Published in final edited form as:

AJOB Prim Res. 2013 ; 4(3): 39–48. doi:10.1080/21507716.2013.807892.

Understanding the Severity of Wrongdoing in Health Care Delivery and Research: Lessons Learned From a Historiometric Study of 100 Cases

James M. DuBois, PhD, DSc,

Albert Gnaegi Center for Health Care Ethics, Saint Louis University, 3545 Lafayette Ave, Salus Building, St. Louis, MO 63104

Emily E. Anderson, PhD, and

Loyola University of Chicago

John T. Chibnall, PhD

Saint Louis University

James M. DuBois: duboisjm@slu.edu

Abstract

Background—Wrongdoing among physicians and researchers causes myriad problems for patients and research participants. While many articles have been published on professional wrongdoing, our literature review found no studies that examined the rich contextual details of large sets of historical cases of wrongdoing.

Methods—We examined 100 cases of wrongdoing in healthcare delivery and research using historiometric methods, which involve the statistical description and analysis of coded historical narratives. We used maximum variation, criterion-based sampling to identify cases involving 29 kinds of wrongdoing contained in a taxonomy of wrongdoing developed for the project. We coded the presence of a variety of environmental and wrongdoer variables and rated the severity of wrongdoing found in each case. This approach enabled us to (a) produce rich descriptions of variables characterizing cases; (b) identify factors influencing the severity of wrongdoing; and (c) test the hypothesis that professional wrongdoing is a unified, relatively homogenous phenomenon such as “organizational deviance.”

Results—Some variables were consistently found across cases (e.g., wrongdoers were male and cases lasted more than 2 years), and some variables were consistently absent across cases (e.g., cases did not involve wrongdoers who were mistreated by institutions or penalized for doing what is right). However, we also found that some variables associated with wrongdoing in research (such as ambiguous legal and ethical norms) differ from those associated with wrongdoing in healthcare delivery (such as wrongdoers with a significant history of professional misbehavior).

Conclusions—Earlier intervention from colleagues might help prevent the pattern we observed of repeated wrongdoing across multiple years. While some variables characterize the vast majority of highly publicized cases of wrongdoing in healthcare delivery and research—regardless of the

kind of wrongdoing—it is important to examine and compare sets of relatively homogenous cases in order to identify factors associated with wrongdoing.

Keywords

Professionalism; medical ethics; research ethics; misconduct; organizational deviance; professional misbehavior

INTRODUCTION

Professional wrongdoing causes myriad problems for society. Wrongdoing by healthcare professionals harms patients (Carr 2003; Hall et al. 2008), wastes scarce dollars (Aldrich 2009; Iglehart 2009), and may contribute to healthcare disparities (Boulware et al. 2002; Williams and Mohammed 2009). Wrongdoing among researchers harms participants (DuBois 2008), pollutes the scientific literature (Titus, Wells, and Rhoades 2008), and damages public trust (Coughlin, Barker, and Dawson 2012). While many articles have been published on professional wrongdoing, our literature review found no studies that examined the rich contextual details of large sets of historical cases of wrongdoing (DuBois et al. 2012).

In this article, we describe our approach to understanding professional wrongdoing in healthcare delivery and research, an approach that has evolved over the past four years as we gathered and analyzed preliminary data. We share findings from Phase I of our study, which involved the analysis of 100 cases of wrongdoing, and we examine a variety of factors that led us to alter fundamentally our methodology in Phase II of the study (which is currently under way). In the introductory section we provide information that is essential to understanding our Methods section, including our general approach, our guiding assumption about professional misbehavior, why we focused on the severity (or seriousness) of wrongdoing, and how we identified predictor variables to track in cases of wrongdoing.

Historiometry: A Novel Approach to Understanding Wrongdoing

While wrongdoing has been studied in many different ways, we adopted a historiometric approach. Historiometry combines qualitative and quantitative methods to study variables that either for logistical or ethical reasons cannot be investigated using prospective or experimental methods. While we have reviewed many studies that use surveys or experimental procedures to understand wrongdoing (DuBois et al. 2012), examination of actual cases provides rich, ecologically valid information that would be difficult to otherwise obtain because it would be unethical to try to induce professional wrongdoing in the actual care of patients or enrollment of research participants.

Historiometric research involves the statistical description and analysis of a group of coded historical narratives. While the nature of the narratives and statistical methods used has varied widely, psychologists have used such methods to understand a broad array of individual and social phenomena, including the decision-making of Supreme Court justices; the leadership styles of successful presidents; the personality traits of destructive charismatic leaders as well as creative geniuses; and the influence of birth order, parental loss, and role

models on human development (Deluga 1997; Mumford 2006; Simonton 2003, 1999; Suedfeld and Bluck 1988). In our Methods section, we explain our particular application of historiometry.

A Focus on “Ethical Disasters”

One obvious limitation to any approach that draws from published accounts of wrongdoing is that the method restricts the kinds of cases that can be sampled; only cases that are fairly serious are likely to be published and described in sufficient detail to enable analysis. While this is generally true, we believe that cases of serious professional wrongdoing—such as data fabrication (Fanelli 2009) and sexual abuse of patients (Carr 2003)—deserve careful study even though they are less common than less serious violations of professional standards such as inappropriately assigning authorship (Martinson, Anderson, and de Vries 2005). This is similar to maintaining that first-degree murders (which are commonly reported in newspapers) deserve study even though petty thefts (which are not commonly reported) are more pervasive in society and cost consumers billions of dollars each year. Moreover, highly publicized cases of professional wrongdoing are presumably the cases that most damage public trust, insofar as they are the cases best known to the general public. We call highly publicized cases of serious professional wrongdoing “ethical disasters” and have deliberately chosen to focus upon such cases. We hope that by identifying factors associated with wrongdoing, we may enhance educational efforts aimed at preventing wrongdoing and facilitate and encourage early intervention in cases of wrongdoing so as to reduce repeat infractions.

