Nature and role of observational studies in public health policy concerning the effects of dietary salt intake on blood pressure$^{1-3}$

R Brian Haynes

ABSTRACT Does dietary sodium in excess of a specified daily consumption substantively elevate blood pressure in normotensive people? Does lowering the daily consumption of sodium reduce blood pressure in normotensive people? Sound observational and interventional studies can address these questions, but there are substantial differences in the ability of various research designs to provide clear, bias-free answers. I summarize established scientific principles for addressing issues of causation and the effects of interventions and compare observational and interventional designs. Observational studies are important in exploring the possible determinants of health problems but are subject to bias and cannot directly assess the effects of interventions. They are superseded by sound interventional studies, particularly randomized controlled trials, in answering the key questions concerning causes and benefits of intervention. Am J Clin Nutr 1997;65(suppl):622S–5S.

KEY WORDS Sodium, diet, etiology, risk, blood pressure, randomized controlled trial, cohort studies, case-control studies, surveys

INTRODUCTION

"The real purpose of the scientific method is to make sure nature hasn't misled you into thinking you know something you don't actually know" (1). Ideally, recommendations to the public about how to avoid natural threats to health should be based on convincing scientific evidence that following the recommended practice will result in appreciably more good than harm, considering the difficulty and expense of doing so. Without such evidence, there are ample lessons from history of recommendations being overturned when better evidence becomes available: promoting low weight gain during pregnancy to ease childbirth, suppressing arrhythmias after myocardial infarction, and so on. Moreover, parables remind us that yelling wolf when there is no wolf results in public inaction when a real wolf invades the community.

Recommendations to the public concerning the consumption of essential nutrients, such as salt, must be quantitative (how much or little is safe to consume) rather than qualitative (whether to consume at all, which is not an option). Quantitative questions about nutrition pose challenges for scientists because it is difficult and expensive to generate precise answers when the effects of a single essential nutrient must be considered in the context of many nutrients and their interactions. There is no simple, single, definitive research method or approach that can provide a clear answer; for some questions even the combined current methods of science may not be sufficient to provide an irrefutable measure of effect. Nevertheless, for many questions about nutrition, modern scientific techniques provide some powerful tools for determining whether individual factors can harm healthy humans and whether manipulating these factors reduces the harm. It is to be hoped that this is the situation for determining whether there are adverse effects to healthy humans of salt consumption above what is practical and essential to sustain normal body function.

There are two main, related questions to be addressed: Does dietary sodium in excess of a specified daily amount cause more harm than good? Does lowering the daily consumption below this specified quantity shift the balance of harm to good? Before proceeding to discuss the roles of observational evidence that might answer these questions, it is important to state that even if these questions are answered with strong evidence, policy recommendations will not be dictated, because evidence is only one part of any decision about recommendations (2). There ought to be at least three parts to any policy decision: evidence; circumstances, including resources and priorities; and preferences and values about what sorts of recommendations and actions are reasonable and appropriate. For example, even though it may be clear that public possession of hand guns is dangerous to human health, values about rights of individuals may outweigh the evidence. I will return to the incorporation of evidence into policy after discussing the types of evidence.

TYPES AND STRENGTHS OF EVIDENCE CONCERNING CAUSING MORE HARM THAN GOOD

Austin Bradford Hill, one of the fathers of modern medical epidemiology, proposed diagnostic tests for establishing causality:

1 From the Health Information Research Unit, Department of Clinical Epidemiology and Biostatistics, McMaster University Faculty of Health Sciences, Hamilton, Canada.

2 Supported in part by a National Health Scientist Award from the National Health Research and Development Program of Health Canada.

3 Address reprint requests to RB Haynes, Room 3H7, McMaster University Medical Center, 1200 Main Street West, Hamilton, Ontario, L8N 3Z5, Canada. E-mail: bhaynes@fhs.mcmaster.ca.
sation that provide a standard to this day (3, 4). These tests, in
descending order of importance, are as follows:

Is there evidence from true experiments in humans?
The strongest, least biased evidence about whether a sub-
stance causes harm comes from randomized controlled trials in
which one group of individuals receives a putatively harmful
dose of the substance and another randomly allocated group
does not. The participants are then followed closely to deter-
mine whether there is any relation between the adverse out-
comes of interest and the treatment groups. If the substance is
potentially highly toxic, a more ethical alternative design
would include people who are naturally exposed to high
amounts of the substance, with random allocation of willing
participants to either continue with their natural exposure or to
switch to an intervention that lowers their exposure.

