OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

8

Safe or Sorry? Cancer Screening and Inductive Risk

Anya Plutynski

Introduction

The focus of this chapter will be on the epistemic and normative questions at issue in debates about cancer screening, with a special focus on mammography as a case study. Such questions include: How do we know who needs to be screened? What are the benefits and harms of cancer screening, and what is the quality of evidence for each? How ought we to measure and compare these benefits and harms? What are the sources of uncertainty about our estimates of benefit and harm? Why are such issues so contested? What are the major drivers of dissent and consensus on the data and their interpretation? How, if at all, do values play a role in debates surrounding mammography screening? In sum: In what ways do inductive risk, broadly conceived,[1] come into play in the science behind cancer screening, and mammography screening in particular?

Problems of underdetermination and thus inductive risk come into play in the science of mammography screening at several points: in assessing epidemiological data on cancer prevalence and mortality, in debating how best

1. Inductive risk, as defined by Heather Douglas (2000), is "risk of error" in inferring to a scientific hypothesis; she argues that non-epistemic values enter into scientific inquiry in cases where risk of error has non-epistemic consequences (i.e., in a scientific inquiry into the effects of dioxin on humans or the environment, or, of course, in medicine). One might speak of "epistemic risk" more broadly (following Biddle and Kukla, this volume) as "any risk of epistemic error that arises anywhere during knowledge practices." This could include not only accepting false hypotheses but also upstream methodological choices, such as characterization or classification of data.

acprof-9780190467715.indd 149 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

to design trials, in evaluating the data from clinical trials, in constructing meta-analyses, and last but not least, in assessing how best to communicate about the risks and benefits of screening. From the design of studies, to the interpretation of results, to their application in the clinic, there are choices that need to be made about which risk trade-offs we are willing to tolerate—whether we are inferring to merely empirical hypotheses (e.g., about the baseline risk of cancer in group X), or more "mixed" evaluative judgments about overall "effectiveness" of mammography. Because these choices involve uncertainty, values arguably play an ineliminable role in the science supporting claims about the relative effectiveness of mammography screening. To be clear, my aim here is not to claim that the science is flawed. Rather, the aim is to explore the dimensions of uncertainty and inductive risk (broadly understood) in our estimates of benefits and costs of screening, and encourage greater transparency among scientists and in the public. This will ideally result in more open and informed decision-making by patients and their families.

What Is Inductive Risk?

According to what has been called the "argument from inductive risk," scientists need to decide whether to accept or reject a hypothesis, based on some limited body of evidence. Such judgments can be in the wrong; one runs a risk of error.[2] When choosing the standards of evidence required for accepting or rejecting hypotheses, scientists thus must weigh the importance of making various sorts of mistakes. This consideration informs scientists' methodological choices. For example, scientists might choose to avoid either type I or type II error, or choose different p values, such that the chance of error is more or less likely (Churchman 1948; Rudner 1953). That considerations of seriousness of error play a role in establishing standards of evidence is one of the main arguments for the role of values in science; insofar as such decisions are value-laden ones, as Richard Rudner puts it, "the scientist qua scientist makes value judgments" (see also Douglas 2000, 2009; Elliott 2011).

2. Some (Bayesian) philosophers may already resist the very idea that scientists must accept or reject a given hypothesis. Thus, they may argue that all that is required is that scientists assign a subjective degree of belief to the hypothesis. We will return to this possibility later.

acprof-9780190467715.indd 150 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

Daniel Steel (2015b) nicely characterizes the argument from inductive risk as follows:

1. One central aim of scientific inference is to decide whether to accept or reject hypotheses.
2. Decisions about whether to accept or reject a scientific hypothesis can have implications for practical action, and when this happens, acceptance decisions should depend in part on nonepistemic value judgments about the costs of error.
3. Therefore, nonepistemic values can legitimately influence scientific inference. (2)

Critics of this argument have rejected either premise 1 or premise 2. Richard Jeffrey (1956) denied premise 1 as follows: scientists are not in the business of accepting or rejecting hypotheses; rather, they simply assign probabilities to hypotheses (or, a range of probabilities), where these probabilities represent degrees of belief. It is up to policymakers to decide what to do with this information. If values enter in, it is only in the context of moving from evidence to practical decisions about what to do. According to Jeffrey, the scientist qua scientist simply reports subjective degrees of belief. If Jeffrey is right, then the argument from inductive risk fails.

Rudner (1953) anticipated and responded to this objection, and several philosophers of science have echoed or elaborated upon Rudner's reply (Douglas 2000, 2009; Steel 2015b; Steele 2012). For instance, Heather Douglas argues that contextual factors influence a much earlier stage of research than the interpretation of data, or reporting of probabilities, namely the characterization of data (Douglas 2000, 569–72). Katie Steele (2012) argues that even if we grant Jeffrey's point, when communicating their findings, scientists have to convert their subjective degrees of belief to some other measure. The translation involves a decision on the part of the scientist about how to report these degrees of belief; and, Steele (2012) argues that this is a value-laden decision. For instance, scientists have to give a confidence interval for some probability distribution, and "a cautionary approach . . . typically amounts to a wider credible interval. . . . This choice involves value judgments" (10). Steel (2015a) has a similar argument:

> acceptance, and thus the argument from inductive risk, is already in the picture at the stage of deciding upon a probability model for the likelihood function and prior distribution. These decisions involve a choice

acprof-9780190467715.indd 151 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

> among probability distributions: uniform, binomial, Poisson, gamma, beta, normal, etc. And these decisions can have consequences for what sorts of errors are more or less likely. Consequently . . . acceptance decisions and the value judgments they entail are there from the start. (5)

In sum, Douglas, Steele, and Steel argue that even were we to grant Jeffrey's view, arriving at an assignment of probabilities requires choices that make appeal to background assumptions, and so are at least potentially value laden. In Steel's words: "probabilistic assessments of evidence or degrees of confirmation themselves depend *on accepting data, background knowledge, and probability models*, and hence are also subject to the argument from inductive risk" (Steel 2015a, 7, emphasis added). We will return to this argument in the context of epidemiology of mammography screening.