A Guiding Assumption in Phase I

The psychological literature commonly treats professional wrongdoing as a meaningful and coherent construct called “organizational deviance” or “professional misbehavior” (Bolin and Heatherly 2001; Sims 2009; Vardi and Weitz 2004; Victor, Trevino, and Shapiro 1993). This position maintains that a wide variety of different kinds of wrongdoing—for example, data fabrication, privacy violations, and abuse of prescribing privileges—might be understood as different forms of the same underlying phenomenon (professional misbehavior) and might be predicted by the same individual and environmental variables. If this assumption is correct, then sampling should be heterogeneous, representing the full range of misbehaviors in our domains of interest (health research and healthcare delivery).

While we adopted this assumption (or working hypothesis) in Phase I of our study, we also tracked carefully the different kinds of wrongdoing and the different domains in which they occurred (specifically health research and healthcare delivery), enabling us to determine whether in fact “professional misbehavior” is a uniform phenomenon.

Developing a Taxonomy to Guide Heterogeneous Sampling

To facilitate heterogeneous sampling, we searched for taxonomies of wrongdoing in health research and healthcare delivery. We needed a taxonomy that was comprehensive, focused on noncontroversial forms of wrongdoing that are punished, and sufficiently clear and specific to enable high levels of inter-rater reliability. We did not find one that met these criteria.

We thus developed a taxonomy by conducting a review of academic literature, ethics codes, regulations, and cases (DuBois, Kraus, and Vasher 2012). We additionally consulted with experts in healthcare ethics and health law and applied our taxonomy to a preliminary set of cases. In the end, we identified 14 categories of wrongdoing in healthcare delivery (including violations of consent, privacy, and professional boundaries) and 15 categories in medical research (including inappropriate management of risk, conflict of interest violations, and research misconduct). In applying the taxonomy to 50 excerpts from our larger database of 300 cases, our three raters obtained a free-marginal multi-rater Kappa coefficient of .85.

Identifying Variables to Extract from Historical Cases

Our next step was to determine what sorts of variables should be extracted when examining historical cases of wrongdoing. Although the literature suggests that some individual personality traits such as cynicism and narcissism may correspond with wrongdoing (Antes et al. 2007; Bandura, Underwood, and Fromson 1975; Judge, LePine, and Rich 2006; Roback et al. 2007), we quickly discovered that published accounts of professional wrongdoing focus more on characteristics of the environment than personality characteristics of the wrongdoer.

Therefore, we conducted a thorough review of environmental factors associated with professional wrongdoing (DuBois et al. 2012). We identified ten variables that data suggest play a contributing role in wrongdoing by providing means, motive, or opportunity (MMO): conflicting roles (Grover 1993; Levine 1992); wrongdoing financially rewarded (Jennings 2006; Rodwin 1993); others benefitting from wrongdoing (Victor, Trevino, and Shapiro 1993); protagonist (wrongdoer) being penalized for right behavior (Hegarty and Sims H 1978; Jansen and Von Glinow 1985); others being penalized for right behavior (Schuchman 2008); the mistreatment of the protagonist or perceived organizational injustice (Greenberg 1993; Keith-Spiegel and Koocher 2005; Martinson et al. 2006); ambiguous professional norms (Davis 2003; DuBois and Dueker 2009; Meyer and Jr. 2002; Shah et al. 2004; Simonton 2003); particularly vulnerable victims (Bandura, Underwood, and Fromson 1975; Zimbardo 2007); oversight failures (Bramstedt and Kassimatis 2004; Marshall 1999); and the protagonist being in a position of authority over co-workers (Milgram 1965; Trevino, Butterfield, and McCabe 1998).

Despite some promising work on the association of some social dynamics such as “pressure to publish” or “competition” with wrongdoing in research (Anderson et al. 2007; Martinson et al. 2009), we did not include such variables in our study. This decision was based on our preliminary research (and experience) that indicated that these variables are ubiquitous and thus of little predictive value when taken separately from the individual’s subjective experience of pressure and coping abilities (with the latter variables being inappropriate for historiometric methods).

For each environmental variable, we were able to identify at least one case that illustrated its presence (and from a commonsense, intuitive perspective, illustrated its causal role within the MMO framework).

Why We Focused on the Severity of Wrongdoing

While a historiometric approach to understanding and predicting professional wrongdoing has some advantages over experimental methods (e.g., strong ecological validity and feasibility), it also presents some serious methodological challenges. It is a research maxim that causality cannot be inferred from statistical association; but detailed description of cases does not even provide evidence of statistical association. Statistical significance testing requires either the comparison of groups or comparison of cases using factors that vary within the group. When studying professional wrongdoing, comparison cases are difficult to find; no one publishes detailed accounts of business as usual. For this reason, we decided to focus on the *severity of wrongdoing* as a continuous variable.

We believed we could create a sample of cases representing broad variation in the severity of wrongdoing if we included in our study not only the “ethical disasters” that were of primary interest (such as performing unnecessary spinal surgeries or enrolling inappropriate patients into a high risk clinical trial), but also cases involving relatively minor compliance failures and false accusations.

