

SPECIAL SERIES: GUIDANCE FOR YOUNG INVESTIGATORS

Forming research questions[☆]

R. Brian Haynes*

*Clinical Epidemiology and Biostatistics, McMaster Health Sciences Centre, McMaster University,
1200 Main Street West, Room 2C10b, Hamilton, Ontario, Canada L8N 3Z5*

Accepted 16 June 2006

This article is based on the first chapter of a new book, *Clinical Epidemiology: How to Do Clinical Practice Research*, which is about designing and conducting practical research in clinical settings. The background for the book is described in another article in this issue [1]. The article begins with a clinical research scenario that provides a behind-the-scenes look at our own research. We've taken this approach in part to elucidate, illustrate, and titillate, but mostly to keep this as an unvarnished account of real research rather than a theoretically pristine but unattainable counsel of perfection.

1. Clinical research scenario

Observational studies of various sorts during the first half of the 20th century had established a relation between cerebrovascular strokes in the anterior part of the brain and narrowing of the carotid arteries in the neck and within the skull. This led to the invention of two surgical procedures. The first, carotid endarterectomy (CE), was introduced in 1954 to remove obstructions in the carotid artery as it passes from the aorta through the neck to the anterior brain. The second, extracranial–intracranial arterial bypass (EC-IC bypass), was developed in the late 1960s for individuals who had partial obstruction in the part of the carotid artery protected by the skull. Because CE cannot be performed in this part of the artery, the EC-IC bypass procedure “bypasses” the obstruction by freeing a branch of the superficial temporal artery on the outside of the skull then creating a hole in the skull over a branch of the middle cerebral artery, and then joining the superficial temporal artery to the middle cerebral artery using microsurgical techniques that include an operating microscope and sutures that are

invisible to the human eye. By the late 1970s, neither of these procedures had been tested in a well-conducted randomized controlled trial (RCT), although there had been an inconclusive trial of CE (in which surgical calamities were excluded from the analysis—a study in how *not* to analyze a clinical trial!) [2], and a second study was abandoned because of a 35% perioperative stroke and death rate among the first 43 patients admitted to the study [3].

In the late 1970s, a group led by Henry Barnett, a neurologist at the University of Western Ontario, David Sackett, a clinical epidemiologist at McMaster University, and Skip Peerless, a neurosurgeon at the University of Western Ontario, set about the task of evaluating surgical interventions intended to prevent strokes in the areas of the brain fed by the carotid artery. The key target at first was the surgical procedure that was by then relatively entrenched, CE. Testing the waters with surgeons who performed this procedure, and whose cooperation would be essential, it proved difficult to arouse enthusiasm for evaluating the procedure.

The target then shifted to the newer surgical approach, EC-IC bypass. This elegant procedure had been developed in Switzerland by Yasargil [4], brought to Canada by Peerless, and spread during the next 10 years to many countries. It was technically feasible, with high rates of bypass patency, but was very expensive, requiring both high surgical expertise and sophisticated equipment. Most surgical teams able to perform EC-IC bypass were in university centers, whereas CE had disseminated widely into the community hospitals as well. Although there were many case reports and case series attesting to the merits of EC-IC bypass, none of these compared it with medical treatment alone. This time it proved possible to recruit enough interested neurosurgeons and neurologists to form a study team. I was just completing my clinical training in internal medicine at the time, having previously completed my research training under Dave Sackett. My role in this process was to help develop the background literature review and the justification, including sample size considerations, a proposal for the study question, and a preliminary outline for the study design. Heady stuff for a young squirt anticipating an academic appointment the following year!

[☆] Based on Chapter 1 of Haynes RB, Sackett DL, Guyatt GH, Tugwell P. *Clinical epidemiology: how to do clinical practice research*. Philadelphia: Lippincott, Williams, Wilkins; 2005:3–14. With permission of the publisher.

* Corresponding author. Tel.: 905-525-9140; fax: 905-577-0017.

E-mail address: bhaynes@mcmaster.ca (R.B. Haynes).

