The print and electronic media have an enormous influence on how the public views health issues. Both health policymakers and scientists recognize journalists’ effect on public understanding. Reporting health stories requires judgment about how to interpret evidence and about the implications of evidence for the public. But most journalists have little formal training in assessing the validity of evidence that bears on health issues, so inaccurate or deceptive reporting seems common. To begin to address this problem, we have built on others’ work and developed a set of guidelines to help journalists understand and interpret health stories.

Many obstacles confront the health journalist, including limitations of time and space, editorial priorities, and the need to create stories that are compelling enough to warrant space in a publication. Our journalist’s guidelines will not help with those issues. However, even given other constraints, understanding principles of scientific inquiry into human health problems will help journalists to produce more informed articles.

**SCENARIOS**

Picture yourself as a journalist with a focus on health and medicine assigned six new topics for possible expansion into full articles. In each case, scientists, physicians, or community groups are making claims that might catch public attention. For each of the six stories, the goal is to sort out the authenticity of the claims. These scenarios are all based on actual media reports, and all address issues of public interest.

- Researchers announce a causal link between a bacterium found in the stomach, Helicobacter pylori, and stomach ulcers. They claim that the bacterium is probably the cause of such ulcers, which have previously been attributed to excess acid.
- A local community radiologist associated with a magnetic resonance imaging (MRI) facility claims that MRI provides breast images far superior to those of conventional mammograms. He argues that all women over the age of 50 should now receive MRI breast screening.
- A researcher presents data suggesting that a low-fat diet decreases breast density. High density is associated with breast cancer. He suggests that women who want to lower their risk of breast cancer should lower the amount of fat in their diet.
- Researchers announce that postmenopausal hormone-replacement therapy prolongs life. They note that the effect is greater than anyone predicted and argue that physicians should be offering hormone-replacement therapy to all postmenopausal women.
- A new study that pooled results on more than 61,000 women in 11 countries suggests that abortion causes breast cancer. Antiabortion groups cite yet another reason to ban abortion.
- The World Wildlife Fund Canada releases a list of brand-name detergents that contain nonylphenol ethoxylates (NPEs), which they say disrupt the action of hormones and cause cancer and birth defects. The group suggests that the public should be made aware of the information.

In the following article, we present a set of guidelines to help journalists and editors assess whether a biomedical study or claim like those above is valid (that is, close to the truth) and how the data bear on people’s concerns about their health. We present these guidelines, which are based on discussions by a team of health journalists and physicians, as a series of questions journalists...
should use when researching a health article (Table 1). A second series of questions will be particularly useful for querying investigators making claims about treatment results (Table 2).

**What Are the Principal Questions or Issues?**

Implicit questions underlie health stories, and the real question a research article addresses may not always be clear. Scientists have some predefined question, presumably addressing a gap in knowledge, when they embark on a study. When they interpret their results, however, they might extrapolate well beyond what is appropriate or justifiable. Health researchers have found it useful to think of three elements in defining their research questions:

- **Who** might be affected?
- **What** is the relevant intervention or exposure, and **what** are the alternatives?
- **What** is the outcome of the intervention or exposure? By outcome, we mean the key effect of potentially beneficial or harmful exposure. Examples of beneficial outcomes include making people feel better, avoiding strokes or heart attacks, or making people live longer; examples of adverse outcomes include disability, cancer, or cognitive impairment.

The underlying question might not always be obvious. Consider the following:

A local community radiologist associated with a magnetic resonance imaging (MRI) facility claims that MRI provides breast images far superior to those of conventional mammograms. He argues that all women over the age of 50 (the age over which randomized trials suggest screening reduces breast cancer) should now receive MRI breast screening.

Here, two questions may arise. Both involve women at risk of breast cancer. The first question has to do with the accuracy of MRI versus standard mammography in the diagnosis of breast cancer. The second question also concerns MRI and women with breast cancer but addresses a more fundamental issue. The outcome associated with this second question is different: death from breast cancer. Identifying the second question allows the journalist to ask the researcher, “What data do you have to show us that women screened with MRI will live longer than women screened with conventional approaches?” When, as in this case, no data are forthcoming, this information will guide journalists in their approach to the story.

Consider the following:

A researcher presents data suggesting that a low-fat diet decreases breast density. High density is associated with breast cancer. The researcher suggests that women who want to lower their risk of breast cancer should lower the amount of fat in their diet.