Most critics of the argument from inductive risk, however, have rejected premise 2. As Steel notes, many have argued that "non-epistemic values should not influence acceptance construed in a cognitive, non-behavioral sense that is appropriate to science (Dorato 2004; Lacey 1999, 2004; Levi 1960, 1962, 1967; McMullin 1982; Mitchell 2004)" (Steel 2015b, 150). Steel offers a reply to this objection that draws upon Jonathan Cohen's (1992) thesis that to accept a proposition p in a context is to decide to treat p as an available premise for reasoning in that context. On Steel's view, this sense of "acceptance" does not require that we interpret acceptance in a behavioral sense. This move raises some interesting questions, for example, about whether inference itself is a kind of behavior. Setting this aside, however, it does seem that in judgments about background beliefs we are willing to treat as premises in arguments that, for example, some medical intervention is more or less effective, there is an opportunity for normative values to influence one's choice, whether such a choice is in terms of how we operationalize "effectiveness," or perhaps that the relevant probability distribution of absolute risk in this case is a normal one, (even with the caveat that we take this to only have a probability of 80%). Steel argues that such choices shape both "upstream" and "downstream" stages of scientific reasoning.

While there are clearly several decision points about what to accept in various stages in the research on cancer screening, it is unclear whether we ought to regard them as merely choices to treat a certain proposition as a premise in an argument. For instance, choices of study design or the choice to exclude data in one's analysis or to treat diagnoses of ductal carcinoma in situ (DCIS) as diagnoses of cancer are not exactly like accepting a premise. Nonetheless, they involve inductive risk or perhaps "epistemic risk" in the broader sense

acprof-9780190467715.indd 152 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

discussed by Kukla and Biddle (this volume). They argue that epistemic risk, broadly conceived, is any risk of epistemic error that arises anywhere during knowledge-making practices. Such practices may involve not only inferences but also practical choices of research design or choices of operational measures and definitions, just to name two. Whether you characterize these more "practical" decisions, as well as more narrowly epistemic ones, as carrying "inductive" or "epistemic" risk, both clearly carry risk of over- or underestimation of the benefit of screening. Both also are at least potentially value-laden. Below, we will consider three ways in which the assessment of mammography's effectiveness involves such risky choices: choice of endpoint, design and assessment of trials, and estimation of one important potential harm of screening: overdiagnosis.

The Science of Cancer Prevention: Starting with Choice of Endpoint

The first step in assessing the effectiveness of any medical intervention is choice of outcome measured. What is it that we wish to be "effective" at? The choice of endpoint, or what we measure when measuring the success of an intervention, and how we measure it, can yield different conclusions about the relative effectiveness of an intervention. Not all choices of endpoint converge on the same assessments of a medical intervention. Indeed, the wrong choice of endpoint can cause one to make very poor assessments of screening effectiveness, a point made especially vivid by a statement NY City Mayor Rudi Giuliani made in a 2007 campaign advertisement. He explained: "I had prostate cancer, 5, 6 years ago. My chance of surviving prostate cancer—and thank God, I was cured of it—in the United States? Eighty-two percent. My chance of surviving prostate cancer in England? Only 44 percent under socialized medicine" (cf. Gigerenzer et al. 2008). Giuliani's argument on behalf of US health care's advantages is founded on a misleading measure of success. Five-year survival rates are the rate of survival of patients diagnosed with disease; for prostate cancer, these are roughly 82% in the United States versus 44% the United Kingdom. This may sound very compelling, until one considers how these rates are measured. Five-year survival is estimated by taking the number of patients diagnosed with cancer who are still alive after five years, and dividing by the total number diagnosed with cancer. This number, however, may be artificially inflated by an increase in diagnosis of early stage cancers, and this is exactly what happened in the case of prostate cancer in the United States (for a discussion, see, e.g., Gigerenzer et al. 2008).

acprof-9780190467715.indd 153 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

Early diagnosis does not necessarily mean one benefits from screening; in fact, early diagnosis may simply increase the amount of time one is aware of a disease. Consider twins John and Bill; imagine that they have the exact same cancer with the exact same prognosis. John is diagnosed at 65 and Bill, at 70. Both die at 75 from cancer. John's early diagnosis makes it look like he benefited from early diagnosis, if we understand "benefit" in terms of years of survival from diagnosis. For, strictly speaking, he "gained" five years of survival, as a result of early diagnosis, over and above Bill's survival from the time of diagnosis, even though they died at the same age. This is a case of lead-time bias. Lead-time bias is when a screening method appears to extend life, but in fact, patients are simply aware of the diagnosis for a longer period of time than their peers who opted out of screening.

The extreme of lead-time bias is overdiagnosis bias, the bias in estimates of effectiveness that comes from the diagnosis of disease that would never have led to symptoms in the lifetime of the patient. Overdiagnosis can occur when someone is diagnosed with an indolent or slow-growing disease, or, they are diagnosed in very old age, when the patient is more likely to die of other causes before the cancer can progress to clinical symptoms. Including such cases in estimates of screening's preventive effectiveness is thus a serious flaw; for, no deaths were in fact prevented, and indeed, screening in this case can cause harm. Some cases of prostate cancer are indolent or relatively slow-growing; many men die with prostate cancer, but not of prostate cancer. Two recent clinical trials, one in the United States and a second in Europe, estimated that the number of men overdiagnosed for prostate cancer may have been as high as 40%–50% (Liong et al. 2012, e45803). This rate of overdiagnosis carries no small cost. The quality of life for men treated for prostate cancer may be decreased, because standard treatment for prostate cancer involves removal of the prostate, which may cause incontinence or impotence, or "chemical castration" (the administration of hormones) which causes weight gain (cf. Shen et al. 2015). These costs of screening would be invisible if five-year survival rates were used as a measure of the relative effectiveness of prostate cancer screening.