Because randomized controlled trials provide the strongest
evidence of causation, all such studies should be considered
before recommendations are made; contradictions between
these trials and other types of studies should be ruled in favor
of these trials if they are free from major flaws. If no adequate
trials have been done but it is feasible to do so, trials should be
done before public recommendations are made, except in some
special circumstances. First, sometimes it is not necessary to
conduct trials because the evidence from lesser studies is
overwhelming. We do not need to do randomized controlled
trials of exposure to land mines or cyanide to determine that
these are harmful to health. Second, sometimes it is not pos-
sible to conduct trials because the exposure cannot be manip-
ulated in a controlled trial (eg, genetic pedigree). Neither of
these circumstances applies to the exposure of humans to salt.
Indeed, there are randomized controlled trials of salt exposure
( albeit only with biochemical and blood pressure outcomes), so
these considerations are academic.

Is the association strong?
The stronger the association between a putative causative
agent and a noxious outcome, the more likely that the associ-
ation is causal. Randomized controlled trials provide the least
biased estimates of strength of association. Other traditional
study designs that provide information on strength of associa-
tion, in descending order of freedom from bias, are nonran-
domized trials, cohort studies, case-comparison studies, and
surveys. Of these, the last three are observational.

In cohort studies, people initially free of the outcome of
interest ( eg, hypertension) are assessed for their exposure to the
putative causative agent (eg, usual daily intake of salt) and then
followed to determine the rate at which the outcome occurs.
Other features that could be related to risk for the outcome are
measured at baseline and adjusted for in the analysis so that the
independent effect of the exposure can be determined. Unlike
the situation for a true experiment ( ie, a randomized controlled
trial), the baseline level of exposure may be due to some
unmeasured factor that is actually the causative agent.

In case-control studies, case subjects with an outcome of
interest ( eg, hypertension) are assessed for their previous ex-
posure to the factor of interest and the same is done for a group
of control subjects who do not have the outcome of interest but
who are otherwise similar. It is often difficult to measure
exposures in the past and to be sure that control subjects are
similar to case subjects in all ways but the outcome of interest.
Thus, there are many sources of bias that can distort the
findings of such studies (5).

In a survey, people are assessed simultaneously for the
exposure and outcome of interest. Although it is also possible
to measure the strength of association in such studies, the result
is subject to more problems with bias than are found with the
study designs discussed previously. This is particularly so
when the exposure has a gentle influence that must be sustained
over a considerable period for an effect to be observed. Para-
doxically, it is also possible to miss a toxic effect in a survey
because people who have experienced an adverse effect may
not be well enough to be surveyed ( they may be dead) or may
have diminished their exposure as a result of the outcome and
therefore may appear to not have been exposed.

Is the association consistent from study to study?
A causal association is supported if different lines of study
and different investigators in different settings come to similar
conclusions. However, false positive and negative conclusions
are predictable in research ( the price paid for economy), so
complete consistency is not expected. Nevertheless, if a sub-
stantial minority of studies disagrees with the majority of
studies, this criterion cannot be satisfied, particularly if there
are studies of equal strength that disagree.

Is the temporal relation correct?
For an exposure to be blamed for causing an outcome, it
must precede the outcome. Experiments and cohort studies can
easily establish this temporal requirement. Some case-control
studies and surveys may have difficulty doing so.

Is there a dose-response gradient?
If higher exposures are associated with higher risk for out-
comes, the case for causation is strengthened. Any of the study
designs described above can contribute such information.

Does the association make epidemiologic sense?
A common source of early information about possible public
health risks comes from the juxtaposition of vital statistics and
other routinely collected data. Thus, it may be noticed that in
countries with higher dietary salt intake there is increased
prevalence of hypertension compared with countries with
lower salt intake. Aside from the measurement problems in
routinely collected data, such studies do not measure the ex-
posure and outcome in the same person at the same time, so the
findings may be coincidental.

Does the association make biological sense?
Agreement with current understanding of biological mecha-
nisms and responses of cells, tissues, and organs to stimuli
supports causal relations. This is the only yardstick that can be
applied to nonhuman data. Major limitations of biological
sense as a basis for judging causation include the innate capa-
city of the human mind to invent—and believe—explanations
of almost any set of phenomena, bolstered by a conven-
nient selection from among an often vast pool of contradictory
biological observations.
SUMMARIZING THE EVIDENCE

Because single studies almost never clearly answer questions of causation or intervention regarding human health, the available evidence must be assembled and carefully weighed. In assembling the evidence, commentators on a given issue often are selective in the evidence they cite, overvaluing evidence that supports their own view. New standards are being set for the collection and summarization of evidence through such organizations as the Cochrane Collaboration (6). Standards for comprehensive reviews of evidence appear in Table 1. These standards may seem to provide a counsel of perfection that is unreasonable to even try to achieve. However, both the ethics and sense of supporting expensive biomedical research could be questioned if the findings of the enterprise are not thoroughly considered when questions of public policy arise that have been addressed by the research.

Some strategies can reduce the work required in summarizing evidence. First, because of the marked differences in the strength of various study designs, it may be reasonable to collect and summarize only the most powerful class of studies. Second, within each study design type it may be reasonable to exclude from detailed consideration studies with major problems in execution, for example, loss of >20% of participants to follow-up.