Using such an approach, we asked the following research question: *Which variables are correlated with the severity of wrongdoing in healthcare delivery and research?*

In this article, we present preliminary data that address this question. More importantly, we indicate why historiometric research on wrongdoing is better suited to explore different research questions. We also explain some of the key methodological lessons we have learned and present our modified approach to conducting historiometric research on wrongdoing.

METHODS

Our method involved identifying appropriate cases, producing a standardized written synopsis of cases, extracting and rating the presence of key variables, rating the severity of wrongdoing using several scales, and examining the relationship between the severity of wrongdoing with variables of interest.

Sampling

We used criterion-based heterogeneity sampling. Criterion sampling required that all cases meet a series of inclusion criteria. Cases were required to have:

1. Involved a credible allegation of professional wrongdoing
2. Occurred in the context of healthcare delivery or research
3. Occurred in the US
4. Occurred after 1900
5. Been described in at least 3 independent published accounts that include a good description of a well-defined environment

6. One key protagonist whose behavior could be predicted. (Note: We use the neutral term “protagonist” to refer to the alleged wrongdoer.)

Cases were identified through literature searches in a wide variety of databases (including PubMed, Lexis-Nexis, and Google Scholar) and examination of the reports of oversight bodies (such as the Office of Research Integrity and state medical boards). To ensure a heterogeneous sample, we investigated at least one case from each of the 29 taxonomy categories. We completed 37 cases involving medical practice (funded by a BF Charitable Foundation pilot grant), and 63 cases involving health research (funded by an NIH R21). Because we aimed to predict the severity of wrongdoing, we also attempted to sample cases representing a wide range of wrongdoing, including cases in which protagonists were found guilty, as well as allegations of wrongdoing that were covered by the press or court records but were eventually found to be unsubstantiated.

Case Research and Production of a Case Synopsis

To enable separate teams to rate the presence of predictor variables and the severity of wrongdoing, we produced a synopsis of each case. This ensured that all raters worked with standard information about each case.

A 10-page procedures manual guided case research and synopsis writing. A research assistant (RA) conducted a thorough literature review for each case using Lexis-Nexis Law (which covers legal cases, newspapers, and scholarly literature), MedLine, PsychInfo, Google, GoogleScholar, Firstgov, and other databases as needed. Our source materials included: newspaper articles by investigative journalists, proceedings of medical boards, findings of investigating bodies such as the Federal Bureau of Investigation (FBI) or the Office of Research Integrity, court proceedings, and occasionally books and other written sources. An average of 33 sources was obtained and read per case (range = 3–165, standard deviation [SD]=29). The variation in the estimated total number of words found in all sources consulted was not correlated with the severity rating of cases.

The mean length of a case synopsis was 1273 words (range=409–3026, SD=561). The variation in length of synopsis was not correlated with expert evaluations in severity ratings. Synopses included all available information pertinent to the variables that we tracked, including: key facts about the protagonist including name, age, degrees, and profession; facts about other main characters including colleagues, collaborators, and whistleblowers; a description of the victim population when applicable; a general description of the professional setting; a description of the event (the wrongdoing), including a description of the outcomes, consequences, and how the wrongdoing was identified and stopped; details on the environmental predictor variables (such as oversight failures, conflicting roles, or financial rewards for wrongdoing); and information relevant to assessing the protagonist’s history of wrongdoing.

The case writing process was iterative and used a team-based approach. Once the RA completed a draft of the case it was sent to a Case Editor, who read at least 3 primary sources on the case, conducted an abbreviated literature review, ensured that sufficient detail was included to enable rating of all environmental and severity variables, and then edited the

case to ensure that it conformed to the guidelines provided in the procedures manual. The RA and Case Editor worked together to revise and finalize each case.

Data Extraction

RAs completed a Case Datasheet with more than 70 different data points, including the years the wrongdoing occurred, an estimation of the number of words consulted in the literature review, the field of the wrongdoer, and the kinds of wrongdoing exemplified in the case. The datasheet additionally contained two scales. The Consequences to the Protagonist scale inquired into whether the protagonist suffered any of the following consequences:

- Publication of credible accusation of wrongdoing during their active career¹
- Loss of job, professional opportunities, funding, or reimbursement eligibility
- Financial penalties (settlements, restitution, fines)
- Prison, criminal probation, house arrest, community service
- Loss of licensure or credentialing (applied only to the domain of medicine)

On the assumption that one was more severely punished if more of these events occurred, we created a Consequences scale of 0 – 4 points in the domain of research and a 0 – 5 points in the domain of healthcare practice. Creation of a scale also allowed us to test the correlation of the Severity of wrongdoing to Consequences.

Second, the datasheet contained a History of Wrongdoing scale, which ranged from 0 – 4 points with 1 point awarded for each of the following:

- Committed multiple kinds of wrongdoing
- Repeated the wrongdoing
- Engaged in wrongdoing in multiple professional environments
- Convicted of a felony crime outside of work setting

Although our methodology did not allow us to capture individual personality traits (such as narcissism or cynicism), this scale allowed us to take into account individual behavioral variation across environments. Scores in the 3 – 4 range indicate that the individual engaged in wrongdoing in a variety of environments, suggesting a stronger individual component.

