

# Use of data from nonrandomized trial designs in evidence reports: An application to treatment of pulmonary disease following spinal cord injury

Gregory P. Samsa, PhD; Joseph Govert, MD; David B. Matchar, MD; Douglas C. McCrory, MD

Center for Clinical Health Policy Research, Duke University Medical Center, Durham, NC; Department of Medicine, Duke University Medical Center, Durham, NC; Department of Community and Family Medicine, Duke University Medical Center, Durham, NC

**Abstract**—Evidence reports summarize the evidence pertaining to various health-related topics. Including evidence from nonrandomized studies into such reports involves a trade-off between availability and bias. We describe a general framework by which information from nonrandomized studies might be integrated reasonably into evidence reports and illustrate its application to a recent evidence report on preventing pulmonary complications among patients with spinal cord injury. The proposed framework, which is based upon the premise that producing a fair summary of the evidence requires only a level of evidence judged by clinical experts to be sufficient to the task at hand, may help focus scarce resources, strengthen the quality and documentation of decisions including evidence from nonrandomized studies, and suggest high-priority areas for future research.

**Key words:** *evidence reports, information synthesis, nonrandomized trials, study design.*

---

This material is based on work supported by the Agency for Healthcare Research and Quality (Contract No. 290-97-0014).

Address all correspondence and requests for reprints to Gregory P. Samsa, PhD, Duke Center for Clinical Health Policy Research, 2200 West Main Street, Suite 230, Durham, NC 27705; email: samsa001@mc.duke.edu.

## INTRODUCTION

### Preface

At the heart of the movement toward evidence-based medicine is the unassailable notion that clinical practices should have the strongest scientific basis possible. Another unassailable notion is that, in most circumstances at least, well-conducted randomized controlled trials (RCTs) provide the strongest possible evidence. Thus, it is readily understandable that the evidence-based evidence movement, as exemplified by the Cochrane Collaboration and the Agency for Healthcare Research and Quality (AHRQ), has in the main begun by addressing those components of medical practice having many moderate- to large-sized randomized trials.

Although certainly a reasonable policy, this approach begs the question of what should be done in those areas of medicine that, for various reasons, do not yet (and perhaps never will) have large randomized trials. Much of rehabilitation medicine falls within this category, as does spinal cord injury (SCI), with its relatively small number of incident cases and multiple clinical issues, making it very difficult to separate the effects of numerous simultaneous interventions.

The Duke Evidence-Based Practice Center (EPC) recently performed an AHRQ-sponsored evidence

review on treatment of pulmonary disease following SCI. We were also awarded a methodological supplement to consider the question of the proper use of evidence from nonrandomized trials. We found that, despite the large scope of medical practice that has not been addressed by randomized trials, very little had been written regarding what to do with the non-RCT literature.

Accordingly, the following report is the quintessential “thought piece”—that is, a conceptual effort describing an approach that, while potentially promising, has yet to be definitively tested. Our EPC is currently experimenting with this method, which is in large part inspired by the philosophy of continuous quality improvement, and we hope to present subsequently a more rigorous assessment. In the interim and in the spirit of continuous improvement, we would be happy to receive comments and suggestions about how the ideas described here might best be implemented.

### Rationale for Development of a Framework

The mission of EPCs is to promote evidence-based decision making by producing reports that summarize the evidence pertaining to various health-related topics. These topics typically involve the *efficacy* (i.e., benefit under ideal conditions) and *effectiveness* (i.e., benefit in usual practice) of one or more interventions pertaining to diagnosis, prevention, and/or treatment of disease. As a rule, the above interventions are selected because the condition in question has a major impact on public health, there is significant uncertainty about the best way to proceed, and/or practice variations are extensive. Important decisions are based upon these evidence reports (e.g., through guidelines formulated with input from evidence reports). Indeed, given that these impor-

tant decisions cannot be delayed, an implication stemming from the criteria for choosing topics is that the evidence report must proceed even if the studies upon which it is based are not definitive.

One of the critical components of an evidence report is an assessment of the strength of its underlying evidence. It is generally agreed that the strength of evidence pertaining to the implications of an intervention depends upon a number of factors, including the *plausibility* of the intervention's mechanism of action, the ability to *extrapolate* from other studies (e.g., of similar interventions, populations and/or conditions), the *strength* of the observed benefit, the *consistency* of this benefit (e.g., across subpopulations within a single study, across studies), and the *design* of the studies in question. For example, **Table 1** presents the grades of evidence used in developing the most recent version of the American College of Chest Physicians guidelines for antithrombotic therapy (Gordon Guyatt, MD, personal communication; February, 2000). All the above elements are included in this scheme. RCTs are, of course, preferred, but information from less definitive nonrandomized designs is accepted as well.