An RCT of EC-IC bypass was conducted, beginning in 1978 and reported in 1985 [5]. The study showed no benefit for surgery; in fact, evaluation of the functional status of patients showed that the surgery delayed natural recovery from stroke for up to 1 year [6]. With this result, skepticism began to grow about whether CE was any more respectable than its downstream cousin, EC-IC bypass. The conditions had now become more favorable for testing CE—under certain conditions. Many surgeons remained opposed to testing CE, and those who were potentially willing to participate in such a trial wanted to ensure that the procedure was given a fair chance to succeed. To them, this meant that only surgeons with a “good track record” for CE would be included in the study, that the obstruction in the carotid artery would be severe enough, that patients would be likely to benefit from its removal (although many surgeons were offering the procedure for lesser degrees of narrowing), and that the patients themselves would be healthy enough to undergo surgery and live long enough thereafter for a benefit from surgery to be observed.

2. Where do researchable questions come from?

Eugene Ionesco, the father of the “theater of the absurd,” once said, “It is not the answer which enlightens, but the question.” This certainly applies to health care research—new knowledge originates from having asked answerable questions. To find new and useful answers to important problems that have not already been resolved, you need to know a lot about the problem and precisely where the boundary between current knowledge and ignorance lies. Without knowing a lot about the problem, it is difficult to imagine that plausible diagnostic tests and interventions will be developed. Without knowing the current state of knowledge, it is difficult to know whether one is headed in the right “next-step” direction. Thus, the first answer to the question introducing this section is that researchable questions come from finding the “cutting edge” of knowledge for a health problem with which you are familiar. This is not as demanding a condition for applied health research as it can be for basic science because good applied research usually builds on basic research. Indeed, it has been said that in applied research, the questions are easy but getting the answers is hard. This may be true—but composing important questions that can be answered validly by current applied research methods is still a considerable challenge.

3. Key considerations in developing a study question

As the clinical research scenario at the beginning of this article illustrates, many factors contribute to the formulation of a study question. Further, particularly in applied research, developing a question is an iterative process, not a “light bulb” phenomenon. To be sure, the light bulb must

come on, but there is much work to be done both before the light will shine and afterward. The iterative components include, to name a few, the basic dimensions of the clinical problem, the plausibility and feasibility of the design, the colleagues you will work with, the other resources you can muster to address the question, and the contingencies that emerge as you conduct the trial. The main interplay will be between what you would really like to do and what is really possible to do. This is anything but a linear process, but we’ll have to present it as such, given the nature of the printed word—forewarned is forearmed—don’t stick to the sequence discussed in subsequent text if your question could benefit from a different sequence. But the principles illustrated in the following sections will usually apply during the course of developing a study question, even if the sequence differs.

3.1. Basic dimensions

The basic dimensions of a problem that lead to the formulation of important research questions include understanding the biology and physiology of the problem, its epidemiology (i.e., determinants and distribution, prevalence, incidence, and prognosis), and frustrations in its clinical management that lead to unsatisfactory results for patients. For example, for strokes, the association of anterior brain infarcts with atherothrombotic narrowing of the carotid arteries fits with the biology of the small clots often found at these narrowings, which can break off and impact in the smaller arteries of the brain, causing a stroke. The occurrence of strokes also fits with the physiology of impairment of blood flow that occurs when the narrowing exceeds 75% of the normal luminal diameter of the carotid artery in the neck. The fact that biology and physiology do not provide an adequate basis for how to deal with the problem is evident from the results of the EC-IC bypass study. Indeed, in this trial, patients with the best surgical results, in terms of increased blood flow to the brain, fared worst for prevention of stroke.

As for the epidemiology, we know that stroke is one of the leading causes of death and major disability and that the risk of recurrence after a minor stroke is considerable, at about 10% in the first year and then about 5% per annum thereafter [7]. No one who deals with stroke victims can escape the conclusion that strokes would be better prevented than treated, if a safe and affordable preventive intervention is available, because the damage caused by a completed stroke is irreversible in the brain and the loss of function strokes incur is often unrecoverable. Case series and hospital surveys have documented that both EC-IC bypass and CE procedures can be performed with a lower perioperative morbidity and mortality than the observed rates of events mentioned earlier, although some studies of the quality of care for CE showed that perioperative rates of morbidity and mortality were higher than the risk of stroke recurrence in some hospitals, especially in community hospitals with

low volumes of cases. Further, in the time frames of the EC-IC bypass and CE trials, these interventions were based on biology, physiology, and anecdotal experience, and they had not been tested in large randomized trials. Thus, the basic elements were in place for an initial study question for this trial along the lines of “Does CE do more good than harm in preventing stroke recurrence in patients with carotid circulation strokes?”

