Teaching and Learning about ethical practice: The case analysis

Rochelle E. Tractenberg, PhD, MPH, PhD

Reprint Requests:
Rochelle E. Tractenberg
Building D, Suite 207
Georgetown University Medical Center
4000 Reservoir Rd. NW
Washington, DC 20057
rochelle.tractenberg@gmail.com

Collaborative for Research on Outcomes and –Metrics; Departments of Neurology, Biostatistics, Bioinformatics & Biomathematics, and Rehabilitation Medicine, Georgetown University Medical Center, Washington, D.C.

Acknowledgement: The author is the Chair (2017-2019) and former Vice-Chair (2014-2016) of the ASA Committee on Professional Ethics; perspectives and opinions are solely those of the author and do not reflect the opinions/perspectives of the Committee or the ASA. There are no actual or potential conflicts of interest.

Running Head: Teaching Ethical Practice with Case Analysis

Keywords: case analysis; stakeholder analysis; ethical statistical practice; ASA Ethical Guidelines; ethics education; statistics and data science

Abstract

Statistics, biostatistics, and data science are unique disciplines/a unique discipline in the sciences: anyone with an Internet connection and computing device can utilize the methods from these disciplines –irrespective of preparation to do so. Most empirical and all experimental sciences require some form of data analysis, including qualitative methods. However, even those in degree or formal educational programs learning about statistics/biostatistics/data science do not receive training in what constitutes "ethical practice". The American Statistical Association (ASA) maintains, and recently (2018) updated, Ethical Guidelines for Statistical Practice. Understanding and being able to utilize these Guidelines (GLs) is relevant for *all* applications of statistical and data science methodologies – whether for true "research" (following the scientific method) or for business or other predictive/decision-making support. Thus, students who will go on to be statisticians and non-statisticians alike need to learn about ethical statistical practice, including those who seek to apply these methods in marketing, policy, and higher education. This article describes how to employ the case study method to teach the ASA GLs, using simple vignettes and a specific tool called a "stakeholder analysis template". The template is introduced as a method for understanding the harms and benefits, as well as the stakeholders, in each of a series of tasks common to the collection, analysis/manipulation, and drawing of inferences or conclusions based on data in any shape or size. The ASA Ethical Guidelines are discussed with respect to their potential to guide data collection and munging (two specific tasks), with three learning objectives: 1. describe how different individuals ("stakeholders") may be affected by decisions and actions; 2. enumerate harms and benefits that are most clearly relevant for each stakeholder with respect to the activity; and 3. identify which ASA GL Principles (and/or specific elements) seem most relevant to this activity. The stakeholder analysis template is intended to facilitate teaching and learning – and the ultimate utility – of the ASA Ethical Guidelines for Statistical Practice.

1. Ethics education and training for statistical practice

Statistics, biostatistics, and data science are unique disciplines/a unique discipline in the sciences: anyone with an Internet connection and computing device can utilize their methods – irrespective of preparation to do so. Moreover, all disciplines with experiments or inference tests must use at least some of the key constructs from the domain. The American Statistical Association (ASA) has developed and promulgated a set of Ethical Guidelines (GLs) for Statistical Practice – most recently revised in 2018 (ASA 2018 – see Appendix)– that describe ethical practice in observable ways. The ASA GLs are relevant for all who engage with data, whether or not they self identify as "statisticians", "biostatisticians", or "data scientists" and whether or not they are members of the American Statistical Association (ASA). Those who use these methods, tools, and techniques have ethical obligations to the profession (ASA GLs A and D), those who contribute data or from whom data are obtained (ASA GLs B and C), and to those with whom (ASA GLs A, E, and F) or *for* whom (ASA GLs C, D, E, F, and H) they are applied. As noted in the Purpose of the Guidelines,

> "The principles expressed here should guide both those whose primary occupation is statistics and those in all other disciplines who use statistical methods in their professional work. Therefore, throughout these Guidelines, the term "statistician" includes all practitioners of statistics and quantitative sciences, regardless of job title or field of degree, comprising statisticians at all levels of the profession and members of other professions who utilize and report statistical analyses and their implications." (ASA 2018)

Failing to inform new practitioners – at all levels - of these obligations does not make the responsibilities go away- but it does ensure that fewer practitioners are able to take responsibility for ethical practice. One way to support the diffusion of the ASA Ethical Guidelines throughout the practice of statistics and data science is to encourage all practitioners (as defined above) to become familiar with them and how they work. The full Guidelines document (as approved by the ASA Board in April 2018) appears in the Appendix.

These obligations are as essential to "ethical statistical practice" as is the knowledge of methods and their appropriate use and reporting[1]. Tractenberg (2016-a) noted that, just as "(g)raduate instruction in statistics requires the presentation of general frameworks and how to reason from these." (Hubert & Wainer, 2011:62), "**instruction in statistics *also* requires the presentation of general *ethical* frameworks, and instruction and practice in how to reason from *these*.**" (emphasis in original; Tractenberg 2016-a: p 517)

However, most if not all statistics education focuses on the latter and the former is left to chance (or, deemed the responsibility of future employers to train into our students). It is especially important to introduce the construct of ethical statistical practice to those who receive formal training in the technical aspects of that practice. While "ethical statistical practice" is obviously a priority for trainers of the next generation of professionals, this training is also essential for those who will not go on to be formally employed as statisticians, biostatisticians, or data scientists; arguably this describes many, if not most, PhD students and all those medical and clinical

---

[1] The ASA actively seeks input on the Ethical Guidelines and the Committee on Professional Ethics (of which this article's author is Chair, 2017-2019) created an ongoing comment and

students who are encouraged to “do research”. Considering the reproducibility crisis in the biomedical sciences (Baker, 2016) and ongoing concerns with data privacy –and the continuing analysis and exploitation of that data – it is difficult to justify *not* committing to preparing data collectors, wranglers, analyzers, and interpreters to do each of these tasks in an ethical manner. In addition to promoting ethical practice whenever data are used, such a commitment can promote both professionalism and intellectual humility throughout technical and professional preparation to engage with data. One barrier to a wider acceptance of this commitment is that the ASA GLs are most often discussed *by* ASA members *with* ASA members – but, ASA members are by no means the only people working with data. To accomplish the Purpose of the Guidelines, that they guide all who work with data, these discussions need to happen among those outside of the ASA. Another barrier is that simply sharing the ASA GLs is not sufficient – training and practice in how they can be used is also needed (see Tractenberg, 2016-a; 2016-b).