This study concerns the effect of a low-fat diet versus a normal diet on women at risk of breast cancer. The outcome of the study was fibrotic disease of the breast, but the outcome that the public is much more concerned about is breast cancer. The investigator is ready to extend the findings to reduction in breast-cancer risk (on the basis of the association between fibrosis and cancer), but that is not actually what was studied.

Those examples make it evident that precise definition of the underlying question a research study is supposed to answer can take a journalist a long way toward understanding a story and deciding on the validity, or authenticity, of the research. Having gained that understanding, the journalist begins to ask specific questions about the methods of a study and its implications for human health.

**Is the Question Asked Really of Interest?**

Once journalists have sorted out the question(s) that the research addresses, they can ask whether there is a gap between the reported data and the inferences that researchers draw. What is the warranted message about the original research that journalists should convey in their health stories? The first issue in judging authenticity is whether the investigators’ question matches the one in which journalists, and their audiences, are interested.

**Who was studied?**

The population studied might not be the population to which we would like to apply the results. The public is interested in health effects in human beings, but the subjects of an experiment might have been laboratory animals or even cells in a culture dish. The concerns of the World Wildlife Fund Canada in the final scenario, it turns out, are based on the effects of NPEs on animals, although human studies are apparently under way. Findings in animals might suggest health effects in human beings, but they seldom prove them. They can also be well worth reporting to the public, but inferences about human effects should be made very cautiously.

In addition, studies in humans might apply only to restricted populations. The results of medical interventions tested in men may not apply to women. Programs aimed at reducing teenage pregnancy have
failed to show a benefit, but many have been conducted largely in poor, inner-city environments; perhaps the interventions would have a greater impact in more affluent environments, in which different cultural forces influence adolescents’ behavior. In studies of serum prostate-specific antigen (PSA) for the detection of early prostatic cancer, different levels of test accuracy in men of different ages have been observed. Journalists should consider exactly who was included in a study and whether and how generalizations to other populations should be limited.

**What was studied?**

Journalists who consider exactly what intervention a study addressed might quickly identify serious limitations that affect whether or how they’ll write their news article.

- If the study’s claim concerns a potentially hazardous exposure, what was the magnitude of the exposure? Investigators studying drugs that cause cancer in animals might administer the medication in doses thousands of times stronger than that which clinicians would offer to humans.

- Is a new treatment feasible or practical? In our second scenario (associating fat ingestion with breast disease), a journalist might question whether the diet offered the women is sufficiently palatable to be realistic for North American women. Two studies of the use of thrombolytic agents to dissolve clots in people with acute stroke involved patients who came to the hospital within 4 hours of the onset of symptoms and had a computed tomographic (CT) scan within that time. Even if the drug proved effective, if most patients are seen after 4 hours, the drug will be useful for only a small proportion of the stroke population.

- Is the treatment being studied one that is currently used in clinical practice? Most of the studies on which the claim of the life-prolonging effects of hormone-replacement therapy are based examined the effects of estrogen, whereas current hormone-replacement regimens include both estrogen and progesterone. It turned out that the makers of captopril had sponsored the trial and chosen two comparison drugs with a high incidence of side effects. Not surprisingly, patients

**Table 1: Exploring a Health Story: Questions to Ask**

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Is the question asked really of interest?</strong></td>
<td>Who was studied?</td>
</tr>
<tr>
<td></td>
<td>What was studied?</td>
</tr>
<tr>
<td></td>
<td>What outcome was studied?</td>
</tr>
<tr>
<td><strong>How strong are the methods?</strong></td>
<td>Are patient groups the same at study’s start and end?</td>
</tr>
<tr>
<td></td>
<td>Do groups stay equivalent after randomization?</td>
</tr>
<tr>
<td></td>
<td>Was the outcome measurement the same?</td>
</tr>
<tr>
<td><strong>How large was the effect?</strong></td>
<td></td>
</tr>
<tr>
<td><strong>What do other studies report?</strong></td>
<td>How strong are the methods? The case of meta-analysis</td>
</tr>
<tr>
<td></td>
<td>Is the question sensible?</td>
</tr>
<tr>
<td></td>
<td>Was the selection of studies unbiased?</td>
</tr>
<tr>
<td></td>
<td>How good were the primary studies?</td>
</tr>
<tr>
<td><strong>Conflict of interest</strong></td>
<td>Are scientists emotionally invested in their results?</td>
</tr>
<tr>
<td></td>
<td>What role does funding play in how the scientists are reporting their results?</td>
</tr>
<tr>
<td><strong>Verification and corroboration</strong></td>
<td>How reliable is the source used to verify or corroborate this information?</td>
</tr>
</tbody>
</table>

34
taking captopril had a better quality of life than patients taking the alternative medications. Journalists can question whether a new treatment is being compared with the best available alternative.