Reduction in age-adjusted cancer mortality as an endpoint, in contrast, is less likely to lead to inflated ideas about the effectiveness of a screening regimen. Indeed, some argue that we should not be measuring cancer-related mortality—that is, lives lost to cancer—but overall mortality. This is because disease-specific reductions in mortality are not necessarily the same as reductions in overall mortality. Indeed, overall mortality may actually increase because of screening; for, screening itself carries some risk, either because of

acprof-9780190467715.indd 154 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

the long-term effects of radiation, or biopsies and other follow-up imaging, following from false alarms (cf. Prasad, Lenzer, and Newman 2016). In sum, a more comprehensive assessment of a screening regimen's effectiveness should, in principle, measure reduction of overall mortality. Indeed, a different choice of endpoint may result in a reversal of assessments of the overall effectiveness of screening. In sum, in defining a measure of "effectiveness," we need to first answer the questions: What do we care about?

In sum, without attention to these forms of bias (lead time, overdiagnosis bias), the choice of measure for estimation of the effectiveness of screening may be misleading. This has bearing on our discussion of inductive risk. As Steele (2012) pointed out, scientists must make a choice about what they plan to measure, as well as how to communicate the results of their research. A scientist or policymaker who hopes to convey the effectiveness of screening thus might only measure five-year survival, or, relative risk reduction versus absolute risk reduction. That is, they may choose to represent the information in a way that suggests that the benefits of screening far outweigh any potential harms. Alternately, a scientist more concerned about the potential harms of screening might choose to report age-adjusted reduction of mortality. A good statistician or epidemiologist understands the difference. But, all too often, such differences are not transparent to many readers of the literature; indeed, even clinicians misunderstand or misinterpret this basic difference between mortality and survival statistics (Gigerenzer et al. 2008). One could argue that choice of endpoint is a matter of science communication, not inductive risk, because the question at issue here is which data to report, not an inference to a hypothesis, as in the standard cases of inductive risk. However, by choosing to define or operationalize "effective" interventions as those which increase rates of five-year survival, one is making a value-laden choice that may result in overestimation of screening's benefits and increase in overdiagnosis. In contrast, taking reduction of age-adjusted mortality as the goal of screening avoids both lead time and overdiagnosis bias.

The assessment of mammography screening's effectiveness does not, of course, end with choice of endpoint. The following sections will discuss the assessment of the clinical trials themselves and contested estimates of one potential harm of screening: overdiagnosis.

Mammography Screening Trials

Governmental organizations, as well as international bodies and professional societies, such as the US Preventive Services Task Force (USPSTF)

acprof-9780190467715.indd 155 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

and the American College of Radiology, disagree quite significantly on how to interpret the evidence for mammography's effectiveness in particular age groups. In all these cases, competing views about the rigor of the research are often very difficult to separate from normative considerations, concerning the relative weighting of competing precautionary considerations. There are also quite serious differences in methodology; for instance, whether, and to what extent, "evidence-based" medicine versus clinical expertise ought to play a role in recommendations for various interventions is clearly at play in the mammography "wars."[3]

There are several points where risk of error comes into play in the assessment of mammography screening: in the matter of deciding how best to select subjects to include in one's study, in the choice of measures of potentially confounding variables to consider in matching cases with controls (e.g., age, SES, etc.), in design of a trial (e.g., how to randomize, how many rounds of mammography to consider, how long to follow patients subsequent to the trial, whether autopsies will be conducted, whether overall mortality or only cancer-related mortality will be measured), and last but not least, in the overall assessment of the evidence (e.g., in assessments of statistical significance, exclusion or inclusion of data, whether DCIS cases will be considered as cancer diagnoses) and reporting of that evidence.

There have been eleven completed mammography screening trials, conducted in Sweden, Norway, the United States, Canada, the United Kingdom, and Singapore, according to the Cochrane collaboration (Gøtzsche and Jørgerson 2013). The studies each arrived at slightly different estimates of the benefit of mammography screening, though this is not altogether surprising, as the trials were conducted at different times, in different places, with slightly different populations (some included women as young as 40 and as old as 75; others only investigated 50–59), and different methods of randomization. These trials have been subject to a host of analyses and critiques, by national health services and international bodies (e.g., the Cochrane review, WHO).

The assessment of overall benefit versus harm of mammography generally involves meta-analyses and systematic reviews. Meta-analyses use quantitative statistical techniques to synthesize the results of several studies, to yield what is called a "summary effect size" or single quantitative measure of the

3. See Solomon (2015, ch. 9), for a thoughtful discussion of how competing methodological stances have informed debates among different organizations regarding mammography between the ages of 40 and 50. For an overview of how "intuitive" thinking about cancer screening has undergone refinement, see Croswell et al. (2010).

acprof-9780190467715.indd 156 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

effect of some intervention (Uman 2011). Meta-analyses begin with a survey of the available literature. Studies that are inconclusive, or, where measure of outcomes is not commensurable, may be excluded from consideration. Thus, there is a risk of error and a potential for values to enter into decision-making even in this relatively "pure" statistical method (cf. Ioannidis 2008; Stegenga 2011). Systematic reviews are also based on searches of the scientific literature, with the goal of "identifying, appraising, and synthesizing all relevant studies on a particular topic" (Uman 2011, 57). Systematic reviews will often include a meta-analysis, but also include other sources of evidence, and often rank the quality of evidence, based on prior agreed-upon standards for assessing the quality of a particular type of study. Randomized clinical trials are by and large regarded as the gold standard of evidence for establishing that various clinical interventions are effective.[4] In cases where randomization is impossible, case-control, cohort studies, and ecological studies are considered as sources of evidence in support of linking an intervention with some outcome. Various organizations have established protocols for assessing the quality of clinical trials and for ranking and compiling such evidence in a systematic review. For instance, Cochrane's acronym PICO (or PICOC)—Population, Intervention, Comparison, Outcomes and Context—denotes a set of components considered essential to the assessment of the quality of evidence in a systematic review. This is a kind of institutionalized attempt to make standards explicit, or exhibit transparency, as well as control the role of competing values in assessment of evidence for effectiveness of various medical interventions.

