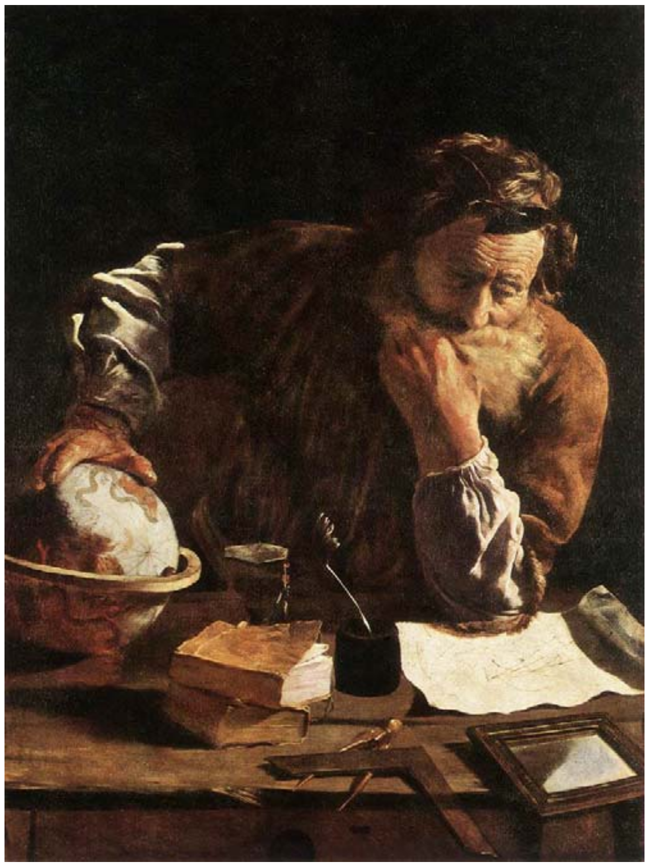
www.vtpi.org
Info@vtpi.org
250-360-1560

Evaluating Research Quality

Guidelines for Scholarship

28 November 2010

By Todd Litman
Victoria Transport Policy Institute



Portrait of a Scholar
by Domenico Feti, Italian painter (b. ca. 1589, Roma, d. 1623, Venezia)
Oil on canvas, 98 x 73,5 cm, Gemäldegalerie, Dresden

Abstract

This paper discusses the importance of good research, discusses common causes of bias, provides guidelines for evaluating research and data quality, and describes examples of bad research.

A version of this paper was presented at the *International Electronic Symposium on Knowledge Communication and Peer Reviewing*, International Institute of Informatics and Systemics (www.iiis.org), 2006.

"Everyone is entitled to his own opinion, but not his own facts."
-attributed to Senator Patrick Moynihan

Introduction

Research (or *scholarship*) investigates ideas and uncovers useful knowledge. It is personally rewarding and socially beneficial. But research can be abused. *Propaganda* (information intended to persuade) is sometimes misrepresented as objective research. This is disturbing to legitimate scholars and harmful to people who unknowingly use such information. It is therefore helpful to have guidelines for evaluating research quality.

Some people have few qualms about manipulating research. They consider it a game, assuming that all sides abuse information equally, or that a desired outcome justifies misrepresentation. But distorted research causes real harm and deserves strong censure.

Good research reflects a sincere desire to determine what is overall true, based on available information; as opposed to bad research that starts with a conclusion and only presents supporting *factoids* (individual facts taken out of context). A good research document empowers readers to reach their own conclusions by including:

- A well-defined question.
- Description of the context and existing information about an issue.
- Consideration of various perspectives.
- Presentation of evidence, with data and analysis in a format that can be replicated by others.
- Discussion of critical assumptions, contrary findings, and alternative interpretations.
- Cautious conclusions and discussion of their implications.
- Adequate references, including original sources, alternative perspectives, and criticism.

Good research requires judgment (or *discernment*) and honesty. It carefully evaluates information sources. It acknowledges possible errors, limitations and contradictory evidence. It identifies excluded factors that may be important. It describes key decisions researchers faced when structuring their analysis and explains the choices made. For example, if various data sets are available, or impacts can be measured in several ways, the different options are discussed. Sometime, multiple analyses are performed using alternative approaches and their results compared.

Good research is cautious about drawing conclusions, careful to identify uncertainties and avoids exaggerated claims. It demands multiple types of evidence to reach a conclusion. It does not assume that *association* (things occur together) proves *causation* (one thing causes another). Bad research often contains jumps in logic, spurious arguments, and *non-sequiturs* ("it does not follow").

Bad research often uses accurate data, but manipulates and misrepresents the information to support a particular conclusion. Questions can be defined, statistics selected and analysis structured to reach a desired outcome. Alternative perspectives and data can be ignored or distorted. Critics of an idea sometimes exaggerate issues of uncertainty. They imply because we don't know everything about an issue, we know nothing about it.

A good research document provides a comprehensive overview of an issue and discusses its context. This can be done by referencing books and websites with suitable background information. This is especially important for documents that may be read by a general audience, which includes just about anything that will be posted on the Internet.

