

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2020 VOLUME 14 NUMBER 2 (52)

DOI: 10.2478/EP-2020-0017

Trying an Accused Serial Sexual Harasser for Libel in a US Civil Court

Jonathan J. Shuster

Professor Emeritus University of Florida, Gainesville, FL, USA
shusterj@ufl.edu

Mark Handler

Director of Professional Services, Converus Inc., Lehi, UT, USA
polygrapheditor@gmail.com

Key words: Accuracy Detection, Barnard Test, MeToo#, Polygraph, Sexual Harassment

Disclosure

Mr. Handler is a fulltime employee of Converus Inc. This work was entirely self-funded. Dr. Shuster does all his consultant work pro-bono and selected Converus Inc. based on its merits. Only after making the choice, did he meet Mr. Handler. While Mr. Handler works for Converus Inc., he has no conflict of interest, since we expect that cases will virtually always be settled out-of-court, meaning that most if not all accuracy detection tests will not actually be done. In 2019, Converus Inc. did over 88,000 EyeDetect tests. Incrementally, serial harassment cases will have zero measurable impact on the business volume of the company.

Abstract

The goal of this article is to provide a class of MeToo# victims of a high-profile serial sexual harasser with a non-invasive method for civil action, when the accused publicly dismisses the victims' claims as lies. When these libelous claims do occur, the victims can be assembled into a class-action libel/defamation case, which in most US states must be mounted within two years of the claim. Because under current civil methods, the plaintiffs would be subject to intense cross-examination in a civil jury trial, class-action lawsuits with small numbers of plaintiffs (e.g. 5–8) have proven impossible to conduct. This article provides a blueprint to create a collaboration amongst the victims, credibility-assessment (lie-detector) experts, statisticians, and MeToo# attorneys to litigate libel suits, which will likely produce out-of-court settlements. Once the first case is successfully completed, precedent will be set to bring other perpetrators to justice, and act as a deterrent to future exploitation. The evidentiary basis would be based on testing the null hypothesis that all plaintiffs are lying, to compare the inferred lying rates of the plaintiffs to similar population controls, who would be known liars, to a “Yes” answer to “Did X sexually harass you?”

Introduction

According to the website <https://www.npr.org/sections/thetwo-way/2018/02/21/587671849/a-new-survey-finds-eighty-percent-of-women-have-experienced-sexual-harassment>, about 80% of adult American women reported being a victim of sexual harassment. There are two implications of this information. First, this implies virtually every American has either been a victim or is close to a victim. Second, these numbers also imply that a substantial number of their perpetrators have committed these acts against multiple women, making them serial harassers. Mounting a criminal case against serial violators is all but impossible because they hire top notch attorneys, who put the victims on trial about their personal lives, making it rare for victims to step forward. Statutes of limitation (usually two years) adds another layer of difficulty for the victims. Very few perpetrators have ever been successfully brought to justice in the US criminal courts. Although reports in other countries, notably Australia (64%) Italy (44%), and Sweden (64%) are lower than the US 80% of adult women reporting sexual harassment, it is clear that we have a world-wide problem.

The central theme of this article is to propose a legal strategy to sue a high-profile serial sexual harasser for libel even though statutes of limitation for criminal prosecution have expired. While this paper deals with US state and federal civil courts, the concepts may well have applicability to other democracies with civil courts that have pretrial meeting requirements in advance of trying the case. At the time of this writing, several US states are attempting to remove such limitations. Such removal may or may not allow grandparenting in statutes of limitations that had already expired under the old rules.

We shall delve into the implications of this in the discussion, but our strategy would only be enhanced by such state rulings, which ultimately might have to be adjudicated in the US Supreme Court.

There are three legal levels of evidence in US law, all founded upon the principle of “Proof by Contradiction”, taught to all law students in logic classes. First, “Beyond a Reasonable Doubt” applies only to criminal cases. Second, “Preponderance of Evidence” applies to inconsequential civil cases involving purely financial disputes, where the jury must decide whether the defendant is more likely than not to have been responsible for some damage to the plaintiff. Finally, “Clear and Convincing Evidence” should apply to civil cases where there are major consequences to both sides in the verdict. Cases such as attribution of harm for drug side-effects or environmental health effects from industry should fall into this designation. Here, we argue that testing the “null hypothesis” that the defendant caused no harm at “ $P < 0.05$ ” (taking multiple jeopardy in time and dimension into account), makes the overall system equitable and uses reproducible criteria, which have thousands of precedents behind them. What the contradiction represents starts out by assuming the defendant is not responsible for damages and then using the evidence to infer whether they rise to the level needed to contradict (i.e. reject) this assumption. Clearly, we should use “Clear and Convincing Evidence” to adjudicate a serial harassment libel case. If in the actual study we reject the null hypothesis at $P < 0.05$, it means that if we repeated the experiment in a new population where the null hypothesis is actually true (all plaintiffs are lying), the probability of incorrectly rejecting this null hypothesis is at most 5% (Clear and Convincing Evidence).

