimize his probability of achieving certain goals. A lawyer over performs when a client receives more knowledge or services than necessary to reach a goal. The consequence of over performance is that the lawyer does too much work, which is generally charged to and paid for by the client.\(^5\)

Suboptimal performance is not always intentional: some lawyers are simply unable to deliver the right performance. For example, the Dunning-Kruger effect dictates that the incompetent usually grossly overestimate their capability and therefore do not recognize their own incompetence and face a credence good problem themselves.\(^6\) Some lawyers are not able to deliver optimal, or even adequate, service levels. Research shows that 72 percent of the briefs in a certain area of labor law lack essential case law and are therefore considered inadequate. 86 percent of inadequate briefs are rejected whereas 44 percent of adequate briefs are rejected.\(^7\) It is also possible that there are lawyers who are unwilling provide optimal service levels and instead pursue their own (short term) interests at the cost of their clients’ interests.

---

\(^1\) Couwenberg et al 2009, p. 53
\(^2\) Dulleck & Kerschbamer 2001, p. 1
\(^3\) Hadfield 2000
\(^4\) Hadfield 2000
\(^5\) Dulleck & Kerschbamer 2001, p. 32
\(^6\) Dunning & Kruger 1999
\(^7\) Moss 2013
There is not a lot of literature on actual suboptimal performance. However, there is some research that suggests that lawyers treat each client differently. For example, lawyers who are paid on a contingency basis perform less well than their colleagues who are paid per hour. In addition, appointed lawyers, who receive a fixed (quite low) fee for each case, tend to visit their clients less often and spend less time on cases than defense counsels, who receive a fixed salary from the government. Therefore, it seems that lawyers do not always place their clients’ interests first. It should be noted that a lawyer’s self-interest does not necessarily revolve around money. Factors like reputation might also influence the behavior of a lawyer.

Another factor that makes it hard to determine optimal service level is that the winner takes all in a legal procedure, so the client with the best lawyer will win, even if the opponent has a very good lawyer. Therefore, the optimal level of service varies per case. In summary, to prevent suboptimal levels of service a client needs a lawyer who is both competent and ethical.

One would expect that market mechanisms take care of bad and mediocre lawyers. However, the market for lawyers is not functioning optimally due to factors such as: complexity, the credence good problem, monopoly, lack of good lawyers, abundance of mediocre lawyers and the winner takes all problems. Furthermore, even if the market functioned perfectly, there are likely to be more bad than excellent lawyers. Even if the bad lawyers are effectively being pushed out of the market, they will be able to provide inadequate services for some time period. Bar associations as well as legal systems tend to use several measures to control the quality of legal services. These measurements can be distinguished in entry controls (e.g. education requirements), process controls (e.g. peer review, continuing education, license revocation) or output controls (e.g. malpractice). However, it seems that these control measures are not sufficient to prevent bad performance. For example, most law systems require a law degree to qualify as a lawyer; however, a great law student does not automatically make a great lawyer. In addition, control measurements such as peer review and malpractice laws tend to suffer from a reluctance to report on colleagues.

---

8 Kitzer 1998
9 Cohen 2012
10 Zarzuela 2015
11 Hadfield 2000
12 Hadfield 2000
13 Moss 2013
14 Carlson 1976
15 Daicoff 1996
16 Lossa & Jullien 2012
So, clients face a serious risk when hiring a lawyer. This risk is a threshold for access to legal services. Consumers tend to have more problem with this risk than business people, who are better accustomed to taking risks. As clients might not know if their lawyer is competent and ethical, they usually use word of mouth references to find a lawyer, combined with information listed in commercial expressions from lawyers and public listing.

3.2. Do lawyers make a difference?

‘In Utopia there is no need for lawyers’, Sir Thomas Moore wrote once. In the rare event that conflict arose, the Utopians merely stated the facts of the case before the judge, who provided a fair and equitable judgement. If the wise Utopians can live without lawyers, why can’t we? In theory, a judge should follow the law or his own judgement. The lawyer should not influence the judge’s decision, apart from using the correct procedure. In practice, that notion is a bit too theoretical. Research does suggest that lawyers do influence the outcome of a case, generally in a positive sense. Clients who are defended by a defense counsel (attorneys who are employed by the government) generally receive lower sentences than clients who receive an appointed lawyer. Private lawyers perform better than appointed lawyers, but on the same level as defense counsels. Japanese women who are represented by a lawyer tend to obtain a higher financial compensation for a divorce and Taiwanese research shows that personal injury victims receive higher compensation when represented by a lawyer.

3.3. The litigation process

This research focusses on which lawyers win or lose in court. Of course, litigation is only a small part of a lawyer’s work and actually filing procedures in court is only a small part of litigation. So, it seems questionable whether the win-rate in court actually says anything about the quality of a lawyer. This research does not aim to determine the actual quality of a lawyer by means of the win-rate. However, it is important to know whether the lawyer can influence the court decision. Intuitively, one would argue that the win-rate of a lawyer is dependent on the quality of the cases. However, lawyers can, in fact, determine whether they accept a case. Even if they accept a less promising case, they can largely influence whether a court procedure is started and if a final verdict is reached.

17 Couwenberg et al 2009, p. 53
18 Morton 1991
19 Couwenberg et al 2009, p. 18-30
20 Anderson & Heaton 2012
21 Cohen 2012
22 Anderson & Heaton 2012
23 Chang et al. 2015
To understand this, it is important to know more about how actual disputes turn into court decisions and the lawyer’s influence on that process. As Galanter argued, the court decision is actually the final step in dispute resolution. In general, a (potential) conflict follows a pattern, which is shown in Figure 1.

![Figure 1: Galanter’s pyramid (Galanter 1983)](image)

As disputes progress through the pyramid, most cases are resolved without the court system. Ellickson, for example, shows that farmers in Shasta County (USA) face a lot of damage caused by trespassing cattle owned by their cattle droving neighbors. Intuitively, a thriving trespass litigation industry is to be expected in the area. In reality, the farmers actually choose to ignore most trespasses. If their neighbor does not take appropriate measures to prevent future damage, the farmers turn to alternatives including: gossiping or refusing help in the future. Litigation is only a last resort. The reason for this behavior is that litigation tends to be very expensive, not only in terms of money, time and energy, but also in the form of relationship costs. Macaulay found similar behavior in the area of contract law. In reality, the dispute pyramid shown in Figure 1 has a much broader base and a very small top. Research in the Netherlands shows that professional help is sought for only 47 percent of the disputes.

24 Galanter 1983
25 Ellickson 1998
26 Macaulay 1963
Only 18 percent of the disputes are actually brought to a court (of which 11 percent is an arbitral institution). A verdict is obtained in eight percent of the cases.\(^27\) A study in 10 different countries shows similar results.\(^28\)

Therefore, only a minority of the conflicts become “lawyer’s problems”. These problems are shown in Figure 1 between the curly brackets as “litigation”. Litigation is defined as the whole process from advising a client who has a problem, through to the execution of a verdict. Usually clients only consult a lawyer when they perceive they are unable to resolve the conflict themselves, either due to lack of skill or knowledge.

The lawyer’s first task is to determine the legal position of the client and advise how to achieve the goals set by their clients. Most of the cases that are handled by a lawyer are settled.\(^29\) Settlement is preferred by both parties because it is faster, less risky and cheaper than a court procedure. In order to achieve the best settlement, both parties engage in a process of strategic bargaining.\(^30\) If the bargaining succeeds, a trial is avoided. The bargaining phase can be seen as a selection process.

This notion was first advanced by Priest and Klein, who are the founders of the selection process theory.\(^31\) In its most simple form, this theory assumes that if parties have symmetric goals (nota bene, asymmetric goals are mostly indirect goals such as gaining a reputation as a litigator or changing case-law), a plaintiff will only sue if the expected value (probability of a win times the value of the win) is higher than the expected costs. However, the defendant, knowing this, can prevent a trial by making a settlement offer equal to the difference between the expected value and the costs. In this perfect world, trials would not occur but, in the real world, parties face difficulties determining their position.