Perhaps the most fundamental concern in using information from nonrandomized designs is that it may be biased. In the extreme form of this putative bias, an intervention that appears to be efficacious in non-RCT studies might have no effect, or even be harmful, when assessed with the use of the more rigorous RCT design. Concerns such as the above have led the Cochrane Collaboration to focus on the summarization of results from RCTs, to the virtual exclusion of other forms of evidence. Reports from EPCs have varied in their use of non-RCT evidence. In any event, a perception that the exclusion

**Table 1.**  
Grades of evidence.\*

Grade	Level of Evidence
1a	Risk/benefit clear from RCT results in other populations, which can be unequivocally extrapolated to the present case, or overwhelming evidence from observational studies
1b	Risk/benefit unclear from randomized trials without important limitations
1c	Risk/benefit clear from RCT results in other populations, which can be unequivocally extrapolated to the present case, or overwhelming evidence from observational studies
2a	Risk/benefit unclear from randomized trials without important limitations
2b	Risk/benefit unclear from randomized trials with important limitations
2c	Risk/benefit unclear from observational studies

RCT = Randomized controlled trial(s)

\* Grades of evidence used in developing recent American College of Chest Physicians guidelines on antithrombotic therapy. Gordon Guyatt, MD, personal communication.

criteria used in evidence reports (whether from EPCs or elsewhere) are overly restrictive has, in turn, led to a backlash from various parties who believe non-RCT data can be useful. In particular, these critics would argue that a “null” evidence report based only on selected RCTs runs the risk of excluding much of the relevant evidence in favor of existing clinical practice, and thus, unfairly characterizes current practice as lacking a sufficient scientific basis. This issue is particularly salient in those areas of medicine that have few or no RCTs available for review.

For purposes of this exposition, we assume that an evidence report is being prepared on a topic for which definitive RCTs are not extant (indeed, this describes the majority of topics of interest to clinicians and policymakers) and that information from non-RCT designs must, of necessity, be included. We first describe a framework by which such information might reasonably be integrated into the presentation. Then, by way of illustration, we use an example from our recent evidence report on preventing pulmonary complications of cervical SCI of traumatic etiology.

## METHODS

### Conceptualization of the Question Under Study

Before proposing an answer to the question “how can non-RCT data be incorporated into an evidence report?” it may be useful to frame more precisely the “question behind the question” under study. Recognizing that circumstances will force us to use imperfect information in any event, we ask which of the following options is most likely to be the case?

1. RCTs are always the ideal design, and the use of non-RCT data is thus a methodologically flawed, yet often necessary, expedient.
2. All designs have advantages and disadvantages and should be considered as equal.
3. Both of the above statements are exaggerated, and the truth lies somewhere in between.

Apart from helping to clarify the underlying issues, recognition of this “question behind the question” may assist in understanding the strong opinions sometimes expressed in the debate about non-RCT evidence. In particular, if the first assertion is true then the position taken by advocates of non-RCT evidence might reasonably be characterized as overly pragmatic, while if the second

assertion is true, the position taken by advocates of limiting evidence to RCTs might reasonably be characterized as unnecessarily narrow.

If one were to recognize that the first two of these options were intentionally stated as extremes, perhaps a more actionable version of the previous question is “under what circumstances is an RCT the only valid option, and under what circumstances is an RCT merely a preferable option and why?” Our discussion is limited to the assessment of the clinical benefits and risks of an intervention (i.e., its “impact”). For purposes of describing the epidemiology of a condition, its public health burden, its cost, and so forth, the superiority of non-RCT designs is not in question.

At this point, it may be helpful to recall the distinction between internal and external validity and between efficacy trials and effectiveness trials. *Internal validity* refers to the defensibility of conclusions within a given study, while *external validity* refers to the degree to which study results can be generalized. An efficacy trial is an RCT where the intervention is administered in highly controlled and otherwise optimal conditions, while an effectiveness trial is an RCT in a less controlled setting intended to approximate typical practice. By its tight experimental control, an *efficacy trial* is designed to maximize both the internal validity of conclusions as well as the benefit of the intervention. *Effectiveness trials* address the construct of the most ultimate interest to the users of an evidence report—namely, the impact of an intervention in actual practice—but tend to have less internal validity (i.e., somewhere between efficacy trials and nonrandomized designs) because of the multiple additional considerations implied by its looser experimental control. Both effectiveness and efficacy trials ultimately rely upon the principle of randomization for their validity. The validity of non-RCT designs increases with the degree to which it can be successfully argued that their results are similar to what would have been observed had randomization been applied.

Given the above, the ideal situation would be for an evidence report to be based upon both efficacy trials and effectiveness trials (and for these trials to be of high quality with consistent results). With their high internal validity, the efficacy trials would provide the most definitive possible proof that the rationale supporting the intervention is sound. With their high external validity, the effectiveness trials would illustrate that the promise of the intervention demonstrated in the efficacy trials can, in fact, be fulfilled in practice. The benefit of interventions

is consistently less when measured with effectiveness trials than with efficacy trials (1); thus the need for effectiveness trials (or their surrogates) is particularly acute.

We now turn to the questions “under what circumstances is an RCT the only valid option? under what circumstances is an RCT merely a preferable option? and why?” as they apply to efficacy and effectiveness trials. One problem that soon becomes apparent is the small number of effectiveness trials in the literature. This points to one possible use of non-RCT designs. If we can be confident of the underlying efficacy of an intervention (i.e., through the analysis of efficacy trials or other means), then observational data (e.g., from prepost designs, cohort studies, and data bases) might be a substitute for the missing effectiveness trials by providing a description of the likely behavior of the intervention in actual practice. This description is subject, potentially, to various well-catalogued biases associated with nonrandomized designs (2). On the other hand, provided the goal is to estimate the impact of the intervention in typical practice, it must be recognized that bias is a characteristic also shared by efficacy trials. In other words, although efficacy trials are unbiased and internally valid for the purpose of assessing efficacy, they tend to be biased optimistically for assessing effectiveness. As substitutes for effectiveness trials, efficacy trials and non-RCT designs can potentially play a complementary role.

