7-22-10

Supplementary materials for Koocher and Keith-Spiegel's "Peers Nip Misconduct in the Bud"

(July 22, 2010) *Nature 466*, 438-440

How Do Researchers Respond to Perceived Scientific Wrongdoing?

Overview, Method and Survey Results

Gerald P. Koocher[1], Patricia Keith-Spiegel[2], Barbara G. Tabachnick[3], Joan E. Sieber[4], & Darrell L. Butler[5]

After a brief overview, we provide a summary list of findings for easier reading. Methodological details and complete data analyses begin on page 7. A complete version of the survey from which the data were derived appears as Appendix A.

I. Project Overview and Implications

Research conducted irresponsibly or fraudulently and left uncorrected eventually touches many scientists in some way. Personal careers can disintegrate. An entire body of otherwise solid work may fall into disrepute. The reputation of the university, hospital, or institute where the research took place can be tainted. The public trust upon which scientists depend for support erodes.

Current institutional and federal policies concerning research wrongdoing rarely consider the roles that colleagues can play to prevent the widespread dissemination of invalid data, even though an alarmingly high percentage of scientists and advanced students confide that they know of research wrongdoing committed by their peers (e.g., *1, 2, 3, 4, 5*). Few cases, by comparison, ultimately come to the attention of the appropriate agencies (e.g., *6, 7, 8*).

Several possibilities may explain the apparent gap between what researchers know and what ultimately may become public knowledge. Federal agencies only pursue projects that they have funded. Furthermore, they only pursue allegations that fall within a narrow definition of scientific misconduct; namely fabrication, falsification, or plagiarism (FFP). Other criteria also apply, namely that the incident must involve a significant departure from accepted research practices, be committed intentionally or irresponsibly. Guilt must be proven by a preponderance of the evidence (8). Some incidents may be adjudicated at the institutional level because they involve research wrongdoing other than FFP or because the associated project had no connection to a funder or agency that directly investigates wrongdoing. Alternatively, individuals who observed or learned of scientific wrongdoing may have intervened informally and kept the outcome private among the parties involved. Or they may have feared taking action, allowing the suspected wrongdoing to remain unexamined and unchallenged. Our study sought to discover what actually happens when researchers suspect scientific wrongdoing.

We take the position that scientists themselves can serve on the front lines to foster integrity in science. Colleagues, supervisors, and assistants who work with, or nearby, or in the same content area,

---

[1] Associate Provost and Professor of Psychology (Keith-Spiegel and Sieber were also affiliated with Simmons College, Boston, MA during the conduct of the research reported here). Funding for this research came from the National Institute of Neurological Disorders and Stroke and Office of Resaerch Integrity via grant no R01 NS04973.. [2] Professor Emerita, Ball State University, [3]Professor Emerita California State University, Northridge, [4] Professor Emerita, California State University, East Bay, [5] Professor, Ball State University and President, nHarmony. We extend our very deepest appreciation to our consultants for their invaluable assistance: Mark S. Frankel, Judith P. Swazey, C. Kristina Gunsalus, Paul J. Friedman, and Judy Beal.

*Address correspondence to Dr. Koocher, email: Koocher@simmons.edu.

constitute a powerful and potentially valuable resource to minimize contamination of the research record due to the purposeful or unintentional commission of scientific errors. Informal resolutions may be the only source of correction for many acts that corrupt science when the researcher may not be engaging in these acts purposely. In contrast to many widely publicized cases of whistle blowing, we know almost nothing about whether successful, gentler, behind-the-scenes interventions are preventing or minimizing research wrongdoing. Anecdote also holds that many researchers feel caught off guard without effective response strategies when they observe colleagues, assistants, or students behaving irresponsibly.[2]

The results of our online survey project provide insights into important and heretofore unanswered questions about what happens when research scientists observe or hear about scientific errors, what they do with this information, and what happens to them after that. As such, it has major advantages over the "what if?" surveys that pose contrived scenarios and ask respondents what they imagine they would do in the actual situation (e.g., *9*). Our work also provides additional detail about researchers' dealings with misconduct and research wrongdoing that has not been examined in the few previous surveys assessing actual experiences.[3] We addressed the following heretofore unanswered questions:

- Who, if anyone, takes action (which can range from private collegial interactions to reporting suspicions to the appropriate institutional office) when they suspect research wrongdoing?
- What are the conditions under which they will take action?
- What motivates informal interventions or action?
- What circumstances enhance the likelihood that scientists will intervene to prevent or correct irresponsible research?
- Under what conditions do researchers fail to take action when they suspect research wrongdoing? Do they experience regrets?
- How do researchers interpret their personal responsibility to actively monitor and act when they suspect research wrongdoing?
- What kinds of interventions or actions do researchers attempt?
- Which action strategies work and which result in unsuccessful or difficult outcomes?
- Do researchers who take action face negative consequences afterwards? If so, what form do they take?

We found that researchers are confronting a variety of types of wrongdoing besides FFP in large numbers and usually not formally reporting what they suspect or know. We should be looking at a broad spectrum of behaviors that can have a detrimental impact on the outcome of research and the validity of the scientific research record. By ignoring these behaviors other than FFP that corrupt science, we may be concealing much of the problem (10). The key remains sensitizing researchers to their importance, which is made more difficult by their lack of standing. Thinking more broadly about the definitions would, in our opinion, contribute substantially to reducing the pollution of data, not just through the raising of consciousnesses about ethics but also by delineating what behaviors are worthy of deserving scrutiny.

II. Quick Summary of Findings

Our results present a very complex picture. Here we provide a listing of the more general and interesting ones, although some are mitigated by the circumstances surrounding specific incidents, such as the personality of a suspected wrongdoer or how serious the alleged infraction is judged to be. As intriguing as many of our findings are, we encountered some devils when attempting to get at some of the details, as can be seen in the results section.

---

[2] A product of this research grant, *Responding to Research Wrongdoing*, is available for distribution free of charge at www.ethicsresearch.com. The *RRW* offers guidance to those who observe or suspect scientific wrongdoing based largely on the findings of this project.

General Findings

1. All but 406 of the 2599 respondents shared personal experiences involving suspected scientific error or wrongdoing by others. The fact that 84% of participants reported at least one case places our incidence rate among the highest of any survey to date, although our study also offered a broad definition of "bad science" options beyond fabrication, falsification or plagiarism (FFP) and placed no time limit on when the incident occurred. Examples of other acts that could corrupt the scientific record include poor supervision of assistants, carelessness, authorship disputes, failure to follow the rules of science, conflicts of interest, incompetence, and hostile work environments that impact on research quality.

2. Almost two-thirds our respondents who provided at least one experience claim to have taken some form of action when they observed or suspected scientific wrongdoing by dealing with the matter informally on their own (or with others present), reporting it to the appropriate office in one's institution or, in rare cases, directly to ORI (Office of Research Integrity, the reporting agency for NIH funded projects) or other funder.

3. The vast majority of respondents who intervened did not formally report the matter initially, preferring to attempt to correct the problem or effect damage control on their own or sometimes in partnership with others. After an attempt at an informal intervention, some incidents were later reported to the appropriate office in their institution, ORI, or the funder.

4. The odds of intervening did not heavily depend on the type of irresponsible action. This finding may feel disconcerting to federal officials overseeing research misconduct in that the acts they consider the most serious (FFP) did not predict a better likelihood of taking any action. Intervention rates proved similar across all types of acts except for publication disputes (excluding plagiarism) where the rate of intervention was lower, perhaps because junior people are usually the ones who see themselves slighted in publication credits and view complaining as too risky. The highest rate (and only slightly higher) of intervention occurred for acts of carelessness, failure to follow the rules of science, and when workplace stressors compromised research quality.

5. Respondents self-selected what experiences they shared, thus our breakdown here does not reflect a true portrait of offender types. Thirty-one percent of the shared incidents involved wrongdoings or errors committed by peers, 26% about senior colleagues, 12% junior colleagues, 24% about post docs and research assistants, and the remaining 7% about supervisors.

6. One third of the respondents who intervened indicated that if they had another chance they would have done something differently. Contrary to conventional wisdom, only one in 10 would have never gotten involved in the first place. Most reported that they would have done more and acted more quickly. .

7. A majority of our respondents who had an incident to share conveyed only one, although the survey was programmed to collect up to 10 incidents. Twenty three percent offered two incidents, and 14% provided three or more. We did attempt to ascertain any differences between those who reported a single incident and those who reported three or more. One significant difference was found. It was not, as might be expected, a matter of how long respondents had been doing research. Nor was there a difference in items asking about the degree of individual responsibility researchers have to uphold scientific integrity. The single difference was found for the item asking about their institutions' willingness to solve problems. Those who believed that their institution was *less* willing to take action when issues of scientific irresponsibility arose reported more incidents.

8. Unless one observes an act of irresponsible science in progress, or the individual admits doing it and why, an observer will have great difficulty knowing with certainty that the act was purposeful, or even if an adverse event actually occurred. There is always a chance that the suspected

wrongdoer's behavior flowed from incompetence, sloppiness, or that the observer misunderstood what they saw or heard. Thus, a certain amount of ambiguity pertains whenever an act of research misconduct is suspected, which surely complicates the decision as to how best to proceed. Only 31% of our respondents felt that the suspected wrongdoers were *definitely* aware of their own wrongdoing and engaging in the act intentionally. Another 27% believed that the wrongdoers were *probably* aware of their own wrongdoing.

9. Slightly more than half the incidents that our respondents shared with us did *not* involve FFP. Although our data do not inform us about the actual rates of these additional types of irresponsible science, it seems clear that many respondents find other adverse events worthy of presentation and discussion. Authorship and publication disputes are particularly salient because publication credits comprise a prime currency of professional advancement. Our respondents offered more cases about scholarly publication clashes than about plagiarism.

10. It is of interest to note that overall findings were similar across many variables, such as work setting, type of research being conducted, and earned academic degrees. Our results appear to be applicable to all research settings.

Who Intervened and Why?

1. We sought more detail about why our respondents chose to get involved and what motivated them. A text box was provided for participants to tell us in their own words why they took action. Ten recurring themes appeared, in the order of frequency presented here:

- It was the right thing to do/an ethical duty.
- Involvement is a responsibility/a professional duty.
- To do otherwise would have been damaging to my reputation/lab/ project/institution/ self-protection.
- I was assigned to intervene.
- I wanted to help/support a colleague.
- The act was too serious to ignore.
- To ignore it would be to violate the standards of science/public trust.
- I wanted to prevent a friend/colleague/student from making a mistake.
- I was concerned that others would be harmed.
- The matter could be easily resolved.
- I was angry over what was going on.

.

2. A binary logistic regression analysis predicted that the most likely researchers to intervene are those holding higher status than that of the suspected individual, having less regular interaction or involvement with the suspected individual, basing their suspicions on strong information (i.e., direct observation or disclosure of wrongdoing by the transgressor) rather than second-hand accounts or hearsay, and believing that individuals have a personal responsibility to maintain scientific responsibility, perceiving the transgression as unintentional, or feeling themselves either victimized or as potential objects of future blame.

3. Our findings suggest that a difficult or stressful work environment increased the chances that an intervention would be attempted. (Perhaps emotions helped to fuel a willingness to take action.)

4. The vast majority of respondents who reported that they felt like victims or who might be blamed reported that they intervened. (However, it is our impression from the text of the stories that these interveners may have been be too angry or self-righteous to be effective in their attempts to remediate the problem.)

Intervention Outcomes

1. Of the respondents who intervened one time or more, 28% achieved at least one correction of the problem (e.g., data were cleaned up, mistake repaired, text rewritten). Even though we wish this percentage were higher, it was encouraging to learn that our respondents had achieved 447 corrections. In another 209 incidents the suspected offenders apparently understood that problems existed but could not correct them because the damage had already occurred or because tracking down those adversely affected was impossible.

2. There was no statistically significant difference in the correction rate between informal intervention and formal reporting although corrections favoring informal intervention almost reached statistical significance.

3. Our results revealed that informal interventions often go well. For overall ratings of incident satisfaction, 39% rated as extremely or generally satisfied whereas 34% felt extremely or generally unsatisfied. The rest, 26%, felt neither satisfied nor dissatisfied.

4. We hypothesized that respondents who intervened informally would feel more satisfied with the outcome than those who reported the incident to a more formal authority, such as the appropriate office at their institution. However, this hypothesis was not supported. The outcome satisfaction rates for both types of actions rated similarly, although the overall ratings for formal reporting were slightly higher.

5. The most satisfied groups of interveners were those who went directly to the suspected violator themselves or with others present and the small number who sent unsigned messages, perhaps feeling they had taken action without subjecting themselves to any personal risk. We do not recommend this latter course, however, because it does not allow the suspected offender due process, may only enhance paranoia, and, if the accusation is unfounded, constitutes a moral failure on the part of the note-writer.

6. The lower the status of the suspected violator relative to that of the intervener, the greater the intervener's satisfaction with the outcome.

7. One of the more intriguing findings was the high outcome satisfaction rate among respondents who intervened when they felt sure, or fairly sure, that the suspected offenders seemed unaware that they had engaged in wrongdoing or made errors. Unfavorable outcomes were clearly more associated with those acts perceived of as intentional than unintentional.

8. Approaching the suspected wrongdoer with others did not differentiate those who felt satisfied from those who did not. (When the prospects looked more favorable or the intervention felt comfortable to do alone, the chances for success may have seemed more favorable.)

9. We found no differences in satisfaction level based on whether the interveners had a close working relationship or distant one.

10. Those who discussed the matter with others felt less satisfied than those who did not discuss the matter with anyone. (This may reflect a kind of "ascertainment bias"--referring to false results from non-random sampling or an atypical sample-- given that interveners mostly consulted with others when the matter seemed more difficult, complex, or ambiguous, or the evidence appeared weaker.)

11. Almost all respondents who shared an incident about their own assistants also reported intervening. When incidents involved someone else's assistant, interveners still took action in about half of these.

12. In 1,169 (42%) incidents, respondents reported experiencing no negative feedback as a result of taking action when they suspected wrongdoing. One in ten incidents (296 incidents or 11%) even

resulted in an elevation of status for the respondent as a result of taking some action.

13. Our encouraging findings should not mask the data revealing that negative experiences (e.g., worry, feeling shunned) accrued to interveners in almost half of the incidents. Often the unwanted result involved emotional distress with no additional fallout. However, some reported serious consequences, such as being forced to leave a job or losing previously close allies or friends. Negative fallout following intervention due to lack of institutional support occurred infrequently. Fears about possible legal action proved rare (and none apparently ever materialized). In short, a decision to get involved is not without risk, although we gained insights into how to avoid many of them (see *Responding to Research Wrongdoing*, a booklet available for free download at www.ethicsresearch.com).

Who Did Not Get Involved and Why?

1. About a third of our sample did not take action on any incident they shared with us. The odds of intervening decreased considerably when the suspected individual held a higher status position. Intervention was also less likely to occur when the problem involved publication issues, if the evidence of wrongdoing seemed less credible, or if the suspected individual worked in close proximity..

2. Individuals who did not intervene also had lower individual responsibility scores than those who had never observed research wrongdoing or those who had intervened.

3. When the respondents who took no action were asked directly why they did not intervene upon suspecting wrongdoing, the most frequently reported reason involved not seeing the matter as their problem. Further analyses revealed that in most of these cases the respondents felt too remotely involved or saw others already taking action. A *post hoc* analysis suggested that non-interveners may have been more willing to state "not my problem to solve" as a reason for non-intervention with colleagues of same or higher status than themselves than for their own supervisees.

