A CRITICAL ANALYSIS OF THE HSURC STUDY –
THE IMPACT OF PREVENTIVE HOME CARE
AND SENIORS HOUSING ON HEALTH OUTCOMES

By
George Maslany

2000

Room 464 Education Building, University of Regina
Regina, SK, Canada S4S 0A2 (306) 585-4117
http://www.uregina.ca/spr
social.policy@uregina.ca
Angels of Mercy or Death

Cutting to the chase, the major finding of this study led to the conclusion that efforts aimed at helping the elderly via preventive home care services may well be ineffective. Indeed it found that this form of helping might inadvertently be proving harmful to these seniors in terms of higher rates of mortality and loss of independence evidenced by those receiving this type of care, when compared to the results of those not receiving it.

No doubt many, especially those with vested interest in this type of intervention such as the service providers and its recipients, will probably experience difficulties in accepting these findings. Their reaction may well be that while the figures may not lie, these results go so much against the grain of what would have been expected that something must be wrong about how the study was done or at least raise the question of whether the results were properly interpreted. The key issue becomes one of the choices one makes about the way to go to figure out what’s going on here: whether the status quo is to be treated as a matter of dogma and corresponding condemnation regarded as heresies or whether we chose the stance of reality checking (via empiricism) and heeding the wakeup call that has been sounded by this study.

Nevertheless Pandora’s Box has been flung open - the cat is out of the bag. Further research will have to be conducted to address the compelling questions arising, most notable of which is whether home care is really as bad as it appears to be.

A Selected History of Harmful Helping

One must recognize that there have been frequent cases occurring over the years where research has shown that conventional wisdom about the value of helping processes has been incorrect. More specifically in the medical domain there have been numerous examples that could be cited of what has been termed by Illich (1976) as ‘iatrogenesis’, the inadvertent straying from the well known medical maxim of \textit{primum non nocere}, (translated ‘first of all do no harm’). As a consequence we have from time to time discovered, to our chagrin, the critical importance of making a distinction between what we believe and what we know, and that the former, no matter how strong the conviction, is not a proxy for the latter.

A rather ancient example having a bearing here is the one of a study conducted by Louis during the latter part of the 19th century related to a study of bloodletting as the treatment of choice for febrile disorders (Eisenberg, 1977). Louis is credited with making an important inaugural effort in the use of what he referred to as the numerical method to illustrate how to achieve advances in understanding of this treatment approach as well as medical treatments in general. Basically his studies entailed an ‘ex post facto’ survey comparing the outcome of those who could afford this intervention with that of those who either couldn’t or weren’t able to avail themselves of it. In short the results of his analyses determined that bloodletting didn't work to arrest the course of these diseases despite overwhelming and longstanding conventional wisdom to the contrary. In fact his number crunching demonstrated that bloodletting appeared to be doing more harm than
good – a finding he personally experienced great difficulty in accepting despite his results. In any event he set the stage for a major belief bubble to be burst and more were to follow, utilizing the empirical technique he pioneered.

More recent research, which is still relatively dated, but is more directly related to studies of the elderly, has shown similar examples of unexpected outcomes associated with interventions intended to do them good. I am referring to the study reported by Blenkner (1967) in which she described the results of her research comparing the impact of intensity of services provided by social workers and public health nurses for a group of seniors. There were three different levels of casework intensity compared, classified as minimal, medium and maximal. Noteworthy here was that her subjects had been assigned to each of these treatment conditions on a random basis making this a true experiment. The findings of a 6 month follow up were rather startling and completely unanticipated with the medium intervention group exhibiting twice the mortality of the minimum intervention group and the maximum intervention group showing a four fold higher mortality than the group receiving minimal services. The expression ‘killing with kindness’ comes to mind to describe these results. This study could also be regarded as foreshadowing the results of the present one under review.