Many institutions that carry out federally funded research offer training in “the responsible conduct of research”, RCR, or “research ethics training”. The overwhelming majority of such training focuses on the ethical treatment of research subjects; but these features are less important for non-experimental work –which may still include hypothesis and inference testing. The Office for Research Integrity perspective has been more global: rather than targeting “the responsible conduct *of* research”, their objective is to promote “responsible conduct *in* research” (Steneck, 2007, emphasis added). The construct of “ethical practice” is relevant for *all* applications of statistical and data science methodologies – whether for true “research” (following the scientific method) or for business or other predictive/decision-making support – including applications in marketing, policy, and higher education. Thus, instructors and mentors should ensure that all our graduates and trainees understand both the relevance of the ASA Ethical Guidelines for Statistical Practice, and how to use these GLs when they are needed but also, to limit the situations in which ethical guidance is needed.

The easiest approach is to distribute the ASA GLs widely, but this is unlikely to have the desired (or any) effect. Instead, this article describes how to employ the case study to teach the ASA GLs. The case study is a method for teaching (NCCSTS site) that has historically been used for business and ethics education. The primary objective for this article is to enable statistics and data science practitioners to understand case analysis and how to use it to teach ethical practice principles, with examples from the ASA GLs (2018). The following sections briefly define the case study method (Section 2), then introduce a tool that can be useful for facilitating case analyses, a stakeholder analysis template (Section 3). The results of the stakeholder analysis identify ASA GL principles that can be informative, so the GLs are introduced next (Section 4). Using the GLs to identify an ethical challenge and formulate and justify a course of action is the next section (Section 5). Finally, summarizing the case analysis makes learning observable, and assessable (Section 6).

## 2. The case analysis and ethics education

“Ethical case analysis is a common exercise for identifying and reasoning about ethical challenges in complex situations” (https://sites.psu.edu/ethicsofdatamanagement/unit1/using-ethical-concepts-to-analyze-case-studies/). In general case analyses teach by “…asking students to devise and defend solutions to the problems presented by each case.”

https://en.wikipedia.org/wiki/Case_method. In ethics education, a case study is typically a story or vignette that describes at least one actor (someone who must decide how to act or whose actions are presented) and at least one stakeholder (someone who has been, or could be, affected by the actor). Ethical principles (or rules) are provided to assist the learners in determining if any rules or ethical guideline has been broken or violated, i.e., if unethical behaviors are identified.

Many different resources exist for stories and vignettes that can be used to seed or stimulate discussions about ethical considerations (e.g., https://www.scu.edu/ethics/ethics-resources/ethics-cases/; https://kenan.ethics.duke.edu/multimedia-publications/case-studieswhitepapers/teaching-caselettes/; http://ethics.iit.edu/eelibrary/node/17816). However, key concepts of interest in RCR training have been identified by the National Institutes of Health, including: conflict of interest; authorship; the roles and responsibilities of mentors and mentees; and the role of the scientist in society (see also Stenneck 2007). Because these topics are fairly specific to the biomedical sciences, and because all were developed before "data science" or even "Big Data" were identified as career path options, case analysis/ethics case materials that have RCR topics list items as their focal learning objectives may have little to no overlap with what a statistician and data scientist might encounter "at work". In fact, "offshoot" resource archives are emerging – with a majority focusing on artificial intelligence (AI) and other aspects of technology that relate to data (e.g., http://ethics.iit.edu/eelibrary/search/site/Responsible%20Conduct%20of%20Research?f%5B0%5D=im_field_subject%3A1309 - with 136 entries as of 31 March 2019). These resources may also not be appropriate for the purposes of teaching ethical statistics and data science practice, as they focus on the creation or implementation of technology, rather than having to do with either statistics or data science (beyond specific computational aspects of either). They also do not feature the ASA or ACM ethical guidelines (e.g., "teaching data ethics"; https://www.scu.edu/ethics/focus-areas/technology-ethics/resources/an-introduction-to-data-ethics/, a 2018 instructional module which mentions *no* professional practice guidelines).

An important aspect of case analysis is that the vignette must engage learners with guidance – rules, laws, or guidelines that can be used to analyze the case. The context in which you practice, i.e. your team/lab, workplace, or industry, may have established "ways of doing things", but that is *only one* source of information as you evaluate the case or consider your options; it may be biased. The definition of "unethical behavior" is difficult to pinpoint when there is no consistent description of what is "ethical". Many organizations (the ASA included) seek to define what an ethical practitioner does, so that others in and outside of the profession can recognize the ethical practitioner. Providing these guidelines does not seek to bias or direct the learner's decision making, but rather, to contextualize their thinking about responsibilities and issues in authentic, applied/practical, contexts. The observability of ethical practice that guidelines present both describes ethical behavior and also supports new practitioners' aligning themselves with that discipline.

While professional guidelines come into existence, in part, to delineate and describe the professionals as a community (Tractenberg et al. 2015), their *existence* alone does not affect practice (without enforcement, such as exists for law and medicine). Professional guidelines must be taught and practiced. There are perhaps no more compelling rationales for teaching professional guidelines than the existence of "norms" that ignore or go against professional practice standards. In the biomedical research context, the norms of p-hacking, hypothesizing after results are known (HARKing), and failing to correct for multiple comparisons –all representing unethical failures to report fully and transparently, contrary to ASA GL Principles A, B, and C - have created a reproducibility crisis (Freedman, 2010; Collins & Tabak, 2014; Baker 2016). Thus, training *all* new practitioners to reconsider what they know or accept as "standard practice" (Stark & Saltelli, 2018) will ultimately benefit the profession *and* those who depend upon its practitioners to be ethical.

Cases can be quite complex (eg. https://aiethics.princeton.edu/case-studies/case-study-pdfs/), but they can also be very simple. The complexity of the cases that are used to teach ethics should be dictated by what the learners can already do and what the instructor hopes to move them towards. In this article, we assume that learning the contents of the GLs, and how to use them in everyday practice, are the main objectives; learning how to do a formal case analysis for its own sake is *not* an objective. With those assumptions in mind, we first consider two aspects of statistical and data science practice: data collection and data munging[2]. The learning objectives associated with our discussion of cases around these activities are:

- describe how different individuals ("stakeholders") may be affected by decisions and actions;
- enumerate harms and benefits that are most clearly relevant for each stakeholder with respect to the activity; and
- identify which ASA GL Principles (and/or specific elements) seem most relevant to this activity.

Without any detail beyond "you are collecting data" or "you are munging data", instruction and discussion can support learning that can support ethical practice throughout a career. However, with additional detail (and for those learners who can do the above-mentioned basic reasoning), a slightly more substantial vignette can be created and used to teach how the GLs can be leveraged to identify whether an ethical problem exists or will be created, and then to make a decision and justify that decision about what to do. The minimum additional information could be "you are asked to collect data using an automated procedure", or "you are given data sets without information about how the data were collected, and instructed to munge them". The additional context is vague – and not necessarily unethical; the objective would be to encourage practice in vague situations specifically to avoid unethical steps. Knowing which GL principles offer guidance on how to respond ethically in these vignettes would be predicated on the earlier learning – deepening the engagement with the construct of "ethical practice". Finally, more elaborate stories can be found or created, challenging the learner to describe and justify decisions that they make, using the GLs. If learners identify gaps in the GLs – noting where they offer insufficient guidance, then such analyses could be forwarded to the ASA to help strengthen the GLs in the future.