4. Other frequently cited reasons for not taking any action included: "the suspected offender was difficult to deal with;" "the suspected offender was my superior;" "I did not know what to do," "evidence was insufficient;" "fear of no institutional support;" "fear of risk to my career status;" and "not deemed a serious enough matter."

5. Forty percent of those who did not get involved *even though they had direct evidence of wrongdoing* still felt misgivings, sometimes years later. Most of these did not act because they were not sure what they should have done.

6. Among those non-interveners who contributed 708 incidents involving *indirect* evidence of wrongdoing, such as gossip and rumor, most had no regrets (79%). Still, for almost a quarter of the incidents reported, misgivings still persisted and most of these involved respondents who remained uncertain what to do and thus had no closure.

7. We wrongly hypothesized that respondents who suspected irresponsible acts but did not intervene would be less likely to believe that their institution would take timely action. We reasoned that this group would feel that had they tried to act they would not have support to back them up, so they felt inclined to not get involved at all. It turns out that, instead, those people who *did* intervene had the *least* amount of confidence in their institution. It appears that those who took action believed that if they didn't do it, no one else would. Those who never heard of an incident appeared the *most* likely to believe that the institution would take care of the matter. These researchers may not have a mindset to "see" wrongdoing and therefore have confidence that the institution is doing its job.

8. We expected that non-interveners would also rate others as less likely to intervene than would interveners, thinking that might provide a way to neutralize feelings of unease or guilt. However, again it was the interveners who thought that others were less likely to get involved. And, interestingly, those who reported *never* observing or hearing of an incident constituted the majority of the respondents who assumed others would or should intervene. It may be that non-observers are largely composed of those who believe that scientists are more ethical than do the other two groups. They may believe that bad science is the work of a very few miscreants and that science is self-correcting, so not to worry. As a result, they may be less vigilant and, additionally, assume that should anything happen the matter will be taken care of by colleagues or by the institution. On the other hand, those who know wrongdoing is out there are more likely to recognize it when it occurs and also believe that the "bystander" effect prevails among their colleagues.

The Role of Consulting with Others

1. We all tend to believe that important decisions, especially complicated ones that carry some risk, would benefit from consultation with reliable others in whom we have confidence. Only about a third of the respondents who intervened also consulted with someone, either a trusted individual other than a supervisor, family or friends or their supervisor. In our study, however, conferring with others did not enhance the perceived feeling of a successful outcome.

2. We find it puzzling to report that the respondents who *did* talk to others about the incident often tended to *not* ultimately take any action. Almost two-thirds of non-interveners talked to others about at least one incident. Because our survey did not ask for detailed information about who spoke with whom and what their discussions entailed, we can only speculate as to the discrepancy between the interveners and non-interveners regarding the nature of their conversations. Did their confidantes discourage non-interveners from taking action? Did non-interveners find advice from more than one person conflicting and then decide to do nothing? Did talking the matter over make non-interveners more nervous than had they kept things to themselves? Did the non-interveners make poor selections of confidants who were unable to help formulate a sound action strategy? Unfortunately our data cannot provide definitive answers.

Findings Related to Taking Personal Responsibility

1. Attitudes for taking personal responsibility were assessed by combining the ratings assigned to several questions. All three groups (i.e., interveners, non-interveners, and those who have never observed wrongdoings) believed that researchers had the *strongest* responsibility to intervene when an act seemed serious and purposeful as opposed to purposeful minor acts and unintentional acts. .

2. Those with strong beliefs that individuals have a primary responsibility to become actively involved in maintaining scientific integrity increased the odds of their intervening by about 25%.

3. As predicted, those who did not intervene had statistically significantly lower personal responsibility scores than those who did intervene or those who knew of no incidents.

4. Contrary to our predictions, interveners rated their institutions as less likely to take action than did noninterveners. The highest ratings for institutional responsibility were by those who had no instances to report.

5. Contrary to our predictions, respondents who observed, suspected, or heard about research wrongdoing but did not intervene rated others are more likely to intervene than did respondents who intervened.

6. Because all of our respondents were principal investigators funded by the National Institutes of Health, they comprised a more sophisticated group than had we selected our population from a more general population of those who conduct research. However, even among members of this high-powered group, the more experienced scientists (defined as number of years spent doing research) did tend to score higher on our measure of individual responsibility. Perhaps they felt the most secure in their positions, saw themselves as having a duty to mentor those less-experienced around them, or recognized how vital it is to the ultimate success of their program that research be conducted validly.

III. Method and Survey Content

The Survey Sample

A large sample[4] of PIs (Principal Investigators) receiving NIH (National Institutes of Health) funding and working in the United States, primarily in research universities and hospital settings, were invited to participate in an online Web-based survey. We chose this population because we knew they were (or recently had been) active in research, more likely because of their senior status to observe what goes on in their research settings, be in a position to supervise others, and probably more knowledgeable about the principles of responsible science than any random group of individuals conducting research.

We drew the sample from the CRISP (Computer Retrieval of Information on Scientific Projects) data base supplied to us by NIH. After eliminating investigators working on projects outside of the United States, we selected a random sample of investigators funded by 15 federal agencies (specifically: NIAAA, NIA, NIAID, NCI, NICHD, NIDA, AHRQ, NIMH, NCMHD, NINDS, NINR, NCCAM, NHLB, NIEHS, and NIGMS). The total number of Principal Investigators (PIs) with research grants from these 15 agencies for fiscal years 2003 and 2004 totaled 21,750, but many had more than one funded project and appeared more than once among the projects listed. Duplicates were omitted, resulting in our potential sample of 11,922 unique scientists, representing a broad selection of research areas and geographical settings.

The Online Survey Method

Although the technology continues to evolve, all available evidence strongly indicates that an online survey was exceptionally well-suited to our particular project, especially because we did innovative programming of our own (described below) and because most of the drawbacks either do not apply or were rather easily minimized.

As with anonymous paper-and-pencil surveys sent through regular mail, participation is completely voluntary. Potential respondents could instantly reject the invitation by pushing the delete button, thus creating a built-in consent mechanism without any attendant experimenter demand characteristics. Despite an upfront cost to design and test the online survey content, the project proved extremely cost-effective given our large sampling population. Additional cost savings resulted from automatic recording and storing of the data. The web-based interface presented a small number of items per screen and could automatically skip questions based on contingencies, which is, at best, awkward on paper questionnaires (e.g., if X go to question Z; if Y, go to next question).

We had access to the total population of federally funded research scientists--making a true random sample possible. We informed potential respondents how and why we obtained their contact information to avoid the suspicions recipients can feel if a project team simply trolls for participants.

We planned in advance for minimizing the disadvantages of using an online survey. We reduced the potential for entry errors by making the directions very clear, using relatively large print on an

[4] Because the survey invitations went out via email, we could not determine exactly how many potential participants actually received, opened, and read our request for their participation. An unknown number may have been held back by firewalls.

uncrowded screen, and, whenever possible, making options forced-choice. The streamlined survey encouraged but did not require written commentary in text boxes. We attempted to minimize dropping out midstream by making the design simple but attractive. We also stressed the significance of the project to the respondents, explained how the data would be used, how we protected their anonymity and guaranteed their confidentiality. They were also informed as to about how long the survey would likely take to complete.

Perhaps the greatest concern to any research modality is the perceived degree of anonymity and confidentiality that respondents can realistically expect, especially when the topic raises sensitive or controversial themes, as ours did. The perceived trustworthiness of our sponsors and reputations of the investigators likely helped reassure respondents that this particular project would conform to the highest ethical standards. A Web-based survey (as opposed to an email survey that allows tracking of respondents), reduces the possibility of participant identification and affords more reassurance to respondents because no authenticated method exists to verify exactly who responded. We did not request any personally identifying information. To further enhance confidence we sought only minimal and very general demographic data.

Additional steps taken to protect against disclosure of the identity of respondents included the investigators and the Web designers keeping each other as “blind” as possible. The researchers maintained the email database on a different server from the Website. The Web designers disabled logging for the website (after debugging) and stored responses to survey items in a database with a time stamp. No other data were stored in the database. No cookies were used. Temporary session variables were employed to allow Website visitors to get to the next (appropriate) page of the survey. However, these temporary session variables were not saved in the database or anywhere else. The Web developers did not provide access to the Website to the researchers except to allow them to see the survey pages just as respondents would, and to access a special Web page that allowed the researchers to download data by dates. The downloading of data allowed researchers to select a date range and download responses (in comma delimited files) for that range.

For this research, specialized survey software was developed by Darrell Butler and his team at *nHarmony*. The software was based on *FormTools*, a copyrighted system that enables web pages to verify input and provide feedback for improper formats and data types (e.g., numbers, but not letters), control which fields are mandatory and which are optional, and provides other helpful characteristics. The survey software defined contingences among and within items, provided a password to respondents so they could start and stop at their convenience, and managed the database. The designers hosted the database on a *Linux* server, and turned off logging and other automatic recording of interactions.

Prior to inviting our sample to participate, the program underwent quality assurance testing with the assistance of our team of consultants and other volunteers. They rated the survey on ease of accessibility, navigation, appearance, timing, overall satisfaction with the experience, and other variables that could affect response and dropout rates. The survey content also underwent an extensive evaluation by paid and volunteer consultants for clarity, readability, and uniformity in understanding, followed by quality assurance testing from at least a dozen off-site computer stations to ensure that the data were captured anonymously and accurately.

The final online survey was then pilot tested by 48 individuals representative of the actual sample. These data were examined carefully to catch any remaining problems. Pilot testing estimated that the survey took from 10 to 40 minutes to complete, depending on how many incidents involving scientific wrongdoing (if any) were reported and how much detail the respondents shared in the flexible text boxes.

Pre-notification invitations were mailed out in waves of around 800 to 900 each. The following week, each wave of recipients with working emails (as far as we could tell) was sent more information about the survey and the web link that allowed them to enter the survey.

The Survey Content

The final selection of misconduct and other forms of irresponsible science included ten categories of acts that could impact on the validity of the scientific record. We asked respondents if they had ever observed or heard about an act that fell into one of our specified categories and, if so, to tell us more about it in open text boxes. We also informed respondents up front that they would have the opportunity to share more than one experience. If the respondent had no experiences observing or hearing about any such acts in the context of his or her work, the survey looped to a page asking a few questions about opinions as to who has responsibility for maintaining integrity in science and a request for some basic demographic data.

Each category is described below in text similar to the description appearing on the survey form. (A complete copy of the survey items appear as Appendix A.)

- *Fabrication/Falsification:* Inventing data that were never actually collected; altering data that were collected; faking records; unjustifiable data removal or treatment of outlying data points.
- *Plagiarism*: The substantial copying of another's work without appropriate attribution; misappropriation of intellectual property.
- *Incompetence:* Examples include poor research design, methodology, or statistical procedure; inappropriate selection or use of a study technique due to insufficient skills or training.
- *Careless work habits*: Examples include sloppy record-keeping; haphazard data collection; cutting corners; inadequate monitoring of the project's progress.
- *Intentional bias*: Examples include: rigging a sample to maximize support for hypotheses; withholding methodology details; deceptive or misleading reporting of data or its interpretation.
- *Questionable publication practices/authorship*: Examples include publishing a paper or parts of the same study in different publication outlets without informing the readers; undeserved "gift" authorships; coerced authorship; omitting someone who deserved an authorship or other form of credit.
- *Inadequate supervision of research assistants.* Examples include giving assistants more responsibility than they are able or willing to handle, insufficient supervision of assistants' work.
- *Failure to follow the regulations of science.* Examples include sidestepping or ignoring the IRB or its directives; circumventing or ignoring human subject requirements with regards to informed consent, confidentiality, or risk assessment; inadequate care of research animals; violating federal research policy.
- *Difficult or stressful work environment that could have a negative impact on the research process*. Examples include mistreatment or disrespectful treatment of subordinates; sexual harassment or other form of exploitation; playing favorites and other factors that create poor morale or acting out by subordinates; using one's position to exploit another; conflicts with the administration or administrative policies.
- *A dishonest act indirectly related to being a researcher*. Examples include unreported conflict-of interest, such as a financial interest in the outcome of an experiment; misuse or misappropriation of grant funds; inflating, distorting, or including bogus accomplishments on a resume.

Those who selected a category were directed to items asking for more detail. These included:

- What position did the individual(s) who committed (or may have committed) this act hold at the time?
- What position relative to individual (or individuals) involved did you hold at the time?
- How did you become aware of possible irresponsible or unethical conduct?
- In your opinion, was the act committed with full knowledge that it was irresponsible or unethical by the individual (or at least one of the individuals, if more than one was involved)?
- Did you attempt to intervene or help address the problem in any way?

At this point, the survey directed respondents to different questions, depending on whether they intervened. Those who checked “yes” to the intervention question were next asked the following, with multiple answer options:

- What did you do?
- How did the intervention turn out?
- Did you personally experience any negative outcomes as a result of getting involved?
- Overall, how did you feel about the final outcome?
- If you had to do it all over again, would you have done anything differently? If yes, what?
- Please describe the event without any identifying information. (open text box)

Those who did offer incidents but did not take any action, the following questions, with multiple options, were asked:

- What prompted you to decide against becoming involved?
- Have you felt misgivings about not getting involved?
- Did you ever discuss the matter with others afterwards?
- Please describe the event including the advice or comments of others, without any identifying information. (open text box)

At this point, any respondent could cycle back and report another incident to us. The survey was programmed to accept up to 10 incidents from a single respondent.

All respondents, including those who reported no incidents, were asked 7 questions about scientific responsibility.

- Generally speaking, do you believe that researchers have an individual responsibility to try to actively correct or minimize problems whenever colleagues appear to have purposefully engaged in a serious form of research misconduct?
- Generally speaking, do you believe that researchers have an individual responsibility to get personally involved in correcting or minimizing problems whenever their colleagues appear to have purposefully committed a minor incident of irresponsible research conduct?
- Generally speaking, do you believe that researchers have an individual responsibility to report problems to the appropriate institutional office whenever their colleagues appear to have purposefully engaged in research misconduct?
- Generally speaking, do you believe that researchers have an individual responsibility to get personally involved in correcting or minimizing problems whenever their colleagues appear to have unintentionally done something that would affect the validity of their data, such as using the wrong statistic or data-gathering technique?
- Do you believe that most researchers in your institution would consider intervening or reporting if they noticed any incident of questionable, unethical, or irresponsible scientific practices?
- If strong evidence of a serious case of research misconduct were discovered and reported, do you think the office in your institution that is responsible for dealing with such matters would take appropriate and timely action?
- Has a colleague or assistant in your institution ever approached you for advice about how to handle an ethical issue related to his/her research?

At the close of the survey, respondents were asked to provide information about very general demographic items (e.g., years doing research, research category, research work setting, and their gender), thanked for their participation, given information about how to obtain the results of the survey when they became available, and queried about their interest in participating in a confidential qualitative interview to provide more information about their experiences or views on the topic of collegial intervention.[5]

IV. Hypotheses

[5] The interview project will be described in a separate report.

To reveal answers to the most essential questions regarding the extent and effectiveness of informal interventions in suspected cases of research wrongdoing, we tested 20 a priori hypotheses and a large number of post hoc hypotheses described in the results section.

A. Suspected Violator Status on Intervention Rate

We expect that intervening upon suspecting that one's own supervisor or senior colleague has committed a scientific error would feel more risky than intervening when the suspected violator has an equal footing or a junior colleague status. The least risky situation would likely involve one's own post-doc or supervisees (e.g., graduate or undergraduate research assistants) because such interventions fit with the legitimate role of supervisors. Furthermore, students and assistants will likely be perceived as less intimidating, and the potential for negative consequences (e.g., damage to one's career) would likely be perceived as lower in most cases.