Even if the intervention is not really the one of primary interest to the journalists’ audience, it need not mean that the study is not worth reporting. But it does mean that practical inferences are weaker, and journalists should point out the limitations.

**What outcome was studied?**

Researchers often play a “substitution game.” Sometimes it is too difficult to study the outcome of real interest, commonly because it takes years to develop or is rare, so scientists use substitutes that are linked to the relevant outcome. For example, instead of examining the effect of an intervention on stroke, death, or breast cancer, researchers might substitute blood pressure, serum cholesterol concentration, or fibrotic breast disease.

The substitution game can be misleading. Cholesterol-lowering drugs have shown surprises when their effects on mortality were actually studied: some cause increases, rather than the expected decreases, in mortality. For years, heart specialists assumed that a drug that suppressed nonlethal disturbances in the rhythm of the heartbeat would also be good for suppressing fatal arrhythmias. Since the finding that a number of drugs that are very effective in reducing nonlethal rhythm disturbances increase death rates, physicians no longer accept this assumption.

If a low-fat diet reduces fibrotic disease, which is linked to breast cancer, does it necessarily follow that the low-fat diet would reduce breast cancer? Perhaps, but the inference is not nearly as strong as if the investigator had examined the effect on breast cancer itself. MRI might produce nicer pictures than conventional mammography and even help to detect smaller lumps, but it does not necessarily follow that its use in a screening program will prolong the lives of women with clinically silent breast cancer. That is an issue that bears on any diagnostic test: a new blood test or imaging procedure might be better at detecting disease, but its application might not result in longer, or better, lives.

Not only should the investigators focus on a key outcome that matters to patients, they should also avoid leaving out other important outcomes. A new drug might be effective, but what about its side effects? Hormone-replacement therapy may prolong life, but how will postmenopausal women feel about intermittent vaginal bleeding? One way to remember whether all relevant outcomes have been included is to think of the acronym SECS: have the investigators considered whether a new treatment compared with current options is Safer, more Efficacious (that is, works better), less Costly, or more Suitable (easier to administer)?

Journalists might also consider whether a study has observed patients for a sufficient length of time. Is follow-up long enough that both beneficial and detrimental effects of the treatment have had time to appear? Adequate follow-up varies widely among medical conditions. In studies of critically ill patients, effects on mortality will be apparent in weeks. In studies preventing distant outcomes like stroke or myocardial infarction, years of follow-up may be required. Adequate follow-up is particularly important in the case of harmful effects that may appear only after long exposure.

**How Strong Are the Methods?**

Once journalists have decided whether the population, the intervention, and the outcome are the ones in which they and their readers are primarily interested and that key outcomes have not been neglected, they can look at the strength of the methods that the investigators have used. The strength of the study methods constitutes the second issue in authenticity.

Both the public and researchers are interested in making causal inferences about health issues. Each of our scenarios suggests a causal question. Does *Helicobacter pylori* cause stomach ulcers? Will MRI scanning or a low-fat diet cause a decrease, or abortion an increase, in breast cancer deaths? Will hormone-replacement therapy lead to a longer life? Will NPEs cause birth defects or cancer?

Human beings intuitively make causal inferences. Many inferences made without use of scientific principles will be spurious. For instance, patients with rheumatoid arthritis and their clinicians are often strongly convinced that patients’ pain varies with the weather. However, a systematic, detailed study has demonstrated no such association.

From the use of bleeding up to the early 19th century, to modern surgical and medical treatments, medical history is littered with enthusiastically endorsed treatments that proved useless or even harmful. Similarly, putatively harmful exposures have ultimately proved benign. Distinguishing between strong and weak causal claims requires an understanding of threats to validity and of how medical researchers can design studies to reduce these threats. We will outline the primary threats and the solutions.
Causal inferences are based on noting differences in outcome between those who are exposed to a putatively causal agent (treatment group) and those who are not (control group). For example, there is a strong association between being in the hospital and dying: the control group, people not in the hospital, die less often. A naive observer of this association would infer a causal connection and suggest that to avoid dying one should stay away from hospitals. Obviously, though, people who come to the hospital are very different from those who do not: they are much sicker. Hospital patients' increased likelihood of dying is not a function of what happens in the hospital, but a function of the increased risk of dying that they have when they enter the hospital.