3.2. Advanced considerations

Once these basic issues have been addressed, and an initial direction for a question seems promising, some additional key questions must be addressed.

Key questions checklist:

- What is the appropriate stage for evaluation?*
- Can internal validity be achieved?*
- To what extent is external validity (generalizability) achievable?*
- What will your circumstances permit?*
- What can you afford?*
- What is the best balance between “idea” and “feasibility”?*

3.2.1. What is the appropriate stage for evaluation?

The suitable stage of evaluation depends mainly on what previous assessments have been made for the question you are most interested in. Most research is incremental, and deliberately so. The less assessment that has been done, the more one can and should consider a less definitive and much less expensive research design (right—it’s about the bottom line).

Most diagnostic tests and treatments, particularly those in current use but incompletely assessed, are evaluated along a spectrum stretching from the explanatory end (can it work under ideal circumstances?) to the management end (does it work under usual clinical circumstances?). Studies for which scientific measures are taken to minimize bias will be somewhere in the middle of this spectrum, but will most often be toward the explanatory end because of the high cost of management studies. No study could be on the extreme of the explanatory end because circumstances of testing are never ideal. Indeed, even if they could be “ideal,” this would differ from the “real world” so much that it would render the results of the study practically meaningless. On the management end, it is not possible or ethical to scientifically and unobtrusively evaluate treatments and tests without introducing so much risk of bias that the results are undependable. This is, admittedly, a matter of debate, with advocates of outcomes research and observational studies claiming that the results of RCTs can be reproducibly achieved in careful observational studies that are based, for example, on

medical records. In our view, the degree of reproducibility in observational studies is unacceptable, and a careful RCT will be substantively better than an observational study at finding the truth.

Studies of causation, prognosis, and clinical prediction should also be staged according to the quality of preceding evidence, using the best study design that you can afford that goes beyond what has been done to date.

3.2.2. Can internal validity be achieved?

Internal validity depends on both study design features (“methods”) and on feasibility. Most study designs are relatively straightforward. Problems with feasibility, however, often stand in the way of success in implementing them. One such problem may be measurement. The basic principle of measurement was espoused by Lord Kelvin long ago (1883 to be exact): “... when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind.” Crudely put, if you can’t measure it, you can’t study it. For example, researchers interested in studying emerging diseases such as severe acute respiratory syndrome or West Nile virus infection first needed to come up with a test or at least a “case definition” before their research could proceed.

A second problem can be follow-up. It is difficult (but not impossible) to do follow-up studies of individuals with addictions or of those who are homeless. One can restrict entry to those individuals who are willing and able to be followed, but this may fundamentally alter the question that is posed because those who will enter may act differently from those who refuse.

Studies based on medical problems that are rare also pose a special challenge: it may require a national or international effort to assemble enough patients. Especially when you are starting out as a researcher, this type of question might be left to someone else or to later years.

3.2.3. To what extent is external validity (generalizability) achievable?

External validity refers to the ability to generalize the results of a study to other settings and patients, whereas internal validity refers to the soundness of the study to answer the exact question that it posed among the participants who began the investigation. A study that is internally invalid should not be undertaken (Period! Full stop!). In contrast, a study with limited external validity may be well justified if it represents a step forward in testing an idea at a reasonable price. Nevertheless, a question that includes a broad spectrum of patients that is similar to the range of presentations one sees in clinical practice has more appeal from a practical perspective than one that doesn’t. The extent to which external validity can be achieved usually comes down to, you guessed it, money: explanatory studies (“ideal circumstances”) generally cost less than

management studies (“usual circumstances”). The choices and their trade-offs are many. The general rule is don’t sacrifice internal validity for generalizability—but pose a question that is as generalizable as you can afford.