Good research may use *anecdotal evidence* (examples selected to illustrate a concept), but does not rely on them to draw conclusions because examples can be found that prove almost anything. More statistically-valid analysis is usually needed for reliable proof.

Peer review (critical assessment by qualified experts, preferably blind so reviewers and authors do not know each others' identify) enhances research quality. This does not mean that only peer reviewed documents are useful (much information is distributed in working papers or reports), or that everything published in professional journals is correct (many published ideas are proven false), but this process encourages open debate about issues.

Consider researchers' ideological and financial interests when evaluating their analysis. Proponents of a perspective may provide *asymmetrical* (one-sided) information, offering evidence that supports their conclusions while ignoring or suppressing other information. Their conclusions are not necessarily false, much legitimate research is supported interest groups, but it is important to evaluate such analysis critically and investigate other information sources that may provide alternative perspectives and results.

Research quality is an *epistemological* issue (related to the study of knowledge). It is important to librarians (who manage information resources), scientists and analysts (who create reliable information), decision-makers (who apply information), jurists (who judge people on evidence) and journalists (who disseminate information to a broad audience). These fields have professional guidance to help maintain quality research. This has become increasingly important as the Internet makes unfiltered information more easily available to a general audience. Guidelines for good research are provided below. Additional information is available from references cited at the end of this paper.

On Bullshit

In his best-selling book, *On Bullshit* (Princeton Press 2005), philosopher Harry G. Frankfort argues that *bullshit* (manipulative misrepresentations) is worse than an actual lie because it denies the value of truth. "A bullshitter's fakery consists not in misrepresenting a state of affairs but in concealing his own indifference to the truth of what he says. The liar, by contrast, is concerned with the truth, in a perverse sort of fashion: he wants to lead us away from it." Truthtellers and liars are playing opposite sides of a game, but bullshitters take pride in ignoring the rules of the game altogether, which is more dangerous because it denies the value of truth and the harm resulting from dishonesty.

People sometimes try to justify their bullshit by citing *relativism*, a philosophy which suggests that objective truth does not exist (Nietzsche stated, "There are no facts, only interpretations"). An issue can certainly be viewed from multiple perspectives, but anybody who claims that justifies misrepresenting information or denies the value of truth and objective analysis is really bullshitting.

"The greatest sin is judgment without knowledge" – Kelsey Grammer

Research Document Evaluation Guidelines

The guidelines below are intended to help evaluate the quality of research reports and articles.

Desirable Practices

1. Attempts to fairly present all perspectives.
2. Provides context information suitable for the intended audience. This can be done with a literature review that summarizes current knowledge, or by referencing relevant documents or websites that offer a comprehensive and balanced overview.
3. Carefully defines research questions and their links to broader issues.
4. Provides data and analysis in a format that can be accessed and replicated by others. Quantitative data should be presented in tables and graphs, and available in database or spreadsheet form on request.
5. Discusses critical assumptions made in the analysis, such as why a particular data set or analysis method is used or rejected. Indicates how results change with different data and analysis. Identifies contrary findings.
6. Presents results in ways that highlight critical findings. Graphs and examples are particularly helpful for this.
7. Discusses the logical links between research results, conclusions and implications. Discusses alternative interpretations, including those with which the researcher disagrees.
8. Describes analysis limitations and cautions. Does not exaggerate implications.
9. Is respectful to people with other perspectives.
10. Provides adequate references.
11. Indicates funding sources, particularly any that may benefit from research results.

Undesirable Practices

1. Issues are defined in ideological terms. "Straw men" reflecting exaggerated or extreme perspectives are use to characterize a debate.
2. Research questions are designed to reach a particular conclusion.
3. Alternative perspectives or contrary findings are ignored or suppressed.
4. Data and analysis methods are biased.
5. Conclusions are based on faulty logic.
6. Limitations of analysis are ignored and the implications of results are exaggerated.
7. Key data and analysis details are unavailable for review by others.
8. Researchers are unqualified and unfamiliar with specialized issues.
9. People with differing perspectives are insulted and ridiculed.
10. Citations are primarily from special interest groups or popular media, rather than from peer reviewed professional and academic organizations.

Making Sense of Information (www.planetizen.com/node/40408)

Professor Ann Forsyth offers the following guidelines to insure that referenced information is true to the content of the sources and allows readers to make independent judgments about the strength of evidence provided:

- *A work that only contains sources available on the internet is likely to give the reader the impression that a writer was not very energetic in his or her investigations.* Planning work often involves looking at physical sites, talking with people, examining historical evidence, using databases, and even understanding technical issues that are documented in reports that don't make their way onto the public internet. Students typically have free access to a large number of such technical documents such as journal articles, historical sources such as historical maps, and expensive databases such as business listings—they should use them.

- *Writers need sources for everything that is not common knowledge* to readers or that is obviously the writer's own opinion. It is not enough to say you found something in multiple places. You need to specifically cite those places.