Many differences in risk between groups that are and are not exposed to a treatment or potentially harmful agent are not nearly so obvious. If patients who receive a treatment are at lower risk, the treatment will look good even if it is useless. How can health researchers ensure that treatment and control patients are similar at the start of a study and so attribute differences at the study’s end to the treatment?

If patients are assigned to treatment or control groups by a chance or random process—analogous to a coin flip—the two groups will be very similar with respect to their risk. Randomization is the term used to describe the process of using chance to determine whether a patient is assigned to the treatment or control group. Randomization is probably the most powerful safeguard researchers have against bias (systematic errors relative to the underlying truth) in medical studies. Randomized controlled trial (RCT) is the term for a study that uses chance to decide whether patients receive a treatment.

Consider the report of the effect of a low-fat diet versus a conventional diet. The investigator randomized women to either the experimental or the control diet, and the fact that this is an RCT enormously strengthens our inference that the diet really did cause the lower rate of fibrosis in the participants' breasts. As we have noted previously, however, we remain less sure about the effect on the frequency of occurrence, or incidence, of breast cancer.

RCTs require that investigators have the power or authority to use chance to determine who gets treatment and who does not. Sometimes, instead of randomly allocating patients to treatment or control, health researchers simply observe patients who did or did not receive a treatment or who were or were not exposed to a putatively harmful agent. This research design is called an observational study.

The disadvantage of an observational study versus an RCT—and it is a big disadvantage—is that no assurances are given that treatment and control groups were the same at the start of the study. Consider the following:

Researchers announce that postmenopausal hormone-replacement therapy prolongs life. They note that the effect is greater than anyone predicted and argue that physicians should be offering hormone-replacement therapy to all postmenopausal women.

All the studies that contributed to that conclusion were observational. It is not difficult to imagine that the risk of dying will be lower for women who are offered, or agree to take, hormone-replacement therapy than for those not offered, or who decline, hormone-replacement therapy. For instance, there could be a social class gradient—affluent women might be more inclined to take hormones. We know that there is a strong association between social class and mortality: poorer people die earlier. Therefore, the real reason for an apparent effect of hormone-replacement therapy might be a difference in social class between those who do and those who do not take hormones.

There are other possible differences. Physicians might be less inclined to give hormones to women who are already sick, who smoke, or who have a family history of heart disease. Any of those factors could create an apparently, but spuriously, lower risk of dying for those receiving hormones. The strength of the inference that hormone-replacement therapy prolongs life is therefore much weaker than would be the case if it were based on RCTs. Fortunately, ongoing large RCTs should give us a more credible answer about the effects of hormone replacement in postmenopausal women.

For many issues, particularly those associated with potentially harmful exposures, randomization is not an option. We could not, for instance, randomize people to smoke or not smoke; we have to be content with observational studies. But the strength of inference in such situations will always be weaker. This applies, for instance, to the association between abortion and breast cancer presented in one of the scenarios.24 Is there any way for researchers to reduce the bias in observational studies? Bias is reduced if investigators are aware of a number of risk factors and ensure they are balanced in the two groups. For instance, consider

---

2. Since we first prepared this article, the first large randomized trial of hormone-replacement therapy found no decrease in cardiovascular risk.
investigators using an observational design to examine the effect of exercise on heart attacks. The investigators could make sure that people who do and do not exercise have similar other risk factors for heart attacks. They would therefore recruit exercising and nonexercising people to make up groups who have approximately the same numbers of men and women with similar education, income, age, smoking history, cholesterol concentrations, and family history of coronary arterial disease. However, even if all the known factors that influence the likelihood of a heart attack are balanced, the many other determinants that are not accounted for may influence patients’ prognoses, that is, their likelihood of suffering an adverse outcome. As a result, RCTs will always provide much stronger evidence than observational studies.

And although a trial is randomized at the start, it does not mean that patients will stay similar at the end. Randomization can be destroyed if, at the end of the study, patients are not available to have their outcome measured or recorded—if they are “lost to follow-up.” It is likely that patients who are lost to follow-up differ, in important ways, from those who are successfully observed. In the extreme, the reason for loss to follow-up could be that a patient dies or suffers the outcome of interest. Patients can be lost to follow-up because of side effects of a medication under study. The lower the rate of loss to follow-up, the more likely that the balance in the groups achieved at the beginning of the study remains at the end. In the study of a low-fat diet to prevent breast fibrosis, for instance, mammographic follow-up was achieved in only 69% of the women enrolled.