3.2.4. *What will your circumstances permit?*

Allowing for the desirability of having our reach exceed our grasp, the natural tendency of us all to “ask the big question” should be tempered by who we are and what circumstances we find ourselves in. The big question of whether CE does more good than harm is too challenging for anyone to tackle, let alone for someone who is just starting out. For example, CE can be and is offered to patients with asymptomatic narrowing of the carotid artery and for all degrees of stenosis. Attempting to answer the question for all indications would be exceedingly difficult. For our CE study, the prevailing clinical conditions meant limiting the question to symptomatic patients for whom surgeons and neurologists felt the procedure was likely to be beneficial, with surgery being done by operators with a record of low perioperative complication rate. As for who would be allowed to conduct the trial, it was very interesting as the junior on the team to see the “politics” of science play out, with the credentials of the senior neurologist, neurosurgeon, and methodologist being on the line before surgeons who were of mixed mind about whether CE “worked” but of a single mind that it was essential to sorting this out that these senior investigators be completely credible and trustworthy.

For CE, the matter of “uncertainty” evolved in a particularly interesting way. The EC-IC study cast enough uncertainty on the biologic–physiologic hypothesis for CE that it became possible to discuss the testing of CE with many surgeons. At one such meeting, Henry Barnett asked exactly the right questions: “Based on the evidence to date, how many of you believe that carotid endarterectomy does more good than harm for patients with stroke and carotid stenosis? And how many of you believe that it doesn’t?” To the amazement of many, the number of hands that went up was about equal for each question, providing sufficient basis for most believers of both persuasions to join forces to settle the matter once and for all. The circumstances were ripe.

3.2.5. *What can you afford?*

If you decide to pursue an investigation, the next consideration is what you can afford. Key aspects of cost include the time to complete the study, the amount of effort required in relation to the expected benefit, the enthusiasm for this effort, and the availability of funds. For time, the longer the study will take, the more important the question needs to be, and the less likely it needs to be that someone else is going to “scoop” you by being in the field ahead of you. Investigations involving large numbers (i.e., of years,

investigators, patients, research, and support staff) generally cost lots of money. Funding agencies and their peer reviewers are generally averse to awarding lots of money—but if a good match exists between their interests and the importance and timeliness of the question you wish to pursue, and if you have a sound plan to answer the question and the resources (i.e., investigators, patients, and commitment) and reputation to do so, then large budgets are at least conceivable. Having said that, if you will need a lot of funds for the question you are posing, it is best for first projects either to be part of a team that is already successful (as was the case for me in the EC-IC study) or to start small, in the form of either a preliminary study to address issues of feasibility for a larger trial or a study that addresses an interesting question (that is not necessarily of earth-shattering importance). In other words, take a small step forward rather than a leap.

A good way to start small is to do a systematic review. Systematic reviews are research studies in themselves and are best done with a protocol that begins with a clear, answerable question and with methods for finding and reviewing articles, minimizing bias, and summarizing and analyzing results. One of the most rigorous ways of conducting such reviews is to prepare such a protocol and to submit it to a funding agency for peer review and funding. Although many systematic reviews are done by voluntary labor, the range of external funding for reviews is as much as \$500,000. No small change! And the real reward from this activity is that it helps define exactly what questions have not yet been answered, setting the stage for next-step original investigations. This is worth considering before doing “first original studies” and, in fact, all major investigations.

3.2.6. *What is the best balance between “idea” and “feasibility”?*

For the CE study, it was believed (but not known at the time) that the degree of carotid stenosis would affect both the risk for stroke and the benefit from surgery. To capture this potentially high-risk, high-response group, the study was set up as two separate trials, one for patients with high-grade stenosis (70%–99%) and one for patients with lower-grade stenosis (30%–69%). Sample sizes were estimated on the basis of a 7% annual event rate for patients with high-grade stenosis and a 4% rate for those with lower-grade stenosis. This estimation assured those who felt strongly that stenosis was correlated with event rates that an early result could be achieved for patients with higher degrees of stenosis and that their results would not be “diluted” by the anticipated larger numbers of patients with lower degrees of stenosis. Further, statistical rules were developed for monitoring the accumulating results so that either of these trials could be stopped early if the results—either better or worse than estimated—warranted. This approach proved not only “political” but also

propitious. Indeed, the risk and the responsiveness for the high-grade group were both underestimated, leading to stopping the trial with a positive result when patients had been in the trial for an average of just 18 months of a planned 60-month trial. These results were quickly conveyed to participating investigators and their patients so that they could be taken into account for subsequent care decisions. For patients in the moderate-grade stenosis group, the trial was continued for its planned duration, and a positive, but less beneficial, result was observed.