- *Sources are needed for both the conceptual framework of the piece* (e.g. levels of public space in squatter settlements, types of planning responses to disasters) *and for the facts and figures you use to support your argument.* Sources are also typically needed for the methods you use to show that you are building on earlier work, even if modifying it in some way.

- *Use sources critically* in a way that respects the reader's needs to be able to judge evidence for herself or himself. Weave material about the source into the text: "According to the XYZ housing advocacy organization...", "based on 150 interviews with clients of CDCs…", "reflecting 10 years of experience working with Russian immigrants". Saying "Harvard professor X claims that…." is not a strong source of evidence. Harvard professors have personal opinions. Readers typically deserve to be told about the evidence.

- *Not all sources are equal***.** Better sources are published by reputable presses (e.g. University Presses), are refereed (blind reviewed articles), or are by reputable organizations. They cite sources and are clear about methods so readers can check their facts (see Booth et al. 2008, *The Craft of Research*, 77-79, for a terrific explanation of this point). Better sources use better methods overall. Of course a writer's own analysis can be a source and they should say that is the case and show, even if briefly, how they did the analysis and why their methods are strong.

- *Wikipedia, ask.com and other similar web sites are typically not appropriate final sources***.** It is possible to start at Wikipedia but scroll straight to the bottom and look at its sources--they are often very useful. For many questions it is better still to use a professional or scholarly dictionary such as dictionaries of geography or of planning terms. The 2001 International Encyclopedia of the Social & Behavioral Sciences (N. Smelser and P. Bates eds.) has substantial sections on planning and urban studies and many university libraries provide free access.

- *One source is frequently not enough*, particularly for controversial or complicated issues. Better writers use multiple sources to allow the reader to see the balance of evidence.

Guidelines For Living With Information (Harris 1997)

These general guidelines are designed to help readers critically evaluate information, particularly from the Internet.

- *Challenge* **-** Challenge information and demand accountability. Stand right up to the information and ask questions. Who says so? Why do they say so? Why was this information created? Why should I believe it? Why should I trust this source? How is it known to be true? Is it the whole truth? Is the argument reasonable? Who supports it?
- *Adapt* **-** Adapt your skepticism and requirements for quality to fit the importance of the information and what is being claimed. Require more credibility and evidence for stronger claims. You are right to be a little skeptical of dramatic information or information that conflicts with commonly accepted ideas. The new information may be true, but you should require a robust amount of evidence from highly credible sources.
- *File* **-** File new information in your mind rather than immediately believing or disbelieving it. Do not jump to a conclusion or come to a decision too quickly. It is fine simply to remember that someone claims XYZ to be the case. Wait until more information comes in, you have time to think about the issue, and you gain more general knowledge.
- *Evaluate* **-** Evaluate and re-evaluate regularly. New information or changing circumstances will affect the accuracy and hence your evaluation of previous information. Recognize the dynamic, fluid nature of information. The saying, “Change is the only constant,” applies to much information, especially in technology, science, medicine, and business.

Association Does Not Prove Causation
A common mistake of bad research is to assume or imply that *association* (two things tend to occur together) proves *causation* (one thing causes or influences another). Below are examples.

- Many people die in hospitals, and there are occasional examples of patients harmed during visits (due to medical care errors or hospital-based infections), so a bad researcher could “prove” that hospitals are dangerous. However, this confuses *causation* (people often go to hospitals when they are at risk of dying), and provides no *base case* (what would happen to those people had they not gone to a hospital) for comparison. It is likely that hospitals significantly reduce death rates compared with what would otherwise occur, despite many examples to the contrary.
- Many dense urban neighborhoods have higher crime and mental illness rates than lower-density suburbs, so people sometimes assume that density causes social problems. But these problems actually reflect poverty and isolation. There is no evidence that for a given demographic group, shifting from lower- to higher-density housing causes social problems (1000 Friends, 1999). It would be more appropriate to conclude that urban social problems are caused by middle-class flight and suburban communities’ exclusionary policies that cause disadvantaged people to concentrate in city neighborhoods.

Detailed Internet Information Evaluation Criteria

Table 1 lists various factors to consider when evaluating the quality of information, particularly from the Internet.

Table 1 **Evaluation Criteria** (Internet Navigator)

Accuracy or credibility	Is the information provided based on proven facts? Is it published in a scholarly or peer-reviewed publication? Have you found similar information in a scholarly or peer-reviewed publication?
Author or authority	Who is the author? Is she or he affiliated with a reputable university or organization? What is the author's educational background or experience? What is their area of expertise? Has the author published in scholarly or peer reviewed publications? Does the author/Web master provide contact information?
Coverage or relevance	Does the information covered meet your information needs? Is the coverage basic or comprehensive? Is there an "About Us" link that explains subject coverage? How relevant is it to your research interests?
Currency	When was the information published? When was the Web site was last updated. Is timeliness important to your information need?
Objectivity or bias	How objective or biased is the information? What do you know about who is publishing this information? Is there a political, social or commercial agenda? Does the information try to inform or persuade? How balanced is the presentation on opposing perspectives? What is the tone of language used (angry, sarcastic, balanced, educated)?
Sources or documentation	Is there a list of references or works cited? Is there a bibliography? Is there information provided to support statements of fact? Can you contact the author or Web master to ask for, and receive, the sources used?
Publication and Web site design	How well designed is the Web site? Is the information clearly focused? How easy to use is the information? How easy is it to find information within the publication or Web site? Are the bibliographic references and links accurate, current, credible and relevant? Are the contact addresses for the author(s) and Web master(s) available from the site?