## 3. The stakeholder analysis template

[2] According to Wikipedia, data munging is defined as "computer jargon for a series of potentially destructive or irrevocable changes to a piece of data or a file. It is sometimes used for vague data transformation steps that are not yet clear to the speaker" https://en.wikipedia.org/wiki/Mung_(computer_term)

Stakeholder is defined "one who is involved in or affected by a course of action" (Merriam-Webster https://www.merriam-webster.com/dictionary/stakeholder); in the context of ethical case analysis, the stakeholder is simply an individual, or group, that might be affected by the outcome of the case. Clearly, identifying *who* might be affected, and the nature of that effect, are essential in understanding risks and benefits that might be associated with any decision or activity. Ethical case analysis implicitly requires that stakeholders are identified; however, this is not often a focus of instruction or practice. Because statistics and data science can have far-reaching implications, consideration of stakeholders warrants more attention than is typical; so to facilitate the learning goals articulated in the previous section, the stakeholder analysis template was created and appears in Figure 1.

Figure 1. Stakeholder Analysis template

| Potential result:<br><br>Stakeholder[1]: | HARM[5] | BENEFIT[5] | UNKNOWN[4] | UNKNOWABLE[3] |
|---|---|---|---|---|
| YOU[2,3] | | | | |
| Your boss/client | | | | |
| Unknown individuals[2] | | | | |
| Employer | | | | |
| Colleagues | | | | |
| Profession | | | | |
| Public/public trust | | | | |

There are two dimensions to the Stakeholder Analysis Template. The first dimension, captured in the columns, represents "*Potential Results*". These capture those effects of a decision or action, summarizing them according to whether or not they may represent net negatives. Potential negative results of any action or decision, or *harms*, include costing money, time, effort; negatively impacting reputations or persons; and other types of conceptual (intangible) or actual (tangible) damage. Like harms, potential positive results, or *benefits*, could be tangible or intangible – and they can have immediate or delayed effects. Benefits could include earning or gaining money; the removal of a harm; saving time or effort; improving reputation; and demonstrating expertise or superiority; among other things.

Since the effects of any action or decision may be negative for one entity, person, or group while positive for another, the potential result must be considered with respect to each "*Potential Stakeholder*" (the second dimension in the template). It may be surprising to realize that one of the potential stakeholders is *you*, the person making the decision. As described earlier, harms (costs in time, effort, and reputation) are easily recognized to yourself, as are benefits. These should be considered first because understanding the potential harms and benefits to yourself can also help you to recognize what they may be for the next stakeholder, your boss or client. If a decision costs *you* time (a harm), this could be a harm to your immediate *boss or cli*ent as well. By contrast, deciding to formally identify a data breach at work could be perceived as a harm to you (you could be punished if your boss does not want others in the company to find out) and a benefit to you (you would be forewarned that maybe your employer will soon be out of business or targeted by authorities). Only when the results of decisions or actions – in terms of both harms and benefits – are recognized can they be balanced against each other to make a *justifiable* decision (Tractenberg & FitzGerald, 2012).

You and your boss/client are fairly clear, recognizable, stakeholders. By contrast, "*unknown individuals*" are not recognizable stakeholders per se, but if you make a programming decision in the creation of an algorithm, or a distributional assumption in an analysis, there could be predictable but not-specific results for these unknown individuals. A simple example is males versus females: prior to 2015 the National Institutes of Health did not require that genetic sex effects should be considered and potentially modeled separately (https://grants.nih.gov/grants/guide/notice-files/not-od-15-102.html) – making the scientific and statistical assumption that all human subjects (after controlling for age, height, and weight in many cases) are exchangeable. The decision to require specification of sex effect hypotheses in proposed research, or to justify not including such specification, represents an acknowledgement that failing to consider sex effects in biomedical science is not appropriate, and is insufficiently rigorous. In terms of harms or benefits, failing to consider that heart attack symptoms differ for men and women (Coventry et al. 2011) for example, and making assumptions in analyses or algorithms that ignore the specific symptoms that women experience, will end up potentially harming "unknown women". The benefit to "unknown men" might be "more is understood about the symptoms prior to a heart attack". Medical professionals may never warn women about what might in fact occur to *them*. Thus, the assumption that "human subjects are exchangeable" has great potential to do tangible and lasting harm to those –unknown individuals - about whom such an assumption is wrong. Even when we know that "women experience heart attack symptoms differently from men", we will not know the specific women for whom the decision to model all humans the same will have this harm, thus the category of stakeholder is "unknown individuals". Unknown individuals also comprises customers – e.g., all those customers who are known to have data/records associated with them, even though the specific ones whose data will be breached is unknown.

Identifying your "*employer*" as a stakeholder may seem more straightforward, but "your employer" could be a person or the corporate entity that is your company's name. Harms and benefits can accrue to both; making your decision about notifying the authorities about a data breach could affect the corporate "employer" without affecting the CEO/owner. The more important aspect of this stakeholder is that harms and benefits could accrue to them based on *your* performance or decision. If you are self-employed, you are the face of your company; so while a benefit may accrue to *you* when you fail to notify the authorities about a data breach "at your company" (e.g., save yourself a lot of paperwork; limit the likelihood that a reporter will publicize the breach and name you specifically, etc.), this would actually constitute a harm to your employer (your personal brand) – demonstrating that your company cannot be trusted by a potential customer.

Few practitioners in statistics and data science work alone, so recognizing that decisions we make about ethical problems can affect our "*colleagues*" should not be surprising. Some statisticians and data scientists are the only professional in that field on a team or in a working group – meaning that our colleagues in those situations are unaware of the ethical practice guidelines we are obliged to follow. If you fail to act on an ethical guideline principle, non-statisticians may never find out; but if they do, their trust in you would be diminished (a harm to you). You might, however, inadvertently create a situation where they might also be faced with ethical dilemmas caused by your action or decision; complicating their jobs creates a harm to *them*.

Professional Guidelines are developed by those in practice, in part, to delineate the qualifications to practice and to encourage public trust in "*the profession*" (Tractenberg et al. 2015). Some fields (e.g., medicine, law) have methods for controlling licenses to practice; all those who violate professional or ethical guidelines harm the profession – by suggesting that these controls do not act to keep "bad actors" out of practice. Those who ignore ethical challenges do not strengthen the profession – even if ignoring data breaches seems to benefit the individual ("keep your head down"; "go along to get along"), they harm the profession overall. When an individual, acting in a professional capacity, behaves unethically or ignores unethical behavior, the profession is impacted negatively.