**DO GROUPS STAY EQUIVALENT AFTER RANDOMIZATION?**

Imagine a randomized trial in which the effect of acupuncture was studied. The control patients receive no intervention. The acupuncture group receives treatment by a pleasant, charming, engaging practitioner who is completely convinced of the effectiveness of his treatment and takes time to explain the rationale, and his previous success, to each participant. If treated patients feel better at the end of the study, we may be skeptical that the benefit was a result of the biological effect of treatment, rather than the psychological impact of patients’ knowing they were receiving what an apparently credible person thought was a very effective intervention. The tendency to feel better because you believe that you are receiving effective treatment is called the *placebo effect.*

Imagine another RCT in which patients who receive an experimental drug return to the clinic more often than those who do not. They are checked carefully and provided with extra advice about how to manage their condition. Their nonstudy medications are adjusted; if there are any new problems, the clinicians deal with them. At the end of the study, when treated patients do better, we are not sure whether the benefit was due to the experimental treatment or to the superior care that patients received.

Those examples show that randomization is not sufficient to ensure an unbiased comparison of treatments. Randomization ensures only that groups are balanced at the start; it gives no assurance that factors that can influence outcome stay balanced as a study progresses.

What can researchers do to ensure the balance of other treatments or factors that influence outcome as their study proceeds? The most powerful technique is to ensure that neither patients nor clinicians know whether they are receiving the active treatment. This masking of treatment is called *double-blinding.* For drugs, it is easy to double-blind. A pill is produced to look, feel, and taste identical to the active treatment; it may contain an alternative medication or have no active ingredient.

Masking acupuncture is more difficult, but still possible. If, in a randomized trial, an equally engaging person uses needles, but puts them in the wrong places in the control group (sham acupuncture), and the treated group still does better, the authenticity of the study in showing the biological effect of acupuncture would be high.

Journalists can ask investigators about the knowledge held by four groups: Did patients know whether they were taking the experimental treatment or the alternative? Did health workers looking after the patients? Did those responsible for measuring or ascertaining patients’ outcomes? Were those who analyzed the data also blinded to which group was receiving the experimental treatment? Although there are four groups blinded here, the term *double-blind* is still used.

What can researchers do if blinding is impossible, as it was, for instance, in the study of different diets to prevent breast cancer? They can reassure us by demonstrating that patients’ care, aside from the experimental treatment, was similar in the intervention and control groups.

**WAS THE OUTCOME MEASUREMENT THE SAME?**

Picture the enthusiastic acupuncturist asking patients, at the end of the study, how much better they were feeling. What if the same person were delegated to determine whether the control patients felt better?

Bias can be introduced not only if patients have different prognoses at the start, or if they are treated differently.
along the way, but also in measurement of the final outcome. Researchers’ best strategy for avoiding bias in measuring outcome is, once again, blinding. If those who measure outcome are unaware of whether patients were in the treatment or control group, we can be confident that the process of determining whether an outcome event occurred was similar for all patients. If that is not the case, one must consider the extent to which bias in measurement of outcome can skew the study results.

**Table 2: Questions To Ask Investigators Making Claims About a Treatment**

- **Question about the study question**
  - Is the alternative treatment the best available?
- **Questions about the study methods**
  - Were patients randomized to experimental and control groups?
  - Were patients in the two (or more) groups similar at the start of the study?
  - How many patients in the groups were “lost to follow-up” along the way?
  - Were patients and caregivers blinded to which group patients were in?
  - Were those measuring outcome blinded to which group patients were in?
  - Were those who conducted the statistical analysis blinded to which group patients were in?
- **Questions about the results**
  - How large is the effect of the exposure or treatment in reducing relative and absolute risk?
  - What were the side effects?
  - How much does the treatment cost?
- **Questions about other studies**
  - Have other studies addressed the question?
  - What were the results in those studies?
- **Questions about conflict of interest**
  - How was the study funded?
  - Has any party interested in this study’s outcome helped pay for your attendance at a scientific meeting?
  - Have you been a paid consultant for any relevant party?
  - Do you have a personal financial stake in the product you have studied?

How large was the effect?