Can a conclusion about effectiveness be drawn confidently even in the absence of an efficacy trial? Two situations are possible: (1) RCTs have been performed in similar populations and/or for similar interventions, and (2) no RCTs are available, but various non-RCTs are. We argue that even if the available information consists entirely of data from non-RCTs, conclusions can still be drawn (indeed, clinical experts and other decision makers are required to draw such conclusions in the face of incomplete information every day); however, these conclusions are less definitive than would be the case were RCTs available.

Whether these conclusions are sufficiently definitive seems to us to be essentially a value-of-information problem. For example, if the stakes are high and the non-RCT evidence is relatively weak, then (a) decision makers should be willing to allocate the resources to perform a RCT, (b) clinical experts should be uncomfortable recommending the intervention without additional data, and (c) the evidence report should reflect the lack of adequate information about the impact of the intervention. On the other hand, if the stakes are low and/or the non-RCT evi-

dence is relatively strong, then (a) it is not worthwhile for the decision maker to allocate the resources for an RCT, (b) clinical experts should be comfortable recommending the intervention, and (c) the evidence report should reflect the current consensus.

### Summation of the Question Under Study

In summary, our argument is as follows. Rarely do we have the luxury of definitive evidence from both efficacy and effectiveness trials performed among the exact population of interest; thus judgment and extrapolation of evidence will be required in any event. As long as the ultimate frame of reference is the behavior of an intervention in typical practice, the choice is not between a pristine analysis of unbiased efficacy trials versus the inclusion of biased data from other sources, but instead, what weight should be applied to imperfect (yet potentially complementary) evidence of various types.

Although non-RCT evidence is in general less definitive than RCT evidence, the question of whether non-RCT-based evidence is sufficient for the purposes at hand is best judged not by a single a priori standard pertaining to study design, but rather on a case-by-case basis by clinical experts. The intent of the method to be described is not to provide a standard by which these clinical experts make the above judgment regarding sufficiency, but rather to provide a framework by which the reasons behind their decisions can be recorded. Our notion is that by formalizing this decision making/recording process, the resulting decisions can be better explained, assessed, updated, and ultimately improved.

### Conceptual Model

For concreteness, we assume that the theoretical underpinning of the evidence report is the *conceptual model* of Woolf (3). This conceptual model uses as background the natural history of the condition in question (e.g., in terms of parameters such as event rates, complication rates, quality of life, utilization, and costs), then appends onto this natural history the impact of the intervention in modifying the above parameters. While Woolf discusses the question of non-RCT evidence, he does not go into great detail on using this evidence most efficiently, nor does he present specific proposals regarding how to make an operational work plan for a report which uses non-RCT evidence (the focus here). In any event, our approach does not depend upon the specific details of Woolf's particular approach to conceptual modeling, but

is presented in such a way as to emphasize consistencies in thinking.

### Analytical Goals

We also note that the goals of evidence reports (and, of course, of most other scientific endeavors) can be described along the axes of descriptive versus inferential, as well as quantitative versus qualitative. Most methodological effort to date has focused on the quantitative component of formal inference. Our proposed method is nonquantitative and primarily descriptive, in the sense that its goal is to describe the reasons supporting various decisions about level, quality, generalizability, and sufficiency of evidence. It is only inferential in the sense that it provides a structure by which these decisions can be made and recorded systematically.

### Proposed Approach: Specification of the Causal Pathway

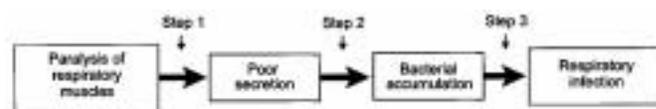
As terminology, the *causal pathway* is the sequence of steps describing the natural history of the condition under consideration, while the *intervention pathway* is the sequence of steps by which the intervention may modify this natural history. Our proposed approach begins with a presentation of the causal pathway as structured in **Table 2**—that is, with the rows corresponding to the steps in the causal pathway and the columns corresponding to various levels of evidence. Levels of evidence are ordered by increasing strength of design, from medical first principles, laboratory studies, observational studies outside the population of interest, observational

studies within the population of interest, and RCTs outside the population of interest to RCTs within the population of interest.

For example, consider the development of pneumonia following SCI. One pathway for the development of this complication is as follows:

1. Because of the paralysis of various muscles involving respiration for patients with SCI, the clearance of secretions tends to be inadequate.
2. In the absence of effective intervention, the inability to clear secretions leads to colonization of mucosa with bacteria.
3. Large numbers of bacterial pathogens tend to lead to respiratory infections, including pneumonia (see **Figure 1**).

Although it is always better to have more data rather than less, the basic idea is to fill in each of the rows in **Table 2** with information that is judged by clinical experts to be at least sufficient to support the step in question. For example, step 1 could be sufficiently well supported by observational studies of secretion clearance for patients with SCI. Step 2 could be sufficiently well supported by case series of patients with impaired secretion clearance (whether due to SCI or not), or even by



**Figure 1.**

Causal pathway: Development of pneumonia in respiratory muscle paralysis.

**Table 2.**

Causal pathway: Development of pneumonia in respiratory muscle paralysis.