In the next sections, we shall present our approach to suing an alleged serial sexual harasser for libel, at a time when the statutes of limitations for criminal prosecution has ended, but a verbal accusation of lying against a set of victims has been made. This can be applied to alleged pedophile priests or high-profile individuals, such as politicians or company executives, accused many years ago of multiple acts of misconduct. A libel class action by plaintiffs against the harasser is currently extremely difficult to pursue, because any trial will subject the complainants to brutal cross-examination, and this makes holding the class together virtually impossible. Real examples of terrifying experiences for those who accuse a powerful individual in non-civil venues include the following: Andrea Constand (Cosby prosecution), Karen Borel (Ghomeshi prosecution), Jessica Mann (Weinstein prosecution), Anita Hill (Thomas US supreme court nomination), and Christine Blasey Ford (Kavanaugh US supreme court nomination). For this reason, cases involving a small number of plaintiffs (5–8) have not yet been tried in a US civil court. Our approach, which will rely on credibility assessment testing of the alleged victims (plaintiffs) and a set of population-based controls, could have an

excellent chance of reaching an out-of-court settlement, thereby sparing the plaintiffs from cross-examination or even from undergoing the accuracy assessment. Jury trials are very risky for defendants in the US, and therefore there is motivation to settle once the defendant's legal team is made aware of the potential scientific investigation.

It is important to note that once the defendant calls the accusations lies, there are three key elements at play: (1) If a plaintiff's honest perception is that s/he told the truth, the defendant's allegation that s/he lied is false; (2) Because the issue at hand deals strictly with the plaintiffs' truthfulness, any credibility assessment test taken by the defendant is irrelevant; and (3) The class of plaintiffs has only to provide evidence that at least one of the plaintiffs was truthful to be entitled to a settlement. However, if the class contains non-harassed plaintiffs, the ability to reach the required level of evidence would be compromised. This approach is analogous to cases where a class of heart attack victims sues a drug maker for excess occurrences over that of a control medication. Whether or not a given individual's attack can be directly attributed to the drug, if the overall level of evidence is sufficient (usually at $P < 0.05$), and there is other supportive mechanistic evidence, the entire class can get a settlement.

Based on an actual potential pro-bono case, the first author investigated a partnership with a credibility assessment (aka lie detector) provider. In this search, he was only able to find one company, Converus Inc. that had published experimental validity data on truthfulness and falseness, essential elements to our legal strategy, making it the sole source at present. These "EyeDetect" assessment tests, as described in Table 4 of Kircher & Raskin (2016), using the five-fold method, are estimated to detect truthful responses 88% of the time and deceptive responses 86% of the time. EyeDetect has the added benefit of being totally non-invasive in that while the actual test questions are being administered, there is no operator in the room with the subject. Two videos showing how this works can be found at <http://converus.com/> under the EyeDetect link. In the future, other potential products may serve the same role as Converus Inc. The only statistical impact company selection would have rests with the "power calculation", i.e. the likelihood of finding Clear and Convincing evidence when all plaintiffs are truthful, since this depends on the validity numbers. Table 1 below maps all possible outcomes into whether we have Clear and Convincing evidence vs. not. The definition as to what is Clear and Convincing evidence would be identical under any Converus competitor. Finally, a proprietary objective algorithm determines the truthfulness inference without human intervention by whether a credibility score is at least 50 (Kircher & Raskin, 2016).

The paper is organized as follows: Study Design: The study design to compare the alleged victims to a similar control group for truthfulness to the key question on harass-

ment is described; Control Selection: How controls might be selected to potentially compare outcomes to the plaintiffs. How the Case might Proceed: This covers how we envision litigation of the US civil case will proceed; and Discussion: This covers the implications and how other countries might find the recommended approach useful.