Formulating a question that strikes a justifiable balance between the idea(s) for your study and the feasibility of answering them is important for success. Early on in the course of testing, this can mean focusing on just those patients who have high risk for adverse outcomes of their condition and who are likely to be highly responsive to the intervention. This restriction clearly limits the number of individuals to whom the results may apply, but if it is relatively easy to find patients who have both of these characteristics, it greatly reduces the cost of initial testing.

4. Composing the final prestudy question

The CE trial study question, as stated by the steering committee [8], was, “The study will determine if carotid endarterectomy is beneficial to patients with carotid stenosis and transient cerebral ischemia or partial stroke by comparing patients randomly assigned to receive carotid endarterectomy in addition to best medical care with those assigned to receive best medical care alone. The study is addressing the following specific questions: 1) Does carotid endarterectomy reduce the risk of subsequent stroke and stroke-related death? 2) Does the degree of carotid stenosis identify patients who will benefit most from carotid endarterectomy? and 3) Will carotid endarterectomy maintain or improve the functional status of patients over time?”

This statement of the study question (or related questions) contains four elements that are captured in the acronym, PICO: Patients, Intervention (for intervention studies only), Comparison group, and Outcomes. For good measure, and to avoid embarrassment in Chile, one could add Time (PICOT). (As one of us discovered after emphasizing the importance of PICO to future researchers in a Catholic university in Chile, “pico” is a slang term for an expansible part of the male anatomy...).

If you have been following the steps above in preparation for a study question of your own, you will have noticed that your question has changed several times. It’s now time to compose the question in a way that will “take charge” and direct the investigation that ensues. This should be a touchstone that you can refer to at times when the study boat hits a log and starts to sink, so that you can plug the hole in a way that suits the purpose of the expedition.

How inclusive should you be in describing the study question? The CE question posed earlier in the text is quite general about all aspects of the study, and one could more

completely describe just one of the two simultaneous CE studies as, “Among competent, consenting patients with recent transient ischemic attacks or partial strokes in the circulation of the carotid artery, and ipsilateral stenosis of 70% to 99%, as judged by expert central review of selective angiograms, who are receiving optimal medical care and do not have elevated surgical risk, does the addition of CE, by surgeons who have an established 30-day perioperative complication rate of less than 6% for persistent stroke or death, reduce the subsequent risk of major stroke and stroke death over a period of 5 years, compared with patients who receive optimal medical care but do not receive CE?” This question could then be iterated for the second study—less than 70% stenosis. Ninety-nine-word questions are difficult to comprehend, so I don’t recommend this much detail in the question itself, but it is important to bear these details in mind when conducting the study and reporting its results, so that the results will not be overgeneralized.

5. Composing secondary questions

It will be obvious from the preceding section that the CE study had several questions. Several basic principles guide the development of additional primary and secondary questions for studies. First, all primary questions must be asked “up front,” at the beginning of the investigation. The same is true, as far as possible, for all secondary questions. This approach ensures that the questions are “hypothesis-driven” (i.e., based on your predictions of what will happen) rather than “data-driven” (i.e., made up after the study results are [partly] in, especially to “explain” findings that may well be simply the play of chance). This approach also allows for proper planning and data collection for these additional questions, including estimates of sample size to determine whether the study is large enough to support reliable answers. These efforts can pay off; it will be less costly to run a study where some of the questions can be answered by data collection from only a subset of patients and where questions for which there can be no chance of a clear result are discarded along with their burden of data collection.

Second, these “add-on” questions should never compromise the primary question. For example, obtrusively measuring the adherence of the patients to their prescribed medications in a management study would undermine the validity of such a study if this measurement is not an intended part of the intervention. As another example, adding greatly to the data collection for a study can compromise the willingness of investigators and patients to participate.

Third, additional questions should not be a large part of the budget because this risks not receiving funding for the major study question. If they do add significantly to the budget (as even some simple measures can), then the secondary questions should be clearly separated in the budget so that reviewers and funding agencies can lop them off

if they are not convinced that they are worth the cost, even if the main study question is.