The final stakeholder to consider is the *public/public trust.* When a representative of a profession is unethical, not only is that a harm to their profession but it also diminishes the public's trust in the profession, how that profession is regulated, and whether or not federal funds should be allocated to support or otherwise engage with that profession. The public – people who are not (yet) your customers, and people who are definitely "unknown individuals", but whose stake in the ethical practice of statistics and data science is very real – represent the cultural context in which we are educated, trained, and employed. The public also influences legislation –for or against our profession. Western culture tends to believe itself to be evidence-based: decisions are supposed to be supported by evidence and data (although see Pencheva, 2019 and McGoughy 2019, both discussing the campaign for Britain to leave the European Union ("Brexit"). Part of that evidence is "evidence that practitioners of statistics and data science can be trusted" – harms accrue to the public trust when, e.g., data breaches happen – the public may seek to require stricter legal controls on data collection (e.g., the General Data Protection Regulation, (EU) 2016/679; see https://en.wikipedia.org/wiki/General_Data_Protection_Regulation). Public sentiment towards federally funded research can also become negative if the public, or the public trust, are harmed.

It is also important for the person completing the stakeholder analysis template to recognize two types of unknown information, which are the final two columns in the table:

- *Unknown*. It is possible for a decision to be required early in a project (for example), before an effect for a given stakeholder can be established as a "harm" or a "benefit". Thought experiments, in which these effects are imagined - rather than observed or remembered from personal experience - (https://plato.stanford.edu/entries/thought-experiment/) or simulations can help to determine which of these is more likely; and an important aspect of the stakeholder analysis template is to document where more information is needed. Additionally, whether a potential stakeholder is – or will become – an actual stakeholder may also need to be determined.

- *Unknowable.* Both early and late in a project, it may be simply impossible to determine whether the effects for a particular stakeholder will be positive (benefit) or negative (harm). Whenever something appears in the "unknowable" column, it suggests that whatever decision is taken currently may need to be revisited in the future. Recognizing something as "unknowable" does not mean it should be ignored, but rather suggests that more thinking or more specification is required –possibly both.

Now that we have discussed every cell (row *x* column intersection) of the stakeholder analysis template, we can consider what information it might hold as a whole.

1. Knowing to whom harms may accrue can guide you to where the professional guidelines can assist in decision making. This is the topic of the next section.
2. Articulating the harms that may accrue to YOU is essential for you to "treat others' data as you would your own" (Loukides, Mason & Patil, 2018: Chapter 3). You need to recognize the harms that can accrue *to you* before you can compare those to you and those to others. Moreover, "others' data" could relate to your boss/client, your employer, unknowable others, or the public. Recognizing whether or not benefits or harms accrue to these different types of "others" is the only way for you to make a decision about how you want other people to treat *your* data: in someone else's table, *you* are the client, an unknowable other, or part of the public.
3. If there are no recognizable harms, and plausibly no "unknowable" harms <for which your decision would be responsible>, then there can be no conflict. It is really important to recognize whether something truly is unknowable or if it is actually something that can be known – but you just don't know it. The key words here are "recognizable" and "plausible" – your failure to recognize something doesn't mean it does not exist. And, beware of straw man[3] or red herring[4] harms!
4. If there are plausible harms (or benefits) that you cannot identify, but you believe/suspect may exist, then there is insufficient information for you to make a decision and you need more information. Recognizing this – instead of making an uninformed decision – is currently *not part of the norm.* Learning how to use this table and complete a case analysis is essential for enabling *informed decisions about ethical challenges* for *current and* future practitioners.
5. All harms are not the same; all the benefits are not the same; and harms and benefits are not exchangeable.

Discussion about stakeholders and harms/benefits that may result from any of the standard activities in statistics and data science can strengthen the learners' engagement with GLs, and also encourage consideration –and acceptance - of the responsibilities to practice ethically that the GLs describe (lest any of the harms befall any of the stakeholders).

## 4. ASA Ethical Guidelines

[3] "Straw Man": defined as "an argument, claim or opponent that is invented in order to win or create an argument", Cambridge English Dictionary.

[4] "Red Herring": defined as "something that takes attention away from a more important subject", Cambridge English Dictionary.

The ASA Ethical Guidelines, revised in 2016 and 2018, include eight principles, each of which has 4-11 elements. The eight general principles for ethical statistical practice (*http://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx)* are listed here; see the Appendix for the specific elements.

**A. Professional Integrity and Accountability**

The ethical statistician uses methodology and data that are relevant and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results. The ethical statistician does not knowingly accept work for which he/she is not sufficiently qualified, is honest with the client about any limitation of expertise, and consults other statisticians when necessary or in doubt. It is essential that statisticians treat others with respect.

**B. Integrity of data and methods**

The ethical statistician is candid about any known or suspected limitations, defects, or biases in the data that may impact the integrity or reliability of the statistical analysis. Objective and valid interpretation of the results requires that the underlying analysis recognizes and acknowledges the degree of reliability and integrity of the data.

**C. Responsibilities to Science/Public/Funder/Client**

The ethical statistician supports valid inferences, transparency, and good science in general, keeping the interests of the public, funder, client, or customer in mind (as well as professional colleagues, patients, the public, and the scientific community).

**D. Responsibilities to Research Subjects**

The ethical statistician protects and respects the rights and interests of human and animal subjects at all stages of their involvement in a project. This includes respondents to the census or to surveys, those whose data are contained in administrative records, and subjects of physically or psychologically invasive research.

**E. Responsibilities to Research Team Colleagues**

Science and statistical practice are often conducted in teams made up of professionals with different professional standards. The statistician must know how to work ethically in this environment.

**F. Responsibilities to Other Statisticians or Statistics Practitioners**

The practice of statistics requires consideration of the entire range of possible explanations for observed phenomena, and distinct observers drawing on their own unique sets of experiences can arrive at different and potentially diverging judgments about the plausibility of different explanations. Even in adversarial settings, discourse tends to be most successful when statisticians treat one another with mutual respect and focus on scientific principles, methodology and the substance of data interpretations.

**G. Responsibilities Regarding Allegations of Misconduct**

The ethical statistician understands the differences between questionable statistical, scientific, or professional practices and practices that constitute misconduct. The ethical statistician avoids all of the above and knows how each should be handled.

**H. Responsibilities of Employers, Including Organizations, Individuals, Attorneys, or Other Clients Employing Statistical Practitioners**

Those employing any person to analyze data are implicitly relying on the profession's reputation for objectivity. However, this creates an obligation on the part of the employer to understand and respect statisticians' obligation of objectivity.