Hypothesis A-1**.** *A negative relationship will exist between the status of suspected colleagues and attempts by respondents to intervene.*

Degree of Closeness of Involvement on Intervention Rate

Opportunities to intervene can involve individuals whose impact is direct and personal, such as close co-workers, victims of someone's irresponsible science, or those who might be blamed for the transgression. Or they can involve individuals one does not know personally, distant coworkers, or situations in which the one had no direct involvement. We expect that risk and unpleasant emotional factors may have less impact when suspected violators have greater personal and administrative distance from a potential intervener.

Hypothesis A-2. *Respondents will be more likely to have intervened when less involvement exists between themselves and the suspected colleagues involved.*

Intervention as a Function of the Level of Credibility of the Information

Previous research suggests that individuals have greater reluctance to take action when the information available to them appears ambiguous or possibly incorrect (e.g., *11, 12*).

Hypothesis A-3**.** *Among respondents who intervened, fewer will have become involved based on indirect or hearsay evidence compared to first hand or strong evidence.*

Intervention with One's Own and Someone Else's Supervisees

We expect that intervening with those over whom one has supervisory responsibilities (e.g., post-docs, graduate and undergraduate research assistants) carries less personal risk. Noticing, promptly correcting errors, and dealing with improper behaviors easily qualify as legitimate roles of those who supervise post-docs and other research assistants. Ignoring potentially serious problems with supervisees represents an abandonment of the supervisor/mentor role. However, it may feel more intrusive or unsafe to intervene with the supervisees of others.

Hypothesis A-4. *A higher rate of reported interventions will occur for respondents' own supervisees as compared to intervening with someone else's supervisees.*

B. Hypotheses Related to Satisfaction with Intervention Outcomes

Satisfaction with Formal Reporting v. Informal Intervention

When concerns about research wrongdoing come to the attention of persons higher in the organization, the matter could quickly escalate into a stressful situation. The administration might ignore, ostracize or retaliate against the "whistleblower."

Hypothesis B-1. *Respondents who reported a suspected colleague to their supervisor, superior, or appropriate office will rate the outcome as less satisfactory than will those who handled the matter informally.*

Informal Intervention Levels of Satisfaction

Anecdotal evidence suggests that informal interventions can produce successful outcomes by avoiding public criticism, censure, and humiliation for the perpetrator. Suspected violators may feel less defensive, more prone to listen, and even appreciative when quick remediation of a problem without undue stress or notoriety seems possible.

Hypothesis B-2. *Overall, informal interventions will lead to more reported favorable outcomes than unfavorable ones*.

C. Hypotheses Related to Misgivings for Failing to Intervene

Misgivings Over Failing to Confront Supervisees

Whereas confronting wrongdoing on the part of supervisees easily qualifies as part of a supervisor's duty, role responsibilities seem less clear when one has no supervisory obligations regarding junior, peer, or senior colleagues suspected of scientific wrongdoing.

Hypothesis C-1. *The rate of respondents' misgivings for failing to intervene will be greater when suspected violators are those over whom one has supervisory responsibility as compared to colleagues over whom one has no supervisory responsibilities.*

Misgivings Over Failing to Confront the Incidents Perceived as More Serious

Misconduct, as discussed earlier, is currently largely defined as premeditated dishonesty involving fabrication and falsification, and plagiarism. The less frequently discussed forms of research wrongdoing, including unintentional behaviors that contaminate the validity of data, have yet to attract wide acknowledgement in the scientific ethics literature. Examples include carelessness, inadequate supervision of research assistants, difficult/stressful work environment, and incompetence.

Hypothesis C-2. *The rate of misgiving for failing to intervene will increase when the infraction qualifies as one traditionally considered more serious (i.e., FFP and failure to follow the rules of science) than for those infractions less often discussed.*

Misgivings Over Failing to Confront Those with Whom One Has Direct Involvement

Avoiding unpleasantness or conflict in the workplace is understandable, and ignoring scientific misbehavior remains one way of maintaining the status quo. Those who turn away may have lingering concerns or twangs of conscience.

Hypothesis C-3. *Those respondents more directly involved with the suspected violator (close co-worker, those who fear becoming a victim or of getting blamed, or serving as the administrator) will express greater misgivings about not intervening than will those who did not get involved but were, at some level, removed from the suspected violator.*

D. Hypotheses Related to the Role of Perceived Personal Responsibility

The literature on bystander intervention research reveals that people get involved when they notice the situation, recognize it as one requiring intervention or assistance, and conclude that they have a responsibility to act (*13*,*14*). Based on selective attention theory (e.g., *15*), those who do not accept a degree of personal responsibility may be less likely to notice irresponsibility. (However, a recent study has concluded that those who do not cheat appear less likely to believe that others cheat, which constitutes a confounding application of selective attention theory (*16*).)

Responsibility Level of Those Who Did Not Intervene

Hypothesis D-1. *Respondents who observed, suspected, or heard about irresponsible acts but did not intervene will rate others as less likely to intervene than will respondents who intervened.*

Hypothesis D-2. *Respondents who believe that the monitoring of research integrity of others is an individual responsibility will likely have intervened more often than those who see themselves as having less individual responsibility.*

The Role of Perceived Institutional Support on Those Who Did Not Intervene

The often ill-fated consequences of whistle-blowers are well-documented and well-publicized. Without a perceived likelihood of support, the inclination to intervene is lower.

Hypothesis D-3. *Respondents who have suspected irresponsible acts but did not intervene will be less likely to believe that the institution would take timely action.*

The Role of Years of Experience on Individual Responsibility

We expect those with more years of research experience will have higher expectations for the integrity of those in their workspace or in their field, based on the assumption that they have seen more and better understand the damage that research wrongdoing can do to the scientific record.

Hypothesis D-4. *More experienced investigators will more likely report a higher level of individual responsibility to intervene than will respondents with less experience.*

Individual Responsibility When Errors Seem Intentional or Unintentional

It seems reasonable to believe that confronting an individual who has made a mistake but does not realize it would prove easier than confronting someone who seems purposely engaged in research wrongdoing. Although ignorance does not always qualify as an excuse, people usually consider it a mitigating factor in making judgments, possibly allowing for a gentler confrontation as well as expecting a more positive response.

Hypothesis D-5. *Respondents' ratings of individual responsibility to intervene will rank higher when the error seems unintentional.*

E. Hypotheses Related to the Number of Events Observed

Generally speaking, biomedical research allows an opportunity for investigators to observe tangible data sources (e.g., tissue samples, medical record transcription inaccuracies, or patient improvement) more often than for social and behavioral research and many other types of research procedures involving such techniques as surveys, interviews, or performance scores. For example, survey takers leave without ever having their actual identities recorded. Such data are ripe for undetectable tampering by malintentioned investigators.

Hypothesis E-1. *Those who define themselves as biomedical researchers will report knowing or hearing about more suspicious acts overall than will other types of researchers.*

Errant behavior may be more difficult to observe first-hand. However, stories about research wrongdoing may circulate widely in the workplace. Because so few cases of research wrongdoing ever find their way to a higher authority compared to the number of reported instances on anonymous surveys, we expect that a major reason could be that the evidence did not seem solid enough to report.

Hypothesis E-2. *Respondents will report more instances of credible secondhand information or hearsay or gossip than instances involving direct observation or strong evidence.*

Research assistants will likely work in more visible ways and settings than one's senior colleagues. As a result, observing assistants' work will typically occur in more close confines and circumstances. Some errors will be expected in students and assistants and those that occur will be more easily noticed, as compared to peer colleagues.

Hypothesis E-3. *A higher rate of incidents will be reported for one's own research assistants and post docs than for other professional colleagues.*

F. Hypotheses Relating to Discussing the Matter with Others

Formal reporting may feel more stressful and involved than seeking an informal solution.

Hypothesis F-1. *Most respondents who intervened with their own colleagues (junior, senior, and post docs) will not have reported the incident to a supervisor or the appropriate office for reporting research wrongdoing*.

Respondents who observe fabrication or falsification--the forms of research wrongdoing considered most serious--may not feel comfortable intervening on their own. We expect that those who intervene in incidents traditionally defined as more serious will be more likely to discuss the situation with others.

Hypothesis F-2. *Those who intervene when the act traditionally discussed as more serious (i.e., fabrication and falsification, plagiarism) are more likely to have discussed the incident with others as compared to those who intervene when the act falls among those less frequently discussed (i.e., carelessness, poor supervision of research assistants, and difficult and stressful work environment).*

We decided to test for sex differences, and hypothesized that women would more likely consult with others about an incident of suspected research wrongdoing than would men.

Hypothesis F-3. *Female respondents will be more likely to have discussed the matter with others than will male respondents.*

V. Results

A. Descriptive Data

This section presents our descriptive data tables and figures. Based on a review of descriptive data, we tested a number of post hoc hypotheses. Post hoc hypotheses and findings are included with the results from our original planned comparisons in the next section.

From here on, we will refer to our respondents who took some form of action as **CRs** (for **"Concerned Researchers"**). Those suspected of purposeful or unintended scientific errors will be referred to as **SVs** (for **"Suspected Violators"**). The term "intervention" generally applies to a CR taking some form of action to prevent, minimize, or report an act perceived to be scientifically irresponsible. Those **respondents who did not intervene will simply be referred to as "non-interveners."**

It must be noted that respondents could often enter more than one incident or check more than one response to a single question. Missing data precluded complete data sets for some tables.

Our Sample

Although invitations were sent electronically to our original sample pool of 11,922, the "bad email" and "bounce rate" notification ran high (58.9%), leaving us to presume that at best only 4900 invitations to participate may have reached valid email accounts. We decided to create another random sample directly from the 2005 and 2006 CRISP online lists, adding 4,357 more unique email addresses of which 1,003 bounced. Thus, the total number of email invitations presumed to have reached their destinations was 8,254. We received 3,309 (40.1%) responses, of which 2,599 (31.5%) contained sufficient data to meaningfully analyze. (One can reasonably assume that true response rate may be higher as firewalls may have blocked an unknown number of invitations to participate, some messages may have reached accounts no longer monitored by the intended recipient, etc.)

Of the 2,599 Principal Investigators who supplied usable data for our online survey, the majority (63.7%) had worked as active researchers for over 15 years, 19.7% for 10 to 15 years, 12.8% for 5-9 years, and 3.8% for 1-4 years. The majority were men (61.2%). The majority of our respondents (67.9%) had earned PhD degrees, 26.6% had MD degrees, and the remainder split among other advanced degrees (e.g., MPH, DSW, DNSc). We asked respondents how they defined themselves professionally: 37.1% defined themselves as biomedical researchers, 31.5% as medical researchers, and 19.8% as social/behavioral researchers. Smaller percentages identified with public health (4.0%), nursing (3.3%) or "other" (4.3%). We asked respondents to categorize their research specialty. The vast majority selected "biomedical" (75.2%), followed by social/behavioral (22.5%), with the small remainder choosing educational, public policy, or human factors/industrial.

The most common work settings were: research universities (59.0%), followed by hospital or medical settings (24.5%), private agencies (5.7%), comprehensive universities (4.1%), private industry (3.8%), with small numbers from other settings (four-year college, community clinic or agency, or private corporation).

Because our respondents were grant-funded PIs, we wanted to learn if their colleagues and assistants consulted with them on ethical matters. When asked if they had been approached for ethics-related advice, 46.2% said "no," 26.9% said "once or twice," "20.9% "occasionally," "and 6.0% "often" or "very "often."

Number of Incidents Reported

All but 406 of our respondents reported at least one incident of perceived scientific wrongdoing or errors. Fifty three percent of those reporting an incident chose to convey a single one, although the survey instrument was programmed to collect up to ten incidents. Frequencies of all respondents, including those who had no incidents to share, appear as Table 1.

Table 1. Numbers of Incidents Reported by Respondents

Number of incidents reported per respondent	Frequency	Percent including all respondents
0	406	15.6
1	1381	53.1
2	498	19.2
3	193	7.4
4	68	2.6
5	30	1.2
6	12	.5
7	11	.4
Total number of respondents	2599	100.0

Types and Prevalence of Irresponsible Activity and Interventions

The first question appearing on the online survey asked respondents if they had an incident of poor scientific practice to describe and, if so, to select one of 10 categories that best represents it. Table 2 presents the categories in the order of the number of shared incidents.

Table 2: Types and Frequencies of Adverse Events

Rank	**Type** (by number of respondents reporting)*	**Frequency of incidents****	**%**
1	Fabrication/falsification (583)	608	17.3
2	Questionable publication practices (562)	601	17.0
3	Plagiarism (444)	464	13.1
4	Difficult or stressful work environment (402)	434	12.3
5	Incompetence (413)	420	11.9
6	Carelessness (328)	334	9.5
7	Dishonesty indirectly related to the conduct of research (178)	185	5.2
8	Intentional bias (169)	176	5.0
9	Failure to follow the rules of science (162)	169	4.8
10	Inadequate supervision of assistants (132)	136	3.9
	*Respondents could report more than one type of incident. (3373)	**Respondents could report more than one incident in each category (3527)	100

Because respondents selected both what types of incidents as well as how many to share with us, we cannot draw reliable conclusions regarding overall prevalence of each type of wrongdoing. However, we expected the most frequently selected categories to be fabrication/falsification and plagiarism, those acts falling under the formal definition of “research misconduct.” These acts did represent about a third of the incidents reported. We found it to be of considerable interest, however, that the rates of less-discussed sources of research errors and wrongdoing hardly proved inconsequential,

confirming an impression that inspired this project.

Intervention Rate

Interveners were defined as those who responded "yes" to the question, "Did you attempt to intervene or help address the problem in any way?" Our most compelling finding: almost two-thirds our respondents who had at least one incident to share (63% or 1,386) claim to have intervened in at least one incident when they suspected wrongdoing, leaving only a little over a third (807) who never intervened on any shared incident.

Who Were the Suspected Violators?

Our data cannot tell us what group offends more than any other because respondents could choose which incidents from their experiences to share. As can be seen in Table 3, respondents often informed on higher status persons: about one third of the incidents involved supervisors and senior colleagues as SVs.

Table 3. Suspected Violators' Position Relative to the Respondents'

Rank	Who committed the act?*	Incident frequencies by position**	%
1	Peer colleague (874)	1144	31.1
2	Senior colleague (702)	939	25.5
3	Junior colleague (375)	433	11.8
4	My supervisor (216)	273	7.4
5	Someone else's grad assistant (230)	247	6.7
6	Someone else's post doc (204)	228	6.2
7	My grad assistant (146)	157	4.3
8	My post doc (129)	141	3.8
9	My undergrad assistant (57)	61	1.6
10	Someone else's undergrad assistant (55)	59	1.6
	*Respondents could report more than one SV per incident (2988)	**Respondents could report more than one incident (3682)	100.00

How Respondents Became Aware of Incidents

As shown in Table 4, awareness that irresponsible science occurred most often (62%) through direct channels (i.e., direct observation or direct evidence). Whereas we might assume that rumor and hearsay are frequent channels of information about wrongdoing, our respondents did not choose to share many such incidents if they knew of any.

Table 4. How Respondents Became Aware of the Adverse Events

Rank	Source of awareness*	Total Number of incidents**	%
1	Direct observation (679)	908	32.2
2	Direct evidence (711)	838	29.8
3	Credible second-hand information (663)	760	27.0
4	Direct disclosure by persons involved (170)	187	6.6
5	Hearsay or rumor (115)	122	4.3
6	Other (4)	4	.1
	*Respondents could report multiple incidents (2563)	**Respondents could select more than one source per incident (2819)	100.0

Respondents' Role in Reported Incidents

The association between the CR respondents and the SVs appears in Table 5.