At issue in all these reviews is which studies to include, and what kinds of bias may have been operating in each trial, as well as whether extrapolating to current practice is warranted. For instance, a major concern, of radiologists in particular, is that there have been significant technological advances in digital mammography since these trials were conducted. Another concern is that the studies' methods and quality were highly variable, and results were not commensurable; or, the outcome measures in different trials were not exactly identical—some include DCIS among cases of diagnosis of cancer, others only measured invasive disease. Given the extent of uncertainty about both the quality and import of the trials, as well as their comparability, different institutions have arrived at different conclusions about mammography's

4. For a critical discussion of RCTs as the "gold standard," see, e.g., Cartwright (2007), Howick (2011), and Worrall (2002).

acprof-9780190467715.indd 157 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

benefits. And some reviews have been hugely contentious (for a point by point history of the Cochrane review, from the perspective of one of the participants, see, e.g., Gøtzsche [2012], one of the co-authors).

Despite disagreement on the upshot, however, everyone who has reviewed the research found several potential sources of bias, even in the best, most well-controlled studies. Sources of potential bias included: (1) how women were invited to participate, (2) how they were randomized in the study, for example, whether the screened individuals were appropriately matched to controls (thus controlling for *selection bias*) (3) whether women excluded from the study either pre- or post-randomization were excluded for reasons that could lead to additional selection bias (e.g., excluding women who had already been diagnosed with breast cancer from the screened versus control group), (4) whether participants or personnel were adequately blinded (*performance bias*), (5) whether autopsies were conducted on participants and whether cause of death assignments were blinded (*detection bias*), (6) whether there were high levels of attrition or lack of participation (*attrition bias*), and last but not least, (7) whether evidence was tampered with or handled inappropriately, or whether records were kept accurately.

Unfortunately, one or more of these biases were evident in even the best of the studies. Some studies counted the same controls twice (Stockholm); others excluded more women from the control than the screened group with previous breast cancers (Malmö); and yet others used different clinics for control versus screened groups, so that there was (arguably) a difference in socioeconomic status between groups (Edinburgh) (selection bias) (Gøtzsche and Jørgensen 2013). While screening appeared to show a benefit, particularly in women from ages 50 to 60, in many of these trials, a particularly contentious result of the Cochrane review was that the benefit of screening women starting as early as 40, was small, if not absent altogether. This should not be altogether surprising; cancer is a disease that increases in prevalence with age, and so, the benefit of screening to younger women is expected to be small.[5]

5. Prevalence of disease is distinct from incidence. Incidence is the number of cases diagnosed in a given population in a specific time frame. Prevalence is the number of people with the condition at a specific point in time. If a disease is very low prevalence, screening is by and large unwarranted. This is because screening a lot of healthy people for a very rare disease, even with a highly sensitive test, is likely to turn up a lot of false positives. A screening test is highly sensitive when it detects most disease. That is, if the disease is present, the test is very likely to be positive. This comes at the expense of lots of false positives. In contrast, a test is specific when has few false positives, but may miss some of the disease. This is good if the condition is uncommon, and the cost of false positives overwhelms the advantage of finding disease. The positive predictive value of a screening test (PPV) is the number of true positives out of the total number of positive diagnoses (true + false positives).

acprof-9780190467715.indd 158 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

One of the striking features of mammography is its relatively low positive predictive value (Saslow et al. 2007). That is, mammography screening can rule a lot of cancers "in," but cannot rule a lot of benign lesions "out." Especially in relatively younger women, or those with fibrocystic or dense breasts, there are frequent "follow-up" mammograms and biopsies for suspicious lesions. This is a well-known cost of screening, one that increases with lower prevalence. Indeed, for any screening test, even one that is highly specific and highly sensitive, if the disease is very rare (i.e., prevalence is low), you run the risk of many false positives. Imagine that the rate of disease in a population is 0.05% (the disease is rare). Out of 10,000 individuals screened, it turns out that even if a test is 95% sensitive and 95% specific, the test will still identify as many as 500 false positives.[6] Thus, not surprisingly, universal screening for cancer is likely to identify many false positive cases in younger populations, where cancer is rare.

One of the main areas of contention in assessing the effectiveness of cancer screening is how "low you go" (i.e., how young should screening start, given that cancer is much less prevalent in the young?). The only way to assess this is to try to assign a measure of the magnitude of the benefit of screening (or reduced risk of mortality from breast cancer) given the baseline (or absolute risk of mortality from breast cancer) in any particular age group. For women ages 40–49, the Cochrane review concluded that the absolute risk reduction was very small or none at all (Gøtzsche and Jørgerson 2013). In other words, given the very low prevalence of disease in this population, the absolute benefit in reduction of age-adjusted mortality to younger women was very small, at least according to the best available trial data.

In systematic reviews determining whether the benefits of cancer screening outweigh the costs, there are very different estimates of the overall harms in terms of number of unnecessary tests—and false positive diagnoses—for different age groups. According to one estimate, "for women between the ages of 40 and 49 years, the false positive rate is quite high, and the expected benefits are quite low: more than 1900 women would need to be invited for screening mammography in order to prevent just one death from breast

6. If the disease has .05% prevalence, then 5 out of 10,000 individuals are expected to have disease. If a test is 95% sensitive and specific, then, 95% of the time, those found to be negative are negative, and 95% of the time, those found to be positive are positive. So, let's say that almost five out of those five found to be positive are positive. But, it is also true that of the 9,995 of those without disease, only 9,495 are found to be negative (.95 × 9,995 = 9,495). But that means that roughly 500 of those found positive are in fact false positives, since 9,995 (*TN* + *FP*) − 9,495 (TN) = 500.

acprof-9780190467715.indd 159 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

cancer during 11 years of follow-up, at the direct cost of more than 20,000 visits for breast imaging and approximately 2000 false positive mammograms" (Quanstrum and Hayward 2010, 1076). Perhaps needless to say, considering these factors at all is contentious; indeed, some contend that such psychological and financial costs should not be considered in assessments of the overall effectiveness of screening. The argument seemed to be that comparing psychological harm to mortality was simply to compare incommensurables. Nonetheless, the USPSTF "modeling" report attempted to quantify the relative risks and benefits of screening, based on six models that were developed independently within the Cancer Intervention and Surveillance Modeling Network (CISNET) of the National Cancer Institute (NCI) (USPSTF 2009). They argued that the overall costs (including harms to women in their forties in terms of repeat mammograms, unnecessary biopsies, associated psychological harm, and overdiagnosis), outweighed the marginal benefits (in terms of mortality reduction) to women. Critics objected to everything from which harms to include, to how to measure them, to whether it was morally objectionable to compare resource costs with lives saved.