The ASA GLs themselves mention stakeholders that also appear in the template; so familiarity with the GLs can help learners understand the variety of stakeholders who can be affected by their actions and decisions. For example, benefits accrue to the profession when principles A and F in particular are followed (i.e., "the profession" is a stakeholder whenever considerations involve these principles), while the public is a clear stakeholder whenever principle C is involved. Moreover, employers have a responsibility to respect the obligations of the statistician to be objective. Since the GLs describe what "the ethical statistician" does – to decrease the likelihood of harms accruing- following the guidelines will support ethical statistical practice. Table 2 outlines a few basic harms and benefits that can arise for each stakeholder, in the process of data collection. ASA GL principles that may inform how to avoid the harms that are identified are also included.

Table 2: Identifying ethical challenges that can arise in **data collection**.

| Stakeholder | HARMS | BENEFITS |
|---|---|---|
| YOU | Potential legal liability if data incorrectly obtained; getting consent is hard<br>Principles A, B, D, E, F, G | Easy to collect=better |
| Your boss/client | Potential legal liability if data incorrectly obtained<br>Principles A, B, C, D, E, F, G | Less time, easy to collect = more $$ |
| Unknown individuals | Potential group biases arise; unknown consent=potentially negative provenance<br>Principles A, B, D | No benefits |
| Employer | Potential legal liability if data incorrectly obtained<br>Principles A, B, D, E, F, G, H | Less time, easy to collect = more $$ |
| Colleagues | Sharing ill-gotten data implicates them as well<br>Principles A, B, D, E, F, G | You spend time & incur liability, colleagues get free |

| | | |
|---|---|---|
| | | data! |
| Profession | Working with ill-gotten data undermines credibility<br>Principles A, B, C, D, E, F, G | Possible innovation using data, or in how to collect more data, makes everyone's jobs easier |
| Public/public trust | Ill-gotten data undermines credibility of those who *collect* it, and possibly also the profession that allows – and/or benefits from -such collection.<br>Principles B, C, D | Public may never find out (not actually a benefit to THEM!) but if they do, maybe there are legal recourses they can benefit from (also not a real benefit! Legal recourse is expensive and time consuming!) |

Note that Table 2 shows multiple GL Principles are applicable to mitigate or avoid harms to all stakeholders. The relationships between the GLs and benefits is less clear; the process *assumes* that ethical practice is an overall benefit. The stakeholder analysis table can be used – with the simple prompt "collecting data" – to encourage consideration (listing, discussing, comparing, ranking) of harms and benefits that accrue to each stakeholder; alternatively the same table can be used to ask learners to list which principles would help ensure data collection avoids harms (and how each accomplishes that). For example, GL Principle B discusses **Integrity of data and methods** (emphasis added):

> "The ethical statistician is candid about any known or suspected limitations, defects, or biases in the data that may impact the integrity or reliability of the statistical analysis. Objective and valid interpretation of the results requires that the underlying analysis recognizes and acknowledges the degree of reliability and integrity of the data."

The ethical statistician:

- Item 5: "Clearly and fully reports the steps taken to preserve data integrity and valid results."
- Item 10: "…exercises due caution **to protect proprietary and confidential data**, including all data that might inappropriately reveal respondent identities." (emphasis added)

In addition to offering guidance about how to *avoid* harms to different stakeholders, GL B.5 and GL B.10 both also suggest how to MEET ethical challenges arising in data that have already been collected:

The ethical statistician:

- Item 5: "Clearly and fully <u>reports</u> the steps taken to preserve data integrity and valid results."
- Item 10: "…<u>exercises due caution</u> **to protect proprietary and confidential data**, including all data that might inappropriately reveal respondent identities." (emphasis added)

The decision/ethical way forward is to document everything you do/have done ("fully report the steps") specifically to meet the challenge (to protect the data and sources). *Then,* make sure all reports go along with all results and data from this collection that you go on to share, publish, or present.

"Case analysis" need not be elaborate – or based on extensive cases/vignettes - to be an effective teaching tool.

5. Using the ASA GLs to identify an ethical challenge, and formulate and justify a course of action.

Sometimes we are asked to perform, or tasked with, activities that are contrary to these GLs, and at other times it can be difficult to fully follow one GL principle. An example would be how to follow Principle C, and respect both the public and a client if the client asks for an analysis or algorithm that might create bias against (harms to) unknown individuals in the public. Thus, in addition to learning the GLs themselves, it can also be helpful to understand how to prioritize them. The stakeholder analysis template can help here, too. There is another approach to ethical problem solving that bears mentioning: ethical reasoning, a learnable skill set that can be taught and improved upon throughout a career (Tractenberg & FitzGerald, 2012). There are six elements of knowledge, skills, and abilities (KSAs) that are required for ethical reasoning:

1. **Identify and 'quantify' prerequisite knowledge**: a practitioner should know the ASA GLs and be able to use them to complete the stakeholder analysis template. This prerequisite knowledge is essential for ethical practice, as well as for deciding what to do when faced with requests to practice unethically.
2. **Identify or recognize the ethical issue**: In Table 2 we saw that the activity "data collection" has the potential to create harms if GL principles are not followed as data are collected. Without more context, "the ethical issue" is recognize as simply "not following the GLs leads to harms". The ASA GLs specify in their statement of purpose that "Above all, professionalism in statistical practice presumes the goal of advancing knowledge while avoiding harm; using statistics in pursuit of unethical ends is inherently unethical." I.e., ethical practice advances knowledge *while avoiding harm,* so statistical and data science practice that does not *seek to avoid harm* will be unethical. We must also act ethically in response to a specific order, request, or situation, particularly when these involve unethical behavior (or requests that we engage in them) by others.
3. **Identify decision-making frameworks**. The ASA Ethical Guidelines for Statistical Practice represent a "*virtue ethics*" decision-making framework, which can be summarized as, "what would the (ideal) ethical practitioner do in this case?" Alternatively, a *utilitarian* framework can be helpful to sort out the positive and negative effects of a decision on each of the stakeholders. The utilitarian perspective can generally be summarized as, "how can benefits be maximized while harms are minimized?" Both of these frameworks are supported by the stakeholder analysis template.
4. **Identify and evaluate alternative actions** (on the ethical issue). There are *always* a set of decisions that can be made in any circumstance: either i) agree to comply with an unethical request, or ii) decline to comply with an unethical request; or, a) do nothing, b) consult or confer with a peer or a supervisor – using the professional guidelines or other resources, or c) report violations of policy, procedure, ethical guidelines, or law. NB: "reporting" is clearly required when law and policy are violated.

5. **Make and justify a decision**. Using the alternative actions outlined in step 4, articulating the decision of what to do in the face of the ethical challenge. Rather than acting, without justifying the act or decision, some discussion of how stakeholder effects were considered will document the reasoning for you at a later point or for others. Stakeholders and benefits/harms should be considered, and as was noted for the stakeholder analysis template, not all benefits are equal, neither are harms, and harms and benefits are not exchangeable. So the justification is not a simple matter of counting up harms or benefits.
6. **Reflect on the decision**. Thinking about what is hard, what additional information would have been helpful, how to get better at these challenging features of ethical reasoning, and how to help create the culture that promotes fluency in ethical reasoning are all reflections that may seem like "an extra step" for those new to this kind of reasoning, but will in fact support a more ethical work place in the future. Such reflection can also serve as case studies to be analyzed by others in the future.