Table 5. Respondents' Role Relative to the Adverse Event

Rank	Number of times each category was reported by respondents*	Incidents frequency**	Percent
1	I was not involved (971)	1216	31.7
2	I was a close co-worker (663)	842	21.9
3	I was a distant co-worker (388)	463	12.1
4	I was a victim (355)	421	11.0
5	Someone close to the project confided in me (237)	271	7.1
6	I didn't know the individual(s) personally (230)	243	6.3
7	I was the administrator (192)	225	5.9
8	I might get blamed (137)	154	4.0
	*Respondents could report more than one incident (3175)	3835* *Respondents could select more than one answer per incident	100.0

Because a high rate of respondents indicated that they had no personal involvement in the incident, we looked at whether this group had reported incidents that they knew about only through hearsay and therefore were not in a position to act. Table 6 parses out those who reported no direct involvement and describes how they learned about the incident.

Table 6. Source of Awareness by Those Not Personally Involved in the Incident
(all reported incidents combined)

Source of Awareness	Percentage of total incidents in which respondents described role as "NOT personally involved"	Total Incidents
Direct (observation and SV confessions)	522 (27%)	1933
Credible Second-hand info	425 (59%)	760
Hearsay/Gossip	70 (57.4%)	122
Other	2 (50%)	4
Total	1019 (36.1%)	2819

Table 6 shows how respondents became aware of reported incidents in which they were not personally involved. Overall, those respondents who reported that they had no direct involvement in the incident were more likely to have come by the information indirectly, split almost evenly between credible second hand and hearsay when compared to the total number of incidents. However, curiously, over a quarter of those who claim they were not personally involved came by the information directly. It is not clear whether these respondents chose not to become involved, or whether they interpreted involvement as directly affecting them (independent of their own response to the incident).

Did the SV Commit the Irresponsible Act with Full Knowledge of Its Inappropriateness?

In 58.2% of the incidents, as seen in Table 7, our respondents felt certain, or reasonably sure, that the SVs knew they were committing an irresponsible act, although only about a third felt "definitely" sure. In only 7% of the events did our respondents believe the alleged SVs remained completely unaware of making a scientific error. Looking at the data another way (and including the "probably yes" numbers), 69% did not feel absolutely sure that they knew if the SVs were fully culpable.

Table 7. Respondents' Opinions Regarding Motive of SVs

Rank	Awareness Level of SVs*	N**	%
1	Definitely Yes (762)	1021	31.4
2	Probably Yes (729)	871	26.8
3	Not sure (586)	663	20.4
4	Probably not (414)	460	14.2
5	No (222)	233	7.2
	* Respondents could report more than one incident and use a different category for each incident (2713)	3248** respondents could report more than one incident	100.0

What Actions Did Respondents Who Intervened (CRs) Take?

Table 8 shows the various actions that CRs took, from most to least frequent. The most frequent action was going straight to the SV. However, if we combine three categories (i.e., discussed with trusted

individual, supervisor, or family and friends) discussions of the matter with others would comprise the most frequent type of action taken. (Remember, respondents could select more than one action.)

Table 8. Intervener Actions

Rank	Actions Taken by CR Respondents (by number of respondents)*	Numbers of each action taken**	%
1	Discussed directly with individual(s) involved (691)	828	23.9
2	Discussed with trusted individual (not supervisor) (555)	665	19.2
3	Thought long and hard first (358)	422	12.2
4	Discussed with family/friends (259)	301	8.7
5	Reported to appropriate administrative office (257)	284	8.2
6	Reported to supervisor/superior (244)	281	8.1
7	Sent a signed message (243)	268	7.7
8	Discussed with involved individual(s) with others present (200)	224	6.4
9	Discussed with supervisor/superior (326)	172	4.9
10	Sent an unsigned message (22)	24	0.7
	* Respondents could take more than one action in each incident (3155)	3469 **Respondents could report more than one incident	100.0

Among 565 (284 + 281) incidents in which the CR reported the incident to a supervisor or administrator, 122 (21.6 %) also approached the SV by themselves, 40 (7 %) spoke with the SV as part of a group, and 42 (7.4 %) spoke with the SV by themselves and also as part of a group. Thus, there were 565 incidents of reporting to a superior and 1052 incidents in which the CR discussed the incident with the SV either alone (828) and/or with others (224). All told, in the 3469 actions about which we have complete data, there were 204 incidents in which CRs reported to an authority as well as discussing the incident with the SV either alone or with others present.

How Did the Intervention Turn Out?

Table 9 reveals 447 incidents resulting in a correction of the problem. Very occasionally (19 incidents) the CR's concern was unwarranted. A smaller percentage of incidents were reported "up the line." Because respondents could report more than one incident as well as more than one answer per incident, the total number of responses (2207) exceeded the total number of incidents in which one of the listed outcomes occurred (1586). Table 9 shows percentages for both responses and incidents, with percentage based on incidents summing to more than 100%.

Table 9. Results of the Intervention

Rank	Outcome* (by number of respondents)	Frequencies of incident outcomes	% of responses	% of incidents
1	SV(s) corrected the problem (394)	447	19.7	28.2
2	SV(s) denied the problem (377)	430	19.0	27.1
3	Event elevated to a local office (285)	321	14.1	20.2
4	SV(s) did nothing to correct the problem (263)	310	13.7	19.5
5	I gave the individual(s) a chance to save face (215)	232	10.2	14.6
6	The SV(s) understood but there was no way to correct it (199)	209	9.2	13.2
7	The individual(s) did not respond (143)	168	7.4	10.6
8	The incident was elevated to the ORI/federal level (128)	134	5.9	8.4
9	My concern turned out to be unwarranted (18)	19	0 .1	0.1
	* Respondents could select more than one answer per incident in which they took action (2022)	2270 **Respondents could report more than one incident	100.0	141.9 (1586)

Table 10 breaks down outcomes for only those interventions where the respondents believed that the suspected SV lacked awareness of making a mistake.

Table 10. Results of Intervention in Cases When Suspected Violators May Have Been Unaware of Wrongdoing (all incidents combined)

Outcome Ranked	Frequencies of incident outcomes	%
SV(s) corrected the problem	204	39.6
The SV(s) understood but there was no way to correct it	72	14.0
SV(s) denied the problem	61	11.9
I gave the individual(s) a chance to save face	50	9.7
SV(s) did nothing to correct the problem	49	9.5
The individual(s) did not respond	28	5.4
Event elevated to a local office	27	5.3
The incident was elevated to the ORI/federal level	12	2.3
My concern turned out to be unwarranted	12	2.3
Total	515	100.0

(Levels of satisfaction among CRs for all of the above categories appears as Table 13.)

Satisfaction With Intervention Outcomes

As seen in Table 11, in a majority of the incidents (64.6%) the CRs did not rank their involvement as “unsatisfactory.” Over a third of the incidents found respondents feeling dissatisfied with the outcome,

although only 11% involved extreme dissatisfaction.

Table 11. Satisfaction with Intervention Outcome

Rank	Outcome Satisfaction Level among CRs*	Frequency of incidents**	%
5	Extremely satisfied (102)	106	5.5
1	Satisfied (554)	643	33.5
2	Neither satisfied nor unsatisfied (445)	491	25.6
3	Dissatisfied (413)	467	24.4
4	Extremely dissatisfied (165)	211	11.0
	*Respondents who reported and rated the satisfaction level of at least one incident (1679)	**1918** **Respondent could report more than one incident	100.0

Tables 12 and 13 present the means and standard deviations for the CR satisfaction levels (1=very dissatisfied to 5= very satisfied) as they relate to the actions that respondents took and how they turned out. Although we offered no a priori hypotheses, these figures give us a clue as to how satisfaction relates to what the CRs did and how it turned out.

Table 12. Satisfaction Level Associated with Each Intervention Action Taken in First Incident[6]

Actions Taken by CRs in First Incident (*N*)	Satisfaction	
	Mean	SD
Sent an unsigned message (11)	3.36	1.07
Discussed with involved individual(s) with others present (106)	3.27	1.06
Discussed directly with individual(s) involved (452)	3.11	1.06
Reported to appropriate administrative office (178)	3.07	1.13
Discussed with trusted individual (not supervisor) (375)	2.92	1.07
Sent a signed message (144)	2.90	1.14
Reported to supervisor/superior (152)	2.81	1.17
Thought long and hard first (230)	2.80	1.06
Discussed with supervisor/superior (208)	2.68	1.08
Discussed with family/friends (157)	2.56	1.11

It is of interest to note that the small group of those most satisfied wrote anonymous notes. This is not recommended, however, as it does not allow the SV due process and, if the suspicion is in error, constitutes a moral failure.

[6] For some data presentations we used only the first incident that respondents shared so that each respondent was represented only one time.

Table 13. Satisfaction Level Association with Each Intervention Outcome for First Incident

Outcome of CR Intervention in First Incident (*N*)	Satisfaction	
	Mean	**SD**
My concern turned out to be unwarranted (12)	4.08	0.67
SV(s) corrected the problem (253)	3.80	0.74
The incident was elevated to the ORI/federal level (105)	3.27	1.03
The SV(s) understood but there was no way to correct it (135)	3.11	0.89
I gave the individual(s) a chance to save face (129)	3.08	1.04
Event elevated to a local office (183)	3.04	1.16
The individual(s) did not respond (84)	2.56	0.93
SV(s) denied the problem (221)	2.53	1.04
SV(s) did nothing to correct the problem (157)	2.17	0.90
Responses are mutually exclusive		

Apparently the CRs felt relief upon learning that their suspicions were incorrect. Yet, as would be expected, greater satisfaction is associated when the concern is remedied.

Negative Fallout Following Intervention

As can be seen in Table 14, in a slim majority of incident outcomes respondents reported experiencing no negative feedback as a result of getting involved or that they gained in stature. The remaining CRs reported negative experiences.

Table 14. Intervention Aftermaths

Rank	Fallout?*	Incidents**	%
1	No negative fallout (923)	1169	42.2
2	I gained respect (261)	296	10.7
3	Disrespect (255)	295	10.7
4	Emotional costs only (264)	293	10.6
5	Career status jeopardized (158)	195	7.0
6	Social costs (146)	169	6.1
7	Loss of reputation (146)	167	6.0
8	No support from institution (118)	140	5.1
9	Feared legal action (39)	43	1.5
	*Respondents could select more than one option when describing a single incident (2310)	2767 **Respondents could report more than one incident	100.0

Would Respondents Who Intervened Do Anything Differently?

We asked CRs if they had to go through the intervention process again, would they have done something else instead. In at least one of the incidents reported, 523 (32.9%) respondents said “yes” and in at least one incident reported 1066 (67.1%) CRs said “no.”

We gave these respondents who checked that they would have done something differently the opportunity to use a text box to tell us more in their own words about what they would have done instead. A content analysis of the 123 who chose to comment revealed several themes, listed here in the order of frequency:

- would have been more prepared,
- would have acted more forcefully earlier,
- would have anticipated problems earlier,
- would have not done it alone; would have reported directly to a proper administrative or oversight office,
- would have left my job,
- still don't know, but would have done something differently,
- would have done nothing,
- would have put preventive measures in place to preclude what happened,
- would have documented the incident better, and
- would have engaged legal counsel.

It is of interest to note that only about 10% of those who supplied reasons would have done nothing whereas most would have been even more proactive.

Reasons for Not Intervening

Table 15 lists the reasons in order of frequency selected for *not* getting involved. We found the most frequently selected reason for not getting involved was, "not my problem." Table 16 examines whether they were somehow involved in the incident, and it appears that most were not.

Table 15. Why Non-intervening Respondents Chose Not to Get Involved

Rank	Reason *	N	%
1	Not my problem to solve (264)	305	12.4
2	Someone else took care of it (253)	276	11.3
3	Individual(s)s were difficult (199)	233	9.5
4	Individual(s) was my superior (168)	194	7.9
5	Evidence was insufficient (156)	169	6.9
6	I feared no institutional support (100)	163	6.7
7	I considered several options but no satisfactory answer presented itself (142)	155	6.3
8	Risk to career status (131)	151	6.2
9	I could not think of how to respond (127)	147	6.0
10	Not a serious enough matter (134	137	5.6
11	I did not need the aggravation (110)	130	5.3
12	Risk to my reputation (94)	114	4.7
13	I might be a target (91)	108	4.4
14	I did not want to risk the social costs (66)	72	2.9
15	Individual(s) was also my friend(s) (61)	70	2.8
16	I feared getting sued (23)	26	1.1
	*Respondents could select more than one reason per incident (2119)	2450 **Respondents could relay more than one incident	100.0

Among those who chose "not a serious enough matter" as a reason for not getting involved, curiously 9 involved incidents of fabrication/falsification. In 12 of the incidents (8.8%) CVs were referring to plagiarism. In the remaining 116 cases CVs were referring to other categories of research wrongdoing.

Table 16. Respondents' Role Among Non-interveners Who Stated the Incident was "Not My Problem to Solve" (all incidents combined, *N* = 305 respondents)

Category of Involvement	Incidents frequency	Percent
I was not involved	203	55.4
I was a distant co-worker	49	13.4
I was a close co-worker	44	12.0
Someone close to the project confided in me	38	10.4
I didn't know the individual(s) personally	26	7.1
I was a victim	5	1.4
I was the administrator	0	0
I might get blamed	1	0.3
Total	366	100.0

Did Noninterveners Regret Not Getting Involved?

Table 17 describes the impact of failing to intervene when research wrongdoing was suspected. Among those respondents who contributed 490 incidents involving *direct* evidence, less than 6% who did not get involved reported having regrets and know what they should have done. However, a third had regrets but did not know what to do. A slim majority felt no misgivings.

Among those respondents who contributed 708 incidents involving indirect evidence of wrongdoing, such as gossip and rumor, most expressed no regrets (79%). Still, for about a quarter of the incidents reported, misgivings were still felt and most of these involved respondents who felt uncertain about what to do.

Table 17. Rate of Misgivings for Not Intervening

		Felt misgivings about not getting involved? (by incident)				Total
		Yes, realize what should have been done	Yes, but still not sure	Only at first	No	
How became aware	Direct evidence	27 (5.5%)	165 (33.7%)	22 (4.5%)	276 (56.3%)	490
	Indirect evidence	7 (1%)	122 (17.2%)	16 (2.3%)	563 (79.5)	708
Total		34	287	38	839	1198

Did Non-interveners Discuss the Adverse Incident with Anyone?

We asked non-interveners if they later discussed the matter with others. Almost two-thirds of the noninterveners, (712 or 62.7%), indicated that they had done so for at least one incident whereas 424 (37.3%) did not discuss any incident with anyone. Unfortunately our survey did not request information about how those discussions turned out. Perhaps many were talked out of taking action.

B. Tested Hypotheses

Predictions of Intervention Rate

Multilevel modeling (MLM) was initially attempted to analyze intervention rate, based on the first two rounds of incidents[7] as the first level of the hierarchy and respondent identification as the second

[7] Respondents could covey more than one incident. Most offered only one. The hypothesis testing was restricted to the first incident conveyed, occasionally extended to the second wave (the second story offered).

level of the hierarchy. However, the intra-class correlation of .04 makes the need for multilevel modeling questionable. Therefore, responses were combined for the first two incidents, ignoring dependencies created by treating all data as if from different respondents in a logistic regression analysis. The results of this logistic regression analysis did not differ substantively from those of MLM, and the statistical assumptions of logistic regression are better met in these data than those of MLM. Therefore, we report a simpler binary logistic regression analysis here.