Despite their differences, in 1997, and yet again in 2009, the USPSTF (2009), and the Cochrane review (Gøtzsche and Jørgerson 2013) arrived at relatively similar recommendations. The USPSTF panel concluded that the evidence suggested that "routine" (i.e., annual or biannual) screening for women under 50 was not worth the overall cost. Instead, patients under 50 ought to consult with a physician to discuss the benefits and risks of routine screening, given their individual risk factors (family history, parity, smoking habits, etc.). Moreover, they argued that the benefits of biannual screening are most evident for populations in the age range of 50–74. In this, they followed the Cochrane's review.

These reviews were met with a firestorm of opposition. The American Cancer Society, the American College of Radiology, the Society for Breast Imaging, and the Radiology Society of North America all rejected these conclusions, and recommended routine screening for women in their forties.[7] In Europe, the Cochrane review process was delayed, and even before the final review was produced, a paper published in the *Lancet* (Gøtzsche and Olsen 2000) created a huge stir. One issue that arose again and again was what to

7. At least until recently: The ACS pulled back from their initial resistance to the USPSTF's results in the summer of 2015 (Oeffinger et al. 2015). However, the Society for Breast Imaging and the American College of Radiology (ACR 2015) continues to recommend cancer screening in women ages 40–50, despite the ACS's reversal.

acprof-9780190467715.indd 160 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

include among overall costs and benefits assessed. In particular, the measure of the extent of overdiagnosis and its consideration as a cost, and relatedly, estimates of baseline prevalence, were a significant point of contention.

Measuring Overdiagnosis

In order to assess screening's effectiveness in reducing mortality from a particular type of cancer, one needs an estimate of background or baseline incidence and mortality from this particular cancer in this population (a group of individuals with a particular age range, or sex). That is, one needs to know how many individuals would have gotten cancer, and how many of these would die from cancer without screening. But, once screening has already become the standard of care, estimates of baseline incidence and mortality are difficult to arrive at. Various indirect sources of evidence of varying quality are thus appealed to, including historical epidemiological data and long-term follow-up data from the original clinical trials. Which such data to trust, and how decisive it is, is a contentious matter. For instance, one estimate of baseline incidence is arrived at by subtracting "catch-up" cancers in unscreened groups from the total cancers in the screened group, as measured in the original clinical trials.

The authors of the Cochrane review used this method to estimate rates of overdiagnosis and arrived at a strikingly high number, 30% (Gøtzsche and Jørgerson 2013). Perhaps not surprisingly, radiologists in particular have been skeptical of Cochrane's estimates of overdiagnosis (Detterbeck 2012; Kopans, Smith, and Duffy 2011). Some argue that there was insufficient "follow-up" time in measures of compensatory drop in mortality in the screened groups, or that variability in overdiagnosis estimates be explained by differences in screening policies and different uptake between programs (see, e.g., Kopans, Smith, and Duffy 2011). This is an instance of how acceptance of very different background beliefs, or what comparison and contrast is considered relevant to assessment of a given outcome, may well be informed by values. In the name of scientific rigor (e.g., challenging that we can extrapolate from baseline rates of cancer incidence and mortality in the past to the present), one can deny that such evidence is relevant to current estimates of screenings' benefits. But it is difficult to separate such epistemic norms from normative values; for, radiologists have a vested interest in insisting on screening's effectiveness (Quanstrum and Hayward 2010).

In addition, many remain skeptical of the possibility that some cancers are simply unlikely to progress. However, the concerns raised about overdiagnosis

acprof-9780190467715.indd 161 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

have led some organizations to rethink how to categorize early stage cancer. An NIH working group recommended that indolent lesions be renamed "IDLE," indolent lesions of epithelial origin (see, e.g., Esserman et al. 2014). This seems a clear case of a shift in priorities regarding inductive risk; the NIH is erring on the side of caution with respect to overdiagnosis. Emphasizing that extremely early stage diseases are unlikely to progress may well prevent overdiagnosis and overtreatment. At issue in estimates of overdiagnosis are thus background assumptions about the possibility of indolent disease, as well as the legitimacy of extrapolation from various sources of evidence about background incidence and mortality. There are three sources of this evidence, all of which are contested: autopsy data, historical RCTs, and historical data.

Consider the evidence for overdiagnosis that comes from autopsies. Autopsy studies have found a significant disease reservoir of subclinical cancers in otherwise healthy individuals. Two studies, one of American men, and the other a study of Greek men (all of whom died from causes other than cancer) determined that the disease reservoir of prostate cancer ranges from 30% to 70% (Welch and Black's 2010 estimate, citing Sakr et al. 1996 and Stamatiou et al. 2006; see also Santen, Yue, and Heitjan 2013). As might be expected, the disease reservoir of cancer was significantly age-dependent. Another study of thyroid cancer found disease reservoirs as high as 100% (Welch and Black's 2010 estimate, citing Harach, Franssila, and Wasenius 1985). A further study of breast cancers in middle-aged women who died from other causes found ranges from 7% to 39% (Welch and Black 1997; see also Santen, Yue, and Heitjan 2013). To be clear, these were all very early stage and in many cases likely slow growing or indolent disease; in contrast, the lifetime risk of metastatic disease is significantly lower than 30%–70%. It follows that screening may detect a significant percentage of cancers (overdiagnoses) that may never have resulted in clinical symptoms, disease or death (cf. Welch and Black 2010), especially in younger patients.