Table 3 is another stakeholder analysis template, filled out for the activity of data munging. Specifically, the vignette (case) for this analysis is given below and the stakeholder analysis follows:

> You have obtained data from multiple sources and provenance information about the data is inconsistent – different people at work describe it differently and there is no real evidence about the provenance of any of the data.

Table 3. Stakeholder analysis for data munging vignette.

| **Stakeholder** | **HARM** | **BENEFIT** |
|---|---|---|
| YOU | Potential legal liability if munging across confidentiality levels breaks anonymity in one set.<br>Principles B, C, D, E, F, G, H | Easy to collect=better; munged data may be harder to trace source of breach (not a REAL benefit!!!) |
| Your boss/client | Potential legal liability if munging makes incorrectly obtained data more widely available<br>Principles B, C, D, E, F, G, H | Munging easy to collect variety of data = better than collecting/getting good data. |
| Unknown individuals | unknown provenance in SOME data can create harms to other data after munging and can inject bias.<br>Principles B, C, D, E, F, | No benefits |
| Employer | Potential legal liability<br>Principles B, D, H | Less time, easy to collect = more $$ |
| Colleagues | Sharing ill-gotten data implicates them if they use it. | Sharing bad data is bad. |

| | | |
|---|---|---|
| | Principles B, C, D, E, F | |
| Profession | Munging that breaks confidentiality undermines credibility<br>Principles A, B, C, D, E, F | Possible innovation makes everyone's jobs easier |
| Public/public trust | Broken confidentiality undermines credibility<br>Principles A, B, C, D, E, F | Public will never find out. (not actually a benefit to THEM!) |

Data munging can violate many of the GL Principles if not done thoughtfully.
Just as in the data collection example, Principle B, Item 10 describes the ethical statistician as one who: "…exercises due caution **to protect proprietary and confidential data**, including all data that might inappropriately reveal respondent identities." (emphasis added) Ethical data munging can be made more likely even if just this ONE principle and item are followed. In data munging, everything possible must be done to ensure that proprietary and confidential data are protected.

In addition to obligations relating to the data, there are also obligations relating to those who contributed to the pool of data you may be munging:
**D. Responsibilities to Research Subjects**

> The ethical statistician protects and respects the rights and interests of human and animal subjects at all stages of their involvement in a project. This includes respondents to the census or to surveys, those whose data are contained in administrative records, and subjects of physically or psychologically invasive research.

Specifically, D.3: "Protects the privacy and confidentiality of research subjects and data concerning them, whether obtained from the subjects directly, other persons, or existing records." And D.6. : "In contemplating whether to participate in an analysis of data from a particular source, refuses to do so if participating in the analysis could reasonably be interpreted by individuals who provided information as sanctioning a violation of their rights."

The combination of the stakeholder analysis table and the two GL principles (B and D) that are identified as relevant to preventing harms to every stakeholder represent prerequisite knowledge for identifying the ethical problem here: You are being asked to munge data from different sources whose origins cannot be confirmed. You cannot determine if the data have been ethically sourced – so, munging that data, and particularly making it available to others, is not ethical. Choosing the virtue ethics framework, D.6 strongly suggests the decision to be made in this vignette: "the ethical statistician does not participate in the analysis of data that does not have sufficient documentation or evidence that the data were not obtained by violation of the rights of those giving the data."

7. Conclusions.

The case study is a method for teaching that is commonly used to teach ethics, and which can be used to teach ethical practice principles in statistics and data sciences. The stakeholder analysis template provides a framework for evaluating one's knowledge about a given case or situation, helping to identify ethical challenges as they occur: subtleties and the non-exchangeability of harms, and of benefits, and the non-equivalence of these in any circumstance can be recognized with this template. The relevant ASA GL principles can be identified and prioritized using this template; it also helps to formulate and justify a course of action.

If new practitioners learn, and learn to apply, the GLs from an early point in their professional training, then the main type of ethical challenge will arise from someone asking for something that is contrary to the GLs. In those cases, the practitioner who knows the GLs – and knows Principle H should protect them from being asked to violate the GLs at work – would be able to use the GLs to justify how and why they cannot comply with a request to practice unethically. Unfortunately, very few new practitioners (to date) have learned the ASA GLs, leading to struggles throughout the career. "The entire community of scientists and engineers benefits from diverse, ongoing options to engage in conversations about the ethical dimensions of research and (practice)," (Kalichman, 2013: 13). This perspective, from the engineering ethics education community, is also true for the community of statisticians and data scientists – whether they engage with "the community of scientists and engineers" or not.

## APPENDIX:

## Ethical Guidelines for Statistical Practice

***Prepared by the Committee on Professional Ethics***
***of the American Statistical Association***
**Revised by the Committee on Professional Ethics**
**Approved by the ASA Board**
**April 2018**

### Purpose of the Guidelines

The American Statistical Association's Ethical Guidelines for Statistical Practice are intended to help statistics practitioners make decisions ethically. Additionally, the Ethical Guidelines aim to promote accountability by informing those who rely on statistical analysis of the standards that they should expect. The discipline of statistics links the capacity to observe with the ability to gather evidence and make decisions, providing a foundation for building a more informed society. Because society depends on informed judgments supported by statistical methods, all practitioners of statistics, regardless of training and occupation or job title, have an obligation to work in a professional, competent, respectful, and ethical manner.

Good statistical practice is fundamentally based on transparent assumptions, reproducible results, and valid interpretations. In some situations, Guideline principles may conflict, requiring individuals to prioritize principles according to context. However, in all cases, stakeholders have an obligation to act in good faith, to act in a manner that is consistent with these Guidelines, and to encourage others to do the same. Above all, professionalism in statistical practice presumes the goal of advancing knowledge while avoiding harm; using statistics in pursuit of unethical ends is inherently unethical.

Ethical statistical practice does not include, promote, or tolerate any type of professional or scientific misconduct, including, but not limited to, bullying; sexual or other harassment; discrimination based on personal characteristics; or other forms of intimidation.

The principles expressed here should guide both those whose primary occupation is statistics and those in all other disciplines who use statistical methods in their professional work. Therefore, throughout these Guidelines, the term "statistician" includes all practitioners of statistics and quantitative sciences, regardless of job title or field of degree, comprising statisticians at all levels of the profession and members of other professions who utilize and report statistical analyses and their implications.

**A. Professional Integrity and Accountability**

The ethical statistician uses methodology and data that are relevant and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results. The ethical statistician does not knowingly accept work for which he/she is not sufficiently qualified, is honest with the client about any limitation of expertise, and consults other statisticians when necessary or in doubt. It is essential that statisticians treat others with respect.