Study Design

The basic design is a close analog of a “case-control” study, which is a widely used study design in cancer research (Breslow & Day, 1986). Consider a class action civil lawsuit with a set number of plaintiffs and a set number of like controls. As we shall discuss later, we advise against direct matching, but collectively, the controls are selected to be demographically similar to the plaintiffs and have never been in contact with the defendant. We know that controls will not be truthful to a “Yes” answer to the question, “Did X sexually harass you?”. All subjects (Plaintiffs and Controls) will be instructed to answer “Yes” to this question. Ideally, the controls and plaintiffs will be tested, blinded to the device operator by group, and randomly mixed in terms of order. Individual test results should not be released, even to the subjects. For those granting consent, anonymized audio recordings could be made for the court. If, and only if a sufficiently higher percentage of plaintiffs are inferred as truthful than controls to the question, would we infer that this aspect of the case is considered meeting the criterion of Clear and Convincing evidence. The inference from such a set of subject outcomes is that under the null hypothesis that all plaintiffs and controls are lying, the false positive rate (inferring that the subject was truthful when in fact s/he was lying) is higher in the plaintiffs than that in a similar general population of subjects. See the next subsection for how controls might be selected. Of course, this need not be the only evidence presented in the case.

In short, good design principles should include (1) randomizing the order of plaintiffs and controls and (2) blinding the operator as to group identity. If indeed the credibility assessment study is done, an agent of the court should monitor the process.

Control Selection

We recommend that state voter rolls be used to select controls. Since these are used in most states for jury selection, the courts are very familiar with them. A large random frame should be selected from the state voter roll and the sample should be roughly frequency-matched on demographic characteristics such as age, gender, city or county, and ethnicity.

Here is a numerical example of how, before any data are collected, we can set up objective criteria to make an inference as to whether the libel claim is considered supported by the data. In this example, we presume there are Six plaintiffs and 13 controls. The six plaintiffs is the actual number of plaintiffs in the potential trial we communicated on with a MeToo# attorney, and 13 controls is the smallest number of controls that make the probability of finding Clear and Convincing evidence per Table 1 at least 95% (our desired level), when all plaintiffs are actually truthful. If four of six plaintiffs are inferred as truthful, then it takes three or fewer controls inferred as truthful to reach clear and convincing evidence. Of course, professional statisticians, guided by the methods and referenced software below can construct like tables for other sample size scenarios, beyond the 5–8 plaintiffs covered in Table 1 below.

Because of the small numbers of subjects, exact statistical methods (no large sample approximations) are required. We employed the Barnard Test (1945), which is based on exact binomial calculations. Suissa and Shuster (1985) showed that this test is much more powerful than the commonly used Fisher’s Exact Conditional Test. Under the null hypothesis that all plaintiffs are lying, there is at most a 5% chance of a finding in the “Clear and Convincing” range, irrespective of the common equal false positive rates in the two groups. The calculations can readily be performed using the commercial software package StatXact, available from <https://www.cytel.com/software/statxact>.

Table 1: Clear and Convincing Evidence for Recommended Designs

Plaintiffs ¹ (N=5)	Controls ¹ (N=19)	
2	≤1	
3	≤3	
4	≤5	
5	≤9	
Total (Z>2.03)		95% power

Plaintiffs ¹ (N=6)	Controls ¹ (N=13)	
2	0	
3	≤1	
4	≤3	
5	≤4	
6	≤7	
Total (Z>1.83)		96% power

Plaintiffs ¹ (N=7)	Controls ¹ (N=12)	
2		
3	0	
4	≤2	
5	≤3	
6	≤5	
7	≤7	
Total (Z>1.83)		97% power

Plaintiffs ¹ (N=8)	Controls ¹ (N=9))	
3	0	
4	0	
5	≤1	
6	≤2	
7	≤4	
8	≤5	
Total (Z>1.83)		95% power

¹ Number of Plaintiffs or Number of Controls inferred as truthful

Z is the Z-statistic with pooled variance per Suissa and Shuster (1985)

If all plaintiffs are truthful, with all controls known liars to yes answers to the implied question, then under Converus accuracy estimates from Kircher and Raskin (2016) of 88% truthfulness and 86% deception, there is at least a 95% probability (Power) of a finding in the set of Clear and Convincing outcomes defined in Table 1 for the various designs we listed. The motivation for the number of controls in each table is that this represents the smallest number need to achieve the desired 95% power for rejection of the null hypothesis when all plaintiffs are truthful.