The second major source of evidence is long-term follow-up studies counting "catch-up" cancers in the unscreened groups, following completion of clinical trials. The difference between the number of catch-up cancers in the unscreened group and the total number discovered through screening is the absolute number of those overdiagnosed. Welch and Black (2010) estimate that this number in breast cancer could be as high as 24% for the Malmo trial (608). Critics of estimates based on this data contend that the authors did not follow the cohort long enough to detect a compensatory decline in mortality in the screened group (see, e.g., Kopans, Smith, and Duffy 2011; Puliti et al. 2012).

acprof-9780190467715.indd 162 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

Critics also argue that the defenders of high estimates of overdiagnosis were unreasonably assuming that background incidence rates are stable and extrapolating forward to current rates. They claim that there may well have been a "natural increase in incidence" (cf. Kopans, Smith, and Duffy 2011) across relevant time period(s), which would confound estimates of the extent of overdiagnosis in current practice. To be sure, this argument is a bit ad hoc; there is no special reason to think that rates of cancer incidence were increasing over this time, in a way that just happened to coincide with increasing rates of cancer screening. Nevertheless, the extent of disagreement in estimates of overdiagnosis is illustrated in a 2012 paper by Donella Puliti et al. in the *Journal of Medical Screening* where they discuss twenty different estimates of overdiagnosis in breast cancer, ranging from less than 10% to as high as 60%. Here is a case where inductive risk comes into play; the source of evidence for background rates of incidence and mortality, and so choice of one means of measuring overdiagnosis versus another may lead to either over- or underestimating the harms of screening. Different estimates of overdiagnosis are based on different estimates of baseline prevalence, and the evidence in support of these estimates of baseline prevalence varies in source and quality. Perhaps needless to say, the skepticism with which some critics regard high estimates of overdiagnosis is at least *prima facie* motivated by value. One can only imagine the strength of disincentive at work against acknowledging that as many as 60% of one's patients were diagnosed and treated unnecessarily for cancers that would never have progressed to symptoms. Perhaps the seriousness of harm involved is what stands behind the vicious tone in many of the exchanges over mammography and overdiagnosis in medical journals; it is not surprising that these debates have been dubbed the "mammography wars."

Concluding Considerations

The ultimate question at issue in the "mammography wars" is twofold: First, what is the best estimate of the actual outcomes of mammography in different age groups? Second, how do we assign values to these outcomes and weigh them against one another in our assessments of whether mammography screening is "effective," especially given the extent of our uncertainty? As a matter of public health, this has the danger of becoming a kind of cost-benefit calculation: how many cancer-related deaths need we prevent to justify screening? One per 1,900 screened? One per 1,300 screened? How ought we to weigh the harms to those screened, and how ought we to weight the relative value of "mere" psychological harms (such as those associated with

acprof-9780190467715.indd 163 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

anxiety over false positives) versus more serious harms (such as overdiagnosis and overtreatment)?

At issue in much of the controversy over mammography is not only how to interpret the evidence, but arguably, a basic disagreement over matters of justice. With any screening regimen in a healthy population (and indeed, with any preventive intervention), Rose's paradox arises (see, e.g., John 2011). Rose's paradox is that it is an inevitable feature of any public health measure that most people screened are not in fact likely to benefit from participating. Thus, screening a healthy population is like a "contract": we ask those involved to participate, in order to reduce overall risk, though few will actually benefit. (Of course, providers of the preventive care will benefit.) So, the question becomes: How much of a cost ought we ask the public to bear for a very small chance of benefiting? If the cost is minor inconvenience or side effects with very little chance of long-term consequences, many people may be willing to bear this burden. Yet, if the cost is overdiagnosis and overtreatment, how many should be asked to bear this burden? This issue is not a novel one, nor is it unique to cancer screening; indeed, it pervades modern "risky" medicine—the treatment of early stage disease or "pre-" disease with various drugs and aggressive interventions. Such aggressive extension of preventive care benefits pharmaceutical companies, but does not in fact benefit most patients. Whether such aggressive preventive care is optimal depends upon whether you think that the aim of medicine is treatment of disease, or risk reduction at the population level (for a discussion, see, e.g., Aronowitz 2015).

Competing views about just this issue are arguably informing very different perspectives on how to assess the evidence regarding the benefits and costs of cancer screening. These perspectives shape both evaluation of evidence regarding benefits of mammography, assessment of whether and which costs are serious enough to be tolerable, and communication of results. Reports and reviews of research on mammography may favor weighing the evidence in one direction versus another, whether because of exclusion or inclusion of evidence, different assessments of quality of evidence, or even simple matters of organization in presentation of data. One might open with estimates of overall mortality reduction rather than age-specific, thus obscuring the important differences among age groups; or, one might bury estimates of costs in the body of the paper, or foreground risks of screening, by including a detailed description in the abstract. These choices are not merely stylistic; they represent both the authors' values, which shape both their estimates of the seriousness of benefit and risk, and the quality of the evidence. In this way, inductive risk is very much at play in this case.

acprof-9780190467715.indd 164 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

Also at stake are norms of medicine and competing views about whether and when "paternalist" medicine is justified. In principle, respect for patient autonomy is important, so women should be informed of risks and benefits of mammography screening; and, both risks and benefits should be communicated as clearly as possible. But, some argue on pragmatic grounds that most members of the public are either unwilling or unable to rationally assess their options, let alone be compliant with recommendations, so "nudging" them toward one or another option is permissible. Yet, if some percentage of those "nudged" are in turn overdiagnosed and overtreated, on the pretense that they will benefit, then arguably, a genuine injustice is done. At issue here then are fundamentally philosophical disagreements about justice, harm, autonomy, and beneficence, and the role of the physician with respect to both individual patients and the patient population more generally.