The ethical statistician:

1. Identifies and mitigates any preferences on the part of the investigators or data providers that might predetermine or influence the analyses/results.
2. Employs selection or sampling methods and analytic approaches appropriate and valid for the specific question to be addressed, so that results extend beyond the sample to a population relevant to the objectives with minimal error under reasonable assumptions.
3. Respects and acknowledges the contributions and intellectual property of others.
4. When establishing authorship order for posters, papers, and other scholarship, strives to make clear the basis for this order, if determined on grounds other than intellectual contribution.
5. Discloses conflicts of interest, financial and otherwise, and manages or resolves them according to established (institutional/regional/local) rules and laws.
6. Accepts full responsibility for his/her professional performance. Provides only expert testimony, written work, and oral presentations that he/she would be willing to have peer reviewed.
7. Exhibits respect for others and, thus, neither engages in nor condones discrimination based on personal characteristics; bullying; unwelcome physical, including sexual, contact; or other forms of harassment or intimidation, and takes appropriate action when aware of such unethical practices by others.

**B. Integrity of data and methods**

The ethical statistician is candid about any known or suspected limitations, defects, or biases in the data that may impact the integrity or reliability of the statistical analysis. Objective and valid interpretation of the results requires that the underlying analysis recognizes and acknowledges the degree of reliability and integrity of the data.

The ethical statistician:

1. Acknowledges statistical and substantive assumptions made in the execution and interpretation of any analysis. When reporting on the validity of data used, acknowledges data editing procedures, including any imputation and missing data mechanisms.
2. Reports the limitations of statistical inference and possible sources of error.

3. In publications, reports, or testimony, identifies who is responsible for the statistical work if it would not otherwise be apparent.
4. Reports the sources and assessed adequacy of the data; accounts for all data considered in a study and explains the sample(s) actually used.
5. Clearly and fully reports the steps taken to preserve data integrity and valid results.
6. Where appropriate, addresses potential confounding variables not included in the study.
7. In publications and reports, conveys the findings in ways that are both honest and meaningful to the user/reader. This includes tables, models, and graphics.
8. In publications or testimony, identifies the ultimate financial sponsor of the study, the stated purpose, and the intended use of the study results.
9. When reporting analyses of volunteer data or other data that may not be representative of a defined population, includes appropriate disclaimers and, if used, appropriate weighting.
10. To aid peer review and replication, shares the data used in the analyses whenever possible/allowable, and exercises due caution to protect proprietary and confidential data, including all data that might inappropriately reveal respondent identities.
11. Strives to promptly correct any errors discovered while producing the final report or after publication. As appropriate, disseminates the correction publicly or to others relying on the results.

**C. Responsibilities to Science/Public/Funder/Client**

The ethical statistician supports valid inferences, transparency, and good science in general, keeping the interests of the public, funder, client, or customer in mind (as well as professional colleagues, patients, the public, and the scientific community).

The ethical statistician:

1. To the extent possible, presents a client or employer with choices among valid alternative statistical approaches that may vary in scope, cost, or precision.
2. Strives to explain any expected adverse consequences of failure to follow through on an agreed-upon sampling or analytic plan.
3. Applies statistical sampling and analysis procedures scientifically, without predetermining the outcome.
4. Strives to make new statistical knowledge widely available to provide benefits to society at large and beyond his/her own scope of applications.
5. Understands and conforms to confidentiality requirements of data collection, release, and dissemination and any restrictions on its use established by the data provider (to the extent legally required), and protects use and disclosure of data accordingly. Guards privileged information of the employer, client, or funder.

**D. Responsibilities to Research Subjects**

The ethical statistician protects and respects the rights and interests of human and animal subjects at all stages of their involvement in a project. This includes respondents to the census or to surveys, those whose data are contained in administrative records, and subjects of physically or psychologically invasive research.

The ethical statistician:

1. Keeps informed about and adheres to applicable rules, approvals, and guidelines for the protection and welfare of human and animal subjects.
2. Strives to avoid the use of excessive or inadequate numbers of research subjects, and excessive risk to research subjects (in terms of health, welfare, privacy, and ownership of their own data), by making informed recommendations for study size.
3. Protects the privacy and confidentiality of research subjects and data concerning them, whether obtained from the subjects directly, other persons, or existing records. Anticipates and solicits approval for secondary and indirect uses of the data, including linkage to other data sets, when obtaining approvals from research subjects, and obtains approvals appropriate to allow for peer review and independent replication of analyses.
4. Knows the legal limitations on privacy and confidentiality assurances and does not over-promise or assume legal privacy and confidentiality protections where they may not apply.
5. Considers whether appropriate research-subject approvals were obtained before participating in a study involving human beings or organizations, before analyzing data from such a study, and while reviewing manuscripts for publication or internal use. The statistician considers the treatment of research subjects (e.g., confidentiality agreements, expectations of privacy, notification, consent, etc.) when evaluating the appropriateness of the data source(s).
6. In contemplating whether to participate in an analysis of data from a particular source, refuses to do so if participating in the analysis could reasonably be interpreted by individuals who provided information as sanctioning a violation of their rights.
7. Recognizes that any statistical descriptions of groups may carry risks of stereotypes and stigmatization. Statisticians should contemplate, and be sensitive to, the manner in which information is framed so as to avoid disproportionate harms to vulnerable groups.

**E. Responsibilities to Research Team Colleagues**

Science and statistical practice are often conducted in teams made up of professionals with different professional standards. The statistician must know how to work ethically in this environment.

The ethical statistician:

1. Recognizes that other professions have standards and obligations, that research practices and standards can differ across disciplines, and that statisticians do not have obligations to standards of other professions that conflict with these Guidelines.

2. Ensures that all discussion and reporting of statistical design and analysis is consistent with these Guidelines.

3. Avoids compromising scientific validity for expediency.

4. Strives to promote transparency in design, execution, and reporting or presenting of all analyses.

**F. Responsibilities to Other Statisticians or Statistics Practitioners**

The practice of statistics requires consideration of the entire range of possible explanations for observed phenomena, and distinct observers drawing on their own unique sets of experiences can arrive at different and potentially diverging judgments about the plausibility of different explanations. Even in adversarial settings, discourse tends to be most successful when statisticians treat one another with mutual respect and focus on scientific principles, methodology and the substance of data interpretations. Out of respect for fellow statistical practitioners, the ethical statistician:

1. Promotes sharing of data and methods as much as possible and as appropriate without compromising propriety. Makes documentation suitable for replicate analyses, metadata studies, and other research by qualified investigators.
2. Helps strengthen the work of others through appropriate peer review; in peer review, respects differences of opinion and assesses methods, not individuals. Strives to complete review assignments thoroughly, thoughtfully, and promptly.
3. Instills in students and non-statisticians an appreciation for the practical value of the concepts and methods they are learning or using.
4. Uses professional qualifications and contributions as the basis for decisions regarding statistical practitioners' hiring, firing, promotion, work assignments, publications and presentations, candidacy for offices and awards, funding or approval of research, and other professional matters.