Associations between exposures and outcomes are more likely to be causal when they are strong. For instance, one reason we believe that smoking causes lung cancer is that heavy smoking is associated with an increase of a factor of 10 in the likelihood of cancer.\(^{25}\) Abortion is associated with a 30% increase in the likelihood of breast cancer.\(^{23}\) Clearly, the latter is much less compelling than the association between smoking and lung cancer. It is also much weaker than the association between the presence of *Helicobacter pylori* and gastric ulcer.\(^{26,27}\) Many discounted the initial reports of a possible causal relationship between these bacteria and ulcers because it seemed biologically implausible, but the fact that the risk was increased by a factor of 20 could have alerted people to the authenticity of the causal claim. RCTs that have since shown that antibiotics directed at *Helicobacter pylori* can cure peptic ulcers have confirmed that the bacteria are a very important cause of the illness.\(^{28}\)

The issue of the magnitude of effects also applies to new treatments. The pharmaceutical industry is adding to our understanding of treatment efficacy by funding huge trials involving thousands of patients. Such trials are able to show very small treatment effects. That a treatment “works,” however, does not mean that it should be widely administered. If the effect is small enough and the economic costs or side effects great enough, the new treatment might best be shelved.

For example, one “clot-busting” drug, tissue plasminogen activator (TPA), cut the number of deaths after myocardial infarction (MI) by 15% more than another agent, streptokinase. The investigators reported that the results were “statistically significant”; that is, chance is an unlikely explanation of the differences in death rate between the TPA- and streptokinase-treated patients. The statistical significance, however, tells us nothing about whether the differences are important. Only about 7% of the streptokinase-treated patients died. The 15% reduction in relative risk (the risk of the adverse outcome for those who received treatment divided by the risk for the control group) therefore translates into only a 1% reduction in absolute risk (the risk of the adverse outcome for the control group minus the risk for the
One must therefore ask whether, across the range of patients included and the range of interventions, more or less effective drugs exist. If all studies consistently show no effect, then we can be confident that neither treatment had an effect. If the studies show different results, or in case of conflicting results, then we must be cautious about concluding anything. For example, some studies show that hormone replacement therapy reduces the risk of breast cancer, while others show no effect or even an increased risk. Similarly, some studies show that aspirin reduces the risk of heart attack, while others show no effect. These differences can be due to the nature of the studies, the patients included, the interventions used, or the outcomes measured. Therefore, it is important to consider whether other (possibly contradictory, possibly confirmatory) data are available. If this is the first study addressing the question, the strength of inference based on the findings will be low. Because meta-analyses have gained such a prominent role in recent years, we will specifically address issues about their validity.

A meta-analysis combines (pools) results from different studies to generate a single estimate of the effect of a treatment. All the issues raised up to now, including whether the question matches the one of interest, and the nature of the results, can be applied to a study using meta-analysis. We will now discuss specific issues relevant to meta-analyses.

**What do Other Studies Report?**

Rarely should the adoption of a new drug or surgical treatment, or the incrimination of a risk factor for a disease, be based on a single study. Often, studies give conflicting results. Prominent examples include the use of clot-busting drugs for acute stroke, the association between electromagnetic radiation and cancer, and the association between hormone replacement therapy and breast cancer. In general, it is only after several studies have demonstrated consistent results that we can be confident that we have a good idea of the underlying truth.

**How Strong are the Methods? The Case of Meta-Analysis**

Two of our scenarios—the hormone-replacement therapy and the abortion and breast cancer examples—refer to studies that summarize the entire literature. This type of study, a meta-analysis, can produce stronger results than a single study alone. When focusing on reports of single studies, journalists should consider whether other (possibly contradictory, possibly confirmatory) data are available. If this is the first study addressing the question, the strength of inference based on the findings will be low. Because meta-analyses have gained such a prominent role in recent years, we will specifically address issues about their validity.

A meta-analysis combines (pools) results from different studies to generate a single estimate of the effect of a treatment. All the issues raised up to now, including whether the question matches the one of interest, and the nature of the results, can be applied to a study using meta-analysis. We will now discuss specific issues relevant to meta-analyses.

**Is the Question Sensible?**

The rules already discussed here will help the journalist understand the structure of a meta-analysis. Again, consider who the patients or participants were in each of the studies that contributed to the meta-analysis, what was the intervention or exposure that the patients experienced, and what outcomes were measured. A meta-analysis that asks whether all types of chemotherapy have any effect on mortality for all patients with cancer would not be sensible. Different therapies are likely to have different, and perhaps even opposite, effects in patients with different sorts of cancer. It would be misleading to perform a meta-analysis that included some harmful and some beneficial treatments, find no overall effect, and conclude that neither treatment had an effect.