Predictors of intervention included ratings of *status of SV* (a 7-point scale in descending order: my supervisor, senior colleague, peer colleague, my post doc, my graduate assistant, my undergraduate assistant); *available information* (a 3-point scale in increasing order: hearsay or rumor, credible second-hand information, directly disclosed or individual disclosed or direct evidence); *perception that the suspected error was committed intentionally* (5-point scale in increasing order: no, probably not, not sure, probably, definitely yes); *distance of respondents' involvement* (distant = not formally involved, distant co-worker, didn't know people; close = close co-worker, victim, might be blamed), and *whether the SV was a colleague or supervisee*. The 10 categories of actions that occurred (fabrication/falsification, plagiarism, incompetence, carelessness, intentional bias, questionable practices, inadequate supervision of assistants, failure to follow the rules of science, difficult/stressful work environment, dishonest act indirectly related to research) also served as a set of predictors. These 10 items were coded into 9 dichotomous variables, such that each was compared with the last item (dishonest act indirectly related to research). The total *N* was 1709 incidents, of which about one-third were second incidents. The large amount of missing data was due primarily to the failure to report whether there was an attempt at intervention (the DV), therefore no attempt at imputation was made.

Table 18 shows the contribution of the individual predictors to the model in terms of regression coefficients and Wald tests of them as well as odds ratios and 99.7% confidence intervals around them. To control for familywise Type I error rate, alpha = .004 for individual predictors. Three of the items reached this criterion, all of them congruent with the hypothesized direction. Odds of intervening decreased with increasing status of SV; the odds of intervening were about 40% less for intervention when SVs held higher status. Table 19 (a) shows intervention rates ranging from about 49% for senior colleagues to about 93% for one's own undergraduate students and post-docs. Hypothesis A-1, predicting a negative relationship between the status of suspected SVs and respondent interventions was confirmed. Table 19 also suggests that respondents may have been more willing to state "not my problem to solve" as a reason for non-intervention with colleagues of same or higher status than themselves than for their own supervisees.

Table 18. Logistic Regression of Intervention (N = 1709)

Predictor	Regression coefficient (B)	S.E.	Wald χ^2 (df = 1)	*p*	Odds Ratio	99.7% C.I. for Odds Ratio	
						Lower	Upper
Fabrication/falsification	-.06	.29	.04	.852	.95	.39	2.26
Plagiarism	-.12	.30	.15	.698	.89	.36	2.18
Incompetence	.14	.31	.21	.645	1.15	.46	2.92
Carelessness	-.03	.33	.01	.931	.97	.36	2.61
Intentional bias	-.30	.35	.72	.396	.74	.26	2.10
Questionable practices	-.37	.28	5.67	.017	.51	.22	1.18
Inadequate supervision of assistants	.26	.38	.47	.491	1.54	.42	4.04
Failure to follow rules of science	.43	.38	1.31	.253	1.72	.50	4.77
Difficult/stressful work environment	.54	.30	3.17	.075	1.83	.69	4.22
Status	-.41	.07	46.77	<.001	.60	.48	.75
Closeness	-1.31	.13	109.98	<.001	.27	.19	.39
Colleague vs. own assistant	.01	.36	<0.01	.989	1.01	.34	94
How became aware	-1.09	.11	90.50	<.001	.34	.24	.47
Done with full knowledge	.04	.05	.56	.455	1.04	.89	1.22
Constant[8]	6.51	.67	94.30	<.001			

There were no prior hypotheses about which actions might result in higher rates of intervention. Table 18 shows that odds of intervening did not differ significantly for any of them at alpha = .004. That is, intervention rate did not depend on the type of wrongdoing when taking into account the other five predictors of intervention.

Hypothesis A-2 predicted that respondents will intervene when less direct involvement exists between them and the SVs. As expected, closeness to the SV was negatively related to intervention; odds increased by about 73% for SVs distant to the respondent. Table 20 indicates about an 80% intervention rate for distant SVs vs. a 43% intervention rate for closer colleagues. Hypothesis A-3 predicted a negative relationship between intervention and credibility of the evidence. As expected, as rounds of information became less credible (indirect or hearsay), odds of intervention decreased by about two-thirds. Table 21 shows that intervention rate decreased from about 74% with direct information to about 20% for hearsay.

[8] The constant basically tests whether there is equal likelihood of intervention or not (in regular regression, the constant tests whether the mean response differs from zero). Its usefulness in logistic regression is in building a model to predict an individual's response on the basis of all the predictors. The constant is added to the sum of products of B weights times the scores on each of the predictors and converted into a probability of intervention.

The next set of four tables (Tables 19-22) presents cross-tabulation with "Intervention for Statistically Significant Predictors" (first and second incidents combined.) *N*s differ because all available data are used for each cross tabulation.

Table 19. Status of Suspected Violator by Attempt to Intervene

Status of SV	Intervention Attempted?			Total (100%)
	No, "Not my problem"	No, other reason	Yes	
my RA undergrad	0 (0%)	3 (7%)	42 (93%)	45
my RA grad	0 (0%)	9 (8.4%)	98 (91.6%)	107
my post doc	1 (1.1%)	5 (5.6%)	84 (93.3%)	90
junior colleague	2 (0%)	73 (24%)	238 (76%)	313
peer colleague	19 (1.4%)	308 (3.8%)	475 (59.2%)	802
senior colleague	16 (2.7%)	284 (48.3%)	288 (49%)	588
my supervisor	3 (2.2%)	48 (35%)	86 (62.8%)	137
Total	41 (2%)	730 (35%)	1311 (63%)	2082

Table 20. Closeness to Suspected Violator by Attempt to Intervene

Closeness to SV	Attempted to intervene?		Total
	No	yes	
distant	223 (29.3%)	878 (79.7%)	1101
close	678 (57.2%)	507 (42.8%)	1185
Total	901	1385	2286

Table 21. How Respondents Became Aware by Attempted to Intervene

How respondents became aware	Attempt to intervene?		Total
	No	yes	
Direct	400 (25.7%)	1158 (74.3%)	1558
Second hand	395 (59.5%)	269 (40.5%)	664
Hearsay	84 (80%)	21 (20%)	105
Total	879	1448	2327

Table 22. One's Own and Another's Supervisee by Attempt to Intervene?

Whose supervisee	Attempt to intervene? no	Attempt to intervene? yes	Total
Other's	208 (46.8%)	236 (53.2%)	444
Own	21 (3.9%)	229 (91.6%)	250
Total	229	465	694

As can be seen in Table 23 intervention became more likely when the respondent believed that the SV was unaware of the irresponsibility of the act than when the respondent believed that the SV was probably unaware, $\chi^2(1, N = 689) = 6.65, p = .01$.

Table 23. Attempt to Intervene When Suspected Violator Was Unaware of Mistake (all incidents combined)

Awareness	Attempted to intervene? No	Attempted to intervene? Yes	Total
Probably not aware	151 (33.1%)	305 (66.9%)	456 (100%)
Not aware	55 (23.6 %)	178 (76.4%)	233 (100%)

Table 24 presents the types of bad science and the rate of intervention for those who did intervene and those who did not.

Table 24. Attempts to Intervene by Type of Wrongdoing

Action	Attempt to intervene? no	Attempt to intervene? yes	Total total
Fabrication/falsification	260 (36.7%)	284 63.3%	544
Plagiarism	140 (32.5%)	291 67.5%	431
Incompetence	106 (33.4%)	211 66.6%	317
Carelessness	67 (27.1%)	180 72.9%	247
Intentional bias	56 (47.1%)	63 52.9%	119
Questionable publication practices	221 (49.4%)	226 50.6%	447
Inadequate supervision of assistants	33 (34.7%)	62 65.3%	95
Failure to follow rules of science	32 (29.9%)	75 70.1%	107
Difficult/stressful work environment	86 (30.6%)	195 (69.4%)	281
Dishonesty unrelated to research	33 (35.7%)	51 64.3%	84
Total	1034	1638	2672

A post hoc test was run to assess whether those who said they were victims or afraid of being blamed intervened or not. The results appear in Table 25.

Table 25. Involvement by Attempt to Intervene (all incidents combined)

Respondent's Role	Attempted to intervene? No	Attempted to intervene? Yes	Total
I was the victim	91 (22%)	323 (78%)	414 (100%)
I might get blamed	23 (15.5%)	125 (84.5%)	148 (100%)

A one-way, two-group chi-square indicated that a vast majority (80%) of those who felt like victims or who might be blamed reported that they intervened than did those who reported that they did not intervene ($p < .001$), suggesting that acts involving direct threat to the CR will more likely lead to action.

Finally, we did a post-hoc analysis to learn if our predictions as to who intervenes and who does not could be improved. A summary logistic regression model was developed on the basis of results of the previous logistic regression model and two additional variables to derive a prediction of intervention. The first five predictors are as described previously. Individual responsibility is the sum of the responses to four items relating to individual responsibility. Institutional responsibility is the rating of whether the

institution would take appropriate and timely action if strong evidence of a serious case of research wrongdoing were discovered and reported.

Table 26. Post Hoc Summary Logistic Regression of Intervention (*N* = 1553)

Predictors	Regression coefficient (B)	S.E.	Wald χ^2 (df = 1)	*p*	Odds Ratio	99.7% C.I. for Odds Ratio	
						Lower	Upper
Questionable publication practices	-0.77	.06	24.29	<.001	.46	.29	.74
Difficult/stressful work environment	0.66	.20	9.77	.002	1.88	1.03	3.43
Status	-0.51	.07	60.07	<.001	.60	.50	.73
Closeness	-1.36	.13	107.65	<.001	.26	.17	.38
How became aware	-1.10	.12	80.04	<.001	.33	.26	.42
Individual responsibility	0.20	.03	50.37	<.001	1.23	1.12	1.34
Institutional responsibility	0.18	.07	7.51	.006	.84	.69	1.02
Constant	4.09	.67	94.30	<.001			

Using a criterion for interpretation of α = .001, changes in the set of predictors permitted questionable publication practices to emerge as a significant predictor in this new model, reducing the odds of intervention by about half. Higher individual responsibility scores increased the odds of intervention by about 25%. There also is some suggestion that creating a difficult/stressful work environment may increase the odds of intervention and rating of institutional responsibility may decrease the odds of intervention.

Table 27 shows the accuracy of classification of intervention on the basis of the set of 7 items in the post hoc logistic regression analysis. About 85% of respondents who attempted to intervene in the incident were correctly classified, however only about 59% percent of classifications were correct for respondents who did not intervene in that incident, producing an overall correct classification rate greater than 75%. Thus, the logistic equation does a good job of predicting who *will* attempt to intervene, but a poor job of predicting those who *will not* attempt to intervene. That is, if non-intervention is predicted, that prediction will be wrong almost half of the time.

Table 27. Classification of Whether Respondent Intervened

Observed		Predicted		
		Attempted to intervene?		
		No	yes	Percent correct
Attempt to intervene?	No	335	230	59.3
Intervene?	Yes	144	825	85.1
Overall Percentage				75.6

One item left out of the foregoing analysis was the comparison of intervention rates for one's own vs. other's assistants (including post docs), because of collinearity with other items related to status of the SVs. This analysis pertains to Hypothesis A-4 that predicts a higher rate of reported interventions for respondents' own supervisees as compared to intervening with someone else's supervisees. A separate logistic regression analysis of intervention for the first two incidents was done using own vs. other's assistant as the sole predictor.

As hypothesized, a positive relationship existed between intervention rate and one's own supervisee, $\chi^2(1, N = 694)$, $p < .001$, Nagelkerke $R^2 = .23$ with 95% confidence limits from .18 to .29. Odds of intervention were 9.6 times greater for one's own supervisee than for a colleague's supervisee (with 99.7% confidence limits for odds ratio from 4.62 to 20.01), B=2.26, Wald $\chi^2(1, N = 394) = 122.25$, $p < .001$. Classification, however, was not improved over chance because rate of intervention was greater than 50% for all supervisees (one's own and those of colleagues); thus all respondents would be classified as interveners. The intervention rate for one's own supervisees was 92% and for colleague's supervisees intervention rate was 53%, as seen in Table 22.

Satisfaction with Intervention Outcomes

We defined satisfaction with outcome as the mean rating of that item on a scale of 1 (extremely dissatisfied) to 5 (extremely satisfied). Separate two-group between-subjects ANOVAs on satisfaction were run for those who intervened in the first or second rounds of shared incidents. Contrary to the hypothesis for the first incident, those who handled the matter informally and directly with the SV on their own or with other peers were a bit *less* satisfied ($M = 3.01$, $SD = 1.06$) than were those who formally reported (to a superior or the office to which research issues are reported) ($M = 3.39$, $SD = 1.07$), $F(1, 899) = 8.07$, $p = .01$, $\eta^2 = .01$ with 95% confidence limits from .00 to .02 However, for the second incident, there were no significant differences in satisfaction between those who handled the matter formally vs. informally, $F(1, 422) = 1.21$, $p = .34$, $\eta^2 < .01$ with 95% confidence limits from .00 to . Hypothesis B-1 was not supported.

We hypothesized that among those who intervened, informal handlers would report more favorable outcomes than those who reported to a superior or a research office. Table 28 shows cross-tabulations for the first two reported incidents. A 2 x 2 chi-square analysis for the first incident revealed no significant difference in favorability ratings, $\chi^2(1, N = 636) = 0.46$ with a lower 95% (one-tailed) confidence limit of zero for chi square. We found a similar result for the second round of incident, $\chi^2(1, N = 308) = 2.73$, $p = .10$, with a lower 95% (one-tailed) confidence limit of zero for chi square. Hypothesis B-2 was not supported.

Table 28. Cross-tabulation of Ratings of Favorability of Outcome by Whether Incident Was Handled Informally or Formally.

Iteration	How handled	Favorability of outcome		Total
		Unfavorable	Favorable	
First incident	Informally	349	365	614
	Formally	12	10	22
Second incident	Informally	143	149	237
	Formally	11	5	12

A variety of post hoc analyses followed the non-confirmation of these hypotheses. Separate two-group post hoc ANOVAs were run on satisfaction with outcome as a function of the favorability of the

SV's response. For the first incident, satisfaction was significantly greater (M = 3.59, SD = 0.85) when the SV responded favorably (understood and corrected the problem, understood the problem but could not correct it, the concern was unwarranted, or the SV was given an opportunity to save face), than for an unfavorable (M = 2.43, SD = 0.99) response (denied, did not respond, or understood but did not correct), $F(1, 799) = 311.54$, $p < .001$, $\eta^2 = .28$ with 99% confidence limits from .22 to .34. For the second incident, satisfaction also was significantly greater (M = 3.51, SD = 0.95) when the SV responded favorably than for an unfavorable response (M = 2.43, SD = 0.99) response, $F(1, 359) = 143.57$, $p < .001$, $\eta^2 = .33$ with 99% confidence limits from .23 to .42.

Separate post hoc two-group ANOVAs evaluated satisfaction with outcome among CRs who intervened as a function of whether the SV was junior (supervisee or junior colleague), a peer colleague, or senior (senior colleague or supervisor). As seen in Figure 1, the lower the status, the greater the respondent's satisfaction with the outcome, $F(2, 864) = 22.96$, $p < .001$, $\eta^2 = .07$ with 99% confidence limits from .02 to .14 for the first round of incidents. For the second round of incidents also, the lower the status, the greater the respondent's satisfaction with the outcome, $F(2, 399) = 31.61$, $p < .001$, $\eta^2 = .07$ with 99% confidence limits from .02 to .14.