Use these questions to critically evaluate print and Web based information.

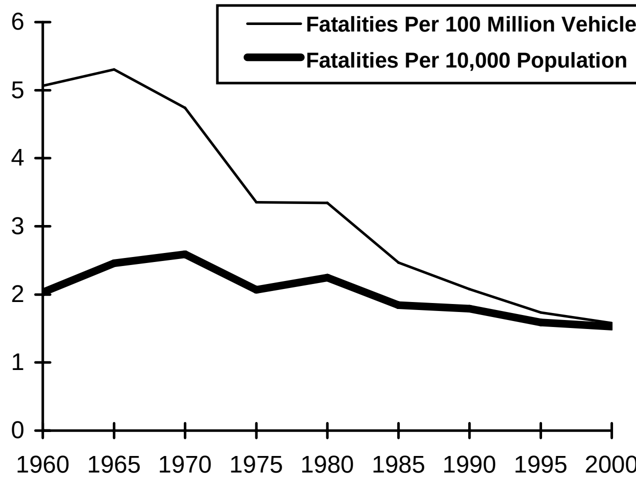
Be particularly skeptical about organizations that misrepresent themselves. For example, the *Sport Utility Vehicle Users Association* is an industry-funded organization established to oppose new vehicle safety and environmental regulations. The *Independent Commission on Environmental Education* is an industry-funded organization established to criticize environmental education. The Competitive Enterprise Institute's *Center for Environmental Education Research* was established to downplay environmental risks such as acid rain and global warming. Such organizations frequently criticize others for providing biased information, yet often do this themselves (Smith, 2000).

Reference Units

Reference units are measurement units normalized to help compare impacts (Litman 2003). Common transportation reference units include per capita, per passenger-mile, and per vehicle-mile. The selection of reference units can affect how problems are defined and solutions selected. For example, if traffic fatality rates are measured per vehicle-mile, traffic risk seems to have declined significantly during the last four decades, suggesting that current safety programs are effective (Figure 1). But measured per capita, as with other health risks, traffic fatality rates have hardly declined during this period despite the implementation of many safety strategies. When viewed in this way, traffic safety programs have failed and other approaches are justified.

This occurred because reduced crash rates per vehicle-mile have been largely offset by increased per capita mileage, and as vehicles feel safer drivers tend to drive more miles and take small additional risks, such as driving slightly faster, leaving less distance between vehicles in traffic, and driving under slightly more hazardous conditions. Measuring crash rates per vehicle-mile implies that automobile travel becomes safer if motorists drive more lower-risk miles (for example, if mileage on grade-separated highways increases), even if this increases per capita traffic deaths.

Figure 1 **U.S. Traffic Fatalities** (Litman and Fitzroy 2005)



Traffic fatality rates declined significantly if measured per vehicle-mile, but not if measured per capita.

There is often no single right or wrong reference unit to use for a particular analysis. Different units reflect different perspectives. However, it is important to consider how reference unit selection may affect analysis results, and it may be useful to perform analysis using various reference units. For example, if somebody claims that vehicle pollution emissions declined 95% in the last few decades, it is useful to inquire which types of emissions (CO, VOCs, NOx, particulates, etc.), how this is measured (per vehicle-mile, per-vehicle, per capita, etc.), and whether this reflects new vehicles, fleet average emissions, optimal driving conditions, average driving conditions, or some another combination of vehicles and conditions.

Good Examples of Bad Research

This section summarizes some examples of poor quality research.

The Dangers of Bread (www.geoffmetcalf.com/bread.html)

The following is a humorous example of how legitimate-sounding statistics can be applied with false logic to support absurd arguments.

A recent headline read, "Smell of baked bread may be health hazard." The article described the dangers of harmful air emissions from baking bread. I was horrified. When are we going to do something about bread-induced pollution? Sure, we attack tobacco companies, but when is the government going to go after Big Bread? Well, I've done a little research, and what I've discovered should make anyone think twice...

1. More than 98% of convicted felons are bread eaters.
2. Fully *half* of all children who grow up in bread-consuming households score below average on standardized tests.
3. In the 18th century, when virtually all bread was baked in the home, the average life expectancy was less than 50 years; infant mortality rates were unacceptably high; many women died in childbirth; and diseases such as typhoid, yellow fever and influenza ravaged whole nations.
4. More than 90% of violent crimes are committed within 24 hours of eating bread.
5. Bread is made from a substance called "dough." It has been proven that as little as one pound of dough can suffocate a mouse. The average American eats more bread than that in one month!
6. Primitive tribal societies that have no bread exhibit a low occurrence of cancer, Alzheimer's, Parkinson's disease and osteoporosis.
7. Bread has been proven to be addictive. Subjects deprived of bread and given only water to eat begged for bread after only two days.
8. Bread is often a "gateway" food item, leading the user to "harder" items such as butter, jelly, peanut butter and even cold cuts.
9. Bread has been proven to absorb water. Since the human body is more than 90% water, consuming bread may lead to dangerous dehydration.
10. Newborn babies can choke on bread.
11. Bread is baked at temperatures as high as 400 degrees Fahrenheit! That kind of heat can kill an adult in less than one minute.
12. Bread baking produces dangerous air pollution, including particulates (flour dust) and VOCs (ethanol).
13. Most American bread eaters are utterly unable to distinguish between significant scientific fact and meaningless statistical babbling.