Steps	Level of Evidence					
	A-first principles	B-laboratory studies	C-observational studies outside SCI	D-observational studies inside SCI	E-RCTs outside SCI	F-RCTs inside SCI
1. Because of paralysis of various muscles involving respiration for patients with SCI, clearance of secretions is poor.	—	—	—	Wang et al. 1999 (4) DeVivo et al. 1999*	—	—
2. In absence of effective intervention, large quantities of bacteria tend to accumulate in lungs.	—	—	—	—	—	—
3. Large numbers of bacterial pathogens tend to lead to respiratory infections, including pneumonia.	—	—	—	DeVivo et al. 1999 (5)	—	—

SCI = Spinal cord injury

RCT = Randomized controlled trial(s)

\* DeVivo 1999 spans steps 1–3 (5)

reliance upon medical first principles. Step 3 could be sufficiently well supported by case series demonstrating that increased colonization with bacterial pathogens tends to precede, and greatly increase the likelihood of, clinically apparent pneumonia.

Alternatively, since step 2 is in such little doubt, step 3 could be sufficiently well supported by observational studies demonstrating an increased risk of pneumonia and other respiratory infections within patients for whom secretion clearance is poor (e.g., SCI, chronic obstructive pulmonary disease, cystic fibrosis, bronchiectasis), thus effectively combining steps 2 and 3. Here, the extrapolation is that, for purposes of preventing pneumonia, what matters is primarily the fact of poor secretion clearance and not its cause. One potential validation of this extrapolation would be the observation that the relationship between poor secretion clearance and respiratory infections is consistent across a number of patient populations. Within the SCI literature, an example of the level of support for step 1 is an observational study by Wang (4), demonstrating that mean peak expiratory flow rate during coughing decreases with increasing injury level. Since the causal pathway uses secretion clearance rather than peak expiratory flow, the proper interpretation of this finding depends upon the extent to which peak expiratory flow (i.e., in contrast to other factors such as peak excitatory force and forced vital capacity) correlates with actual mobilization and clearing of secretions. This relationship is relatively unstudied.

Much of the information in the SCI literature addresses not a single step in the causal pathway, but rather multiple steps at once. For example, various observational studies, particularly those from the Model Systems, demonstrate that the rate of pneumonia among persons with SCI is significantly higher than among the general population, with the risk of pneumonia rising with increasing injury level (5). The extrapolation from these findings, as often made in the SCI literature, is that such information is not only sufficient to demonstrate an overall relationship between SCI and pneumonia but is also adequate to support steps 1, 2, and 3 individually. That is, it is assumed that observing an overall relationship between SCI and pneumonia, as implied by the above causal pathway, is sufficient to establish the validity of this pathway, even without evidence directly supporting its individual steps. One of the roles of expert judgment is to indicate the degree to which such extrapolations are likely to be valid.

More specifically, we propose that expert input be applied at the following stages in the above process:

1. Elucidating the causal pathway (i.e., identifying the steps leading from “root causes” to clinically relevant outcomes).
2. Determining the lowest acceptable level of evidence sufficient to verify each step. This task includes (2a) determining the *lowest-ranking study design* (i.e., left-most column) which would be acceptable and (2b) determining the *least amount of evidence required for studies using this design* (e.g., would a single citation suffice or is a literature search needed?).
3. For low-ranking designs (e.g., laboratory evidence), nominating a representative citation or citations.
4. Regardless of design, assessing the assumptions required to extrapolate from the data to the conclusions. This task includes (4a) clearly stating what assumptions are required to extrapolate from the data to the conclusions, (4b) specifying the degree to which such assumptions are likely to be valid, and (4c) describing the evidence that would be required for these assumptions to be sufficiently well supported.
5. When a study spans a number of steps, assessing the assumptions required to extrapolate to individual steps. This task includes (5a) clearly stating what assumptions are required to extrapolate to individual steps, (5b) specifying the degree to which such assumptions are likely to be valid, and (5c) describing the evidence that would be required for these assumptions to be sufficiently well supported.

A number of points regarding the above process may be made. First, EPC literature reviews are noteworthy for their comprehensiveness and attention to detail. Certainly, it would be prohibitively expensive to perform a comprehensive review of the evidence pertaining to each step in the above pathway. We do not propose this. For example, as discussed above, step 2 in the pathway (patients with poor secretion clearance will tend to develop large quantities of bacteria in the lung) is medically trivial and not at all in dispute. Indeed, there is little doubt about any of the steps in the above causal pathway. The clinical experts would most likely simply nominate a single representative study to support this step, thus obviating the need for a literature search and comprehensive review. The reason why step 2 is specified at all will be apparent from the discussion of the intervention pathway.

Second, even though many of the steps in the causal pathway might appear to be trivially reductionist,

specifying the steps and having experts make an explicit decision as to the level of evidence required to support these steps serves the educational function of informing the user of the evidence report about current thinking in the field. It might also serve to uncover unexpected disagreements, although in the above example, such disagreement would be quite surprising.

Third, for describing the causal pathway, few/none of the steps requires RCTs for validation. Essentially, RCTs are only relevant to the description of the intervention pathway.

Fourth, it has been our experience that discussing the question of whether it is appropriate, as a general rule, to extrapolate results from non-RCT designs and runs the risk of becoming bogged down in fruitless debates about methodological first principles. However, in a specific circumstance, the question of which assumptions are required for an extrapolation to be valid is usually more actionable as well as less controversial. Clinical experts might still disagree on whether these assumptions hold in a given case, but stating the assumptions themselves is a task that is usually quite manageable, as is the task of describing the evidence that would be required to support these assumptions.