One of these difficulties is information asymmetry, where one party has less (factual or legal) information than the other. Another difficulty is uncertainty about the outcome of the case. There is always some degree of uncertainty about the outcome of a procedure. The uncertainty increases when the legal consequences of the facts are not clear. If it is clear that someone has acted unlawfully, there is less uncertainty and vice versa. However, if the case is less clear, the uncertainty increases. Furthermore, factors like strategic behavior and risk aversion complicate the matter more.\(^32\)

\(^27\) Ter Voert & Klein Haarhuis 2015
\(^28\) Balmer & Pascoe 2014
\(^29\) Ter voert & Klein Haarhuis 2015
\(^30\) Daughety & Reinganum 2014
\(^31\) Priest & Klein 1984
\(^32\) Hylton & Naizhen 2009
Hylton and Naizhen fine-tuned and tested the trial selection model in 2009. They found that the original model only holds under rather special conditions: no information asymmetry, no uncertainty and an even distribution of the guilt.\textsuperscript{33}

This implies that, with perfect knowledge of the uncertainty and the factual information, most court procedures could be prevented. Court procedures therefore mostly occur due to lack of knowledge, or, more specifically, if one party has an unrealistic perception of their legal position.\textsuperscript{34}

In this research, the ultimate lawyer is seen as a legal professional who has the ability and the obligation to fully inform clients in order to prevent them from getting involved in an unwinnable procedure. The ability to do so will vary from lawyer to lawyer based on individual skills and knowledge. For example, some lawyers are significantly better at predicting case outcomes than others.\textsuperscript{35} Thus, in general there is no good excuse for losing cases.

The ability and obligation to inform a client is meaningless if the client does not follow the advice. There is some empirical evidence that suggests that lawyers actually can influence whether clients go to court. For example, lawyers who are paid based on contingency fees tend to settle more cases and settle them earlier, most likely because the time invested in a procedure does not outweigh the increased reward.\textsuperscript{36} Also, it seems that criminal defense lawyers in the UK and US are able to structurally induce their clients to plead guilty.\textsuperscript{37} Naturally, a client is not required to follow the advice of their lawyer. However, as explained above, clients are often unable to determine whether the advice is in their best interest, so they are likely to follow the advice from the lawyer. Furthermore, lawyers have several techniques to influence their clients such as: deflating expectations, emphasizing uncertainty, and selling settlement proposals.\textsuperscript{38} In addition, lawyers are not required to start a disadvantageous procedure just because their client wants it. On the contrary, Dutch lawyers are considered “dominus litus”, which means master of litigation: they are solely responsible for their actions in court and cannot justify them by hiding behind their client. If the lawyer and the client do not agree on the treatment of the case, the lawyer has to withdraw.\textsuperscript{39} This is obviously best-practice as it is often very hard for a lawyer to turn down a well-paying but less promising case.\textsuperscript{40}

When a case is brought to court, the lawyer is likely to influence the case outcome. Research shows that higher quality lawyers are able to provide better information to the judge and draft convincing
legal arguments, influencing the legal decision making process.\textsuperscript{41} Furthermore, in a more subtle way, they have intangible information about how a judge will react to certain arguments and claims.

Further, there is a reputation bias. Judges, especially the less capable ones, are more likely to trust the lawyers with better reputations.\textsuperscript{42} A study consisting of interviews with almost 700 judges shows that judges believe that better lawyers influence case outcomes and that a better lawyer is more likely to obtain a favorable jury decision (in the USA). However, if the judges face a weaker counsel, almost 90 percent respond by conducting more research themselves, therefore acting as a balancing force.\textsuperscript{43}

So, there is evidence that suggests that lawyers can, in fact, determine: (I) whether a case goes to court and (II) how the court will decide. This means that if a lawyer has an average high win-rate, this is either caused by superior selection skills or by superior trial skills.

3.4. Defining the “winner” of a case
An important part of this study is determining who is the winner of a procedure. It is therefore useful to identify a definition of the concept “winner” in this context. In the literature, several attempts were made to determine the winner in a court case. Anderson & Heaton determined the winner in criminal cases by comparing the length of the sentences in the criminal prosecution guidelines to the actual verdict. Using this method, it is possible to measure whether a sentence was lower than ordinary.\textsuperscript{44} Although this approach provides a precise measurement of the “winner”, its downside is that it is not generally applicable to civil law where there are no prosecution guidelines. Furthermore, it is very time consuming to determine the regulative sentence for each judgement, which also makes it very time consuming and not easily usable for large numbers of sentences. Another approach is determining how much money the lawyer won for his client(s).\textsuperscript{45} Using this approach, the value of a judgement is accounted for. However, this method can only be used on cases in which a monetary outcome is expected, thus excluding a lot of cases, and is therefore only useable in specific areas of law such as medical malpractice. Also, this method does not discount for the actual value of the claim.

\begin{itemize}
\item \textsuperscript{41} Haire et al. 1999
\item \textsuperscript{42} Lossa & Jullien 2012
\item \textsuperscript{43} Posner & Yoon 2010
\item \textsuperscript{44} Anderson & Heaton 2012
\item \textsuperscript{45} Hyman et al 2016; Kritzer 1998; Murayama 1999
\end{itemize}
and determine the ability of a lawyer to win a case.\textsuperscript{46} Such simulations are not of much use in terms of this research as this study aims to measure the winner of real cases. As shown above, the literature does not show consensus on the concept of “winner”. Instead, the “winner” is defined based on the aim of the research, the available data and methodology.

3. Academic and practical relevance

As explained previously, it is very difficult for clients to determine the quality of the legal services because of the credence good problem. Therefore, clients mostly confine themselves to word-of-mouth referrals.\textsuperscript{47} Ironically, these referrals are given by former clients who themselves cannot determine the quality of legal advice. This research aims to determine whether there are certain lawyers who win significantly more than others. This would confirm the belief that finding the right lawyer is important. A court case often has a significant impact on the lives of a client as it mostly involves large amounts of money or even imprisonment. Further, most people will only go to court once or twice in their life. The win-rate might also form a benchmark regarding the quality of a lawyer. This can be used by recruiters, lawyers and bar associations. The research also aims to determine whether there are some objective parameters that indicate the lawyers’ probability of winning.

Predictive parameters like \textit{experience, office size, specialization, gender and hometown advantage} are used. These parameters are also available to the (potential) clients and can be objectively determined. That is, contrary to catchy advertisements, a lawyer cannot easily window dress the size of their office. These parameters should help clients to find the best lawyer.

The research is also expected to contribute to the existing literature. Currently, there are some studies that argue that lawyers do influence the outcome of a procedure. However, most of these studies are either based on small case-studies or based on the opinions of clients or judges.\textsuperscript{48} Some research has been done on the influence of lawyers on payments in medical malpractice cases. In that study it was shown that a certain law office collects half of all money paid in medical malpractice cases in a certain region.\textsuperscript{49} These results indicate that the lawyer influences the outcome of a case, however it does not indicate which type of lawyer does so. This research uses objectively determined parameters to determine which group of lawyers are more likely to win based on actual court cases. The research is also done in the US and UK, which both have common-law systems. It will be interesting to discover whether the results differ in the Dutch civil law system.

\textsuperscript{46} Posner & Yoon 2010; Moss 2013; Goodman-Delahunty 2010  
\textsuperscript{47} Couwenberg et al. 2009  
\textsuperscript{48} Chang et al. 2015; Couwenberg et al. 2009  
\textsuperscript{49} Hyman et al. 2016
Furthermore, there is some literature that indicates there are some objective predictors like age and office size. However, these findings are mostly collateral to other findings and no study was found that focusses solely on these performance-predictors. This research also contributes to the literature in a more methodological area. The data in this research is collected automatically by a crawler, which searches the available online court documents. The software is also able to determine the winner of a court-procedure. Although there is ample research regarding computerized decision models and the classification of legal data, actual determination of the winner is still believed to be manual labor. This research shows to what extent it is possible to enable a computer program to decide who won the case in the Dutch legal system.
4. Performance predictors

As shown, it is likely that certain lawyers are better than others and that these lawyers will win more cases. Logically, one wonders which type of lawyers do win more: are there factors that predict whether a lawyer will win, and to what extent? The literature on this topic is rather thin and often contradicts itself, as the definition of “winners” can differ greatly. Moreover, some scholars are simply not convinced of the value of winning. Some of these predictors are discussed below.