Similarly, the *Dihydrogen Monoxide Research Division* (www.dhmo.org) explores the risks presented by Dihydrogen Monoxide (water), demonstrates its association with many illnesses and accidents, and describes a conspiracy by government agencies to cover up these risks.

They Say…

They say all sorts of things. For example, *they* say that Eskimos (Inuit) have 23 words for snow, although they never offer a specific citation from an Eskimo dictionary (for discussion see www.derose.net/steve/guides/snowwords). They say that you lose 80% of your body heat from your head (possibly for well-dressed, bare-headed people, and even that is probably an exaggeration), and that you shouldn't go into the water for an hour after eating (probably good advice for swimming in rough conditions after a heavy meal, but not for relaxing in the water after a snack), and that people have more accidents when the moon is full (which probably only applies to people who believe this myth). They say that "you only live once," but they also say that people are reincarnated.

Much of what "they" say may have some factual basis, but people invoke "them" to validate ideas without bothering to define the issues or test their validity. Just because a concept is frequently repeated does not mean it should be accepted without question.

*Cold Reading Tricks **(www.ianrowland.com)***

Magician Ian Rowland (1998) describes various ways that unverifiable, contradictory and ambiguous language is often used by psychics, astrologers and other tricksters to impress audiences with knowledge about a person they just met (called "cold reading"). To an unskeptical audience, such tricks can give the impression of real knowledge and insight.

For example, the *Rainbow Ruse* is a statement which credits a person with both a trait and its opposite ("I would say that on the whole you can be a rather modest person but when the circumstances are right you are the life of the party"). The *Fuzzy Fact* involves a statement that leaves plenty of scope to be developed into something more specific ("I can see a connection with Europe, possibly Britain, or it could be the warmer, Mediterranean part?"). Using the *Vanishing Negative*, Rowland will ask his subject a question such as, "You don't work with children, do you?" If the answer is, "No, I don't" his reply is, "No, I thought not. That's not really your role." If the subject answers, "I do, actually" his reply is, "Yes, I thought so."

Dissident AIDS Research (www.rethinking.org)

The *Rethinking AIDS Society* (www.rethinking.org) is convinced that AIDS is not an infectious disease, that HIV is at worst a harmless passenger virus, and that most AIDS deaths result from anti-viral drugs given to patients. Their research consists of:

- Showing an association between people who take anti-viral drugs and death from AIDS. It is not surprising that in some cases these drugs fail. However, such analysis does not indicate the number of deaths that would have occurred without the drugs.
- Quotes taken out of context concerning the problems associated with anti-viral drugs.
- Highlighting research ambiguities, much of which is outdated, and ignoring evidence of drug treatment success.

Rail Transit Not Cost Effective (www.vtpi.org/railcrit.pdf)

Reports by O'Toole (2004 and 2005) claim that that rail transit systems fail to achieve their objectives and are not cost effective. However, the analysis is biased (Litman 2004). Below are some of the major errors in O'Toole's reports:

- Lack of with-and-without analysis. There is virtually no comparison between cities with rail and those without, or comparisons with national trends. It is therefore impossible to identify rail transit impacts.
- Failing to differentiate between cities with relatively large, well-established rail systems and those with smaller and newer systems that cannot be expected to have significant impacts on regional transportation performance (total transit ridership, congestion, etc.).
- Failing to account for additional factors that affect transportation and urban development conditions, such as city size, changes in population and employment.
- Ignoring significant costs. Vehicle expenses are included when calculating transit costs, but vehicle and parking expenses are ignored when calculating automobile costs.
- Exaggerating transit development costs. Claims, such as "Regions that emphasize rail transit typically spend 30 to 80 percent of their transportation capital budgets on transit" are unverified and generally only true for a few regions and years.
- Presenting data and examples that are many years or even decades old as current.
- Ignoring other benefits of rail transit, such as parking cost savings, consumer cost savings and increased property values in areas with rail transit systems.
- Failing to reference documents that reflect current best practices in transit evaluation, or that provide alternative perspectives.