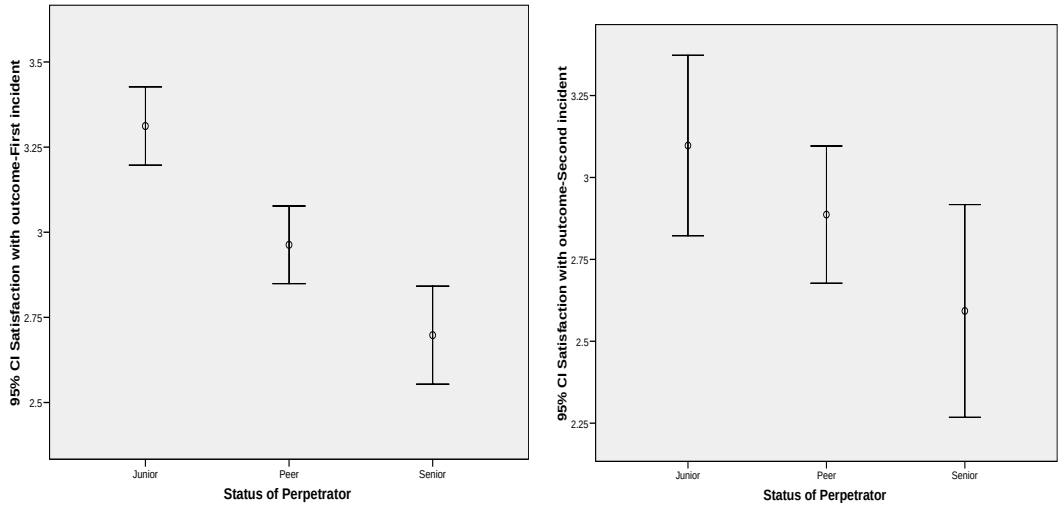

Figure 1. Mean rating (with 95% error bars) of satisfaction with outcome as a function of SV status relative to respondent.

Separate post hoc two-group ANOVAs were done to evaluate satisfaction with outcome among CRs with SVs who were distant vs. CRs with SVs who were close in the first or second response rounds. Neither analysis revealed a statistically significant difference between outcome satisfaction with distant vs. close SVs, $F(1, 927) = 1.04$, $p = .31$, $\eta^2 < .00$ with 99% confidence limits from .00 to .02 for the first incident and $F(1, 410) = 3.74$, $p = .05$, $\eta^2 < .01$ with 95% confidence limits from .00 to .02 for the second incident.

Separate post hoc two-group ANOVAs examined whether those CRs who had discussed the matter with trusted others felt more satisfied with the outcome of the intervention. On the contrary, there was greater satisfaction with the outcome for those incidents in which there was no discussion with trusted others. For the first round of incidents, discussers (M = 2.86, SD = 0.05) were less satisfied than those who did not discuss (M = 3.27. SD = 1.02) the matter with others, $F(1, 1098) = 42.47$, $p < .001$, $\eta^2 = .04$ with 99% confidence limits from .02 to .07. For the second round of incidents also, discussers (M = 2.05, SD = 0.21) also felt less satisfied than those who did not discuss (M = 3.00. SD = 1.14) the matter with others, $F(1, 501) = 14.35$, $p < .001$, $\eta^2 = .03$ with 99% confidence limits from .00 to .07.

Separate post hoc two-group ANOVAs evaluated whether those who approached the SV as part of a group felt more satisfied with the outcome than other interveners. There was no statistically significant difference between the group approach vs. other methods of intervention for the first round of incidents, F (1, 1098) = 3.07, p = .06, η^2 = .00 with 99% confidence limits from .00 to .02. Similarly, the methods of intervention did not produce a difference for the second round of incidents, F (1, 501) = 1.76, p = .19, η^2 = .00 with 99% confidence limits from .00 to .03.

Post-hoc multinomial logistic regression analyses were run to explore differences between those interveners who declared they were "neither satisfied nor dissatisfied" with the outcome and the other two types of responders (collapsed into satisfied and dissatisfied). Four of the outcomes significantly (α = .001) discriminated among the three groups: (1) when SVs denied that there was a problem, (2) when SVs were unresponsive to the intervention, (3) when SVs did nothing to correct the problem, and (4) when SVs corrected the problem. Table 29 shows frequencies for the first and second rounds of incidents for three categories of satisfaction and four outcomes. In general, the respondent was more likely to report dissatisfaction when the outcome was denial, unresponsiveness, or doing nothing to correct the problem, whereas satisfaction was more likely to be expressed by the respondent when the SV corrected the problem.

Table 29. Satisfaction Category Associated with Some Intervention Outcomes.

Iteration (incident rounds)	Outcome	Satisfaction level		
		Dissatisfied or Extremely Dissatisfied	Neither Satisfied nor Dissatisfied	Satisfied or Extremely Satisfied
First incident	Denied	114 (51.6%)*	62 (28.1%)	45 (20.4%)*
	Unresponsive	41 (48.8%)*	30 (35.7%)	13 (15.5%)
	Corrected by SV	11 (4.3%)	52 (20.6%)	190 (75.1%)*
	Did nothing to correct	111 (70.7%)*	34 (21.7%)	12 (7.6%)*
Second incident	Denied	68 (61.6%)*	22 (19.8%)	21 (8.9%)
	Unresponsive	25 (67.6%)	8 (21.6)	4 (10.8%)*
	Corrected by SV	7 (5.7%)*	27 (22%)	89 (72.4%)*
	Did nothing to correct	47 (61%)*	19 (24.7%)	11 (14.3%)*

* Significantly different from "neither" category at $\alpha < .001$.

Separate post hoc 2 x 2 χ^2 analyses examined the relationship between whether there were negative outcomes and whether the respondent would do anything differently (Table 30). For the first round of incidents, those with negative outcomes were more likely to report they would do something differently (35.7%) than those who had no negative outcomes (13%), χ^2 (1, N = 703) = 32.78, $p < .001$, with a lower 99% confidence limit of 9.92. For the second round of incidents, also, doing something differently was more likely to be associated with negative outcomes (37.3%) than without negative outcomes (33.6%), χ^2 (1, N = 330) = 8.59, p = .003, with a lower 99% confidence limit of 0.10.

Table 30. Association Between Desire to Do Things Differently and Favorability of Outcome.

Iteration	Negative outcome?	Would do anything differently?		Total
		No	Yes	
First incident	Yes	258 (64.3%)	143 (35.7%)	401 (100%)
	No	253 (83.8%)	49 (13.2%)	302 (100%)
Second incident	Yes	104 (62.7%)	62 (37.3%)	166 (100%)
	No	127 (77.4%)	37 (33.6%)	164 (100%)

Separate post hoc two-group ANOVAs tested the relationship between satisfaction and favorability of outcome. For the first round of incidents, as might be expected, satisfaction was greater when there were favorable outcomes (M = 3.29, SD = 0.98) than when there was negative fallout (M = 2.49, SD = 1.07) $F(1, 1098) = 135.72$, $p < .001$, $\eta^2 = .11$ with 99% confidence limits from .07 to .16. For the second round of incidents also satisfaction was greater when there were favorable outcomes (M = 3.57, SD = 0.95) than when there was negative fallout (M =2.26 , SD = 1.05 $F(1, 328) = 141.10$, $p < .001$, $\eta^2 = .30$ with 99% confidence limits from .20 to .40.

Final post hoc χ^2 analyses looked at the relationship between favorability of outcome among interveners and whether the error committed by the perpetrator was intentional (probably, definitely yes) or unintentional (probably not, no) errors.

Table 31. Association Between CR's Desire to Have Done Things Differently and Favorability of Outcome.

Iteration	Intentional error?	Unfavorable outcome?		Total
		Yes	No	
First incident	No	62 (27.6%)	163 (72.4%)	225
	Yes	280 (63.9%)	158 (36.1%)	438
Second incident	No	22 (22.2%)	77 (77.8%)	99
	Yes	110 (56.7%)	84 (43.3%)	194

For the first round of incidents, unfavorable outcomes were clearly more associated with intentional errors (63.9%) than with unintentional errors (27.6%), χ^2 (1, N = 1146) = 78.73, $p < .001$, with a lower 99% confidence limit of 39.67. For the second round of incidents, as well, unfavorable outcomes were more highly associated with intentional errors (56.7%) than with unintentional errors (43.3%), χ^2 (1, N = 293) = 31.48, $p < .001$, with a lower 99% confidence limit of 9.22.

Misgivings For Not Getting Involved

A logistic regression analysis tested the prediction of misgivings among those who observed or heard about an irresponsible act but who did *not* intervene in the first round of incidents reported by seriousness of the infraction and whether the SV was a colleague or one's own assistant (including post doc). (There were too few who felt misgivings about not intervening to analyze the second round of incidents) Misgivings were dichotomized into 1 (yes, I have misgivings, yes and know what should have been done, yes and still don't know what should have been done) and 0 (no misgivings). Seriousness was coded into 1 (the lesser discussed forms of research wrongdoing, namely carelessness, inadequate supervision of research assistants, incompetence, or creating a difficult/stressful environment) and 2

(actions usually described as more serious, namely fabrication, falsification, plagiarism, and failure to follow the rules of science).[9]

Neither seriousness of the infraction nor whether SV was a colleague vs. one's own assistant predicted whether the respondent felt misgivings about failing to intervene, $\chi^2(2, N=304) = 2..17$, $p = .34$, with a lower 95% (one-tailed) confidence limit of zero for chi square. Hypotheses C-1 and C-2 were not supported. However, there were only 10 incidents in which respondents failed to get involved when the SV was a supervisee of their own; among those, only two led to misgivings. Thus, there was little power to test Hypothesis C-1.

A 3 x 4 χ^2 analysis tested the hypothesis that a positive association exists between directness of involvement and misgivings about nonintervention. We defined "direct involvement" as when the SV is a close co-worker or administrator, or when the respondent is the victim or expressed fear of getting blamed. We defined indirect involvement as a distant co-worker, someone the researcher didn't know, or a SV about whom someone close the project confided in the participant. The third category was noninvolvement. Table 32 shows cross-tabulations of involvement by misgivings for the first incident.

Table 32. Directness of Involvement in Incident by Misgivings about Nonintervention

Directness of involvement in incident	Felt misgivings about nonintervention?				Total
	Yes, and I realize what I should have done	Yes, but I am still not sure	Only at first	No	
Not involved	6	55	15	250	326
	1.8%	16.9%	4.6%	76.7%	100.0%
Indirectly involved	5	42	2	205	254
	2.0%	16.5%	0.8%	80.7%	100.0%
Directly involved	5	45	5	55	110
	4.5%	40.9%	4.5%	50.0%	100.0%
Total	16	142	22	510	690
	2.3%	20.6%	3.2%	73.9%	100.0%

An analysis of the ordinal by ordinal trend was statistically significant, $t(689) = -3.45$, $p = .001$, Somers' $d = -.11$ with misgivings as the dependent variable. The more direct the involvement, the greater the misgivings about nonintervention. Hypothesis C-3 was confirmed.

A post hoc 2 x 2 χ^2 analysis examined the relationship between misgivings and directness of evidence of wrongdoing. As seen in Table 33, misgivings were less likely when the evidence was indirect than when directly observed, χ^2 (1, $N = 591$) = 23.96, $p < .001$, with a lower 99% confidence limit of 5.38.

[9] Some of our categories of research wrongdoing were omitted from this analysis because their definitions are extremely broad or could result from legitimate disagreement, such as authorship disputes.

Table 33. Non-Intervener Misgivings as a Function of Directness of Evidence of Wrongdoing

Directness of Evidence	Misgivings?		Total
	No or not lasting	Yes	
Direct	152 (65%)	82 (35%)	234 (100%)
Indirect	295(82.6%)	62 (17.4%)	357 (100%)
Total	447 (75.6%)	144 (24.4%)	591 (100%)

Taking Individual Responsibility

A three-group one-way between-subjects MANOVA with trend analysis was performed on three ratings: belief that researchers in your institution would consider intervening or reporting infractions; belief that the institution would take timely action; and the sum of four ratings of individual responsibility (See Table 34). All three measures were highly skewed, resulting in multivariate outliers. Groups were (1) those who reported never having observed or heard about an infraction, (2) those who observed at least once but did not intervene, and (3) those who intervened at least once. Although variances were sufficiently equal, all three DVs were negatively skewed, producing univariate and multivariate outliers in all groups, $p < .001$. After reflection and logarithmic transformation, all assumptions of MANOVA were met; $N = 2054$.

Table 34. Means and standard deviations for individual responsibility items ($N = 2066$).[10]

Item	Mean	Std. Deviation
Do you believe that researchers have an individual responsibility to try to actively correct or minimize problems whenever colleagues appear to have purposefully engaged in a serious form of research misconduct?	4.69	0.713
Do you believe that researchers have an individual responsibility to get personally involved in correcting or minimizing problems whenever their colleagues appear to have purposefully committed in a minor incident of irresponsible research conduct?	4.00	0.985
Generally speaking, do you believe that researchers have an individual responsibility to report problems to the appropriate institutional office whenever their colleagues appear to have purposefully engaged in research wrongdoing?	4.43	0.842
Generally speaking, do you believe that researchers have an individual responsibility to get personally involved in correcting or minimizing problems whenever their colleagues appear to have unintentionally done something that would affect the validity of their data, such as using the wrong statistic or data-gathering technique?	4.04	1.064

Even though all groups were skewed towards promoting responsibility, there was a statistically significant main effect for groups over the combined DVs, multivariate $F(6, 4098) = 19.63$, $p < .001$, partial $\eta^2 = .03$ with 95% confidence limits from .02 to .04. Stepdown analysis showed that all three DVs

[10] Respondents rated each question on a 1-5 scale with 1= “no” and 5= “definitely yes.”

significantly differed among the three groups ($p < .004$). Table 35 shows pooled within-group correlations for the three DVs. Note the relatively large correlations among the measures.

Table 35. Pooled Within-Cell Correlations among Ratings of Responsibility

Groups	1 Believe others would intervene	2 Believe own Institution would take action	3 Sum of individual responsibility ratings
1.Believe that other researchers would intervene	--	.45	.36
2. Believe that own institution would take timely action		--	.22
3. Sum of four individual responsibility ratings			--

Figure 2 shows means and 95% confidence interval for logarithm of rating of belief in other's likelihood of intervention. The significant linear ($p < .001$) and quadratic ($p = .038$) trends show that lower ratings of others were associated with interveners and those who had heard about at least one incident but not intervened. Contrary to hypothesis D-1 respondents who observed, suspected, or heard about irresponsible acts but did not intervene rated others as *more* likely to intervene than did respondents who intervened.

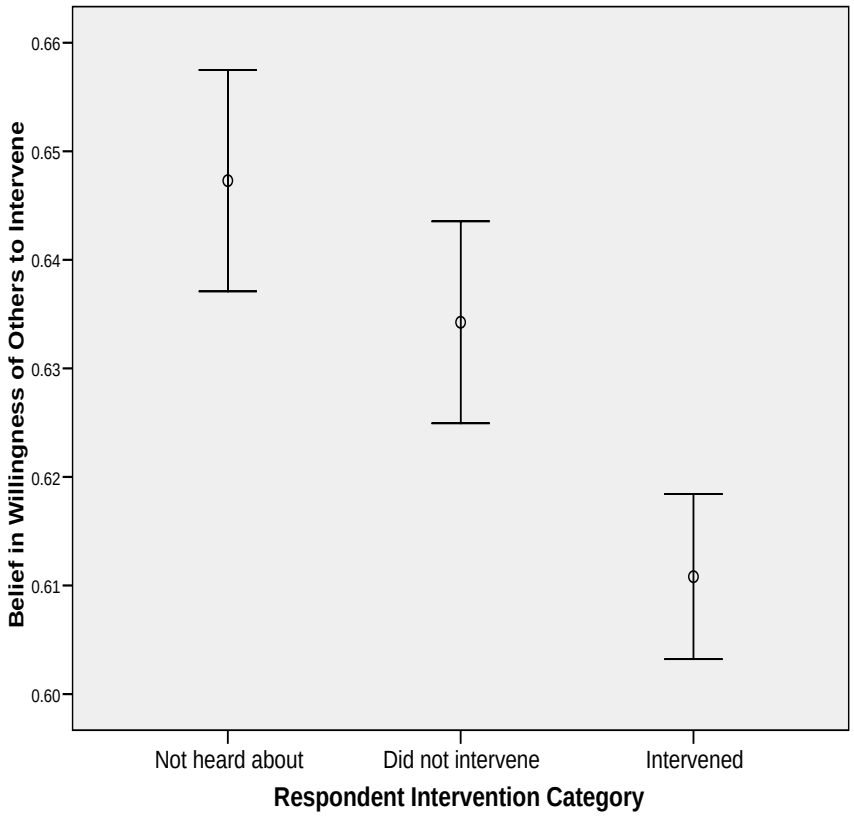

Figure 2. Mean logarithm of rating of belief that others in one's own institution would be likely to intervene, with 95% error bars.