6. Dealing with contingencies

The CE study was originally designed for four separate study groups delineated by the stenosis levels defined in the preceding text and by the presence or absence of ulcerated plaque in the area of the stenosis for each of these two grades of stenosis. It was estimated that 3,000 patients, distributed among the four study groups, would be needed to provide separate answers concerning the benefit of surgery for each level of stenosis and presence or absence of plaque. Early on in the course of the trial, it was determined through central review of the reports from surgeons at the various study sites that the presence or absence of plaque could not be reliably determined. Thus, the question concerning plaque vs. no plaque could not be answered (remember Lord Kelvin!). The sample size estimates were altered to fit the two remaining study cohorts, 600 for the high-grade stenosis group and 1,300 for the moderate-grade stenosis group.

During any trial, you can expect that contingencies will arise that require modification of the protocol and changes in the question that is addressed. Sometimes, as in the example from the CE trial, the contingency will be profound enough that a study question will have to be dropped—if this is the main study question, then the trial may have to be abandoned entirely. Fortunately for the CE trial, there was more than one question, and the early detection of problems in reporting plaque led to a timely reduction in the sample size required. One could easily point fingers and say that the trial should not have proceeded in the first place if this measurement issue had not been addressed, but that is a different matter!

Most of the contingencies that arise will not sink the study if you keep a close eye on the process of the study (e.g., whether patients are being recruited at the anticipated rate) and if adjustments are made that counter the problem without compromising the basic intent of the study. For example, in the CE trial, because of slow recruitment, the upper limit of 80 years for patient age was relaxed if the surgeon judged that the perioperative risk was acceptable. In any event, only patients who were mentally competent and gave their informed consent were included.

The leaks in the protocol that become apparent as the study enters the water, and those that occur once it is under way, need to be plugged. You can plug the low recruitment leak (a very common one!) by recruiting more investigators or by relaxing entry criteria, but these changes need to be recorded and their effect, if any, on the study question needs to be described in reports of the investigation. For example, during the CE trial, standards for care for hypertension, for cholesterol lowering, and for antiplatelet treatment changed because of new evidence. The latter, in particular, had the potential for lowering the risk of stroke, the major study outcome measure. In each instance when major new findings and recommendations came out, they had to be considered by the study's steering committee and a decision had to be made about incorporating them into the protocol in a way that preserved the integrity of the study, if possible—or not, if need be. Although none of these factors changed the course of the trial for CE, the CE study results led to one other major trial being aborted.

References

- [1] Haynes RB, Sackett DL, Guyatt GH, Tugwell P. How to do clinical practice research: a new book and a new series in the *Journal of Clinical Epidemiology*. *J Clin Epidemiol* 2006;59:873–5 [this issue].
- [2] Fields WS, Maslenikov V, Meyer JS, Hass WK, Remington RD, Macdonald M. Joint study of extracranial arterial occlusion. V. Progress report of prognosis following surgery or nonsurgical treatment for transient cerebral ischemic attacks and cervical carotid artery lesions. *JAMA* 1970;211:1993–2003.
- [3] Shaw DA, Venables GS, Cartlidge NE, Bates D, Dickinson PH. Carotid endarterectomy in patients with transient cerebral ischaemia. *J Neurol Sci* 1984;64:45–53.
- [4] Yasargil MG, editor. *Microsurgery applied to neurosurgery*. Stuttgart: Georg Thieme; 1969. 105–115.
- [5] The EC/IC Bypass Study Group. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke: results of an international randomized trial. *N Engl J Med* 1985;313:1191–200.
- [6] Haynes RB, Mukherjee J, Sackett DL, Taylor DW, Barnett HJ, Peerless SJ. Functional status changes following medical or surgical treatment for cerebral ischemia: results in the EC/IC Bypass Study. *JAMA* 1987;257:2043–6.
- [7] Dennis MS, Burn JP, Sandercock PA, Bamford JM, Wade DT, Warlow CP. Long-term survival after first-ever stroke: the Oxfordshire Community Stroke Project. *Stroke* 1993;24:796–800.
- [8] North American Symptomatic Carotid Endarterectomy Trial (NAS-CET) Steering Committee. North American Symptomatic Carotid Endarterectomy Trial: methods, patient characteristics, and progress. *Stroke* 1991;22:711–20.