Coding the Environmental Variables

An Environmental Factors Score Sheet was developed based upon (1) an extensive review of the psychological and social science literature on factors that correlate with or predict professional wrongdoing (DuBois et al. 2012) and (2) qualitative analysis of the first 30 cases by the research team to identify any additional environmental variables that might predict the wrongdoing. It was modeled upon “benchmark” scoring guides used in other historiometric studies (O’Connor et al. 1995). The Case Editor and 2 RAs independently

¹All cases were eventually published, but we did not count the publication as a negative career consequence if publication occurred after retirement or death.

rated the presence of 10 environmental variables in the case (e.g., conflicting roles, oversight failures, or financial rewards for wrongdoing). At least monthly, the 3 raters held a scoring calibration meeting by phone to review their scores and discuss widely discrepant scores. However, raters were not allowed to change their scores unless they overlooked something or a clarification was made to the scoring guide. Raters followed a 9-page scoring guide. An excerpt from the scoring guide is provided in Table 2.

For the raters on all 10 environmental variables, intraclass correlation coefficients, which measure reliability, ranged from .84 to 1.0 (no variance). These scores are extremely high. They are due to extensive training, regular scoring calibration meetings, refinements of the instrument, and the simple 3-point scale.

Rating the Severity of the Wrongdoing

A team of five expert consultants (including a social psychologist, two philosopher-bioethicists, and two physician-bioethicists)—all of who have experience on institutional review boards (IRBs) and ethics committees—read each case and then completed a 6-item score sheet (see Table 1). Three items evaluated the Overall Severity of the wrongdoing, and three items evaluated the degree of Mistreatment of Clients (human patients or research participants). These two scales were used separately in statistical analyses; the Mistreatment scale was only used in cases involving human patients or research participants ($n=64$). Raters followed a scoring guide that was developed by the principal investigator (PI) and revised after pilot testing the measures with the team of raters. To reduce the possibility of bias, the team of severity raters worked independently from the team that scored predictor variables; both teams were blind to the scores of the other team.

To validate the Overall Severity and Mistreatment of Clients scales, we developed a 12-item Violation of Principles scale (focused on the professional principles of autonomy, beneficence, nonmaleficence, and justice). The violation of each principle was assessed with two Likert-scale items. For example, nonmaleficence was assessed with the statements “The misbehavior harmed or risked harming others in an unjustified manner (e.g., without both consent and a reasonable hope of benefit)” and “The professional did not take reasonable measures to minimize the risk of harm to others” using a 7-point Likert scale ranging from 1 (not at all) to 7 (extremely). Raters completed all 3 scales for the first 40 cases.

Regarding inter-rater reliability estimates, the single measures ICC for the severity scale was .68, indicating excellent reliability. For the mistreatment of humans scale the ICC was .86. For the Violations of Principles scale the ICC was .59.

Data Analysis Plan

Data analysis focused on generating basic descriptive data to provide a profile of cases; establishing the validity of the Severity of Wrongdoing scale by testing for correlation with related measures (e.g., the Mistreatment of Clients and Violation of Ethical Principles); examining the correlation of the severity of wrongdoing with a variety of predictor variables; and testing professional domains (research and practice) for differences.

RESULTS

Descriptive Data

The mean year cases occurred was 1991 with a standard deviation of 20 years, which enabled us to track changes in severity of wrongdoing across time. Eighty-seven percent of protagonists were male; 70% were between the ages of 40 and 69. In 77% of cases the wrongdoing persisted for more than 2 years. Seventy-six percent of cases involved more than one kind of wrongdoing.

The mean Overall Severity score was 5.5 on a 7-point scale with a standard deviation of 1.2; the mean Mistreatment of Clients score was 5.2 on a 7-point scale with a standard deviation of 1.5.

Figure 1 presents the mean rating for each of the 10 environmental variables and the protagonist's History of Wrongdoing in both domains (practice and research). Some factors were consistently absent from cases (occurring in less than 5%) and clearly played no causal role; these included institutional mistreatment of the wrongdoer and penalizing the protagonist for doing the right thing. That is, our wrongdoers were not simply victims of toxic environments.

Other variables appeared to a moderate degree across domains, including financial rewards for wrongdoing (research $M=1.6$, practice $M=1.9$), others benefitting from wrongdoing (research $M=1.7$, practice $M=1.8$), and oversight failures (research $M=1.9$, practice $M=1.9$).

Validation of the Severity Scales

As a measure of convergent validity, we tested the correlation of our Overall Severity scale with several related measures, including the Mistreatment of Clients, the Violation of Principles, and the Consequences to the Protagonist scales. We expected the strongest correlation with the Violation of Principles scale because it is the only other scale that aimed to provide an overall picture of the wrongdoing (albeit by assessing "component parts"—specific ways in which it was wrong—rather requesting a global assessment).

The correlation of the Overall Severity scale with the Mistreatment of Clients scale was $r_s=.72$ ($p<.001$, $n=64$). The correlation of the Violation of Principles scale with the Overall Severity scale was $r_s=.81$ ($p<.001$, $n=30$) and $r_s=.68$ ($p<.001$, $n=30$) with the Mistreatment of Clients Scale.

Consequences scales were developed separately for the domains of research and practice. The correlation of the Overall Severity scale with the Consequences to the Protagonist scale was $r_s=.58$ ($p<.000$) in the domain of research and $r_s=.60$ ($p<.000$) in the domain of practice.

These strong, positive correlations provide evidence of construct validity of the Overall Severity scale.

Correlation of Severity with Predictor Variables

Table 3 presents the correlation of Overall Severity with the 10 environmental variables and the wrongdoer's History of Wrongdoing. The protagonist's History of Wrongdoing was the strongest positive correlation with both the Overall Severity scale ($r_s=.41$, $p<.01$) and the Mistreatment of Clients scale ($r_s=.32$, $p<.01$), followed by Financial Rewards for wrongdoing ($r_s=.25$ and $.26$, respectively, $p<.05$). Ambiguous Norms was the strongest negative correlation with Overall Severity ($r_s=-.60$, $p<.01$) and Mistreatment of Clients ($r_s=-.28$, $p<.05$).