For example, in step 1, the assumption required to extrapolate from Wang's study (4) was that peak expiratory flow is sufficient as a surrogate measure for secretion clearance. As either a "thought experiment" or a proposal for a new study, the ideal would be, under appropriate conditions, to measure directly both con-

structs. In the interim, the assumption/extrapolation involves our confidence in what such a study would show, were it to be performed.

As another example, one of the most problematic issues in the SCI literature is the small sample size of many of its studies. This is in large part because SCI's relative rarity implies that few facilities will encounter large numbers of incident cases. However, a much larger number of persons suffer paralysis or weakness of the respiratory muscles, for example, because of conditions such as muscular dystrophy and multiple sclerosis. In some cases, what is important is the fact of the paralysis/weakness, regardless of its cause. In other cases, the cause of the paralysis/weakness predominates, for example, because SCI implies critical differences in physiology and pathophysiology and/or because of the implications of SCI's sudden and unexpected onset. As discussed previously, in the present case, the important consideration is not SCI, per se, but rather impaired secretion clearance. Thus, a number of conditions are potentially appropriate for extrapolation.

### Proposed approach: Specification of the Intervention Pathway

Continuing the example, **Table 3** describes the structure of the intervention pathway, which is identical to the structure of the causal pathway. That is, rows represent steps in the pathway, and columns represent levels of evidence. The role of the clinical experts is essentially as described in the previous section.

**Table 3.**

Intervention pathway: Prevention of pneumonia in respiratory muscle paralysis.

Step	Levels of Evidence					
	A. first principles	B. laboratory studies	C. observational studies outside SCI	D. observational studies inside SCI	E. RCTs outside SCI	F. RCTs inside SCI
1. Assisted cough will improve secretion clearance.	—	—	—	—	?*	Jaeger 1993 (6)**
2. Improvement in secretion clearance reduces number of bacteria in lungs.	—	—	—	—	?	—
3. Reduced accumulation of bacteria reduces risk of respiratory infections, including pneumonia.	—	—	—	—	?	—
4. Assisted cough has minimal risks.	—	—	—	Balshi 1989 (7)	—	—

SCI = Spinal cord injury.

RCT = Randomized controlled trial(s).

\* As discussed in the text, clinical experts might decide that RCTs of assisted cough within patient populations, such as COPD, cystic fibrosis, and bronchiectasis, are relevant, as well as RCTs of other secretion-clearance-enhancing interventions such as chest physiotherapy.

? Denotes this uncertainty.

\*\* Jaeger 1993 is a within-subject experiment, with a validity similar to RCT.

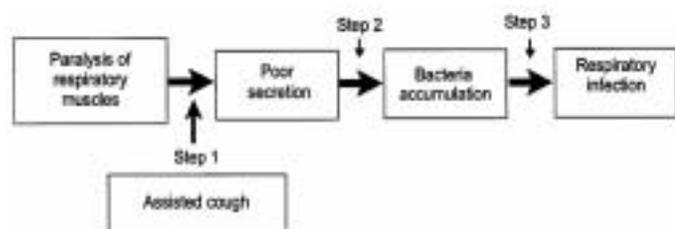
For example, consider the intervention of assisted cough, intended to help prevent the development of pneumonia (and other respiratory infections). The intervention pathway is defined in concert with the causal pathway:

1. In order to assist the coughing function, which is no longer effective among persons with sufficiently high-level SCIs, the maneuver of an assisted cough will improve secretion clearance.
2. This improvement will be sufficiently great as to reduce the quantities of bacteria in the lung.
3. This reduction in bacteria will be sufficiently great as to reduce the incidence of pneumonia (and other respiratory infections).
4. In addition, the maneuver has minimal risks in comparison with its benefits (see **Figure 2**).

The same procedure is followed as before, that is, determining the level of evidence that is sufficient for the task at hand. As a rule, a higher level of evidence will be required to support the steps in the intervention pathway than will be required for the causal pathway.

Regarding step 1 of the intervention pathway, some of the evidence in the SCI literature is represented by Jaeger's experimental study of 14 patients (6). There, using each patient as his or her own control, mean peak expiratory flow rates were higher for assisted cough in comparison with unassisted cough (238 L/min versus 203 L/min). As with the previously cited study by Wang (4), the proper interpretation of this finding depends upon the extent to which peak expiratory flow correlates with the ability to clear secretions.

Steps 2 and 3 have been specified, in detail, to make explicit one of the main potential pitfalls of using non-RCT evidence; namely, even if the mechanism of action for an intervention is specified correctly, its effect may be too weak to make a practical difference for the patient. For example, an assisted cough might reduce the number of bacteria in the lung, but not sufficiently so as to reduce



**Figure 2.**

Causal pathway with intervention: Prevention of pneumonia in respiratory muscle paralysis.

the likelihood of infection. Even though our proposed approach is nonquantitative, the clinical experts must, nevertheless, make a judgment not only about the plausibility of the mechanism of action of the intervention (i.e., the intervention pathway), but also about the magnitude of its likely effects.

Regarding step 4, Balshi's case series study (7) describes possible complications of assisted cough for the subgroup of patients having Greenfield filters. We recognize that it is probably not appropriate to extrapolate these data to all patients with SCI; nevertheless, this study is given as an example of the kind of cohort-based information available in the literature. In addition to cohort-based data, results from case-control studies would also be relevant here.