Figure 3 shows means and 95% confidence intervals for the logarithm of rating of belief that one's institution would take timely action in the face of serious evidence of an infraction. The statistically significant linear ($p < .001$) but not quadratic ($p = .733$) trend shows that, contrary to hypothesis D-3, the highest ratings were associated with respondents who had never observed or heard about infractions rather than with those who intervened. Conversely, lowest ratings were associated with interveners rather than with respondents who had neither intervened nor heard about infractions.

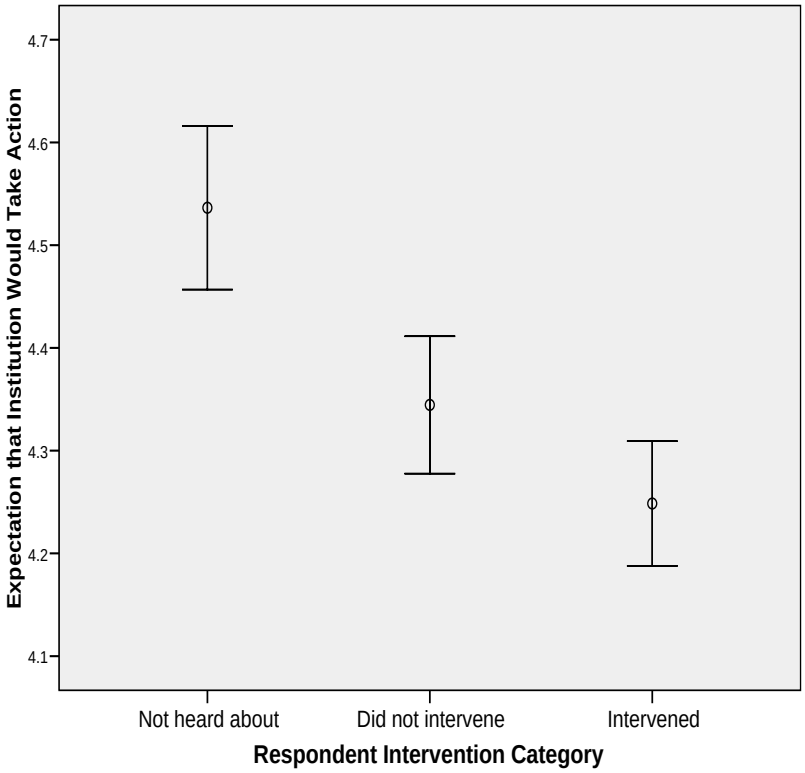

Figure 3. Mean rating of logarithm of belief that one's institution would be likely to take timely action with strong evidence, with 95% error bars.

Figure 4 shows means and 95% confidence interval for the logarithm of the sum of the four ratings describing beliefs in individual responsibility. There was no statistically significant linear trend ($p = .697$), however the quadratic trend ($p < .01$) shows that those respondents who had observed or heard about an incident but did not intervene had the lowest individual responsibility score, lower than those who hadn't observed or heard about any incidents. There is overall support for Hypothesis D-2.

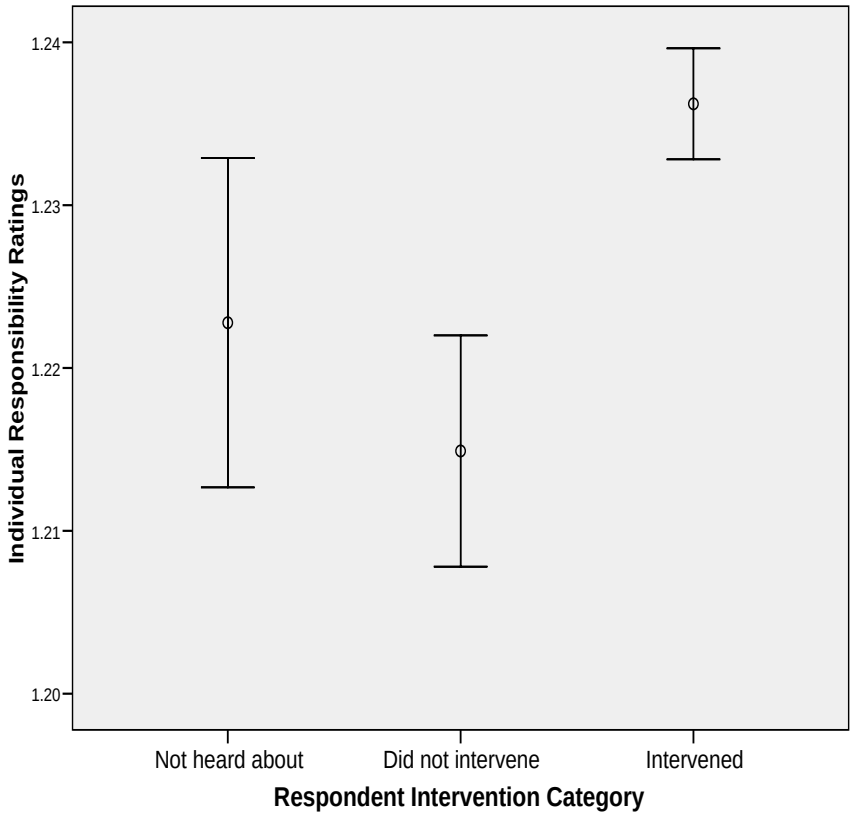

Figure 4. Mean logarithm of sum of individual responsibility rating as a function of intervention with 95% error bars.

As hypothesized, there was a positive correlation, although small, between years of experience and the sum of ratings of the four individual responsibility scores, Spearman *rho* (2048) = .06, p = .008. See Table 36. Hypothesis D-4 was supported. As PIs, most of our respondents were very experienced

A post hoc binary logistic regression analysis (using α = .01) examined differences between those who were never consulted by colleagues on ethical issues compared to those who were consulted often or very often in the sum of four ratings of individual responsibility and satisfaction with the outcome of the first incident in which the respondent intervened. Respondents who were often consulted are more likely to give higher ratings to individual responsibility than those who were never consulted, $\chi^2(2, N = 418) = 9.41$, p =.004, Nagelkerke R^2 = .04 with 99% confidence limits from .00 to .09 using Steiger and Fouladi's (2002) R2 software (*17*).

Table 36. Logistic Regression for Consultation on Ethical Matters (*N* = 418)

	Regression coefficient (B)	S.E.	Wald χ^2 (df = 1)	*p*	Odds Ratio	99% C.I. for Odds Ratio	
						Lower	Upper
Individual responsibility	.20	.07	8.10	.615	1.22	1.02	1.46
Years of research since highest degree	.02	.11	.14	.842	1.02	.76	1.37
Constant	-5.00	1.31	14.58	<.001			

A repeated-measures ANOVA treated multivariately was performed on responsibility ratings. Three groups were formed on the basis of the first round of incidents: those who reported never having observed or heard about an infraction; those who observed but did not intervene; and those who intervened. The three levels of the within-subject factor analyzed multivariately were (1) responsibility for getting involved when there was purposeful serious research wrongdoing, (2) purposeful minor research wrongdoing, and (3) responsibility for getting involved when there was unintentional error. Both main effects and the interaction were statistically significant.

Figure 5 shows that individual responsibility ratings varied with group (as also seen in Figure 4), with the highest rating for those who intervened and the lowest for those who reported no incidents at all, $F(2, 2063) = 19.62$, $p < .001$, partial $\eta^2 = .02$ with 95% confidence interval from .01 to .03. Highest responsibility rating was given for purposeful serious acts, with lowest ratings for purposeful minor infractions, lower on average than those for unintentional errors, $F(2, 2062) = 545.29$. $p < .001$, partial $\eta^2 = .35$ with 95% confidence interval from .31 to .38. However, the interaction, as seen in Figure 5, suggests that ratings of responsibility might be higher for purposeful minor infractions than for unintentional errors among those who intervened, $F(4, 4126) = 5.81$, $p < .001$, partial $\eta^2 = .01$ with 59% confidence interval from .00 to .01. Further, it appears that differences among groups may be smaller when the type of research wrongdoing is both purposeful and serious, and for those acts only, there is little if any difference between those who did not intervene and those who reported no incidents.

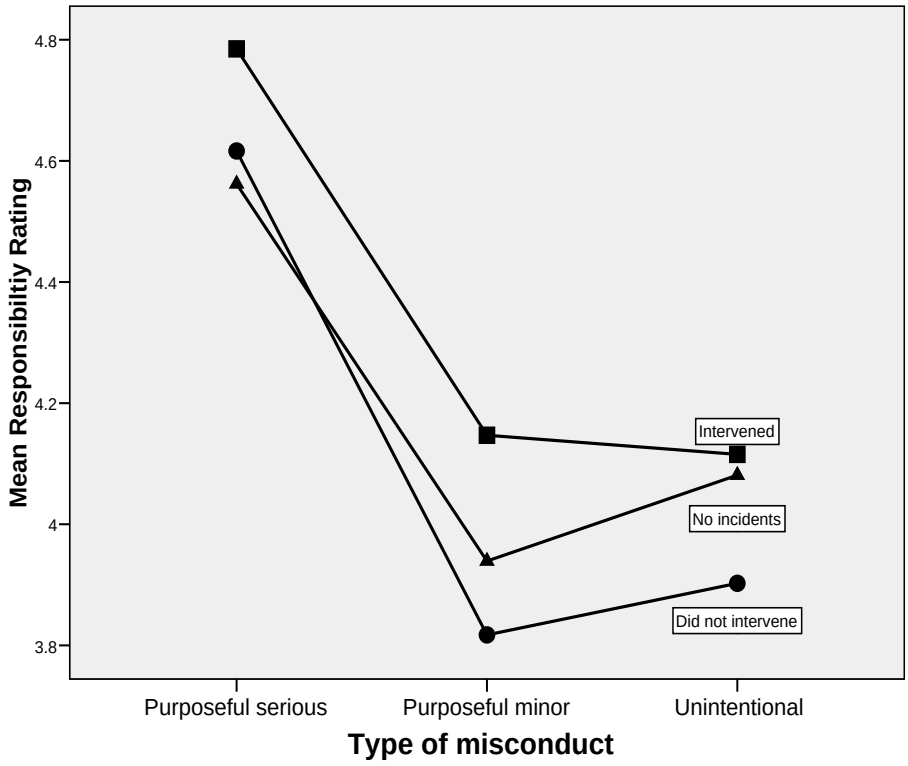

Figure 5. Mean rating of responsibility as a function of type of research wrongdoing and intervention for first incident.

Post hoc simple comparisons were made between ratings for purposeful minor research wrongdoing and unintentional errors for those who did not intervene, separately for those who observed or heard of no incidents and those who did encounter incidents but did not intervene. There was a significant difference between purposeful minor and unintentional error among those who reported no incidents, $F(1, 656) = 6.66$, $p = .01$, $\eta^2 = .01$ with 99% confidence limits from .00 to .04. However, among those who reported an incident but did not intervene, the difference was not statistically significant, $F(1, 656) = 3.59$, $p = .06$, $\eta^2 = .01$ with 99% confidence limits from .00 to .04. Thus, the only (weak) support for Hypothesis D-5 is among researchers who had not encountered research wrongdoing on the part of colleagues, and then only if the hypothetical purposeful research wrongdoing was minor rather than serious.

Number of Events Observed

We predicted that biomedical investigators[11] would report more events than would researchers in other settings. The number of events observed was, highly skewed. Therefore, a Mann-Whitney U test of ranks was performed on total number of events observed as a function of biomedical researchers vs. all others. There was no significant difference between the two groups, $Z(1991) = -1.23$, $p = .22$. Hypothesis E-1 was not supported.

A series of post-hoc χ^2 analyses examined differences between MDs and PhDs for the first iteration of reported research wrongdoing at $\alpha = .00$. No statistically significant differences were found between the two groups of respondents for intervention rate ($p = .770$), favorability of outcome ($p = .060$)

[11] Nursing investigators were deleted from this analysis because they do both biomedical and other types of research and might confound the results.

or satisfaction level (p = .480). There also were no statistically significant differences between MDs and PhDs on any of the nine individual outcome items (all p > .001). A one-way post-hoc ANOVA found no significant differences between MDs and PhDs in ratings of individual responsibility to deal with research wrongdoing on the part of colleagues (sum of four individual responsibility items), p = .282.

A post hoc Mann-Whitney U test of ranks was performed on total number of events observed as a function of research setting: research University vs. hospital. Again, there was no significant difference between the two groups, $Z(1692) = -0.19$, p = .85.

A post-hoc Mann-Whitney U test of ranks also was performed on total number of events observed as a function of sex, indicating significantly higher numbers of incidents reported by female than by male CRs (mean ranks = 1087.25 and 993.70. respectively), $Z(1692) = -3.69$, p < .001.

A Wilcoxon paired-sample signed ranks test was performed on number of direct vs. indirect observations of incidents. Contrary to expectations, respondents reported more direct observations (mean = 1.48, SD = 0.83) than indirect incidents (mean = 1.16, SD = 0.42), $Z(2057) = -6.85$, p < .001. Thus, Hypothesis E-2 was not supported. Because respondents could choose what stories to share, they may have preferred those with which they were more familiar.

A Wilcoxon paired-sample signed ranks test was performed on number of peer vs. assistant incidents. Contrary again to expectations, respondents reported more peer (mean = 1.51, SD = 0.86) than assistant incidents (mean = 1.11, SD = 0.34), $Z(1564) = -5.57$, p < .001. Hypothesis E-3 was not supported. We cannot conclude that peers engage in more wrongdoing, but rather that respondents choose to share more incidents about the wrongdoing of their peers.

A post hoc binary logistic regression was run on characteristics associated with those who observed only one incident (N = 1381) as compared with those who observed four or more incidents (N =121). Characteristics included institutional willingness to solve problems, the sum of four items about individual responsibility, and number of years doing research since highest degree. As seen in Table 35, at α = .01, the only significant difference between the groups was institutional willingness to solve problems, $\chi^2(3, N = 1046) = 22.51$, p < .001, Nagelkerke R^2 = .04 with 99% confidence limits from .02 to .08 using Steiger and Fouladi's (2002) R2 software (*16*). Those who observed only one incident gave a *higher* scores (M = 4.39*, SD* = 0.88) to institutional willingness to solve problems than those who observed four or more incidents (M = 3.90*, SD* = 1.12).