One must therefore ask whether, across the range of patients included and the range of interventions, more or less effective drugs exist. If all studies consistently show no effect, then we can be confident that neither treatment had an effect. If the studies show different results, or in case of conflicting results, then we must be cautious about concluding anything. For example, some studies show that hormone replacement therapy reduces the risk of breast cancer, while others show no effect or even an increased risk. Similarly, some studies show that aspirin reduces the risk of heart attack, while others show no effect. These differences can be due to the nature of the studies, the patients included, the interventions used, or the outcomes measured. Therefore, it is important to consider whether other (possibly contradictory, possibly confirmatory) data are available. If this is the first study addressing the question, the strength of inference based on the findings will be low. Because meta-analyses have gained such a prominent role in recent years, we will specifically address issues about their validity.

A meta-analysis combines (pools) results from different studies to generate a single estimate of the effect of a treatment. All the issues raised up to now, including whether the question matches the one of interest, and the nature of the results, can be applied to a study using meta-analysis. We will now discuss specific issues relevant to meta-analyses.

**Is the Question Sensible?**

The rules already discussed here will help the journalist understand the structure of a meta-analysis. Again, consider who the patients or participants were in each of the studies that contributed to the meta-analysis, what was the intervention or exposure that the patients experienced, and what outcomes were measured. A meta-analysis that asks whether all types of chemotherapy have any effect on mortality for all patients with cancer would not be sensible. Different therapies are likely to have different, and perhaps even opposite, effects in patients with different sorts of cancer. It would be misleading to perform a meta-analysis that included some harmful and some beneficial treatments, find no overall effect, and conclude that neither treatment had an effect.

One must therefore ask whether, across the range of patients included and the range of interventions, more or less effective drugs exist. If all studies consistently show no effect, then we can be confident that neither treatment had an effect. If the studies show different results, or in case of conflicting results, then we must be cautious about concluding anything. For example, some studies show that hormone replacement therapy reduces the risk of breast cancer, while others show no effect or even an increased risk. Similarly, some studies show that aspirin reduces the risk of heart attack, while others show no effect. These differences can be due to the nature of the studies, the patients included, the interventions used, or the outcomes measured. Therefore, it is important to consider whether other (possibly contradictory, possibly confirmatory) data are available. If this is the first study addressing the question, the strength of inference based on the findings will be low. Because meta-analyses have gained such a prominent role in recent years, we will specifically address issues about their validity.
less the same outcome or effect is expected. The meta-
analysis of the effect of abortion on breast cancer
included women from different countries who had
abotions at different ages and after different durations of
pregnancy. The abortion techniques differed, as did the
women's subsequent use of birth control. Whether it
makes sense to pool data across such varied populations
and exposures is a matter of scientific judgment.

**Was the selection of studies unbiased?**

If one wishes to show that a treatment works, an easy way
is to consider only studies with a trend showing efficacy;
even with a useless treatment, an average of half the
studies will show treated patients doing better than
untreated. Thus, a key question a journalist should
consider in a meta-analysis is whether the investigators
obtained a representative, unbiased sample of the
available studies.

This issue can be addressed by asking whether the
investigators had explicit, well-defined rules to decide
which studies to include and which to exclude—*inclusion
criteria and exclusion criteria*. Having established their
criteria, meta-analysts should ensure that they have
searched thoroughly for all possible studies. In particular,
they have to worry about the fact that a “positive” study
showing that a treatment is beneficial or that a putative
exposure (such as electromagnetic radiation or abortion)
is associated with adverse health effects, is more likely to
be published or receive attention than a “negative” study
failing to show benefit or harm. This selective
dissemination of results is called *publication bias*. If the
researchers conducting the meta-analysis have not made
efforts to seek unpublished studies (through sources such
as abstracts presented at scientific meetings, registries of
theses and granting agency awards, or personal contacts
with investigators), their results are suspect, particularly if
they conclude that a treatment is beneficial or that a
putative risk factor is dangerous (as did the researchers
suggesting that abortion is associated with breast cancer).

**How good were the primary studies?**

The adage “garbage in, garbage out” is very applicable to
a meta-analysis. If the original studies used weak designs,
pooling their results does not make them any stronger.
The meta-analyses of abortion and breast cancer and of
hormone-replacement therapy and life span relied on
observational studies. As noted earlier, that weakens the
strength of any inferences they might make. We cannot
decide whether a woman has an abortion by
randomization (and therefore, for that question,
observational studies are the best evidence we are going
to get), but we certainly can randomize women to receive
or not receive hormone-replacement therapy, and such
RCTs are going on now.