5.1. Years of experience.

Intuitively, one would expect that more experienced lawyers tend to win more cases. After all, they know most of the pitfalls and have experience in formulating legal arguments and when considering that even the best university graduates often lack the skills and knowledge that are necessary to be a good lawyer. In the Netherlands, fresh graduates are allowed to represent people in court from the first day of their training as a lawyer, although they are formally under supervision of a patron. However, research contradicts the assumption that experienced lawyers tend to win more cases. Lawyers with between one and five years of experience tend to perform better in court than their more experienced counterparts.\textsuperscript{50} The same holds for fund managers, who also face largely uncertain outcomes.\textsuperscript{51} This effect is generally attributed to the fact that young lawyers tend to be better educated, work harder, and to the survivor bias: only the best young lawyers survive. More experienced lawyers, on the other hand, tend to have more management tasks and less litigation tasks, which also reduces their win-rate.\textsuperscript{52} Contrarily, in certain areas like negotiation experienced lawyers are more successful, mainly because they dare to ask more, which has an anchoring effect.\textsuperscript{53} Although it is likely that less experienced lawyers tend to win more cases, the literature is not strong enough to justify these one-sided hypotheses. Therefore, the following hypotheses are formulated:

H1a: there is no effect between the years of experience of a lawyer and the likelihood that the lawyer wins a case.

H0a: there is no effect between being a sole practitioner and the likelihood of the lawyer to win a case.

\textsuperscript{50} Ota 2014; Wright & Peeples 2013
\textsuperscript{51} Chevalier & Ellison 1999
\textsuperscript{52} Chevalier & Ellison 1999
\textsuperscript{53} Chang et al 2015
5.2. Size of the office

Law offices vary in size. Two hemispheres are distinguished in the literature. First, there are large law offices that focus on big corporate clients. Their clients are mostly repeat players with large budgets. These law offices only recruit lawyers from top-tier law schools. Most lawyers, however, work in smaller offices that offer more generic legal services for one-shotters.\(^{54}\) These type of lawyers earn significantly less.\(^{55}\)

The composition of the Dutch Bar is shown in Table 1.

<table>
<thead>
<tr>
<th>Office size</th>
<th>No of lawyers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 lawyer</td>
<td>2,847</td>
<td>16 %</td>
</tr>
<tr>
<td>2-6 lawyers</td>
<td>5,630</td>
<td>32 %</td>
</tr>
<tr>
<td>6-20 lawyers</td>
<td>4,670</td>
<td>27 %</td>
</tr>
<tr>
<td>20-60 lawyers</td>
<td>2,183</td>
<td>12 %</td>
</tr>
<tr>
<td>60+ lawyers</td>
<td>2,337</td>
<td>13 %</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>17,343</strong></td>
<td><strong>100 %</strong></td>
</tr>
</tbody>
</table>

*Table 1: Office size*

The Dutch legal labor market differs in some ways from the United States. For example, in the Netherlands all universities are considered more or less equal, which means the big law firms recruit, just like the other law firms, from all universities. However, big law firms tend to be a bit pickier, trying to get only the students with the highest grades, although grades seem to be no strong indicator for lawyer effectiveness.\(^{57}\) Big law offices also attract lawyers who are willing to work harder and are more ambitious. In addition, their lawyers are expected to make more billable hours, which means that they should gain experience faster.\(^{58}\) More important than the quality of the lawyers could be the quality of the clients and cases. Big clients often choose big law offices. This means that the largest offices will get the clients with the biggest budgets. These clients have the possibility of settling less promising cases.\(^{59}\) Also, these cases are often more valuable, which means that more time can be spent on legal representation, which means that the lawyer can be better prepared.\(^{60}\) However, Hyman et al. found that small boutique offices collect the most medical malpractice judgments, which contradicts this notion.\(^{61}\) The American legal data company Premonition found that, in fact, big law firms have a higher win rate. However, on average the difference is barely six percent. At the other end of the spectrum are the sole practitioners. Some

---

\(^{(54)}\) One-shotters are persons or entities which only rarely go to court.
\(^{(55)}\) Lossa & Jullien 2012
\(^{(56)}\) NOvA 2015, p. 37
\(^{(57)}\) Schultz & Sedeck 2011
\(^{(58)}\) Forstenlechner & Lettice 2008
\(^{(59)}\) Cane & Kritzner p. 463
\(^{(60)}\) Lossa & Jullien 2012
\(^{(61)}\) Hyman et al. 2016
research suggests that sole practitioners offer a lower service quality, which might be caused by their inability to ask collegial advice and their lack of a sounding board.62

Research on the correlation between office size and the likelihood of winning therefore suggests that there is a correlation, however it is uncertain whether the large or small offices have the benefits, partly because the literature on this topic is rather sparse. Therefore, it is expected that there is a correlation, however it is unsure whether this correlation benefits larger or smaller offices. Also, it is expected that sole practitioners perform differently to lawyers who operate from offices with two or more lawyers.

Therefore, the following hypotheses are formulated:

H1b: There is a correlation between being a sole practitioner and the likelihood of the lawyer winning a case.

H0b: There is no effect on the likelihood of a lawyer winning a case if the lawyer is a sole practitioner.

H1c: If a lawyer operates in an office with two or more lawyers, there is a correlation between the number of lawyers working at this office and the likelihood that the lawyer will win a case.

H0c: If the lawyer operates in an office of two or more lawyers, there is no effect on the likelihood that the lawyer will win a case due to the number of lawyers working at this office.

5.3. Membership of a specialization association

In the Netherlands lawyers are able to join specialization associations that mostly aim to increase the quality of their members in a certain area of the law. Admission qualifications vary greatly among these associations. The Dutch Bar Association (NOvA) certifies certain specialization associations that meet the following requirements: their members have at least five years of experience; they require at least 500 hours annually in a specific area of law; the members have knowledge and experience in the specific area of law; and their members undertake at least ten hours of education yearly.63 In this research only the certified specialization associations will be considered because it is certain that these meet the standards mentioned above.

It is expected that members of the certified specialization associations have a higher win-rate. Not only because they have to meet educational requirements and have experience in their specific area of law, but research suggest that membership of a select groups of lawyers, like the Queen’s

---

62 Anderson & Heaton 2012
63 NOvA 2016
Counsels in commonwealth countries, helps attain better clients and better cases, increasing the win-rate.\(^{64}\)

<table>
<thead>
<tr>
<th>Association</th>
<th>Admission requirements</th>
<th>Yearly requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Familierecht advocaten en echtscheidingsbemiddelaars (vFAS) (family law)</strong></td>
<td>- 3 years of experience</td>
<td>- 736 hours yearly spent on family law</td>
</tr>
<tr>
<td></td>
<td>- For successive two years at least 736 hours yearly spent on family law</td>
<td>- 10 hours of post academic education on family law</td>
</tr>
<tr>
<td><strong>Vereniging arbeidsrecht advocaten Nederland (VAAN) (labor law)</strong></td>
<td>- 50% of their time is spent on labor law</td>
<td>- 16 hours of post academic education on labor law</td>
</tr>
<tr>
<td></td>
<td>- Grotius education (post academic education of at least 50 hours)</td>
<td>- 50% of their time needs to be spent on labor law.</td>
</tr>
<tr>
<td></td>
<td>- 5 years of experience as a lawyer</td>
<td></td>
</tr>
<tr>
<td><strong>Vereniging voor Letselschade Advocaten (VLA) (personal injury law)</strong></td>
<td>- 5 years of experience</td>
<td>- 10 hours of post academic education on injury law</td>
</tr>
<tr>
<td></td>
<td>- Grotius</td>
<td>- 500 hours yearly spent on personal injury law yearly.</td>
</tr>
<tr>
<td></td>
<td>- 500 hours</td>
<td>- Audits</td>
</tr>
<tr>
<td><strong>Vereniging van advocaten voor slachtoffers van personenschade (ASP) (personal injury law II)</strong></td>
<td>- 50% of their time needs to be spent on personal injury law.</td>
<td></td>
</tr>
<tr>
<td><strong>Vereniging van Huurrecht Advocaten (rent law)</strong></td>
<td>- 3 years of experience</td>
<td>- 10 hours of post academic education on rent law yearly</td>
</tr>
<tr>
<td></td>
<td>- 15 hours of legal training in rent law</td>
<td>- 500 hours yearly spent on rent law</td>
</tr>
<tr>
<td></td>
<td>- 500 hours yearly spent on rent law</td>
<td></td>
</tr>
</tbody>
</table>

*Table 2: Specialization organizations*

To the author’s best knowledge, no earlier studies have been conducted regarding membership of specialization organizations and the likelihood of winning a case. As described above, logic suggests that members of these organizations are advantaged. However, there is no strong evidence that this is the case. Therefore, the following two-sided hypotheses are formulated:

H1d: There is a correlation between membership of a certified specialization association and the likelihood of a lawyer winning a case.