Smart Growth Savings (www.vtpi.org/sg_save.pdf)

Various studies show that Smart Growth can reduce public service and infrastructure costs. A study by Cox and Utt (2004) claims that such savings are insignificant, but it contains the following research errors (Litman 2005):

- It incorrectly defines Smart Growth as simply increased density or slower growth. In fact, Smart Growth refers to a number of development factors, including land use clustering, mix, roadway connectivity and multi-modalism.
- It measures density at a municipal scale, which is too large to reflect Smart Growth.
- It only compares differences between municipalities, ignoring differences between development within and outside of municipal boundaries, and between conventional and clustered development within municipal boundaries.
- It only considers a small portion of total costs (municipal, water and sewage expenditures), ignoring other savings resulting from more accessible land use patterns.
- It ignored costs of services provided directly by households in lower-density areas, such as well water, septic systems and garbage disposal.
- It ignores differences in service quality.
- It treats higher municipal employee wage in higher-density cities as a cost and an inefficiency, ignoring differences in average overall wages in such areas.

References and Information Resources

1000 Friends of Oregon (1999), "The Debate Over Density: Do Four-Plexes Cause Cannibalism" *Landmark*, 1000 Friends of Oregon (www.friends.org), Winter 1999.

AEA (2004), *Guiding Principles for Evaluators*, American Evaluation Association (www.eval.org); at www.eval.org/Publications/GuidingPrinciples.asp.

Susan Beck (2004), *The Good, The Bad & The Ugly: or, Why It's a Good Idea to Evaluate Web Sources*, New Mexico State University Library (http://lib.nmsu.edu/instruction/evalcrit.html).

Simon Blackburn (2005), *Truth: A Guide for the Perplexed*, Oxford Univ. Press (www.oup.co.uk).

BSAI (2001), *The Virtual Chase: Evaluating The Quality Of Information On The Internet*, Ballard, Spahr, Andres & Ingersoll (www.virtualchase.com/quality).

Wendell Cox and Joshua Utt (2004), *The Costs of Sprawl Reconsidered: What the Data Really Show*, Backgrounder 1770, The Heritage Foundation (www.heritage.org).

CUDLRC (2009); *A Guide To Doing Research Online*, Cornell University Digital Literary Resource Center (www.digitalliteracy.cornell.edu).

Dihydrogen Monoxide Research Division (www.dhmo.org).

Ann Forsyth (2009) *A Guide For Students Preparing Written Theses, Research Papers, Or Planning Projects*, www.annforsyth.net/forstudents.html.

Harry G. Frankfurt (2005), *On Bullshit*, Princeton University Press (www.pupress.princeton.edu).

Peter Gleick (2008), *Hummer vs. Prius 2008 Redux*, Pacific Institute (www.pacinst.org); at www.pacinst.org/topics/integrity_of_science/case_studies/hummer_vs_prius_redux.pdf.

Robert Harris (1997), *Evaluating Internet Research Sources*, VirtualSalt (www.virtualsalt.com/evalu8it.htm).

David Huron (2000), *Sixty Methodological Potholes*, Ohio State University (http://dactyl.som.ohio-state.edu/Music829C/methodological.potholes.html); see appendix below.

Internet Navigator (2001), *Critically Evaluating Information*, Information Navigator (http://medstat.med.utah.edu/navigator/module3/evaluation.htm).

Todd Litman (2003), "Measuring Transportation: Traffic, Mobility and Accessibility," *ITE Journal* (www.ite.org), Vol. 73, No. 10, October 2003, pp. 28-32; at www.vtpi.org/measure.pdf.

Todd Litman (2005), *Evaluating Rail Transit Criticism*, Victoria Transport Policy Institute (www.vtpi.org); at www.vtpi.org/railcrit.pdf.

Todd Litman (2005), *Evaluating Criticism of Smart Growth*, VTPI (www.vtpi.org); at www.vtpi.org/sgcritics.pdf.

Todd Litman and Steven Fitzroy (2005), *Safe Travels: Evaluating Mobility Management Traffic Safety Impacts*, VTPI (www.vtpi.org).

Ian Rowland (1998), *The Full Facts Book of Cold Reading*, Ian Rowland Publishing (www.ianrowland.com).

Randal O'Toole (2004 and 2005), *Great Rail Disasters; The Impact Of Rail Transit On Urban Livability*, Reason Public Policy Institute (www.rppi.org).

Quack Watch (www.quackwatch.com), is a guide to quackery, health fraud and intelligent decisionmaking.

Skeptic Society (www.skeptic.com) applies critical thinking to scientific and popular issues ranging from UFOs and paranormal to the evidence of evolution and unusual medical claims.

Alastair Smith (2003), *Evaluation of Information Sources*, Information Quality WWW Virtual Library (www.ciolek.com/WWWVL-InfoQuality.html).

Gregory A. Smith (2000), *Defusing Environmental Education: An Evaluation Of The Critique Of The Environmental Education Movement*, Center for Education Research, University of Wisconsin-Milwaukee (www.asu.edu/educ/epsl/EPRU/documents/cerai-00-11.htm).

Wikipedia, *Logical Fallacies* (http://en.wikipedia.org/wiki/Category:Logical_fallacies). This website lists and categorizes various analytic errors.