It should be noted that none of the above studies uses an RCT design (although Jaeger's study is a within-subject experiment and thus has a similar level of methodological validity). Indeed, an RCT of assisted cough (or other form of aggressive pulmonary care) versus passive pulmonary care would, at this point, likely be unethical. The inference from such a study would simultaneously span steps 1–3 and 4 of the intervention pathway. Nevertheless, the evidence in favor of each of the steps in the intervention pathway (whether individually or based upon studies which span steps) can be codified, and SCI experts can then judge its sufficiency as well as state the assumptions required for its most appropriate interpretation.

If the SCI-literature-based evidence is deemed insufficient, then various extrapolations can be considered. Here, following the insight that what is most essential is the improvement of secretion clearance, one might argue not only that the patient population could include conditions such as COPD, cystic fibrosis, and bronchiectasis, but also that other secretion-clearance-enhancing interventions (such as chest physiotherapy) can provide information about the likely magnitude of the effects that can be expected for assisted cough. Once the opportunities for appropriate extrapolation have been exhausted, any steps in the intervention pathway still having insufficient information would remain as having the highest priorities for future research. For assisted cough, presuming that the overall evidence is deemed satisfactory, the more interesting questions would likely pertain to comparisons of various techniques for manually assisted cough, comparisons between manually assisted cough versus electrical stimulation of the abdominal muscles, and so forth.

### Another Example: High Versus Low Tidal Volume for Weaning from the Respirator

To illustrate use of the above method in identifying high-priority research areas, consider the more controversial question of whether high tidal volumes can aid in speeding weaning from the ventilator and reduce the incidence of respiratory complications. A simplified model of the causal pathway is that in order to increase the likelihood for successful weaning, the following conditions should hold:

1. Airways and alveoli should be open.
2. Pulmonary muscles should have adequate strength.
3. Lung structures should not be compromised (e.g., by conditions such as barotrauma, bronchopleural air leak, parenchymal damage, altered hemodynamics, etc.).

None of these points is in question, and documentation of the causal pathway could be based upon easily available evidence, as in the previous example.

A simplified model of the intervention pathway is:

1. High tidal volumes will be more successful in keeping the airways and alveoli open.
2. Keeping the airways and alveoli open will be sufficient to significantly reduce the likelihood of atelectasis and pneumonia (these conditions not only delay weaning but also have the potential effect of leading to a permanent reduction in vital capacity).
3. Recognizing that increased tidal volumes also have risks such as barotrauma. In practice, these risks are minimal, relative to the benefit of the procedure (**Figure 3**).

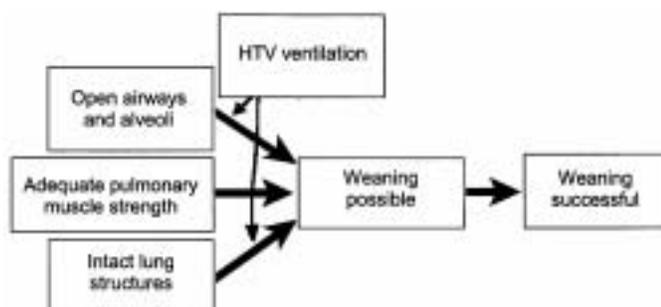
Within the SCI literature, perhaps the most extensive data regarding the impact of high tidal volumes have been reported by Peterson (8). In a nonrandomized retrospective study of 42 patients with injuries at C3 to C4, 23

of these patients had been weaned from the ventilator with the use of low tidal volumes while 19 patients had been weaned with high tidal volumes. Patients with high tidal volumes had lower rates of atelectasis and were able to wean more rapidly from the ventilator.

In contrast to the SCI literature, the literature studying the impact of tidal volumes for other conditions (e.g., acute respiratory distress syndrome (ARDS), acute lung injury, surgical patients under anesthesia, etc.) is much larger and includes RCTs. Indeed, results from the ARDS literature are sometimes cited, perhaps incorrectly, as relevant to patients with SCI. How should the trade-off be made between higher quality information (i.e., larger studies with randomized designs) from the non-SCI literature and more directly relevant data from patients with SCI? Although the results are not entirely consistent, this question becomes particularly salient when one considers that, although one of the five major RCTs with ARDS patients (9–13) favored high-volume management, the preponderance of evidence (including the results of the largest RCT to date) favored the low-volume alternative (9).

Our proposed approach provides a systematic way to proceed. Two questions are fundamental. First, in the opinion of the clinical experts, is the evidence within the SCI literature definitive? For purposes of illustration, we assume that the answer to the question regarding tidal volume and weaning is “no.” Second, what must be assumed in order to extrapolate from the non-SCI literature and are these assumptions defensible? If the assumptions are controversial, what data would be needed to determine whether these assumptions actually hold? This is the point where the tidal volume example diverges from that of assisted cough, since it can be argued that the intervention pathway is fundamentally different in SCI as compared with, for example, ARDS (8).

More specifically, studies dating back to the 1970s have shown that healthy surgical patients under general anesthesia receiving mechanical ventilation are prone to the development of atelectasis and worsened ventilation perfusion mismatch (14,15). The development of atelectasis was worse when patients were supine, as opposed to prone, and was proportional to the duration of the anesthesia. Based on this observation, it was routine for patients receiving mechanical ventilation in Intensive Care Units (ICU) to be administered relatively large tidal volumes (10–15 cc/kg), so that almost normal ventilation (i.e., normal  $\text{CO}_2$  and pH) was maintained and atelectasis was prevented. As might be predicted, not all



**Figure 3.**

Causal pathway: High tidal volume ventilation to facilitate weaning from mechanical ventilation.

critically ill patients who require mechanical ventilation have similar lung pathologies.