\(^{64}\) Lossa & Jullien 2012
H0d: There is no effect on the likelihood of a lawyer winning a case due to their membership of a certified specialization association.

5.4. Gender advantage
Women are becoming better educated than men. In the Netherlands, more females enrolled in higher education in 2014 than men. It is therefore not unimaginable that there will be more female lawyers in the future. It is established that women and men tend to behave differently in a professional environment. However, there has not been a lot of research on the differences between female and male lawyers’ performance in the courtroom. There are some indications that female lawyers are fighting an uphill battle. For example, female lawyers show a higher dropout rate. In addition, female lawyers tend to make less billable hours and provide less new client revenue, which is partly caused by the structure of their organizations. Also, female lawyers tend to spend less time on their work than their male counterparts and earn significantly less. Although the literature seems to suggest an advantage for male lawyers, it is rather thin, which does not justify a one-sided hypothesis. Therefore, the following hypotheses are formulated:

H1e: There is a correlation between the gender of a lawyer and the likelihood of that lawyer winning a case.

H0e: There is no effect on the likelihood of a lawyer winning a case based on the gender of the lawyer.

5.5. Home field advantage
The Dutch Bar Association is divided in local bar associations. Until 2008, Dutch law required formal representation by a local lawyer because each court had different procedures. However, since 2008 lawyers can represent cases in each court in the Netherlands, with the exception of the Supreme Court. However, it stands to reason that a local lawyer still has a home field advantage as they are familiar with the judges, have to travel less and know the local situation better. No research was found on the home field advantage, probably because, in most jurisdictions, only local lawyers are allowed to the bar.

65 Merens & Van den Brakel 2014
66 Faccio & Marchica 2016
67 Alarie & Adjei 2016
68 Azmat & Ferrer 2015
69 Wallace & Young 2010
The lack of research makes it appropriate draw two-sided hypotheses, which are described below:

**H1f:** If a lawyer acts before a court that is based in the district in which their office is located, that lawyer is either more likely or less likely to win a case than lawyers who act before a court that is based in a district where their office is not located.

**H0f:** If a lawyer acts before a court that is based in the district in which their office is located, that lawyer is as likely to win a case than lawyers who act before a court that is based in a district where their office is not located.
5. The model

This study aims to determine if there is a correlation between several predictors and the likelihood of winning a case. To achieve this goal, a model is used in which the “plaintiff winning a case” is the dependent variable and in which the different predictors are independent variables. The goal is to determine whether there is a correlation between the independent variables and the dependent variable. The model is described graphically in Figure 2.

The dependent variable is “Plaintiff winning a case”. In this model, either the plaintiff or the defendant wins a case, there are no draws. This means that the dependent variable is binary and if the plaintiff loses, the defendant wins automatically. Furthermore, a simple approach is taken to define the winner. The plaintiff is labeled as “winner” if any claims, excluding process costs, are granted; in all other cases the defendant wins the procedure.

For the independent variables, a distinction is made between lawyers representing a plaintiff and defending parties. This distinction is justified because it is possible that there is some dependency between the type of plaintiff lawyer and type of defendant lawyer. For example, large organizations mostly hire lawyers from large offices. It is therefore more likely that both the plaintiff and defendant lawyer are from a large office. The possible correlation between type of plaintiff and defendant lawyer could lead to overestimating the plaintiff’s effect if the defendant’s characteristics are not taken into account. In order to discount individual behavior, the plaintiff lawyers are clustered, which means that if a plaintiff’s lawyer has several cases, the model views this lawyer as a single individual. The observations are clustered because multiple cases from the same lawyer could correlate due to the specific character of the lawyer, which is not representative of the group to which the lawyer belongs. Note that it is not logical to cluster the defendant’s lawyers as the plaintiff can always decide whether a case will go to court, which means that the plaintiff’s choice of lawyer might affect which lawyer the defendant picks.

The years of experience are measured up to the judgment of the case. So, the experience of each lawyer increases over time. Note that the crawler software can only determine in what year the judgment was made. The software arranges all judgments artificially in January in a given year. The real experience can therefore be eleven months less than measured.

---

70 Cane & Kritzner 2010 p. 468
Figure 2: Model
6. Data collection

In order to test the hypotheses, data was collected from published court records in the Netherlands. Crawler software was used to find these documents and extract useful data automatically. The obvious benefit of this approach is its scalability. Its Achilles heel is that it still lacks the “human touch” and has problems identifying slight variations in sentence construction. The data collection process is presented in Figure 3.

Figure 3: Data extraction process
7.1. Extracting list of cases

In the Netherlands, all court cases are published on www.rechtspraak.nl. In 1997, the Dutch government decided that basic information about the government, lawmaking, court decisions and democracy should become more transparent and accessible.\textsuperscript{71} More specifically, a representative image of all court decisions should be provided. For the definition of “representative image”, the Dutch government sought connection to recommendation R(95) 11 from the Committee of Ministers of the Council of Europe.\textsuperscript{72} In order to reach this goal, the cases are published online.

For the majority of cases, only a reference is published: in practical terms, no data can be extracted from these references. In approximately 1–2 percent of the cases, the decision is published in full, though anonymized.\textsuperscript{73} The courts decide which cases are published fully according to some uniform selection criteria. The selection criteria have been used by all courts since 2012. According to these criteria, there are four types of cases. First, cases from the highest courts are always published unless they contain only standard decisions. Second, cases from the lower courts that concern certain topics, such as collective labor law, treaty law and EU law, are always published unless they contain only standard decisions. Third, cases from the lower courts that might be of interest and meet criteria such as media attention, discussion in journals or alter a line in jurisprudence are always published. Fourth, all other decisions are published as much as possible.\textsuperscript{74} Given their aim, it is unlikely that the selection criteria have an influence on the outcome of this research. The fully published cases serve as an important source for case law in the Netherlands and beyond and are extensively used as a source of data for legal research.

Until recently, this data extraction was mostly done manually. In order to facilitate legal research, the Dutch courts launched the Open Data service, which allows direct searches in the database for court decisions using queries. These documents are provided in an XML-formatted list of documents that match the search criteria. The relevant documents are sorted by their European Case Law Identifier (ECLI) number.\textsuperscript{75}

The automatic crawler software first retrieves the list of relevant cases by using the three queries. Only civilian law cases are selected because these cases differ significantly from criminal or

\textsuperscript{71} Kamerstukken vergaderjaar 1996-1997, 20644, ‘Naar toegankelijkheid van overheidsinformatie’
\textsuperscript{72} Committee of Ministers of the Council of Europe, Recommendation no. (95) 11, 11-09-1995, ‘Concerning the selection, processing, presentation and archiving of court decisions in legal information retrieval systems.’
\textsuperscript{73} Rechtspraak 2014
\textsuperscript{74} Besluit selectiecriteria uitsprakendatabank rechtspraak.nl 2012, accessibele via https://www.rechtspraak.nl/Uitspraken-en-nieuws/Uitspraken/Paginas/Selectiecriteria.aspx
administrative cases, which makes a comparison not very useful. Also, only cases from the last three years were selected, mainly because older cases are less likely to contain relevant data about the current lawyers.

7.2. Extracting data from individual court decisions

The list of relevant cases contains the ECLI each specific case. Using the ECLI, a specific court decision can be accessed. The court documents are also published in an XML format, which contains certain metadata and enables easy access by software crawlers.

From the court decisions, the following data was extracted by the crawler software: (a) ECLI, (b) date of decision, (c) court, (d) area of law, (e) lawyers representing both parties, (f) winning lawyer. Data (a) to (d) are listed in the metadata and can be quite easily and accurately extracted. As this data is always in the same place and in the same format, there is little risk of the software making an error. Information about the lawyers acting in a certain case is found in the actual text of the decision and is a little bit harder to extract. Fortunately, all court decisions are written in the same format. The name of the lawyer is always preceded by the phrase “advocaat:” (meaning lawyer). Further, the plaintiff lawyer always precedes the defending lawyer. It is therefore relatively easy to determine which lawyer was on which side.