Appendix 1 **Sixty-Four Methodological Potholes** (Based on Huron 2000)

	Problem	Remedy/Advice
Ad hominem argument	Criticizing the person rather than their argument.	Focus on the quality of arguments. Be prepared to learn from people you dislike. Avoid personalizing debates.
Know-nothing	Implies that because some issues are unknown, nothing is known.	Clearly identify what is known and unknown, and the scope of uncertainty.
Discovery fallacy	Criticizing an idea because of its origin (for example, from a religious text).	Criticize the justifications offered in support of an idea rather than how the idea originated.
Ipse dixit	Appealing to authority figures in support of an argument.	Cite published research rather than just “authorities.” Learn to judge research quality.
Ad baculum	Using physical or psychological threats.	Do not threaten.
Egocentric bias	The tendency to assume that other people experience things the same way we do.	Listen carefully to other people. Be cautious generalizing from personal experiences.
Cultural bias	The inappropriate application of a concept to people from another culture.	Look for cultural biases. Perform cross-cultural experiments.
Cultural ignorance	The failure to make a distinction that people in another culture readily make.	Talk with culturally knowledgeable people. Listen carefully in post-experiment debriefings.
Over-generalization	Assuming that a research result generalizes to a wide variety of real-world situations.	Be cautious interpreting results. Investigate other sources. Perform more experiments.
Inertia fallacy	Assumption that evidence supporting a conclusion will grow in the future (e.g., “Research increasingly shows that...”).	Describe research results in the past tense (“Research has shown ...”). Avoid claiming trends or when describing evidence.
Relativist fallacy	The belief that no idea, hypothesis, theory or belief is better than another.	Avoid “absolute” relativism. Don’t mistake relativism for pluralism (considering multiple perspectives).
Universalist phobia	A prejudice against the possibility of cross-cultural universals.	Learn about other cultures. Use cross-cultural surveys or experiments where appropriate.
Problem of induction	The problem that no number of particular observations can prove a general conclusion.	Avoid claiming you know the truth. Present research results as “consistent” or “inconsistent” with a particular theory or hypothesis.
Positivist fallacy	Something deemed not to exist because evidence is available: “Absence of evidence interpreted as evidence of absence.”	Recognize that not all phenomena leave obvious evidence of their existence.
Confirmation bias	The tendency to see events as confirming a theory while viewing falsifying events as “exceptions”.	Be systematic in observations. Do not change the counting or selection criteria to exclude contradicting instances.
Hindsight bias	The ease with which people confidently interpret or explain any set of existing data.	Try to predict observations in advance. Aim to test ideas rather than to look for confirmation.
Unfalsifiable hypothesis	The formulation of a theory, hypothesis or interpretation which cannot be falsified.	Formulate falsifiable theories and interpretations. Identify observations inconsistent with expectations.

	Problem	Remedy/Advice
Smorgasbord thinking	Having enough hypotheses to explain all possibilities.	Do not assume just one prediction. Ask whether you have alternative explanations if data show a reverse trend.
Ad-hoc hypothesis	Proposing a supplementary hypothesis to explain why a favorite theory or interpretation failed a test.	Open to grave abuse. Try to avoid. Test the ad hoc hypothesis in a separate follow-up study.
Sensitivity syndrome	Attempting to interpret every perturbation in a data set; failing to recognize that data contains "noise".	Use test-retest and other techniques to estimate data margin of error. Report chance levels, p values, effect sizes. Beware of hindsight bias.
Positive results bias	Tendency to only publish studies with positive results (data and theory agree).	Seek replications for suspect phenomena. Be aware of possible "bottom-drawer effect".
Bottom-drawer effect	Unawareness of unpublished negative results of earlier studies; reflecting positive results bias.	Ask other scholars whether they have performed a given analysis, survey or experiment. Widely report negative results.
Head-in-the-sand syndrome	Failure to test important theories, assumptions, or hypotheses that are readily testable.	Be willing to test ideas everyone presumes are true. Ignore criticisms that you are merely confirming the obvious. Collect pertinent data.
Data neglect	Tendency to ignore available information when assessing theories or hypotheses.	Don't ignore existing resources. Test your hypotheses using other available data sets.
Research hoarding	The failure to make the fruits of your scholarship available to others.	Publish often. Write short research articles rather than books. Make data available to others.
Double-use data	Using a single data set both to formulate and to "independently" test a theory.	Avoid. Collect new data.
Skills neglect	The human disposition to resist learning new methods that may be pertinent to research.	Engage in continuing education to fill in knowledge gaps.
Control failure	Failure to contrast experimental group with a control group.	Add a control group.
Third variable problem	Presumption that two correlated variables are causally linked; overlooking a third variable.	Avoid interpreting correlation as causality. Carry out an experiment where manipulating variables can test notions of probable causality.
Reification	Falsely concretizing an abstract concept.	Take care with terminology.
Validity problem	When a variable's operational definition fails to accurately reflect its true theoretical meaning.	Think carefully when forming operational definitions. Use more than one operational definition. Seek converging evidence.
Anti-operationalizing	The tendency to raise perpetual objections to all operational definitions.	Propose better operational definitions. Seek converging evidence using several alternative operational definitions.
Ecological validity	Problem generalizing controlled experiment results to real-world contexts.	Seek converging evidence between controlled and real-world experiments.
Naturalist fallacy	The belief that what *is* is what *ought* to be.	Imagine desirable alternatives.