For example, patients with ARDS have an excess of lung water, meaning that there is a striking decrease in the compliance of the lung that makes it necessary to use high ventilator pressures to attain “standard” tidal volumes of 10–15 cc/kg. Furthermore, ARDS is a “patchy” process, meaning that some lung areas that have relatively little damage are, consequently, easily inflated. This means inspired gas, under high pressure, is often distributed unevenly, leading potentially to overinflation of the relatively spared regions of lung. This overinflation, in animal models, leads to lung injury that is indistinguishable from ARDS.

On balance, recent RCTs in ARDS patients (9–13) show that traditional tidal volumes (10–15 cc/kg) have poorer patient outcomes, including increased mortality. Because of these trials, low tidal volume ventilation (4–6 cc/kg) is now considered to be the standard of care for patients with ARDS. Data from these trials are being applied to patients with respiratory failure from causes other than ARDS, including SCI. However, patients with SCI often have relatively normal lungs, which means physiologically they may be much more similar to “normal” subjects who are undergoing general anesthesia, and will be prone to development of atelectasis if ventilated with low tidal volumes. Because SCI patients also may develop aspiration of gastric contents or pneumonia, they are also at risk for ARDS, so perhaps a low tidal volume ventilation strategy in some SCI patients is advisable. Given this conflicted picture, one could argue that RCTs comparing low versus high tidal volume ventilation in SCI patients are indicated, and indeed, may be the only recourse.

The goal of this example was not to recommend a specific approach to weaning from mechanical ventilation but to illustrate the translation of a clinical debate into actionable items.

## DISCUSSION

Whereas the methodological debate about whether to use and how to use non-RCT data for quantitative inference shows no signs of resolution, far less controversial is the use of such data for the purpose of description and qualitative insight. It is important to emphasize that what is being proposed here is not an inferential method. For

example, we do not speculate how information pertaining to the various steps in the causal and intervention pathways, typically available at different levels of detail and from different designs, should be combined to form conclusions. We do not recommend what weight, if any, should be attached to non-RCT evidence, nor do we take a position regarding if, when, or to what degree such evidence is likely to be biased.

Instead, we have proposed a systematic and highly explicit *process* intended to achieve the following goals: (1) to describe what the natural history of the condition in question is likely to be without intervention and why, (2) to describe how the intervention in question is likely to alter this natural history and why, (3) to describe the evidence supporting each of the steps in the above pathways and to pragmatically assess its adequacy, and (4) when this evidence is less than definitive, to describe what assumptions and extrapolations are needed to proceed and to list what evidence currently exists (or will subsequently be needed) to support these assumptions and extrapolations.

Although the proposed process may at first glance seem simplistic, it is intended to use the following insights: (1) producing a fair summary of the evidence only requires a level of information which is sufficient to the task at hand, (2) using information sufficient to the task at hand helps focus the scarce resources of time and money so they can be expended for the maximum benefit, (3) determining that the people best able to decide what constitutes sufficient evidence are clinical experts, who should not decide this according to a fixed rule but, instead, on a case-by-case basis, (4) applying a systematic process should strengthen both the quality and documentation of the above decisions, (5) realizing the documentation function is particularly crucial, as it both communicates current thinking in the area under study and also provides a degree of quality assurance by helping to ensure that the above decisions are as systematic and evidence-based as possible, and (6) realizing one of the byproducts of this process is information about priorities for future research.

Our proposed process has a number of potential limitations. For example, we recognize that, in practice, the decisions necessitated by the process are by no means trivial. For example, within the topic area of SCI, patients may not be comparable either within or across studies (e.g., because of the level of injury, completeness of injury, and presence of multiple other injuries). SCI patients receive numerous interventions; therefore,

contributing either risk or benefit to a single treatment is difficult, both conceptually and practically. Sample sizes tend to be small. One must apply careful judgment when determining what assumptions, extrapolations, and interpretations are appropriate.

A more fundamental conceptual difficulty is the presence of multiple, often related, pathways. Attempting to simultaneously model these pathways would greatly complicate the presentation, while dealing with each possible pathway as if it were independent may be so clinically unrealistic as to lead to potentially inappropriate conclusions. We know of no ideal solution to this problem, but we do note that since the primary purpose of this approach is description and documentation, some degree of simplification (even if not fully consistent with clinical reality) is to be expected and, indeed, perhaps welcomed.

Another issue involves the conflicting goals of minimalism and completeness. This is perhaps most succinctly stated by the question, “In **Tables 2** and **3**, should we operate from right to left (i.e., beginning with the most valid designs) or from left to right?” The process should work either way and, in practice, the decision would probably be made best by the clinical experts, depending upon the level of evidence they deem necessary. For example, in documenting the uncontroversial hypothesis that for SCI patients, in the absence of effective intervention, large quantities of bacteria tend to develop in the lung, we could proceed from left to right. Conversely, for the more controversial question of high versus low tidal volumes for weaning from the respirator, the most appropriate approach would likely be to begin by searching for RCTs.

Similarly, in areas of controversy much more attention would be given to documentation of the empirical evidence behind conflicting conceptualizations of the intervention pathway. For example, relatively little evidence needs to be given in support of the uncontroversial mechanism by which assisted cough improves secretion clearance, thus helping to prevent pneumonia. On the other hand, high tidal volumes may or may not work differently in SCI patients as compared to ARDS patients. In such cases, individual steps may need to be broken down into substeps, and the evidence in favor of each of the substeps marshaled and judged by the clinical experts.