Several variables were significantly correlated with either the Overall Severity or Mistreatment of Clients scale in only one domain. Overall Severity and Mistreatment of Clients were positively correlated with Others Benefitting from the wrongdoing ($r_s=.38$ and $.39$ respectively, $p<.05$) and Others Being Penalized for doing what is right ($r_s=.41$ and $.40$, respectively, $p<.01$) only in the domain of healthcare practice. Vulnerable Victims was significantly correlated with the Mistreatment of Clients only in the domain of research ($r_s=.45$, $p<.01$).

Further Domain Differences

Three variables were present to statistically significantly different degrees in the domains of research and practice. In the realm of research, wrongdoers were significantly more likely to be in a position of authority (research $M=2.4$, $SD=0.66$; practice $M=1.38$, $SD=0.76$; $t=7.09$, $p<.001$)—typically serving as the PI of a study—and the norms were more likely to be ambiguous than in healthcare practice (research $M=1.49$, $SD=0.58$; practice $M=1.12$, $SD=0.32$; $t=3.54$, $p<.001$). History of Wrongdoing scores were higher in the healthcare delivery context than in the research context (research $M=1.52$, $SD=0.84$; practice $M=1.86$, $SD=0.86$; $t=-1.95$, $p<.05$). See Figure 1.

DISCUSSION

Our study of 100 cases of wrongdoing in healthcare delivery and research produced several substantive findings. It also enabled us to identify a series of serious limitations with our approach in Phase I of the project, which involved predicting the severity of professional wrongdoing.

Substantive Findings

Our descriptive data indicate that cases typically involve multiple kinds of wrongdoing and that wrongdoing is typically repeated across two or more years. This reinforces the need for colleagues to intervene early and for institutions to respond in ways that are likely to reduce repeat offenses (e.g., through termination of employment or intensive remediation training).

Although experts may offer slightly different reasons why they believe an action is wrong, we found that experts are able to reach an impressive level of agreement in their overall assessment of the severity of wrongdoing. We also found a positive correlation between the severity of wrongdoing as assessed by our team of experts and the actual consequences to the wrongdoers in cases ($r_s=.58 - .60$, $p<.001$). This reflects well on the ability of

disciplinary bodies to arrive at a meaningful consensus on the severity of infractions and to determine penalties that are proportionate to the degree of wrongdoing.

We also learned that conceiving of professional wrongdoing as one uniform construct such as “organizational deviance” is misguided. Professional wrongdoing is not a coherent phenomenon that should be studied through heterogeneous, maximum variation sampling; rather, the domain of wrongdoing matters. For example, we found that in healthcare delivery the role of “others” in the wrongdoing—whether they are rewarded for the wrongdoing or penalized for doing what is right—is much greater than in research.

Methodological Weaknesses

On the one hand, reinforcing our idea that professional wrongdoing is not a homogenous construct, it seems that it may be important to consider the severity of wrongdoing when trying to understand its causes. In our sample of 100 cases, the severity was relatively high with low variation (5.5/7 with a standard deviation of 1.2). We believe this helps to account for the relative absence of certain variables from our set of cases—variables that other studies led us to expect as predictors. For example, several studies suggest that when employees feel they are being treated unfairly by their institutions or by review boards, they are more likely to engage in misbehavior (Colquitt, Noe, and Jackson 2002; Keith-Spiegel and Koocher 2005; Martinson et al. 2010). Yet, as noted above, our study—which drew from an average of 33 published accounts per case—nearly never found mention of a protagonist who felt wronged by the institution nor of colleagues who spoke of unfair systems within the institution—despite the fact that protagonists frequently offered explanations of their behavior.

On the other hand, focusing on predicting the severity of wrongdoing is problematic in several regards. Consider the example of authority over peers. In the domain of research, we found a negative correlation—the higher the level of authority, the lower the severity. However, there are two problems with this finding. First, the variation in severity is low, and therefore the results may not be particularly practical. In other words, we found some variables may distinguish very bad behavior from very, very bad behavior. Just as no one publishes accounts of “business as usual” in medicine and research, few people publish accounts of mild misbehavior. Second, and much more problematic, these findings sometimes mask a much more valuable fact: Occupying a position of authority actually appears to predict the occurrence of wrongdoing in research. On a three-point scale, the mean level of authority was 2.4, which is very high. This was significantly higher than the level found in medical practice ($M=1.4$). That is to say, in our sample of cases that attracted sufficient attention to become public, occupying a position of authority is correlated with wrongdoing in research.

Plans for Further Research

These observations have led us to modify our fundamental approach to understanding and predicting wrongdoing. With our current dataset of 100 cases, we will use cluster analysis to identify natural groups of non-overlapping cases (e.g., human subjects research violations vs. other research violations). Once we determine that cases do not involve overlapping

kinds of wrongdoing, we can identify the variables that correlate with general kinds or domains of wrongdoing. Moving forward, using our taxonomy, we will identify large sets of specific forms of non-overlapping wrongdoing (e.g., fraudulent medical procedures and improper prescribing or informed consent violations in research and data fabrication). The aim will be to predict the occurrence of a specific form of wrongdoing, rather than the severity of generic professional wrongdoing. With sufficiently large sample sizes, we will be able to develop causal models using regression analyses.