Deciding which party won the case is the difficult part. When looking on a case level, each court decision is different and depends mostly on the claim that was laid. Further, each decision is written by individual judges who use their own grammar and phrases. Another major problem is that there is not always a single winner. For example, who is the winner if one claims €1 million and gets awarded €10,000? In our model, a simple division between winner and loser is chosen. If any of the claims are granted, the plaintiff has won the case, otherwise the defendant has won the case. The (dis-)advantages of this simple division are discussed in the discussion section. In order to determine whether the claims are granted, a disposition list is used. Basically, a disposition list contains a number of phrases that predict whether a claim is granted. The crawler searches for these phrases and if it finds one, points out the winner and loser.

The disposition list was composed manually by analyzing 15 cases from each district court, 25 cases from each high court and 25 cases from the Supreme Court. In total, 270 decisions were analyzed. This resulted in a list containing 79 dispositions.

With this list, a small test sample was generated, which was used to improve the disposition list, after which the process was repeated. When a certain level of “hits” was reached, a larger test sample,
containing 1000 cases, was generated. This was used to further improve the quality of the disposition list. Ultimately, the crawler was used to generate all court cases.

### 7.3. Generating data from lawyers

The data about the lawyers was crawled via the Dutch Bar Association’s (NOvA) website. The crawling process for lawyer information is much more straightforward than the process for court cases. Basically, all the information is on the website, neatly categorized in a table, which leaves little room for error. The gender of the lawyer is defined with the words “de heer” (meaning “Mr.”) or “mevrouw” (meaning: “Mrs.”). The district in which the lawyer resides is determined using the zip-code. The zip-codes were paired with the districts of the eleven lower courts (rechtbanken).

The districts of the four courts of appeal were determined by combining the relevant lower courts. (for example, the Court of Appeal of ‘s Hertogenbosch has jurisdiction over the courts of Zeeland-West Brabant, Limburg and Oost-Brabant.) The Supreme Court has jurisdiction over the whole country and in this model all lawyers are located in their district. This means that there cannot be a home field advantage in Supreme Court cases.
7. The data

8.1. Data on the court cases
A total of 15,217 cases were analyzed using the crawler software. Of these, 13,212 were used. The data is not perfect: 1,080 cases were dropped because they are not in the right format. Furthermore, as this study focusses on lawyers from the Dutch Bar Association (NOvA), 731 cases were dropped because they were decided by courts in overseas territories. The conclusions from the Attorney-General (194) were also dropped because they are no true judgments, leaving 13,212 cases in the data. The plaintiff won 3,420 cases and the defendant won 3,999 cases. 1,171 cases were undecided. The group of undecided cases is mainly composed of court orders which do not contain a final judgement, for instance an order to provide proof of a stated fact. Also a part of the undecided cases consists of decisions involving counterclaims. The crawling software cannot process counterclaims because a ‘winning’ disposition in a counterclaim is essentially a victory for the defendant. This means that the software would create a false positive win for the plaintiff. So, the crawler was able to identify the outcome of the case in 8,469 cases which means that it has a recall of 0.65. The undecided cases (1,150) were dropped because this research focusses only on winning and losing. The sample therefore consists of the cases in which either the plaintiff or defendant won the case (7,421).

Both the lawyer database and the court case database were combined. The spelling of the names of the lawyers was standardized. In some cases multiple lawyers operated for one party. If the lawyers succeeded each other only the last lawyer was counted, as the last lawyer has the option to withdraw a case, preventing a negative outcome. Furthermore, the last operating lawyer has consciously adopted the case, which means that the last lawyer must have believed in the feasibility of the case. If a lawyer has withdrawn without the appointment of a new lawyer, the last (withdrawing) lawyer is counted. If two lawyers operated simultaneously only the first named lawyer was chosen as it is conventional in the Netherlands to name the most senior lawyer first, the most senior lawyer is most likely to be in charge of a procedure. In 3,420 cases out of 7,419 (46 %) it was possible to match the plaintiff lawyer with one of the lawyers in our lawyer database. In 3,999 cases out of 7,419 (53 %) the defendant lawyer was matched with the lawyer in our lawyer database.
The higher percentage of plaintiff lawyer’s matching could be explained because the defendant can choose not to appoint a lawyer, so it is possible that there are less defending lawyers in general. Only 1,269 cases (17%) provided a match for both the plaintiff and the defendant lawyer, which is called our regression sample. Note that the expected match was: 46% (plaintiff lawyers) * 54% (defendant lawyers) = 24% if the variables are completely independent. That the actual match is a lower indicates some dependency, which can possibly be explained because a lawyer cannot represent both parties in a conflict.

The matching percentages are fairly low, this can be blamed on the low number of lawyers in our lawyer database. It is likely that if all lawyers would have been included there would be a higher number of matched. Also, the system only counts perfect matches so typing errors can result in a mismatch, which also explains some of the mismatches. A third reason for the low matching-rate is that sometimes parties are represented by non-lawyers, these cases would never result in a match.

Table 3 shows the size of the samples and total data.

<table>
<thead>
<tr>
<th></th>
<th>15,217</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrong format</td>
<td>1080</td>
</tr>
<tr>
<td>Overseas</td>
<td>731</td>
</tr>
<tr>
<td>AG</td>
<td>194</td>
</tr>
<tr>
<td>Cases</td>
<td>13,212</td>
</tr>
<tr>
<td>No recall</td>
<td>4,743</td>
</tr>
<tr>
<td>Recall</td>
<td>8,469</td>
</tr>
<tr>
<td>Plaintiff</td>
<td>3,420</td>
</tr>
<tr>
<td>(40%)</td>
<td></td>
</tr>
<tr>
<td>Defendant</td>
<td>3,999</td>
</tr>
<tr>
<td>(47%)</td>
<td></td>
</tr>
<tr>
<td>Undecided</td>
<td>1,050</td>
</tr>
<tr>
<td>(13%)</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>7,419</td>
</tr>
<tr>
<td>Plaintiff</td>
<td>3,420</td>
</tr>
<tr>
<td>(46%)</td>
<td></td>
</tr>
<tr>
<td>Defendant</td>
<td>3,999</td>
</tr>
<tr>
<td>(54%)</td>
<td></td>
</tr>
<tr>
<td>Regression Sample</td>
<td>1,269</td>
</tr>
<tr>
<td>Plaintiff</td>
<td>579</td>
</tr>
<tr>
<td>(46%)</td>
<td></td>
</tr>
<tr>
<td>Defendant</td>
<td>690</td>
</tr>
<tr>
<td>(54%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Court cases samples

To check whether the sample and regression samples of court cases are random samples of the data the distribution of cases before either the Supreme Court, courts of appeal, and lower courts was measured. A chi-squared test is performed with the null hypothesis that the proportion of cases from all courts are the same among the samples and the data.
The chi-squared test statistic is 91.92, with a p-value <.001, indicating that the distribution among courts is not equal for the samples and the data. However, as the Supreme Court cases cannot be undecided and because it is easier to determine the winner in Supreme Court cases, it can be expected that relatively more cases from the Supreme Court carry over in the samples. Therefore, a second chi-squared test is performed to test whether the proportion of lower courts and courts of appeal cases is equal in the data and the samples. The resulting chi-squared test statistic is 4.00, with a p-value of 0.14, indicating the null hypothesis that the proportions are equal among the samples does not have to be rejected, indicating that the samples are truly random.

<table>
<thead>
<tr>
<th>Court</th>
<th>Data</th>
<th>Sample</th>
<th>Regression sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower courts</td>
<td>5,998</td>
<td>3,392</td>
<td>554</td>
</tr>
<tr>
<td>Courts of appeal</td>
<td>6,764</td>
<td>3,609</td>
<td>619</td>
</tr>
<tr>
<td>Supreme Court</td>
<td>450</td>
<td>418</td>
<td>96</td>
</tr>
<tr>
<td>Total</td>
<td>13,212</td>
<td>7,419</td>
<td>1,269</td>
</tr>
</tbody>
</table>

Table 4: composition of court case samples

8.2. Data on the lawyers

The crawler was able to collect information about 3,567 lawyers, out of the 17,343 registered lawyers on the first of January 2016. Most of the lawyers in the sample are male (62.25%), the average office-size is 6.53 lawyers and the average years of experience is 17 years. Approximately 20% of the lawyers in our sample are member of a certified specialization organization. To verify whether the lawyer’s sample is randomly drawn from the total population a comparison is made on two characteristics of both the sample and the population. The annual report of the Dutch Bar Association (NOvA) of 2015 states the number of lawyers in each district as well as the gender of those lawyers. The comparison is shown in Table 5, with differences larger than five percentage point in bold. It is clear that the composition of the sample differs from the total population. This means that the sample is not representable for the whole population. It should be noted that not all lawyers litigate and there is no data on which lawyers out of the whole population do actually litigate. This means that it is not possible to determine the actual population of litigating lawyers. Also, only 2,324 lawyers in the sample litigate. Hence, even though the sample differs from the total population, it is likely but not certain that the litigating lawyers from the sample differ of composition from the actual population of litigating lawyers.