Table 37. Logistic Regression of Characteristics by Number of Incidents (N = 1046)

	Regression coefficient (B)	S.E.	Wald χ^2 (df = 1)	p	Odds Ratio	99% C.I. for Odds Ratio	
						Lower	Upper
Institutional willingness to solve problems	-.45	.09	23.67	<.001	.64	.50	.81
Individual responsibility	.02	.04	.25	.615	1.02	.91	1.14
Years of research since highest degree	.09	.12	.12	.442	1.10	.80	1.50
Constant	-.92	.87	1.12	.289			

Approaching Suspected Violators Informally and Discussions With Others

We predicted that most researchers who intervened did so informally rather than reporting the matter to superiors. Among those who intervened for first round of incidents, the form of the intervention was far more likely to be informal (93.4%) than to have been reported to a supervisor or the appropriate office for reporting scientific conduct (6.6%) as tested by a two-group chi-square analysis of goodness of fit for the first incident, $\chi^2(1, N = 1146) = 279.43$, $p < .001$ with 95% confidence limits from 217.84 to 348.92.

For the second round of incidents, 91.5% were informally handled and 8.5% reported to a supervisor or the research office, $\chi^2(1, N = 507) = 95.55$, $p < .001$ with 95% confidence limits from 61.07 to 137.71. Hypothesis F-1 was confirmed.

We hypothesized that CRs who intervened when the incident was one most often thought of as more serious (e.g., falsification, fabrication, and plagiarism) would discuss the matter with others more frequently than for less often considered incidents and that women would discuss the matter more often than would men. Among CRs with peer colleagues and post docs, there was no statistically significant prediction of whether the incident was discussed with a trusted individual from either seriousness of infraction or gender, as tested through direct binary logistic regression. For the first and second rounds of incidents, respectively, $\chi^2(2, N = 1146) = 3.55$, $p = .17$ and, $\chi^2(2, N = 507) = 1.35$, $p = .51$, both with lower 95% (one-tailed) confidence limit of zero for chi square. Hypothesis F-2 and F-3 were not supported.

References

1. S. Eastwood, P. Derish, E. Leash, S. Ordway, *Scientific and Engineering Ethics, **2***, 89 (1996).

2. J. P. Swazey, M. S. Anderson, K. S. Lewis, *American Scientist, **81***, 542 (1993).

3. N. H. Steneck. *Science and Engineering Ethics, **5***, 161 (1999).

4. S. L. Titus, J. A. Wells, L. J. Rhoades, *Nature, **453***, 980 (2008).

5. E. B. Zweibel, R. Goldstein, *Academic Medicine*, ***76***, 337 (2001).

6. J. Glick, *Accountability in Research, 1,* **77** (1989).

7. J. P. Tagney, *New Scientist, **115***, 62 (1987)

8. ORI Annual Report. Available at http://ori.dhhs.gov/documents/annual_reports/ori_annual_report_2007.pdf (2007)

9 N. S. Wenger, N.S., Korenman, S. G., Berk, R., & Liu, H. *Evaluation Review, **23***, 553-570 (1999).

10. 17. B. C. Martinson, M. S. Anderson, R. de Vries. *Nature, **435***, 737 (2005).

11. M. P. Miceli, J. P. Near, *Personnel Psychology*, ***38***, 525 (1985).

12. P. Keith-Spiegel, B. G. Tabachnick, B. E. Whitley, J. Washburn, *Ethics & Behavior*, ***8***, 215 (1998).

13. V. Bart, L. K. Trevino, D. Shapiro, *Journal of Business Ethics*, ***12***, 87 (1993).

14. J. M. Darley, B. Latane, B. *Journal of Personality and Social Psychology*. ***8***, 377 (1968).

15. C. L. Huang,-Pollack, T. H. Carr, J. T. Nigg, *Developmental Psychology*, ***38***, 363 (2002).

16. S. Staats, J. Hupp, H. Wallace, H. Downloaded on November 14, 2008, from http://www.sciencedaily.com/releases/2008/08/080817223646.htm (2008)

17. J. H. Steiger, J. H., R. T. Fouladi, *Behavior Research Methods, Instruments, and Computers*, ***4***, 581 (1992).

Appendix A

SURVEY CONTENT

Content for Introductory Pages

Thank you for considering our anonymous survey. You are among a large, randomly generated sample of experienced research scientists who have previously had or currently hold federal research funding. We understand how valuable your time is and have designed the survey to move along quickly.

We will ask about acts committed by others that raised your concerns about whether their research was being conducted responsibly. These may range from minor concerns and mere suspicions to major infractions and clear evidence. They could involve actions committed by anyone involved in a research project. (Even if you think you have not observed any such actions, we still have a few questions to ask.)

This project's results will ultimately provide valuable information to all scientists via a manual (available online) containing suggestions and best practices for handling concerns. Your input will be very beneficial to other investigators, and is very important to us.

We need your input and suggestions! Previous research suggests that the vast majority of researchers have at least one story to tell. The average time required to complete this survey is about 15 minutes for those who share one incident, somewhat longer for those who choose to share more.

Your participation is completely anonymous. Please do not offer any identifying information.
See the side menu for more information about how we ensure anonymity as well as more information about the project, its sponsor, and the research team.

If you might not be able to complete the survey in one sitting, please read this page. If you leave and return later, your anonymity will not be compromised. (See details about ensuring your anonymity on the left menu.)

Please note this survey is not programmed to allow you to go back to a previous screen, this is to protect your anonymity (see below). However, if you decide to leave the survey temporarily, you may return even if you have closed your browser and turned off your computer.

If you leave for a short time, and do not close your browser, you should automatically return to the page where you left off.

If you return to the survey and do not arrive at the page where you left off, then select "Return to Survey" in the navigation bar. On the next page enter your Survey Key: [all unique numbers and letters]. (Please write down this code exactly as it appears here. The software is case sensitive)

This survey key is algorithmically generated based on the time you started the survey and is not used for tracking identities. It is unique to you and will allow you to continue where you left off. And remember, even if someone gets your code, they cannot see any of the answers you have given to the survey.

Directions
The next screen lists 10 categories of acts that can have a negative impact on science and the research record. The person(s) involved in committing any of the acts could be anyone from an administrator to an undergraduate research assistant.

We are interested in acts you have witnessed or heard about that correspond to any of the 10 categories and how you reacted to them.

If a specific action or event you know about fits with one of the 10 categories, you will be asked several questions related to that specific act.

Each time you have completed the questions about that one act or event, you will be given the opportunity to return to the list to select another event from the same or a different category. Or you can proceed to the conclusion of the survey.

Items designated with a green asterisk (*) must be completed for the program to move forward.

Survey Questions

Have you observed or heard about an act that generally fits one of the 10 categories below? If so, please consider only one event at a time (and none other). Keep this single act in mind as you answer questions about it. (You will have an opportunity to describe additional events later if you choose.) You may want to jot down reminders of additional events now because you will have an opportunity to describe more than one.

1. Fabrication/Falsification
(Inventing data that were never actually collected; altering data that were collected; faking records; unjustifiable data removal or treatment of outlying data points)

2. Plagiarism
(The substantial copying of another's work without appropriate attribution; misappropriation of intellectual property)

3. Incompetence
(Examples: poor research design, methodology, or statistical procedure; inappropriate selection or use of a study technique due to insufficient skills or training).

4. Carelessness
(Examples: sloppy record-keeping; haphazard data collection; cutting corners; inadequate monitoring of the project's progress)

5. Intentional bias
(Examples: rigging a sample to maximize support for hypotheses; withholding methodology details; deceptive or misleading interpretation or reporting of data or its interpretation)

6. Questionable publication practices/authorship
(Examples: Publishing a paper or parts of the same study in different publication outlets without informing the readers; undeserved "gift" authorships; coerced authorship; omitting someone who deserved an authorship or other form of credit.)

7. Inadequate supervision of research assistants
(Examples: giving assistants more responsibility than they are able or willing to handle, insufficient supervision of assistants' work.)

8. Failure to follow the regulations of science
(Examples: sidestepping or ignoring the IRB or its directives; circumventing or ignoring human subject requirements with regards to informed consent, confidentiality, risk assessment, etc; inadequate care of research animals; violating federal research policy.)

9. Difficult or stressful work environment that could have a negative impact on the research process
(Examples: mistreatment or disrespectful treatment of subordinates; sexual harassment or other form of

exploitation; playing favorites and other factors that create poor morale or acting out by subordinates; using one's position to exploit another; conflicts with the administration or administrative policies.)

10. A dishonest act indirectly related to being a researcher
(Examples: unreported conflict-of interest, such as a financial interest in the outcome of an experiment; misuse or misappropriation of grant funds; inflating, distorting, or including bogus accomplishments in a resume.)

11. I have never observed or heard about any act that could fit into any of these categories.
[Respondents directed to last multiple choice questions on responsibility for integrity in science, demographics, and message of appreciation for participating.]

What was the position of the individual(s) who committed (or may have committed) this act? (Check as many as apply.)

A colleague (senior to me)
A colleague (peer)
A colleague (junior to me)
My Supervisor
My post-doc
Someone else's post-doc
My research assistant (graduate level)
Someone else's research assistant (graduate level)
My research assistant (undergraduate)
Someone else's research assistant (undergraduate)
Other

What was your position relative to individual (or individuals) involved?

I was not formally involved in the individual's (or individuals') project or situation.
I was a distant co-worker with the individual (or individuals).
I was a close co-worker with the individual (or individuals).
I was a victim of the act Interested in more victim stories
I was the person who might get blamed for the act.
Someone involved in the project confided in me.
I was an administrator involved in the situation.
I didn't know the individual(s) personally. How did these stories come to their attention?
Other

How did you become aware of possible irresponsible or unethical conduct?

I directly observed the act
An individual who committed, contemplated, or was involved in committing the act disclosed it to me
Direct evidence (for example, notes, records, or physical confirmation)
Credible second-hand information (for example, a trustworthy colleague or assistant observed the act and communicated the information to me)
Hearsay or rumor being passed around
Other

In your opinion, was the act committed with full knowledge that it was irresponsible or unethical by the individual (or at least one of the individuals, if more than one was involved)?

Definitely yes
Probably
Not sure
Probably not
No

Did you attempt to intervene or help address the problem in any way?
Yes
No

[For those who did intervene] What did you do?
I thought long and hard for some time before doing anything.
I discussed the matter with another trusted individual (not a supervisor or superior).
I informally discussed the matter with my supervisor or superior.
I discussed the matter with a friend or family member.
I went directly, by myself, to one (or more) of the individual(s) involved to discuss the matter face-to-face.
I went directly, accompanied by one or more others, to one (or more) of the individual(s) involved to discuss the matter face-to-face.
I sent the one (or those) involved a signed message.
I formally reported the matter to my supervisor or superior.
I reported the incident to the appropriate administrative office that deals with research wrongdoing.
Other

How did the intervention turn out?
After learning more, it appears that my original concern about irresponsible conduct was unwarranted.
The individual(s) denied that any problem existed.
The individual(s) did not respond to me.
The individual(s) understood the problem when it was explained and corrected it.
The individual(s) appeared to have understood the problem when it was explained, but there was no way to correct it.
The individual(s) understood but did nothing to correct it.
I gave the individual(s) an opportunity to save face so that they could solve the problem without acknowledging that it was irresponsible or unethical.
I sent the matter upward within the institution.
The matter was ultimately reported to the Office of Research Integrity or other federal level office.
Other

Did you personally experience any negative outcomes as a result of getting involved? (Check all that apply.)
Yes, the individual(s) involved treated me poorly (such as being rude or disrespectful).
Yes, my career status was put at risk.
Yes, my institution did not fully support my attempts to intervene.
Yes, I experienced some overall social costs (for example, being ostracized by some in the workplace, losing a friend).
Yes, I did (or may) face administrative or legal action
Yes, my overall reputation was affected (for example, now seen as a trouble-maker, tattle-tale, or too self-righteous).
Although I can't say that anything specific was done to me, I have experienced a lot of emotional turmoil. Popular answer—stories here.
No, no negative fallout

Overall, how did you feel about the final outcome?
Extremely satisfied
Satisfied
Neither satisfied nor unsatisfied
Dissatisfied Common—stories as to why needed
Extremely dissatisfied

If you had to do it all over again, would you have done anything differently?
Yes [describe]
No

NOW, please briefly describe the event and how you responded (Please do not include any identifying information.)

[For those who did NOT get involved]
What prompted you to decide against becoming involved?
I could not think of a way to respond.
I considered several options, but no satisfactory solution presented itself.
It was not a serious enough matter.
It was not my problem to solve.
The individual was my superior.
I did not need the aggravation.
I did not want to risk my career status.
The evidence was insufficient.
I might have become a target.
I feared the institution would not support me if the matter escalated.
The individual(s) involved is/are difficult (such as arrogant, malicious).
The individual(s) is/are also my friend(s).
I did not want to risk the social costs (such as ostracism in my workplace).
I did not want to risk getting sued or some other legal action.
I was concerned about risk to my reputation (such as being seen as a trouble-maker or tattle-tale or too self-righteous).
Someone else took care of it.
Other [If someone else took care of it, what was this person's role? text]

Have you felt misgivings about not getting involved?
Yes, and I realize what I should have done.
Yes, but I am still not sure about what I should have done.
Only at first
No

Did you ever discuss the matter with others afterwards?
Yes
No

Text Box. Please briefly describe the event (including the advice or comments of others, without any identifying information)

[All respondents answered from here on down—even those with no incidents to report]

<u>Who is Responsible for Maintaining Responsible Science?</u>

The survey concludes with a few questions about your opinions regarding responsible science and a few demographic questions to assist with data analysis.

Generally speaking, do you believe that researchers have an individual responsibility to try to actively correct or minimize problems whenever colleagues appear to have purposefully engaged in a serious form of scientific misconduct?
Definitely yes Probably Not sure Probably not No

Generally speaking, do you believe that researchers have an individual responsibility to get personally involved in correcting or minimizing problems whenever their colleagues appear to have purposefully committed a minor incident of irresponsible research conduct?

Definitely yes Probably Not sure Probably not No

Generally speaking, do you believe that researchers have an individual responsibility to report problems to the appropriate institutional office whenever their colleagues appear to have purposefully engaged in scientific misconduct?

Definitely yes Probably Not sure Probably not No

Generally speaking, do you believe that researchers have an individual responsibility to get personally involved in correcting or minimizing problems whenever their colleagues appear to have unintentionally done something that would affect the validity of their data, such as using the wrong statistic or data-gathering technique?

Definitely yes Probably Not sure Probably not No

Do you believe that most researchers in your institution would consider intervening or reporting if they noticed any incident of questionable, unethical, or irresponsible scientific practices?

Definitely yes Probably Not sure Probably not No

If strong evidence of a serious case of scientific misconduct were discovered and reported, do you think the office in your institution that is responsible for dealing with such matters would take appropriate and timely action?

Definitely yes Probably Not sure Probably not No

Has a colleague or assistant in your institution ever approached you for advice about how to handle an ethical issue related to his/her research?

No
Yes, once or twice
Yes, occasionally
Yes, fairly often
Yes, often

<u>Requested Demographic Data</u>

Your highest earned degree

MD. Phd. DSW DNSc MPH DSC M.A. or M.S. No Degree Other____

Number of years conducting research after completing your highest degree:

1-4 5-9 10-15 over 15 years.

Sex: Male Female

Which of the following best describes the setting where most of your research takes place?

Research university
Hospital or medical center
Comprehensive university
Four-year college
Community clinic or agency
Private agency, foundation or institute
Private Industry
Private consulting
Government Agency
No research done
Other_____

Which of the following best described the primary type of research you conducted?

Biomedical	Educational
Human Factors/Industrial Organizational	Policy
Social/Behavioral	Other_____

What is your primary professional identity?

Medicine	Public policy	Epidemiology
Public health	Social/behavioral science	Other_____
Biomedical	Nursing	

We welcome any comments you wish to make anonymously about this survey or project.

[Participants were then thanked for their participation.]