Our earliest attempts at this new approach indicate that it is feasible. We have identified sufficient numbers of published cases of specific kinds of wrongdoing. We have also found that by eliminating the need to rate the severity of wrongdoing, we eliminate the need to produce a case synopsis; we are able to extract data from original publications and enter them directly into our datasheets, thereby significantly reducing the time needed to investigate.

While historiometric methods have proven challenging to use in the study of wrongdoing, we believe the approach is feasible and will produce rich, ecologically valid data on an important and understudied topic in healthcare ethics (Brewer 2000).

Acknowledgments

This paper was supported by grants UL1RR024992 and 1R21RR026313 from the NIH-National Center for Research Resources (NCRR) and a seed grant from the BF Charitable Foundation.

We thank Michael Mumford for discussions contributing to the initial approach taken in this study. We thank the following research assistants for writing cases and other tasks: Kelly Carroll, Tyler Gibb, Elena Kraus, Shane Levesque, Andrew Plunk, Timothy Rubbelke, Meghan Vasher, and Rebecca Volpe. We thank the following colleagues for rating the severity of wrongdoing found in cases: Ana Iltis, Tracy Koogler, Sahana Misra, and Lisa Parker.

References

- Aldrich, N. Medicare Fraud Estimates: A Moving Target?; The Sentinel. 2009. p. 1-4.http://www.smpresource.org/Content/NavigationMenu/AboutSMPs/MedicareFraudEstimatesAMovingTarget/Medicare_Fraud_Estimates.pdf
- Anderson MS, Ronning EA, De Vries R, Martinson BC. The perverse effects of competition on scientists' work and relationships. *Science and Engineering Ethics*. 2007; 13(4):437–61. [PubMed: 18030595]
- Antes AL, Brown RP, Murphy ST, Waples EP, Mumford MD, Connelly S, Devenport LD. Personality and ethical decision-making in research: The role of perceptions of self and others. *Journal of Empirical Research on Human Research Ethics*. 2007; 2(4):15–34. [PubMed: 19385805]
- Arnold, Sandra R.; Straus, Sharon E. Interventions to improve antibiotic prescribing practices in ambulatory care. *Cochrane Database of Systematic Reviews*. 2005. <http://www.mrw.interscience.wiley.com/cochrane/clsystrev/articles/CD003539/frame.html>
- Bandura A, Underwood B, Fromson M. Disinhibition of aggression through diffusion of responsibility and dehumanization of victims. *Journal of Research in Personality*. 1975; 9:253–269.
- Bolin A, Heatherly L. Predictors of employee deviance: The relationship between bad attitudes and bad behavior. *Journal of Business and Psychology*. 2001; 15(3):405–418.
- Boulware LE, Ratner LE, Cooper LA, Sosa JA, LaVeist TA, Powe NR. Understanding disparities in donor behavior: Race and gender differences in willingness to donate blood and cadaveric organs. *Medical Care*. 2002; 40(2):85–95. [PubMed: 11802081]

- Bramstedt KA, Kassimatis K. A study of warning letters issued to institutional review boards by the United States Food and Drug Administration. *Clinical and Investigative Medicine*. 2004; 27(6): 316–23. [PubMed: 15675112]
- Brewer, M. Research design and issues of validity. In: Reis, H.; Judd, C., editors. *Handbook of Research Methods in Social and Personality Psychology*. Cambridge: Cambridge University Press; 2000. p. 3-16.
- Carr GD. Professional sexual misconduct--an overview. *Journal of the Mississippi State Medical Association*. 2003; 44(9):283–300. [PubMed: 14564742]
- Colquitt JA, Noe RA, Jackson CL. Justice in teams: Antecedents and consequences of procedural justice climate. *Personnel Psychology*. 2002; 55(1):83–109.
- Cooper C, Selwood A, Livingston G. The prevalence of elder abuse and neglect: A systematic review. *Age and Ageing*. 2008; 37(2):151–60. [PubMed: 18349012]
- Coughlin SS, Barker A, Dawson A. Ethics and scientific integrity in public health, epidemiology and clinical research. *Public Health Reviews*. 2012; 34(1):1–13. [PubMed: 26236074]
- Davis, Mark S. The role of culture in research misconduct. *Accountability in Research*. 2003; 10(3): 189–201. [PubMed: 14979320]
- Deluga RJ. Relationship among American presidential charismatic leadership, narcissism, and rated performance. *Leadership Quarterly*. 1997; 8(1):49–65.
- DuBois, JM. *Ethics in Mental Health Research: Principles, Guidance, and Cases*. New York: Oxford University Press; 2008.
- DuBois JM, Anderson EE, Carroll K, Gibb T, Kraus E, Rubbelke T, Vasher M. Environmental factors contributing to wrongdoing in medicine: A criterion-based review of studies and cases. *Ethics and Behavior*. 2012; 22(3):163–188. [PubMed: 23226933]
- DuBois JM, Dueker JM. Teaching and assessing the responsible conduct of research: A delphi consensus panel report. *The Journal of Research Administration*. 2009; XL(1):49–70. [PubMed: 22500145]
- DuBois JM, Kraus E, Vasher M. The development of a taxonomy of wrongdoing in medical practice and research. *American Journal of Preventive Medicine*. 2012; 42(1):89–98. [PubMed: 22176853]
- Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*. 2009; 4(5):e5738. [PubMed: 19478950]
- Grant D, Alfred KC. Sanctions and recidivism: An evaluation of physician discipline by state medical boards. *Journal of Health Politics, Policy and Law*. 2007; 32(5):867–85.
- Greenberg J. Stealing in the name of justice: Informational and interpersonal moderators of theft reactions to underpayment inequity. *Organizational Behavior and Human Decision Processes*. 1993; 54(1):81–103.
- Grover SL. Why professionals lie: The impact of professional role conflict on reporting accuracy. *Organizational Behavior and Human Decision Processes*. 1993; 55(2):251–272.
- Hall AJ, Logan JE, Toblin RL, Kaplan JA, Kraner JC, Bixler D, Crosby AE, Paulozzi LJ. Patterns of abuse among unintentional pharmaceutical overdose fatalities. *JAMA: Journal of the American Medical Association*. 2008; 300(22):2613–20. [PubMed: 19066381]
- Hegarty WH, Sims PH. Some determinants of unethical decision behavior: An experiment. *Journal of Applied Psychology*. 1978; 63(4):451–457.
- Iglehart JK. Finding money for health care reform--rooting out waste, fraud, and abuse. *New England Journal of Medicine*. 2009; 361(3):229–31. [PubMed: 19516025]
- Jansen E, Von Glinow MA. Ethical ambivalence and organizational reward systems. *The Academy of Management Review*. 1985; 10(4):814–22.
- Jennings, M. *The Seven Signs of Ethical Collapse: How to Spot Moral Meltdowns in Companies... Before It's Too Late*. New York, NY: St. Martins Press; 2006.
- Judge TA, LePine JA, Rich BL. Loving yourself abundantly: Relationship of the narcissistic personality to self- and other perceptions of workplace deviance, leadership, and task and contextual performance. *Journal of Applied Psychology*. 2006; 91(4):762–76. [PubMed: 16834504]