---

76 NOvA 2015
77 NOvA 2015
<table>
<thead>
<tr>
<th>District</th>
<th>No. of Lawyers Population, n=17343</th>
<th>Sample, n=3567</th>
<th>No. of Males Population, n=17343</th>
<th>Sample, n=3567</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amsterdam</td>
<td>30%</td>
<td>23%</td>
<td>59%</td>
<td>67%</td>
</tr>
<tr>
<td>Den Haag</td>
<td>11%</td>
<td>3%</td>
<td>54%</td>
<td>62%</td>
</tr>
<tr>
<td>Gelderland</td>
<td>7%</td>
<td>6%</td>
<td>57%</td>
<td>62%</td>
</tr>
<tr>
<td>Limburg</td>
<td>5%</td>
<td>7%</td>
<td>59%</td>
<td>67%</td>
</tr>
<tr>
<td>Midden-Nederland</td>
<td>11%</td>
<td>11%</td>
<td>37%</td>
<td>58%</td>
</tr>
<tr>
<td>Noord-Holland</td>
<td>5%</td>
<td>6%</td>
<td>55%</td>
<td>61%</td>
</tr>
<tr>
<td>Noord-Nederland</td>
<td>5%</td>
<td>6%</td>
<td>59%</td>
<td>61%</td>
</tr>
<tr>
<td>Oost-Brabant</td>
<td>7%</td>
<td>4%</td>
<td>56%</td>
<td>65%</td>
</tr>
<tr>
<td>Overijssel</td>
<td>4%</td>
<td>5%</td>
<td>58%</td>
<td>60%</td>
</tr>
<tr>
<td>Rotterdam</td>
<td>11%</td>
<td>14%</td>
<td>57%</td>
<td>60%</td>
</tr>
<tr>
<td>Zeeland-West Brabant</td>
<td>6%</td>
<td>15%</td>
<td>57%</td>
<td>61%</td>
</tr>
<tr>
<td><strong>Totaal</strong></td>
<td>100%</td>
<td>100%</td>
<td>55%</td>
<td>63%</td>
</tr>
</tbody>
</table>

A distinction is made between plaintiff and defendant lawyer’s observations. As shown in Table 6 these groups have roughly the same characteristics as the complete population. Note that the standard deviation is a bit higher, which might be caused by the smaller size of the plaintiff and defendant groups. The plaintiff and defendant groups are taken solely from cases in which both a plaintiff and defendant lawyer is known, which accounts for the small sample size (1,269). As only observations are counted some lawyers appear multiple times in our database.
### Lawyers Sample

| Male | 2,225 |
| Average office size | 6.25  |
| Experience years (average) | 17 |
| Specialization organization | 715 |

### Plaintiff lawyers in regression sample

| Male | 914 |
| Average office size | 6.95 |
| Experience years (average) | 16.93 |
| Specialization organization | 244 |

### Defendant lawyers in regression sample

| Male | 861 |
| Average office size | 8.82 |
| Experience years (average) | 18.31 |
| Specialization organization | 261 |

### Home field advantage

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Plaintiff lawyers in regression sample</td>
<td>1,269</td>
</tr>
<tr>
<td>Defendant lawyers in regression sample</td>
<td>1,269</td>
</tr>
</tbody>
</table>

8. **The results**

As this study wishes to identify performance predictors it is necessary to control for other variables when studying another. This can be done using multivariable regression, in which the effect of one variable is estimated keeping all other variables constant. As the dependent variable is binary (the plaintiff either wins or loses), an ordinary least squares (OLS) regression is not appropriate. One of the Gauss-Markov assumptions needed for OLS is that of homoscedasticity, which means that the difference between the predicted and actual value of the dependent variable does not depend on the independent variables. However, OLS estimates can lie outside the probability range and the error term therefore depends on the independent variables. As a result, the t-statistics are biased and one could draw the wrong conclusion.

The probit model estimates the probability that the dependent variable is equal to one given the independent variables, in the following way:
Pr(y_i = 1|x_i) = \Phi(x\beta)

\Phi = cumulative normal distribution

\chi = a vector of independent variables and the constant

\beta = the coefficients associated with these independent variables

As the cumulative normal distribution ranges from zero to one, no predictions are made outside the probability region.\(^79\)

The probit-model holds as a disadvantage that it does not counts the R-squared variable, which is seen as a measurement for the predicting value of the model. This variable can be estimated using McFadden’s pseudo R-squared, which is calculated using the following formula:

\[
R^2_{adj} = 1 - \frac{\ln(\hat{L}(M_{Full}))}{\ln(\hat{L}(M_{Intercept}))}
\]

\(M_{Full}\) = the model with predictors

\(M_{Intercept}\) = the model without predictors,

\(\hat{L}\) = the estimated likelihood

\(R^2_{adj}\) = the pseudo R-squared.\(^80\)

For our model McFadden’s pseudo R-squared = \(1 - \frac{\ln(\hat{L}(861))}{\ln(2200)}\) = 0.128935.

McFadden’s pseudo R-squared is generally lower than the classical R-squared coefficient. According to McFadden a score between .2 and .4 is considered an excellent fit.\(^81\) Our model does provide some predicting value, however is not considered very strong. Also a model with separate plaintiff and defendant variables was tried. As predicted these separate models have little predictive value, based on McFadden’s pseudo R-squared, although the number of observations was a lot higher (approximately 4,000) than in our combined model.

The results of the model are shown in Table 7, significant results are in bold and noted with asterisk.

| Plaintiff winner | Coef. | P>|z| |
|------------------|-------|-------|

\(^79\) Domencich & McFadden 1974
\(^80\) Domencich & McFadden 1974 p. 102 - 124
\(^81\) Hensher & Stopher 1979 p. 306
Table 7: Random effects probit regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plaintiff male</td>
<td>-0.201</td>
<td>0.100</td>
<td>0.044**</td>
</tr>
<tr>
<td>Plaintiff years of experience</td>
<td>0.005</td>
<td>0.005</td>
<td>0.253</td>
</tr>
<tr>
<td>Plaintiff spec. organization</td>
<td>0.117</td>
<td>0.112</td>
<td>0.295</td>
</tr>
<tr>
<td>Plaintiff home field advantage</td>
<td>0.165</td>
<td>0.085</td>
<td>0.053*</td>
</tr>
<tr>
<td>Plaintiff office size</td>
<td>0.002</td>
<td>0.004</td>
<td>0.693</td>
</tr>
<tr>
<td>Def. male</td>
<td>-0.039</td>
<td>0.091</td>
<td>0.671</td>
</tr>
<tr>
<td>Def. years of experience</td>
<td>0.005</td>
<td>0.004</td>
<td>0.169</td>
</tr>
<tr>
<td>Def. spec. organization</td>
<td>0.070</td>
<td>0.102</td>
<td>0.488</td>
</tr>
<tr>
<td>Def. home field advantage</td>
<td>0.027</td>
<td>0.083</td>
<td>0.743</td>
</tr>
<tr>
<td>Def. office size</td>
<td>-0.004</td>
<td>0.004</td>
<td>0.308</td>
</tr>
<tr>
<td>Constant</td>
<td>0.226</td>
<td>0.134</td>
<td>0.092</td>
</tr>
</tbody>
</table>

Note that the model predicts the probability that the plaintiff wins a case. The defendant variables should therefore be read inversely. E.g. a high coefficient for “defendant specialization organization” would imply that if the defendant lawyer is a member of a specialization organization the plaintiff has a higher probability of winning. So, low or even negative values for defendant variables predict an advantage for defending lawyers. The gender variables indicate the probability that a male will win, so a negative coefficient actually predicts an advantage for female lawyers.