Given the extent of disagreement, and the rapidity with which medical evidence for and against various screening methods arrive on the scene, rather than adopt a uniform program of screening by age, it may be advisable to adopt a much more pluralistic approach, one which attends to the fact that patients are variable, that evidence is defeasible, and that novel technology, methods, and sources of evidence are likely to challenge our standards for what works. To be sure, adapting screening recommendations regularly, as new information arises about biomarkers for aggressive cancers, or organismic and developmental factors that indicate a risk of progression, will become necessary. Indeed, perhaps for most interventions (according to a recent paper by Quanstrum and Hayward 2010), rather than seeking a single, universal threshold for intervention, we should be arguing over a minimum of two distinct thresholds: one above which benefit clearly outweighs the risk of harm, in which case clinicians should recommend a treatment; and one below which concern about harm clearly dominates, in which case clinicians should recommend against that treatment. This approach is similar to Mitchell's (2009) adaptive management approach to risk, where appreciation of uncertainty, and willingness to update in light of new evidence, might be a better model for medical decision-making. Putting such policy into practice requires, however, that clinicians acknowledge the extent of uncertainty in estimates of the effectiveness of cancer screening. This is a new paradigm for medicine and medical communication, however, one that is very difficult to adapt to by patients (and clinicians) who want medical decision-making to be black and white. Whether and how we are best prepared for this new paradigm is an open question.

acprof-9780190467715.indd 165 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

In sum, there are several points at which inductive risk enters into cancer epidemiology, not simply at the end of the investigation when deciding upon policies of mammography screening, but at several stages in research: in choices about which endpoints to investigate and report, in design and assessment of the quality of research, and in disputes over the best means to measure one particular harm, overdiagnosis. Moreover, the matter of how to best communicate about risk itself depends on background assumptions about human psychology, as well as competing normative intuitions.

References

ACR (American College of Radiology). 2015. "ACR and SBI Continue to Recommend Regular Mammography Starting at Age 40." http://www.acr.org/About-Us/Media-Center/Press-Releases/2015-Press-Releases/20151020-ACR-SBI-Recommend-Mammography-at-Age-40.

Aronowitz, Robert. 2015. *Risky Medicine: Our Quest to Cure Fear and Uncertainty*. Chicago: University of Chicago Press.

Cartwright, Nancy. 2007. "Are RCTs the Gold Standard?" *Biosocieties* 2(1): 11–20.

Churchman, C. West. 1948. "Statistics, Pragmatics, Induction." *Philosophy of Science* 15(3): 249–68.

Cohen, L. Jonathan. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon Press.

Croswell, Jennifer M., David F. Ransohoff, and Barnett S. Kramer. 2010. "Principles of Cancer Screening: Lessons from History and Study Design Issues." *Seminars in Oncology* 37(3): 202–15.

Detterbeck, Frank C. 2012. "Cancer, Concepts, Cohorts and Complexity: Avoiding Oversimplification of Overdiagnosis." *Thorax* 67: 842–5.

Dorato, Mauro. 2004. "Epistemic and Nonepistemic Values in Science." In *Science, Values, and Objectivity*, edited by Peter Machamer and Gereon Wolters, 52–77 . Pittsburgh: University of Pittsburgh Press.

Douglas, Heather E. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67(4): 559–79.

Douglas, Heather E. 2009. *Science, Policy and the Value-Free Ideal*. Pittsburgh, PA: University of Pittsburgh Press.

Elliott, Kevin C. 2011. *Is a Little Pollution Good for You?: Incorporating Societal Values in Environmental Research*. New York: Oxford University Press.

Esserman, Laura J., Ian M. Thompson, Brian Reid, Peter Nelson, David F. Ransohoff, H. Gilbert Welch, Shelley Hwang, et al. 2014. "Addressing Overdiagnosis and Overtreatment in Cancer: A Prescription for Change." *Lancet Oncology* 15(6): e234–e242.

Gigerenzer, Gerd, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M. Schwartz, and Steven Woloshin. 2008. "Helping Doctors and Patients Make Sense of Health Statistics." *Psychological Science in the Public Interest* 8(2): 53–96.

acprof-9780190467715.indd 166 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

Gøtzsche, Peter C. 2012. *Mammography Screening: Truth, Lies, and Controversy*. Boca Raton, FL: CRC Press.

Gøtzsche, Peter C., and K. Jørgensen. 2013. "Screening for Breast Cancer with Mammography." *Cochrane Database of Systematic Reviews* 6: CD001877. doi: 10.1002/14651858.CD001877.pub5.

Gøtzsche, Peter C., and Ole Olsen. 2000. "Is Screening for Breast Cancer with Mammography Justifiable?" *The Lancet* 355:129–34.

Harach, Hector Ruben, Kaarle O. Franssila, and Veli-Matti Wasenius. 1985. "Occult Papillary Carcinoma of the Thyroid: A 'Normal' Finding in Finland." *Cancer* 56(3): 531–8.

Howick, Jeremy H. 2011. *The Philosophy of Evidence-Based Medicine*. Oxford: John Wiley & Sons.

Ioannidis, John P. A. 2008. "Effectiveness of Antidepressants: An Evidence Myth Constructed from a Thousand Randomized Trials?." *Philosophy, Ethics, and Humanities in Medicine* 3(1): 14.

Jeffrey, Richard C. 1956. "Valuation and Acceptance of Scientific Hypotheses." *Philosophy of Science* 23(3): 237–46.

John, Stephen. 2011. "Why the Prevention Paradox Is a Paradox, and Why We Should Solve It: A Philosophical View." *Preventive Medicine* 53(4): 250–2.

Kopans, Daniel B., Robert A. Smith, and Stephen W. Duffy. 2011. "Mammographic Screening and 'Overdiagnosis.'" *Radiology* 260(3): 616–20.

Lacey, Hugh. 1999. *Is Science Value Free? Values and Scientific Understanding*. London: Routledge.

Lacey, Hugh. 2004. "Is There a Significant Distinction Between Cognative and Social Values?" In *Science, Values, and Objectivity*, edited by Peter Machamer and Gereon Wolters, 24–51. Pittsburgh: University of Pittsburgh Press.