- Keith-Spiegel P, Koocher GP. The IRB paradox: Could the protectors also encourage deceit? *Ethics and Behavior*. 2005; 15(4):339–49. [PubMed: 16578924]
- Levine RJ. Clinical trials and physicians as double agents. *Yale Journal of Biology and Medicine*. 1992; 65(2):65–74. [PubMed: 1519378]
- Marshall E. Clinical research - shutdown of research at Duke sends a message. *Science*. 1999; 284(5418):1246–1246. [PubMed: 10383299]
- Martinson B, Crain AL, Anderson M, DeVries R. Institutions' expectations for researchers' self-funding, federal grant holding, and private industry involvement: Manifold drivers of self-interest and researcher behavior. *Academic Medicine*. 2009; 84(11):1491–1499. [PubMed: 19858802]
- Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. *Nature*. 2005; 435(7043):737–8. [PubMed: 15944677]
- Martinson BC, Crain AL, De Vries R, Anderson MS. The importance of organizational justice in ensuring research integrity. *Journal of Empirical Research on Human Research Ethics*. 2010; 5(3): 67–83. [PubMed: 20831422]
- Martinson BC, Anderson MS, Crain AL, De Vries R. Scientists' perceptions of organizational justice and self-reported misbehaviors. *Journal of Empirical Research on Human Research Ethics*. 2006; 1(1):51–66. [PubMed: 16810337]
- Meyer, WM.; Bernier, GM, Jr. Potential cultural factors in scientific misconduct allegations. In: Steneck, N.; Scheetz, M., editors. *Investigating Research Integrity: Proceedings of the First ORI Research Conference on Research Integrity*. Washington, DC: Department of Health and Human Services; 2002. p. 163-166.
- Milgram S. Some conditions of obedience and disobedience to authority. *Human Relations*. 1965; 18(1):57–76.
- Mumford, MD. *A Comparative Analysis of Charismatic, Ideological, and Pragmatic Leaders*. Mahwah, NJ: Lawrence Erlbaum Associates; 2006. *Pathways to Outstanding Leadership*.
- O'Connor J, Mumford MD, Clifton TC, Gessner TL, Connelly MS. Charismatic leaders and destructiveness: An historiometric study. *Leadership Quarterly*. 1995; 6(4):529–555.
- Roback HB, Strassberg D, Iannelli RJ, Finlayson AJ, Blanco M, Neufeld R. Problematic physicians: A comparison of personality profiles by offence type. *Canadian Journal of Psychiatry*. 2007; 52(5): 315–22.
- Rodwin, MA. *Medicine, Money, and Morals: Physicians' Conflicts of Interest*. New York: Oxford University Press; 1993.
- Schuchman M. Medical whistle-blower protection lacking. *Canadian Medical Association Journal*. 2008; (12):1529. [PubMed: 18519891]
- Shah S, Whittle A, Wilfond B, Gensler G, Wendler D. How do institutional review boards apply the federal risk and benefit standards for pediatric research? *JAMA: Journal of the American Medical Association*. 2004; 291(4):476–82. [PubMed: 14747505]
- Simonton DK. Significant samples: The psychological study of eminent individuals. *Psychological Methods*. 1999; 4(4):425–451.
- Simonton DK. Qualitative and quantitative analyses of historical data. *Annual Review of Psychology*. 2003; 54(54):617–40.
- Sims RL. A study of deviance as a retaliatory response to organizational power. *Journal of Business Ethics*. 2009; 92(4):553–563.
- Suedfeld P, Bluck S. Changes in integrative complexity prior to surprise attacks. *Journal of Conflict Resolution*. 1988; 32:626–635.
- Titus SL, Wells JA, Rhoades LJ. Repairing research integrity. *Nature*. 2008; 453(7198):980–2. [PubMed: 18563131]
- Trevino LK, Butterfield K, McCabe D. The ethical context in organizations: Influences on employee attitudes and behaviors. *Business Ethics Quarterly*. 1998; 8(3):447–476.
- Vardi, Y.; Weitz, E. *Misbehavior in Organizations: Theory, Research, and Management*. Mahwah, N.J.: L. Erlbaum; 2004.
- Victor B, Trevino LK, Shapiro DL. Peer reporting of unethical behavior: The influence of justice evaluations and social context factors. *Journal of Business Ethics*. 1993; 12(4):253–263.