How the case might proceed

Pretrial Meeting and Motions

Before any trial is conducted, a statistician and credibility assessment expert would set up the case-control design, including the number of plaintiffs and number of controls, along with exact criteria to establish Clear and Convincing evidence, producing an objective Table, analogous to Table 1. Along with a description of the testing methods,

these will be known to both sides prior to the meeting. A power analysis will also be provided to both sides. The two sides will meet with the judge/attorneys to lay out the ground rules. At least in the very first libel case, the defense will try to argue that the accuracy testing is inadmissible. It needs to be noted that because there is never an inference that any specific subject is truthful, this challenge will likely be unsuccessful. In fact, credibility assessment testing has been allowed. Even in a minority of cases where a single polygraph has been used, polygraphs have been pivotal in verdicts. For example, in *State v. Dorsey* (1975) 88 N.M. 184 [539 P.2d 204] the court reversed a [53 Cal. App. 3d 115] criminal conviction that had held polygraph evidence to be inadmissible in US court. Next, the defense might try to make the following argument: “In the event that the plaintiffs demonstrate a higher false positive rate under the null, it can be explained by the possibility that they are just better liars”. The presentation of Honts & Thurber (2019) looked at moderators of validity in a large meta-analysis and were unable to uncover any significant ones with meaningful effect sizes. Remember that no tests have yet been conducted. At this point, the defense seems to be grasping at two straws, and the judge seems unlikely to buy in, given the published validity and error rates for these credibility assessment tests per Kircher & Raskin (2016). Since visually, the bar for Clear and Convincing Evidence in Table 1 does not seem very high, a settlement is now extremely likely, and the judge might ask both sides to negotiate one. A civil trial represents a huge financial risk to the defendant, as jury awards are typically very generous to winning plaintiffs. Should indeed the credibility assessments be done, there is no going back for either side. Many true victims want acknowledgement of harassment on the part of the defendant as opposed to a large monetary award.

Actual Trial by Jury

If an out-of-court settlement is not reached, and if admissibility of the accuracy assessment for collective inference is established, the case-control study would proceed as designed in the previous sections. The court would supervise the selection of controls and assure blinding of the defense and plaintiffs, except for a final summary table of outcomes. A neutral statistician should do the analysis which may include secondary inference, such as comparing the estimated probabilities of untruthfulness, since most credibility assessment tests not only provide a yes/no answer but also an estimated probability of truthfulness for each subject.

Discussion

We have provided methodology that might be applied to help MeToo# victims litigate their case against a sexual predator in US civil court, thereby avoiding the daunting ex-

pectation of being cross-examined on every aspect of their lives. This strategy may also motivate civil litigation in countries where the cases are tried before judges. Credibility assessment methodology is certainly imperfect, but even the most severe skeptic may consider that we will tend to see a higher inferred truthful rate (Yes answer) among truly self-perceived harassed women than in control women to the question: "Did X sexually harass you?"

The importance of having a control group cannot be overemphasized. Without one, we would have to rely on validity numbers for truthfulness of the accuracy assessment methods. Although they were extensively tested, they may or may not apply in the harassment context. The only place where validity figures are used is in the power analysis. The study's objective rejection region, established prior to the data collection, has its P-value calculated under equal target population proportions of truthful inferences to the "Yes" answer to the question, "Did X sexually harass you?". This would logically be true under the null hypothesis that all plaintiffs were lying to this question. To illustrate what can happen without a control group, suppose that four of six plaintiffs were inferred to be truthful. Assuming they were all in fact liars, all you could say is that the test validity for detecting falseness ranged from 22% to 96% with exact 95% confidence. This would be fodder to the defense.

There are two notes of caution we need to consider. First, there is no relevance to a defense's credibility assessment of the accused libeler. The defendant has accused all plaintiffs of lying, not whether the accused's honest perception is that s/he did not sexually harass anyone. Second, the plaintiffs' attorneys must avoid cherry-picking and should therefore refrain from selecting the set of plaintiffs on the basis of pretrial accuracy assessment tests, which would be discoverable by the defense in any case.

Ironically, the availability of a powerful tool, that can scientifically evaluate the evidence, may suggest that it will virtually never be applied. If the case goes to trial and if there is a finding for the plaintiff side, the award will be in the hands of a jury of the defendant's peers. Jury awards are generally much larger than out-of-court settlements. Rather than "roll the dice" (in our numerical examples the power is 95%), the case would almost certainly be settled out-of-court.