**G. Responsibilities Regarding Allegations of Misconduct**

The ethical statistician understands the differences between questionable statistical, scientific, or professional practices and practices that constitute misconduct. The ethical statistician avoids all of the above and knows how each should be handled.

The ethical statistician:

1. Avoids condoning or appearing to condone statistical, scientific, or professional misconduct.

2. Recognizes that differences of opinion and honest error do not constitute misconduct; they warrant discussion, but not accusation.

3. Knows the definitions of, and procedures relating to, misconduct. If involved in a misconduct investigation, follows prescribed procedures.

4. Maintains confidentiality during an investigation, but discloses the investigation results honestly to appropriate parties and stakeholders once they are available.

5. Following an investigation of misconduct, supports the appropriate efforts of all involved, including those reporting the possible scientific error or misconduct, to resume their careers in as normal a manner as possible.

6. Avoids, and acts to discourage, retaliation against or damage to the employability of those who responsibly call attention to possible scientific error or to scientific or other professional misconduct.

**H. Responsibilities of Employers, Including Organizations, Individuals, Attorneys, or Other Clients Employing Statistical Practitioners**

Those employing any person to analyze data are implicitly relying on the profession's reputation for objectivity. However, this creates an obligation on the part of the employer to understand and respect statisticians' obligation of objectivity.

Those employing statisticians are expected to:

1. Recognize that the Ethical Guidelines exist, and were instituted, for the protection and support of the statistician and the consumer alike.

2. Maintain a working environment free from intimidation, including discrimination based on personal characteristics; bullying; coercion; unwelcome physical (including sexual) contact; and other forms of harassment.

3. Recognize that valid findings result from competent work in a moral environment. Employers, funders, or those who commission statistical analysis have an obligation to rely on the expertise and judgment of qualified statisticians for any data analysis. This obligation may be especially relevant in analyses that are known or anticipated to have tangible physical, financial, or psychological impacts.

4. Recognize that the results of valid statistical studies cannot be guaranteed to conform to the expectations or desires of those commissioning the study or the statistical practitioner(s).

5. Recognize that it is contrary to these Guidelines to report or follow only those results that conform to expectations without explicitly acknowledging competing findings and the basis for choices regarding which results to report, use, and/or cite.

6. Recognize that the inclusion of statistical practitioners as authors, or acknowledgement of their contributions to projects or publications, requires their explicit permission because it implies endorsement of the work.

7. Support sound statistical analysis and expose incompetent or corrupt statistical practice.

8. Strive to protect the professional freedom and responsibility of statistical practitioners who comply with these Guidelines.

References

American Statistical Association (2018). *American Statistical Association (ASA) Ethical Guidelines for Statistical Practice*. Downloaded from on 21 April 2018.
Baker M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454 (26 May 2016) doi:10.1038/533452a
Collins FS & Tabak LA. (2014). Policy: NIH Plans to enhance reproducibility. *Nature* 505 (30 Jan 2014): 612-613. Downloaded from http://www.nature.com/news/policy-nih-plans-to-enhance-reproducibility-1.14586 on 10 March 2016
Coventry LL, Finn J, Bremner AP. (2011). Sex differences in symptom presentation in acute myocardial infarction: a systematic review and meta-analysis. *Heart Lung* 40(6):477-91. doi: 10.1016/j.hrtlng.2011.05.001.
Freedman DH. (2010). Lies, damned lies, and medical science. The Atlantic November 2010 issue; downloaded from http://www.theatlantic.com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/308269/ on 7 November 2015.
Hubert L & Wainer H. (2011). A statistical guide for the ethically perplexed. In AT Panter & SK Sterba (Eds)., *Handbook of ethics in quantitative methodology.* New York, NY: Routledge. pp 61-124.
Loukides M, Mason H, Patil DJ. (10 July 2018). *Doing good data science.* Downloaded from https://www.oreilly.com/ideas/doing-good-data-science on 15 July 2018
McGoughy E. (2019). Could Brexit be void? *29(3) King's Law Journal 331-343*
*King's College London Law School Research Paper No. 2018-29*. Downloaded from https://poseidon01.ssrn.com/delivery.php?ID=872119091124113026024101002124089087049073070064025021111113077102070094101118124000120032062006006027002112106111097066026007000029026049019127069086031094091017124018052042067100001064104114067072026004117079095122125109075120125086104097074116086104&EXT=pdf
on 31 March 2019.
Kalichman M. (2013). Why teach research ethics? In, National Academy of Engineering (Eds). Practical Guidance on Science and Engineering Ethics Education for Instructors and Administrators. Washington, DC: National Academies Press. Pp. 5-16.
Penchava D. (2019). *Brexit and migration: our new research highlights fact-free news coverage*. Downloaded from https://theconversation.com/brexit-and-migration-our-new-research-highlights-fact-free-news-coverage-109309 on 31 March 2019.
Stark PB & Saltelli A. (2018). https://www.significancemagazine.com/2-uncategorised/593-cargo-cult-statistics-and-scientific-crisis Downloaded from https://www.significancemagazine.com/2-uncategorised/593-cargo-cult-statistics-and-scientific-crisis on 18 July 2018.
Steneck NH (2007). *Introduction to the Responsible Conduct of Research – Revised Edition*. Washington, DC: US Department of Health and Human Services. Downloaded from http://ori.dhhs.gov/documents/rcrintro.pdf on 1 January 2010.
Tractenberg RE, Russell AL, Morgan G, FitzGerald KT, Collmann J, Vinsel L, Steinmann M, Dolling LM. (2015) Amplifying the reach and resonance of ethical codes of conduct through ethical reasoning: preparation of Big Data users for professional practice. *Science and Engineering Ethics* 21(6): 1485-1507. DOI 10.1007/s11948-014-9613-1

Tractenberg RE & FitzGerald KT. (2012). A Mastery Rubric for the design and evaluation of an institutional curriculum in the responsible conduct of research. *Assessment and Evaluation in Higher Education*. 37(7-8): 1003-21. DOI 10.1080/02602938.2011.596923.

Tractenberg RE. (2016-a). Institutionalizing ethical reasoning: integrating the ASA's Ethical Guidelines for Professional Practice into course, program, and curriculum. In, J. Collmann & S. Matei (Eds)., *Ethical Reasoning in Big Data: An Exploratory Analysis*. New York: Springer. 115-139.

Tractenberg RE. (2016-b). *Why* and *How* the ASA Ethical Guidelines Should Be Integrated into Every Quantitative Course. *Proceedings of the 2016 Joint Statistical Meetings*, Chicago, IL. Pp. 517-535.