Once the evidence is assembled, findings from research must be weighed according to their merits. A semiquantitative scale for weighing different types of evidence is provided in Table 2. The contribution from observational studies, including cohort studies, case-control studies, and surveys, is relatively less than that from true experiments. This is clearly in keeping with the standards for grading evidence established by the Canadian Task Force on the Periodic Health Examination (7). As shown in Table 3, sound observational studies would provide level 2–2 or lower evidence compared with the level 1 evidence from randomized controlled trials.

INCORPORATING EVIDENCE INTO POLICY

The decision-making model mentioned earlier has three major elements: evidence, circumstances, and the values of the decision makers or those they represent. It is beyond the scope of this article to delve into the details of circumstances and values that go into policymaking. Briefly, the circumstances that must be considered include issues of effect, feasibility, cost, and alternative demands on resources; values include consideration of both society's current ethical standards and preferences and of the vested interests of key stakeholders.

Thinking and acting strategically are important in seeking a policy. If vested interests have called for a review of the evidence, this should be acknowledged from the outset. If systematic review of the evidence shows that there is no cause for concern (in this case, that salt is a public health risk), then the rest of the policy exercise should be suspended. If the review of evidence is not convincingly systematic, considering fairly all pertinent evidence, then a systematic review should be set up before public policy is considered. If debate erupts during the consideration of evidence or policy, the debaters should be required to specify evidence supporting their position and the audience should assess whether the evidence applies and how much weight to afford it.

CONCLUSIONS ABOUT THE ROLE OF OBSERVATIONAL STUDIES IN POLICY CONCERNING DIETARY SALT INTAKE

Observational studies most appropriately lead to hypotheses that must be tested more rigorously, preferably in randomized controlled trials. In the case of dietary salt and blood pressure, these more rigorous studies have been performed and the interventional studies supersede the observational studies.

### Table 1

<table>
<thead>
<tr>
<th>Standards for systematic reviews of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Clearly stated questions for review and specification of eligibility criteria for content and nature of studies that will be considered.</td>
</tr>
<tr>
<td>2) Comprehensive search for potentially pertinent published and unpublished studies.</td>
</tr>
<tr>
<td>3) Selection of pertinent studies according to explicit criteria, with documentation of reproducibility of selection.</td>
</tr>
<tr>
<td>4) Appraisal of the scientific quality of pertinent studies with documentation of reproducibility of appraisal and justification for exclusions of any studies.</td>
</tr>
<tr>
<td>5) Extraction of quantitative data from studies with individual representation. If possible and appropriate, quantitative summary of evidence through use of statistical techniques of meta-analysis.</td>
</tr>
<tr>
<td>6) Performance of a sensitivity analysis, if appropriate and possible.</td>
</tr>
<tr>
<td>7) Detailed report of all steps and results.</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Guide to importance of individual diagnostic tests for causation: effect of test result on causal decision</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnostic test</strong></td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>RCT (in humans)</td>
</tr>
<tr>
<td>Strength of association</td>
</tr>
<tr>
<td>From RCT</td>
</tr>
<tr>
<td>From cohort study</td>
</tr>
<tr>
<td>From case-control study</td>
</tr>
<tr>
<td>From surveys</td>
</tr>
<tr>
<td>Consistency</td>
</tr>
<tr>
<td>Temporality</td>
</tr>
<tr>
<td>Gradient</td>
</tr>
<tr>
<td>Epidemiologic sense</td>
</tr>
<tr>
<td>Biological sense</td>
</tr>
</tbody>
</table>

* +, Causation supported; 0, causal association not affected; -, causation rejected. RCT, randomized controlled trial. Modified from reference 4.

* Listed in decreasing order of importance.

### Table 3

<table>
<thead>
<tr>
<th>Grading the contributions of evidence from observational studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of evidence (and type of observational study)</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>1 Evidence from randomized controlled trials</td>
</tr>
<tr>
<td>2-1 Evidence from nonrandomized controlled trials</td>
</tr>
<tr>
<td>2-2 Evidence from cohort or case-control studies (cohort study, case-control study)</td>
</tr>
<tr>
<td>2-3 Evidence from comparisons between times or places with or without intervention (surveys)</td>
</tr>
<tr>
<td>3 Opinions of authorities, based on clinical experience, descriptive studies, or reports of expert committees</td>
</tr>
</tbody>
</table>

* From Goldblum et al (7).
Observational studies should form the foundation of any policy concerning the nature of the effects of salt on blood pressure, the merit of preemptive interventions, and the advice that should be offered to nonhypertensive citizens to alter dietary salt intake. If those studies show no substantive effects on blood pressure, then the policy debate should be ended. If there is an effect on blood pressure, larger studies would perhaps be needed to show whether the effects are sustained and lead to important clinical consequences.

REFERENCES