There are limitations to this approach. First, will there be major non-participation rates for controls selected from the voter role?. We believe that society is well in-tune with the MeToo# movement and the potential controls will be sympathetic to the investigation. Second, the plaintiffs may be reluctant to undergo a credibility assessment test. Considering an alternative that involves cross-examination, we believe plaintiffs will be highly motivated to do so. If they understand that the case is about the collective responses of the class of plaintiffs, and not about their individual responses, this should

be at most a minor issue. Third, the independent binomial distribution requires independence of the observations, which is supported by our design which can be enhanced by requiring no crosstalk among the plaintiffs or controls. Most often, the plaintiffs are strangers to each other, something that also would support independence. However, in situations where a company executive is the defendant, one might be suspicious of the independence. For this reason, we strongly recommend that attorneys meet strictly one-on-one with the plaintiffs. Yet, the key outcome is the inferred error rate of the test, and the design assures these are operationally independent. In short, lack of independence should not be a major issue even when the defendant is a company executive.

Credibility assessment tests also can provide estimates for the probability that the subject is truthful given their results. We considered, but rejected, the use of something akin to a randomization t-test to assess the overall outcome. This approach might gain somewhat in power but loses the ability to completely lay out in advance what is concluded from every possible overall outcome. This t-approach might be viable when the number of plaintiffs in the class is small (such as 3 or 4). Power analysis would require actual data from past validity studies on the distributions of the probability of truthfulness from both known truthful and lying subjects.

We also strongly advise against direct matching. This will lead to confusion at the pre-trial meeting. In the analogy to our example, with six plaintiffs and 3 matched controls directly matched to each plaintiff (18 total controls), the data would be laid as a two by four layout for each subject. Instead of one dependent variable in the unmatched (inferred truthful vs. not), the data would have eight counts for the number of positives (0 or 1 for inferred truthful Plaintiff) each matched to 0,1,2,3 inferred truthful Controls (See Table 2). Tabulating which potential outcomes represent Clear and Convincing evidence is much more complex than Table 1. Further, we do not expect a major power advantage, especially if the matching criteria are not a major factor for predicting confidence of the accuracy detection inference.

Table 2: Hypothetical Matched Design Outcomes for 3:1 Matching.

Controls + → Plaintiffs + ↓	0	1	2	3
0				
1				

Entries are number of Plaintiffs deemed lying (0) or deemed truthful (1) vs. Number of the three matched controls deemed truthful (0, 1, 2, or 3). The total counts in occurrences would be 6, one for each matched set.

Note that unlike side-effects civil lawsuits, there is no ability of serial harassment cases to cherry pick the outcome on the basis of dimension (selecting just one outcome from a collection of several, such as cancer, heart disease, liver damage, etc.), or time (repeatedly looking at the data and pouncing when they become significant).

A very encouraging new development is that several states are pushing to eliminate statutes of limitations for sexual misconduct in criminal cases. In all likelihood, this would not affect situations where the statutes of limitations could be grandparented in. But once this capability gets through the court systems, victims will have more options to sue in criminal court or civil court. The lower evidence requirement in civil could still make our approach attractive, as it is unlikely that the use of accuracy detection would be permitted in criminal actions.

Potential Applications in Other Counties

First, Converus Inc is now available in over 40 major languages. Second, some countries have lower standards for criminal convictions than the US beyond a reasonable doubt standard, making it possible in some countries to prosecute serial harassers under the nation's criminal code. Third, in civil cases in other countries, so long as there are pre-trial meetings for mediation purposes, the procedures can likely be adopted to gain a negotiated settlement, which gives both sides a potential benefit. A guilty party might be able to offer a lower financial settlement than would be the case if it went to trial. The victims would get public recognition of wrongdoing by their perpetrators without having to undergo cross examination of their personal lives.

References

- Barnard, G.A. (1945), A new test for 2×2 tables. *Nature*, 156, 177–177.
- Breslow, N.E., & Day N.E. (1986), *Statistical Methods in Cancer Research*. IARC Scientific Publication, France: Lyon.
- Honts, C.R., & Thurber, S. (2019), *A Comprehensive Meta-Analysis of the Comparison Question Polygraph Test*. Paper presented at the annual meeting of the American Psychology Law Society, Portland, Oregon.
- Kircher, J.C., & Raskin D.C. (2016), Laboratory and Field Research on the Ocular-Motor Deception Test. *European Polygraph*, 10, 159–172.
- Suissa, S., & Shuster, J.J. (1985), Exact Unconditional Sample Sizes for the 2×2 Binomial Trial. *Journal of the Royal Statistical Society*, Ser. A, 148, 317–327.