Levi, Isaac. 1960. "Must Scientists Make Value Judgements?" *Journal of Philosophy* 57:345–57.

Levi, Isaac. 1962. "On the Seriousness of Mistakes." *Philosophy of Science* 29:47–65.

Levi, Isaac. 1967. *Gambling with Truth*. London: Routledge & Kegan Paul.

Liong, Men Long, Chun Ren Lim, Hengxuan Yang, Samuel Chao, Chin Wei Bong, Wing Seng Leong, Prashanta Kumar Das, et al. 2012. "Blood-Based Biomarkers of Aggressive Prostate Cancer." PLoS ONE 7(9): e45802. doi:10.1371/journal.pone.0045802.

McMullin, Ernan. 1982. "Values in Science." In *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, vol. 1, edited by P. Asquith and D. Nickles, 3–28. East Lansing, MI: Philosophy of Science Association.

Mitchell, Sandra D. 2004. "The Prescribed and Proscribed Values in Science Policy." In *Science, Values, and Objectivity*, edited by Peter Machamer and Gereon Wolters, 245–55. Pittsburgh: University of Pittsburgh Press.

Mitchell, Sandra D. 2009. *Unsimple Truths: Science, Complexity, and Policy*. Chicago: University of Chicago Press.

Oeffinger, Kevin C., Elizabeth T. H. Fontham, Ruth Etzioni, Abbe Herzig, James S. Michaelson, Ya-Chen Tina Shih, Louise C. Walter, et al. 2015. "Breast Cancer

acprof-9780190467715.indd 167 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

Screening for Women at Average Risk: 2015 Guideline Update from the American Cancer Society." *JAMA* 314(15): 1599–1614. doi:10.1001/jama.2015.12783.

Prasad, Vinay, Jeanne Lenzer, and David H. Newman. 2016. "Why Cancer Screening Has Never Been Shown to 'Save Lives'—And What We Can Do about It." *BMJ* 352: h6080.

Puliti, Donella, Stephen W. Duffy, Guido Miccinesi, Harry de Koning, Elsebeth Lynge, Marco Zappa, and Eugenio Paci. 2012. "Overdiagnosis in Mammographic Screening for Breast Cancer in Europe: A Literature Review." *Journal of Medical Screening* 19(suppl. 1): 42–56.

Quanstrum, Kerianne H., and Rodney A. Hayward. 2010. "Lessons from the Mammography Wars." *New England Journal of Medicine* 363(11): 1076–9.

Rudner, Richard. 1953. "The Scientist qua Scientist Makes Value Judgments." *Philosophy of Science* 20(1): 1–6.

Sakr, W. A., D. J. Grignon, G. P. Haas, L. K. Heilbrun, J. E. Pontes, and J. D. Crissman. 1996. "Age and Racial Distribution of Prostatic Intraepithelial Neoplasia." *European Urology* 30(2): 138–44.

Santen, Richard J., Wei Yue, and Daniel F. Heitjan. 2013. "Occult Breast Tumor Reservoir: Biological Properties and Clinical Significance." *Hormones and Cancer* 4(4): 195–207.

Saslow, Debbie, Carla Boetes, Wylie Burke, Steven Harms, Martin O. Leach, Constance D. Lehman, Elizabeth Morris, et al. 2007. "American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography." *CA: A Cancer Journal for Clinicians* 57(2): 75–89.

Shen, Megan Johnson, Christian J. Nelson, Ellen Peters, Susan F. Slovin, Simon J. Hall, Matt Hall, Phapichaya Chaoprang Herrera, et al. 2015. "Decision-Making Processes among Prostate Cancer Survivors with Rising PSA Levels: Results from a Qualitative Analysis." *Medical Decision Making* 35(4): 477–86.

Solomon, Miriam. 2015. *Making Medical Knowledge*. Oxford: Oxford University Press.

Stamatiou, Konstantinos, A. Alevizos, E. Agapitos, and F. Sofras. 2006. "Incidence of Impalpable Carcinoma of the Prostate and of Non-Malignant and Precarcinomatous Lesions in Greek Male Population: An Autopsy Study." *The Prostate* 66(12): 1319–28.

Steel, Daniel. 2015a. "Acceptance, Values, and Probability." *Studies in History and Philosophy of Science Part A* 53: 81–8.

Steel, Daniel. 2015b. *Philosophy and the Precautionary Principle: Science, Evidence, and Environmental Policy*. Cambridge: Cambridge University Press.

Steele, Katie. 2012. "The Scientist qua Policy Advisor Makes Value Judgments." *Philosophy of Science* 79(5): 893–904.

Stegenga, Jacob. 2011. "Is Meta-Analysis the Platinum Standard of Evidence?." *Studies in History and Philosophy of Science Part C* 42(4): 497–507.

Uman, Lindsay S. 2011. "Systematic Reviews and Meta-Analyses." *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 20(1): 57–9.

acprof-9780190467715.indd 168 2/20/2017 6:19:34 PM

OUP UNCORRECTED PROOF – FIRSTPROOFS, Mon Feb 20 2017, NEWGEN

US Preventive Services Task Force. 2009. "Screening for Breast Cancer: US Preventive Services Task Force Recommendation Statement." *Annals of Internal Medicine* 151(10): 716.

Welch, H. Gilbert, and William C. Black. 1997. "Using Autopsy Series to Estimate the Disease "Reservoir" for Ductal Carcinoma in situ of the Breast: How Much More Breast Cancer Can We Find?" *Annals of Internal Medicine* 127(11): 1023–8.

Welch, H. Gilbert, and William C. Black. 2010. "Overdiagnosis in Cancer." *Journal of the National Cancer Institute* 102(9): 605–13.

Worrall, John. 2002. "*What* Evidence in Evidence-Based Medicine?" *Philosophy of Science* 69(S3): S316–S330.

acprof-9780190467715.indd 169 2/20/2017 6:19:34 PM