**Conflict of Interest**

It is impossible to conduct research without conflict of
interest. Researchers find that their careers tend to be
made by positive, exciting results rather than by negative,
disappointing results. Naturally, investigators study
treatments in which they believe passionately. Even if
they start in a relatively objective frame of mind,
investigators tend to become emotionally invested in the
particular treatments they are studying. Realizing
personal gain can overcome the admirable goal of
scientific truth.

Journalists need to be aware of these conflicts. However,
they should also be aware of types of conflict of interest
that are even more problematic. The pharmaceutical
industry plays an increasingly important role in funding
medical research and in supporting investigators. This
role has increased in the last decade. Investigators
regularly receive funding for their studies from the
industry. Even when their studies are not industry
funded, investigators often attend conferences and
meetings at company expense. Investigators can also act
as paid consultants for the industry, receive honorariums
for acting as advisors, or attend meetings at which they
listen to company presentations. Investigators regularly
attend elegant dinners at industry expense.

These relationships inevitably affect scientists’ points of
view. In one notorious example, editorialists writing for
a major journal withheld information about their
funding from a company while presenting a rosy picture
of the benefit:risk ratio of one of the company’s drugs.
The scientists responsible for recommending the use of
anti-AIDS drugs are regular recipients of industry
funding, including gifts to attend conferences, and
regularly sit shoulder to shoulder with company
representatives in deliberations about the role of drugs in
AIDS treatment. Drug companies have put public
pressure on investigators to close trials prematurely and
publish findings before the data are complete. Industry
prefers that unpalatable results not be published and at
times has made aggressive efforts to prevent publication.
In one example, the industry threatened legal action
against investigators who had found no superiority of a
brand-name over a generic thyroid-replacement drug and
backed off only after embarrassing public exposure. In
another, funding was withdrawn and legal action
threatened to avoid publication of negative results.

Investigators strive to act with integrity and to design
studies to minimize bias. When it comes to interpretation
of results, however, considerable subjectivity is inevitable. At times, industry sponsors press their views on how results should be presented or interpreted, and investigators might find it difficult to evaluate the study’s findings objectively.

An even worse source of conflict of interest occurs when investigators have a personal financial stake in the success of an innovation. More and more often, investigators own shares of the patents in their discoveries or own shares in companies that produce the technologies they are studying or advocating.

**Verification and Corroboration**

How reliable is the source used to verify or corroborate this information? The guidelines we have provided will allow journalists considerable independence in assessing the authenticity and importance of health innovations and developments. Journalists will be able to move beyond both these positions: If it’s a drug company-sponsored press conference, it must have problems; if it’s published in a peer-reviewed journal (in which articles get published only after review by other scientists), it must be authentic.

The peer-review process has important limitations. Many reviewers are not trained in assessing the methods of a study, and journals seldom use explicit methodologic criteria to decide whether a study is acceptable. In addition, most reviewers are influenced by whether the message of a study appeals to their personal interests.

Because of those limitations, reviewers tend to disagree with one another more often than they agree. Our impression is that editors are influenced not only by the scientific quality of an investigation, but also by its newsworthiness. Editors of journals, like those of newspapers, like to publish exciting findings. Even if a study’s methods are strong, investigators or sponsors might draw inferences that go far beyond the data. All in all, journalists need tools to sort out the wheat from the chaff in even the top peer-reviewed journals.

Journalists’ content knowledge might be sufficiently limited and the topic sufficiently complex for them to need additional expert opinion for full insight. When that situation arises, many reporters refer to a “hit list” or “little black book” of experts in a field. The experts can comment on the credibility of the researcher and the validity and clinical importance of the research. Credible experts can be found in university departments, government agencies, professional associations, disease-related organizations, and services on the Internet—Media Resource Service (mediaresource@sigmaxi.org) and Profnets (profnets@profnets.com, http://www.profnets.com).

Journalists’ instinctive caution becomes important in dealing with experts. Experts might be unable to comment realistically on the application of the new therapy or test to patients if they have not seen a patient since graduating from medical school. Even (and perhaps especially) when experts are highly knowledgeable, they are subject to the same conflicts of interest as the investigators. The research may be on a pet subject of the external expert, or the expert might have trained the investigator; thus, the expert could be well disposed to the innovation. Or the expert could be a competitor in the subject of investigation and tend to discount others’ work.

**Conclusion**

Skilled health journalists will be familiar with the principles that we have outlined. Journalists can transform these principles into a checklist of questions to ask scientists making health claims about a new therapy (Tables 1 and 2). Obtaining answers to these questions will allow journalists to present a more reliable and useful analysis to their readers.

**References**