Williams DR, Mohammed SA. Discrimination and racial disparities in health: Evidence and needed research. *Journal of Behavioral Medicine*. 2009; 32(1):20–47. [PubMed: 19030981]

Zimbardo, PG. *The Lucifer Effect*. New York: Random House, Inc; 2007.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Mean Scores of Predictor Variables in Research vs Practice

Table 1

Reporting, Investigation, and Outcomes Variables: Comparison of Frequencies

	FFP	Other Research	Medical Practice	χ^2
Failed reporting attempts	28%	35%	43%	1.98
Whistleblower Description ^a				
Patient/participant/family	0% _a	18% _{a,b}	45% _b	38.36***
Subordinate	23% _a	10% _{a,b}	5% _b	
Peer: institutional	15% _a	3% _b	3% _b	
Peer: external	5%	8%	8%	
Oversight personnel: institutional	15% _a	8% _{a,b}	3% _b	
Oversight personnel: external	5%	13%	10%	
Others (e.g., reporter)	18%	25%	15%	
Unknown	20%	18%	13%	

Table 2

Wrongdoer and Setting Variables: Comparison of Frequencies

Dependent Variable	FFP	Other Research	Medical Practice	χ^2
Setting variables				
Academic medical setting	90% _a	75% _a	3% _b	71.03***
Government funding	70% _a	58% _a	0% _b	45.63***
Private funding	33% _{a,b}	63% _a	0% _b	36.12***
Had an accomplice	10% _a	40% _b	35% _b	10.18**
Others found guilty	20%	33%	35%	2.50
Wrongdoer Variables				
Male	75% _a	90% _b	93% _b	5.89*
Born outside US	10%	20%	20%	1.92
Trained outside US	15%	18%	25%	1.40
Plea of insanity	0%	0%	5%	4.07
Found unfit to stand trial	0%	0%	0%	--
Evidence of addiction	0% _a	0% _a	10% _b	8.28*
Significant personal problems	5% _a	0% _a	23% _b	13.41***
Claimed following orders/policy	3%	10%	0%	5.43
History of Wrongdoing				
Wrongdoing was repeated	68% _a	80% _{a,b}	95% _b	9.79**
Different kinds of wrongdoing	43% _a	65% _{a,b}	75% _b	9.30**
Wrongdoing in multiple institutions	18%	20%	28%	1.28
Felony arrests in personal life	3% _a	0% _a	13% _b	7.37*

* p<.05,

** p<.01,

*** p<.001.

N = 120 cases: 40 FFP, 40 Other Research, 40 Medical Practice. When chi-squared is significant, percentages that do not share subscripts are significantly different by standardized residual.

Table 3

Predictor Variables: Comparison of Mean Scores (ANOVA)

Predictor Variables	FFP ⁺	Other Research	Medical Practice	F
Conflicting roles ⁺⁺⁺	1.15 _a (.53)	1.75 _b (.95)	1.05 _a (.32)	13.28***
Financial reward ⁺⁺⁺	1.33 _a (.69)	2.00 _b (.93)	2.23 _b (.95)	11.70***
Others benefit ⁺⁺⁺	1.30 _a (.61)	1.98 _b (.80)	1.83 _b (.90)	8.26***
Penalized for doing right	1.00 (.00)	1.00 (.00)	1.02 (.16)	n/a
Others penalized ⁺⁺⁺	1.30 (.65)	1.18 (.55)	1.20 (.46)	.56
Mistreatment of wrongdoer	1.08 (.27)	1.02 (.16)	1.00 (.00)	n/a
Ambiguous norms ⁺⁺⁺	1.02 _a (.15)	1.43 _b (.55)	1.05 _a (.22)	16.04***
Vulnerable victims ⁺⁺⁺	1.07 _a (.35)	1.70 _b (.79)	1.68 _b (.69)	12.22***
Oversight failures ⁺⁺⁺	1.65 _a (.77)	2.13 _b (.79)	1.78 _{a,b} (.77)	4.03*
Position of authority ⁺⁺⁺	1.90 _a (.74)	2.45 _b (.68)	1.55 _a (.81)	14.73***
Collaboration	1.20 (.46)	1.33 (.47)	1.00 (.00)	n/a
History of wrongdoing ⁺⁺⁺⁺	1.30 _a (.88)	1.65 _{a,b} (.86)	2.10 _b (1.03)	7.45***
Other Variables				
Duration	3.83 (1.28)	4.30 (1.47)	4.22 (1.48)	1.31
Number of violations ⁺⁺⁺	2.03 _a (1.48)	3.63 _b (1.64)	3.13 _b (1.49)	11.33***
Number of Sources ⁺⁺	19.30 _a (17.52)	32.23 _b (23.98)	34.38 _b (31.19)	4.30*

* p<.05,

** p<.01,

*** p<.001.

N = 120 cases: 40 FFP, 40 Other Research, 40 Medical Practice. When ANOVA is significant, means that do not share subscripts are significantly different by Tukey post hoc pairwise comparison.

⁺ Scores are means with standard deviations in parentheses⁺⁺ All cases had a minimum of 5 sources⁺⁺⁺ MANOVA including these 10 variables was statistically significant, Wilks' Lambda = 0.309, p<.001.