We also note that, even though the columns in **Tables 2** and **3** are arranged in roughly ordinal fashion, as illustrated in the tidal volume example, it is not immediately clear whether RCTs in non-SCI populations should take

preference over observational studies within the SCI literature, or vice versa. A similar question pertains to the placement of within-subject designs relative to RCTs. This is another decision probably best made by clinical experts on a case-by-case basis.

## CONCLUSION

In conclusion, even in fields with large numbers of excellent RCTs, production of a comprehensive evidence report will always require the tasks of extrapolation, assumption, and interpretation. These same tasks also apply to non-RCT designs, making differences between using RCTs and non-RCTs primarily those of degree rather than kind. What can reasonably be expected is not that definitive evidence will always be available on a given topic—indeed, such evidence is by definition illusory—but instead that what is currently known (and believed) will be recorded as explicitly as possible. What can be stated explicitly becomes testable, and this subsequent testing is what allows medical science to continue its advance.

Before accepting the task order on preventing pulmonary complications of SCI, we understood that few RCTs had been performed among patients with this condition. To a certain extent, we wondered whether a comprehensive literature review might be an exercise in futility. Or, instead, would this literature contain numerous insights, just not in the ideal format of large RCTs? Or would we find something in between?

Although the SCI literature is not definitive, we were ultimately very encouraged by the potential for drawing meaningful conclusions. In fact, much is known about the care of patients with SCI, using as the criterion for acceptability the level of evidence which experts in the field judge to be *sufficient to the task at hand*. Admittedly, reviewing a literature such as SCI does require more emphasis on various integrative tasks such as (1) determining minimal acceptable levels of evidence; (2) stating explicitly the assumptions required to interpret the available data and what information, in turn, supports the assumptions; and (3) synthesizing the resulting information into a fair summary of the strength of the evidence. Accomplishing these tasks not only accrues costs in terms of time, budget, and complexity but also gains numerous benefits in terms of understanding more intimately the clinical issues under study.

We conclude that even in the absence of extensive RCTs, useful evidence reports can be produced. One of the products of these evidence reports should be a list of the topics for which RCTs are needed urgently. In this way, we can close the loop between making the best of the current literature and strengthening the evidence base for future reviews.

## ACKNOWLEDGMENTS

This article is based on work performed by the Duke University Evidence-Based Practice Center. The authors of this article are responsible for its contents, including any clinical or treatment recommendations. No statement in this article should be construed as an official position of the Agency for Healthcare Research and Quality of the United States Department of Health and Human Services.

## REFERENCES

1. Goldberg HI. Building healthcare quality: if the future were easy, it would be here by now [commentary]. *Front Health Serv Manage* 1998;15:40–2.
2. Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. 1963. Rand McNally College Publishing Company. Chicago IL.
3. Woolf SH. *Interim manual for clinical practice guideline development*. Rockville, MD: US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1991 (May). Available from the National Technical Information Service. Agency for Health Care Policy and Research. Publication No. 91–0018.
4. Wang AY, Jaeger RJ, Yarkony GM, Turba RM. Cough in spinal cord injured patients: the relationship between motor level and peak expiratory flow. *Spinal Cord* 1997;35:299–302.
5. DeVivo MJ, Krause JS, Lammertse DP. Recent trends in mortality and causes of death among persons with spinal cord injury. *Arch Phys Med Rehabil* 1999;80:1411–419.
6. Jaeger RJ, Turba RM, Yarkony GM, Roth EJ. Cough in spinal cord injured patients: comparison of three methods to produce cough. *Arch Phys Med Rehabil* 1993;74:1358–61.
7. Balshi JD, Cantelmo NL, Menzoian JO. Complications of caval interruption by Greenfield filter in quadriplegics. *J Vasc Surg* 1989;9:558–62.
8. Peterson WP, Barbalata L, Brooks CA, Gerhart KA, Mellick DC, Whiteneck GG. The effect of tidal volumes on the time to wean persons with high tetraplegia from ventilators. *Spinal Cord* 1999;37:284–8.
9. Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med* 2000;342:1301–8.
10. Amato MBP, Barbas CSV, Medeiros DM et al. Effect of a protective ventilation strategy on mortality in the acute respiratory distress syndrome. *N Engl J Med* 1998;338:347–54.
11. Brochard L, Roudot-Thorval F, Roupie E et al. Tidal volume reduction for prevention of ventilator-induced lung injury in acute respiratory distress syndrome. *Am J Respir Crit Care Med* 1998;158:1831–8.
12. Browner RG, Shanholtz CB, Fessler HE et al. Prospective, randomized, controlled clinical trial comparing traditional versus reduced tidal volume ventilation in acute respiratory distress syndrome patients. *Crit Care Med* 1999;27:1492–8.
13. Stewart TE, Meade MO, Cook DJ et al. Evaluation of a ventilation strategy to prevent barotrauma in patients at high risk for acute respiratory distress syndrome. *N Engl J Med* 1998;338:355–61.
14. Froese AB, Bryan AC. Effects of anesthesia and paralysis on diaphragmatic mechanics in man. *Anesthesia* 1974;41:242–55.
15. Pelosi P, Croci M, Calappi E et al. The prone positioning during general anesthesia minimally affects respiratory mechanics while improving functional residual capacity and increasing oxygen tension. *Anesth Analg* 1995;80:955–60.

Submitted for publication December 19, 2000. Accepted in revised form February 3, 2001.