The “office size” and “years of experience” are cardinal variables, which means that their effect could be non-linear. The data was tested for a possible non-linear effect, however, no significant results were found. Also the data was tested for sole practitioners, without any significant results.

With a threshold value of $P>|0.1|$ the data indicates acceptance of hypotheses as shown in Table 8.

| # | Hypothesis                        | (weakly) significant $P>|0.1|$ |
|---|-----------------------------------|-------------------------------|
| H0a | there is no effect between the years of experience of a | Not rejected |

34
lawyer and the likelihood that the lawyer wins a case.

H0b there is no effect between being a sole practitioner and the likelihood of the lawyer to win a case. Not rejected

H0c If the lawyer operates in an office of two or more lawyers there is no effect between the numbers of lawyers working at this office and the likelihood that the lawyer wins a case. Not rejected

H0d there is no effect between membership of a certified specialization association and the likelihood of a lawyer winning a case. Not rejected

H0e there is no effect between the gender of a lawyer and the likelihood of that lawyer to win a case. Rejected for plaintiff lawyers (Coefficient: -.20) 
(P>|Z| 0.044) (95 % conf. interval: -.3961 - -.0056)

H0f If a lawyer acts before a court which is resided in the district in which their office is located that lawyer is as likely to win a case than lawyers who act before a court which is resided in a district where their office is not located. Rejected for plaintiff lawyers (Coefficient: .16) 
(P>|Z| 0.053) (95 % conf. interval: -.0018 - .3321)

Table 8 hypothesis

So, according to the data male lawyers have a distinct disadvantage when they act for the plaintiff party. With a 95 % confidence interval our data suggests that there is a negative correlation between the gender male and the probability to win a case (95 % confidence interval - .39, - .005). The probit analysis coefficients indicate an increase in Z-score for each increase in a certain variable. As the gender variable is binominal the Z-score increases with approximately -.20 from the mean when the plaintiff lawyer is female, which leads to an decrease of probability of approximately 8 percentage points. So, the male plaintiff lawyer is approximately 8 percentage points less likely to win than his female counterpart. Logically this holds vice versa for female plaintiff’s lawyers.

Our data also suggest a home-field-advantage for the plaintiff. Although it must be noted that this advantage cannot be established with 95 % confidence (95 % confidence interval -.002, .332). However, the lower end of the confidence interval is just below zero, so it is at least likely that the home-field-advantage has some positive effect on the win-rate. According to our data plaintiff lawyers who act in their own district win on average 6.5 percentage points more cases.
9. Discussion

The first goals of this research is to explore the application of big-data techniques to research the win-rate of lawyers. The techniques which are used will be discussed first, with their advantages and pitfalls. Then, the actual results will be discussed.

10.1. The number of observations

Our analysis rests on only 1,269 observations, out of 15,000 cases. The low amount of useful observations make it hard to verify significant relationships. It is likely that with more data, more significant results would have been found. For future research it should be possible to collect more data. According to rechtspraak.nl there are approximately 30,000 published court decisions between 2013 and 2016, so the amount of initial court documents could potentially be doubled if the crawling software is fine tuned. Also the disposition list could be improved, to increase the recall rate. A huge improvement in observations could be obtained by crawling all the lawyers, instead of a sample. The sample of lawyers is not a random sample of the population of lawyers, so it is possible that the observations do not hold for the whole population. This doubt can be taken away by repeating this study with the whole population of lawyers. The samples of court data are likely to be random, however with a larger (regression) sample, more significant results might be found.

10.2. The concept of winning

In this research a fairly rudimentary concept of winner is used. If the judge grants a claim the plaintiff wins, otherwise the defendant wins. The advantage is that it provides a clear dividing line, which makes it easy for the software to determine the winner. The disadvantage is that a lot of information is lost, such as the amount of the claim which was granted. When discussing this research a lot of people doubted whether it is possible to objectively determine who the winner is. Often an example like this is used: “who has won if someone is sued for one million and the judge decides that thousand euro have to be paid to the defendant?”. As shown there are a lot of ways to determine the winner, which could imply that the model in this paper has little external validity. There are several arguments to debunk this notion. First, most variables such as: office size and home-field-advantage are unlikely to be influenced by the concept of “winner”. Another concept of winner should benefit these groups of lawyers equally. As is discussed below, females could benefit from our straightforward model. Secondly, our data shows that the plaintiff won in 3,420 cases and the defendant in 3,999 cases, so it seems that the model is pretty evenly balanced. If our definition of winner is flawed it is expected to favor the plaintiff. After all, even the smallest claim granted means a “win”.

36
Although no information is available about the actual win-rate of the whole population, so conclusions cannot be drawn upon the win-rate of the sample. Last, it could be possible that in reality the judge mostly rules completely in favor of one of the parties, not diminishing claims significantly. A lot of judgments are after all basically binary i.e. the labor contract is dissolved or not. This holds more true for courts of appeal and the Supreme Court, which basically can only nullify lower court decisions.\textsuperscript{82} It stands to reason that a judge will favor one party as taking the middle ground will lead to two dissatisfied parties, which could increase the probability for appeal. It would be interesting to further investigate whether judges actually rule completely in favor of one party, perhaps by comparing the percentage of claims granted.

10.3. Females are more likely to win
Surprisingly, according to our data plaintiff lawyers are more likely to win cases if they are female. This effect was not predicted and seems to be pretty strong, i.e. an 8 percentage point increase in probability. There are two types of explanations for this effect. Either females just win more cases or our methodology might favor female lawyers.

There are several explanations in literature why female lawyers could win more. One might be that females (litigants) tend to be judged differently than males. Male judges are more lenient on female litigants, possibly because of chivalrous notions. Female judges tend to judge other females the same or even more lenient.\textsuperscript{83} So female lawyers might also profit from this effect, either because they represent more female clients, or because they profit from the same effect. A second explanation might be that females tend to react differently to risk, although the type of reaction varies depending on the type of risk and experiment.\textsuperscript{84} As explained in the literature review lawyers are believed to make an educated guess about the legal risks and possible beneficial outcomes of a case. This guess in turn is used to determine legal decisions. Risk aversion has obviously a large effect on this decision, which could mean that women make different decisions in the case selection process.\textsuperscript{85} In fact Goodman-Delahunty’s data suggests that women tend to underestimate their legal position, which could lead to a more cautious litigation strategy, explaining a higher win rate.\textsuperscript{86} Females could also profit from a selection bias. The minority of our sample (38 \%) is female. As Alarie and Ajei argue

\textsuperscript{82} A partial nullification is possible but more rare.
\textsuperscript{83} Gill et al. 2015
\textsuperscript{84} Eeckel & Grossman 2008
\textsuperscript{85} Hylton & Naizhen 2009
\textsuperscript{86} Goodman-Delahunty et al 2010
a lot of female lawyers actually enter the bar, however their dropout rate is a lot higher than for men. The remaining female lawyers could just be better selected than their male counterparts. Also, it might be possible that the model in this study favors female specific threats, like risk behavior, as it only rewards complete victory for the defendant i.e. if a plaintiff’s lawyer is more cautious and more precise in stating claims it is more likely that these claims are granted by the judge. It seems strange that the gender benefit only applies for plaintiff lawyers. A possible explanation could be that defendant characteristics seem to matter less than plaintiff characteristics (as shown below). Also it could be that the female specific treats are less rewarded by our model in the defendant role. I.e. the defendant has less room to behave risky.

10.4.  Home field advantage
There is also a strong indication that the plaintiff lawyer at home is more likely to win. It is unlikely that this effect can be blamed on the model as the model should favor the visiting and home party equally. As explained in chapter 4 no literature was found on the home field advantage in litigation. The lack of literature can perhaps be explained because most legal research is performed in the USA, where lawyers are not allowed to plead in another district. Due to the lack of literature it is hard to explain why the home field advantage seems to exist. Although, it should be noted though that the home-field-advantage is strongly recognized in sports and also in the legal world, as the home-field-advantage is generally granted to the defending party. It is possible that lawyers are more comfortable in their own district or have more knowledge about the judges and procedures. Although it must be noted that the Dutch courts generally use the same procedures. Also, lawyers in a certain district can gain a certain reputation among the judges, which might influence judgments. Lossa and Jullien argued that Queen’s Counsels profit from a better reputation and more thrust among judges, this might also apply for lawyers who operate in their own district. The home field advantage is, in any case, an interesting topic for future research.