	Problem	Remedy/Advice
Presumptive representation	The practice of representing others to themselves.	Be cautious when portraying or summarizing others' views, especially disadvantaged groups.
Exclusion problem	Tendency to prematurely exclude competing views.	Remember that "no theory is every truly dead." (Popper)
Post-hoc hypothesis	Following data collection, formulation and testing of additional hypotheses not previously envisaged.	Limit. Beware of hindsight bias and multiple tests. Collect new data; analyze additional works.
Contradiction blindness	The failure to take contradictions seriously.	Attend to possible contradictions.
Multiple tests	If a statistical test relies on a 0.05 confidence level, spurious results will occur each 20 tests performed.	Avoid excessive numbers of tests for a given data set. Use statistical techniques to compensate for multiple tests.
Overfitting	Excessive fine-tuning of a hypotheses or theory to one particular data set or group of observations.	Recognize that samples or observations typically differ in detail. In forming theories, continue to collect new data sets and observations.
Magnitude blindness	Preoccupation with statistically significant results that have small magnitude effects.	Aim to uncover the most important factors influencing a phenomenon first.
Regression artifacts	The tendency to interpret regression toward the mean as an experimental phenomenon.	Don't use extreme values as sampling criterion. Compare control group with an experimental group.
Range restriction effect	Failure to vary independent variables over sufficient range, so effects look small.	Decide what range of a variable or what effect size is of interest. Run a pilot study.
Ceiling effect	When a task is so easy that the experimental manipulation shows little/no effect.	Make the task more difficult. Run a pilot study.
Floor effect	When a task is so difficult that experimental manipulation shows little/no effect.	Make the task easier. Run a pilot study.
Sampling bias	Any confound that causes the sample to be unrepresentative of the pertinent population.	Use random sampling. If sub-groups are identifiable use a stratified random sample. Avoid "convenience" or haphazard sampling.
Homogeneity bias	Failure to recognize that sample sub-groups respond differently, such as between males and females.	Use descriptive methods and data exploration methods to examine the experimental results. Use cluster analysis methods where appropriate.
Cohort bias or cohort effect	Differences between age groups in a cross-sectional study due to generational differences rather than experimental factors.	Use a more narrow range of ages. Use a longitudinal design instead of a cross-sectional design.
Expectancy effect	Conscious or unconscious cues that convey to the subject the experimenter's desired response.	Use standardized interactions with subjects. Use automated data-gathering methods. Use double-blind protocol.
Placebo effect	Positive or negative response arising from the subject's expectations of an effect.	Use a placebo control group.
Demand characteristics	Any aspect of an experiment that might inform subjects of the purpose of the study.	Control experimental conditions using deception or field observation. Prevent subjects from learning about experimental conditions.

	Problem	Remedy/Advice
History effect	Any change between a pretest measure and posttest measure not attributable to the experimental factors.	Isolate subjects from external information. Use post-experiment debriefing to identify possible confounds.
Maturation confounds	Changes in responses due to factors not related to experiment, such as boredom, fatigue, hunger, etc.	Prefer short experiments. Provide breaks. Run a pilot study.
Reactivity problem	When the act of measuring something changes the measurement itself.	Use clandestine measurement methods.
Testing effect	In a pretest-posttest design, where a pre-test causes subjects to behave differently.	Use clandestine measurement methods. Use a control group with no manipulation between pre- and post-test.
Carry-over effect	When the effects of one treatment are still present when the next treatment is given.	Leave lots of time between treatments. Use between-subjects design.
Order effect	In a repeated measures, the effect that the order of introducing treatment has on the dependent variable.	Randomize or counter-balance treatment order. Use between-subjects design.
Mortality problem	In a longitudinal study, the bias introduced by some subjects disappearing from the sample.	Convince subjects to continue; investigate possible differences between continuing and non-continuing subjects.
Premature reduction	Tendency to rush into an experiment without first familiarizing yourself with complex phenomena.	Use descriptive and qualitative methods to explore complex phenomena. Use explorative information to help form testable hypotheses and identify confounds that need to be controlled.
Spelunking	Exploring a phenomenon without ever testing a proper hypothesis or theory.	Don’t just describe. Look for underlying patterns that might lead to “generalized” knowledge. Formulate and test hypotheses.
Shifting population problem	Tendency to reconceive of a sample as representing a different population than originally thought.	Write-down in advance what you think is the population.
Instrument decay	Measurement changes due to fatigue, increased observational skill, etc.	Use a pilot study to establish observational standards and develop skill.
Reliability problem	When various measures or judgments are inconsistent.	Carefully train experimenter; attention to instrumentation; measure reliability; and avoid interpreting affects smaller than error bars.
Hypocrisy	Holding others to a higher methodological standard than oneself.	Hold yourself to higher standards than others. Apply self-criticism. Follow your own advice.

www.vtpi.org/resqual.pdf