10.5.  Defending lawyer’s influence
It is remarkable that the defending lawyer’s characteristics do not seem to influence the probability of winning the procedure to a great extent. As the results are not significant no hard conclusions can be drawn. However, the coefficients of the defending characteristics are structurally lower than the plaintiff’s lawyers characteristics. This could imply that the defending lawyer has less influence on the verdict. Perhaps because the defending lawyer has a limited choice of legal defenses. Also the lower

87 Alarie & Adjei 2016
88 Lossa & Jullien 2012
effect of the defense lawyers characteristics can be explained because the model only rewards complete victories for the defendant, which lowers the number of viable strategies for defending lawyers and therefore their influence on the outcome of the case.

10.6. Low predictability
The model in this research features a low predictability. This could be caused by the low amount of observations. However, another explanation could be that the influence of a lawyer on the outcome of a procedure is small, but noticeable. Even the best lawyer can only do so much in a procedure and might only have a small edge over worse lawyers. This explanation makes sense as in this study only judgments are measured.

As explained the trial selection model assumes that a final judgement only occurs if both parties believe that they have at least 50 % probability to win. As both parties are represented by trained lawyers it is normally not likely that both estimations are very far off. So the model tends to only measure cases in which both parties have almost the same probability to win.

Another explanation could be the role of the judge. In the Netherlands there is a trend towards a more active judge, who directs the court procedure. If the judge is more active there is less room for lawyers to influence the procedure. In fact, the author asked a Dutch judge how she reacted when a lawyer was clearly messing it up. The judge told that she would do more legal research and even tried to ‘help’ the less gifted lawyer. Posner and Yoon found similar reactions in their interviews with 700 judges. Furthermore, Dutch law is becoming more rigid, like the new labour law (Wet werk en zekerheid) as well as require the judge to apply certain rules mandatory, such as European consumer protection laws. Both tend to decrease the influence of the lawyer.

---

89 Lokin 2000; Den Tonkelaar 2015; Wolfhagen & De Zanger 2010
90 Posner & Yoon 2010
10. Conclusion

According to our research, it is best to have a female lawyer, at least if one knows for certain that a court case needs to be filed. In addition, the research shows that plaintiffs’ lawyers do not travel very well. This might be a confidence-booster for female lawyers and could be a reason to review the classical notion of the weaker gender.

Mr. X can sleep sound, for now. It is unlikely he is singled out as a losing lawyer based on his years of experience, membership of a specialization organization or office size. However, this research does show to what extent it is possible to automatically generate information about the win-rate of a lawyer from public court records. The research is a first step in unveiling information about the win-rate of lawyers. So, there is a lot of room for future researchers to improve on, such as: the model, data collection method, and explanations.

Prior to undertaking this study, information about the win-rate of lawyers was not available. If the win-rate of individual lawyers becomes publically available, it could drastically change the legal market as clients have a solid benchmark on the performance of their lawyer, which could increase transparency. In addition, NOvA could potentially use this information to monitor lawyers. Another future use could be for recruitment and human research departments. It should be noted, however, that more research is needed on the relationship between win-rate and quality of the lawyer.

In summary, a losing lawyer is not necessarily a bad one, although it can be an indication. Keeping track of the win-rate of lawyers could also change the behavior of lawyers themselves. As a lawyer told the author: “if I had known that someone was keeping track of my win-rate, I would refrain from certain litigation cases.”
11. Literature

Alschuler 1975

Anderson & Heaton 2012

Azmat & Ferrer 2015

Balmer & Pascoe 2014

Blumberg 1979

Cane & Kritzer 2010

Carlson 1976

Chang et al. 2015

Chevalier & Ellison 1999

Cohen 2012

Couwenberg et al. 2009

Daicoff 1996

Daughety & Reinganum 2014

Den Tonkelaar 2015

Domenchich & McFadden 1975
T.A. Domenchich and D. McFadden, ‘Urban Travel Demand’, Amsterdam 1974

Dulleck & Kerschbamer 2001

Eeckel & Grossman 2008

Hensher & Stopher 1979

Kruger & Dunning 1999


Ellickson 1986

Forstenlechner & Lettice 2008

Faccio & Marchica 2016

Gill et al. 2015

Goodman-Delahunty et al 2010
J. Goodman-Delahunty et al., ‘Insightful or wishful: Lawyers' ability to predict case outcomes’ Psychology, Public Policy, and Law, 16(2), 133. (2010)

Hadfield 2000

Haire et al. 1999

Heinz & Laumann 1982

Helland & Tabarrok 2003

Hilton & Naizhen 2009

Hyman et al 2016

Kritzer 1998

Leliveld et al 2013
J.T.C. Leliveld et al, ‘Dit is een advocaat, eindrapport Commissie Wat is een advocaat’, Den Haag, oktober 2013

Lokin 2000

Lossa & Jullien 2012

Macaulay 1963
Merens & Van den Brakel 2014

Morton 1991

Moss 2013

Murayama 1999

NOvA 2015

NOvA 2016
Nederlandse Orde van Advocaten, ‘Regelement keurmerk specialisatieverenigingen 2016

Ota 2014

Posner & Yoon 2010

Premonition 2014
Premonition, ‘Everything you know about lawyer selection is wrong – Big Data Analyzes Litigation’, Miami, 2014

Premonition 2015

Priest & Klein 1984

Rechtspraak 2014
Rechtspraak, ‘Kengetallen gerechten 2014, via rechtspraak.nl

Shultz & Sedeck 2011

Ter Voert & Klein Haarhuis 2015

Tomasson 1991

Wallace & Young 2010

Wolfhagen & De Zanger 2010

Wright & Peeples 2013

Zarzuela 2015
Client Report

Premonition’s Client Report contains in-depth analytics on a company’s litigation history and attorney performance. Law Firms use this report to improve Business Development through better Client Intelligence. It is used by General Counsel to review their Panel Counsel.

- Litigation History
- Case Types
- Legal Representation
- Attorney/Firm performance
- Case Duration
  (overall, case type, judge, client)

Useful for: Law Firms, General Counsel, Claims Managers

Attorney Report

Premonitions’ Attorney Report contains in-depth analytics on an attorney’s performance and litigation experience. Law Firms use this report to gather intelligence on opposing counsel and identify lateral recruiting opportunities. General Counsel use this report to review the performance of their external counsel.

- Litigation Experience
- Case Types
- Client Representation
- Performance (overall, case type, judge, client)
- Case Duration (overall, case type, judge, client)

Useful for: Law Firms, General Counsel, Claims Managers, Litigation Finance, Pre-Settlement Funding
Judge Report
Premonition’s Judge Report contains in-depth analytics on the outcomes of attorneys in front of a judge. It is used by General Counsel to identify the most effective and efficient attorneys in front of a judge. Law Firms use this report for gathering judicial intelligence, hiring co-counsel and identifying the strength of opposing counsel.

- Attorney Performance by case type
- Case Durations by case type and attorney
- Plaintiff/Defendant outcomes by case type
- Attorney Experience in front of a judge
- Performance Outliers

Useful for: Law Firms, General Counsel, Claims Managers, Litigation Finance

Law Firm Report
Premonition’s Law Firm Report provides in-depth analysis of a Law Firm’s performance and efficiency. General Counsel use this during their Panel Counsel Review process to identify top-performing firms. Law Firm’s use this report to track their litigation experience/performance and to improve their business development by pitching to their strengths.

- Litigation History and Experience
- Client Representation
- Overall Firm Performance
- Specific Attorney Performance
- Average Case Durations (overall, case type, client)
- Case Types
- Benchmarking to Competitors

Useful for: Law Firms, General Counsel, Claims Managers

Court Report
Premonition’s Court Report reveals the quantity of cases and performance of all the lawyers in a particular court. This report is used to identify top performing attorneys or identify potential venues in which to file a case.

- Attorney Experience (Number of Cases in a Court)
- Attorney Performance by Case Type
- Benchmarking to peer performance

Useful for: Law Firms, General Counsel, Claims Managers, Litigation Finance

Top Litigants Report
Premonition’s Top Litigants Report ranks companies by volume of litigation in a court, state, or country. Law Firms use this report to identify potential hidden clients. Litigation Finance Firms use this to identify potential leads.

Useful for: Law Firms, General Counsel, Claims Managers, Litigation Finance