Two important design considerations need to be kept in mind with respect to Blenkner’s study. First of all since her research involved a rather small sample size, large as these effects may have seemed at first blush, they failed to achieve statistical significance. Secondly the effects (of the higher mortality) observed appeared to her to illustrate an indirect as opposed to a direct, detrimental effect of the level of service provided. Specifically it was suggested that the more intensive the service that was provided, the more likely the person receiving it ended up in institutional care (i.e. more intensive service in this case meaning offering them the best service available despite the expense). It was the increased chance of institutionalization for these groups that was hypothesized as more likely the culprit to account for the higher levels of mortality among their sample than that their caseworkers might have been working with them too much. In short the feature of casework deemed most damaging was not the service component but the referral aspect to more intensive levels of care.

The major point to keep in mind here is that while experiments have definite advantages over correlational research designs in terms of establishing cause and effect relationships between independent and dependent variables, the findings associated with experiments are themselves not necessarily definitive. Even had the results cited above achieved statistical significance and been subject to replication, there would still be considerable guess work involved in trying to sort out what aspect of the independent variable (i.e. type of care or its intensity) was responsible for the effect. More will be said about this point later.

More Recent Research Supporting the Foregoing

Various studies in the past have suggested detrimental effects of interventions including those involving the elderly and more recent research has also upheld this position. For
example Segall and Chappell (20000) contend that ‘despite growing medical intervention, seniors do not appear to be better off; indeed, some evidence suggests they may be worse off (p240)’. They go on to cite studies by Havens and Black (1993) and Ford, et al. (1992) in which these researchers report the apparent worsening of the health status of the elderly in recent times despite increasing investments in efforts to enhance their health care. Once again it should be noted that these conclusions were apparently based on surveys not experiments leaving their findings open to alternate explanations.

What remains to be determined is that while the research findings of the study under review are consistent with these previous ones cited above, whether the conclusions arrived at unequivocally follow from the results obtained and furthermore whether the recommendations arrived at also flow from these findings.

Data Analysis

Although various commendable efforts were made in the HSURC study under scrutiny here to exercise statistical control over possible confounding factors that might obscure the relationship between the interventions in question (preventive home care and seniors housing) and outcome, by the authors’ own admission there could still have been something going on that accounts for the differences between the groups compared that may have had an important bearing on the results. Simply put, if a possible critical variable wasn’t included in the database, it certainly couldn’t be controlled. Those sorts of considerations will always haunt a correlational design since you can only get out of a database what’s been put into it.

A variety of information that we would like to have had about the characteristics of their elderly cohort group simply was not available. Examples of variables not available include their relevant history or indicators of their quality of life and adjustment. While the latter variables can’t yet be measured with the same precision as mortality and loss of independence, they are still important domains to consider. So while hospitalization is used as a relatively objective and reliable operational definition for frailty in this study, there are other indicators that may not be as objective and reliable but that more than make up for this shortcoming by being more relevant to the concept of concern. I am specifically referring to the instituting of comprehensive geriatric assessments conducted by interdisciplinary teams (Gallo et al. 1995) which approaches have also been highly endorsed by various reports provided over the years in this province. These would certainly shed more light on adjustment than the binary variable (in or out of hospital) used to assess this domain in this study.

Admittedly one has no choice but to do the best one can with what you’ve got. At the same one must keep in mind that when it comes to making sense out of the results that one has to be careful about making interpretations that go beyond what the data available permits because crucial information may be missing. Also databases continually need to be updated to incorporate state of the art measures even though it makes comparisons over time problematic or impossible.
To be fair to the authors this issue is one that was raised by them and I am merely reiterating it here. Where I differ with the authors is on the importance of this caveat in terms of how it might have a bearing on their conclusions and recommendations. In short I would have expected a more cautious approach at this time with respect to claiming that home care might be harmful and how it should be curtailed as a result. Once again we will have to await the completion of more research on the topic to tell which of us is correct.

A study such as the present one can also serve the purpose of identifying possible variables for inclusion into a database to enhance what future research might discover as could be the case with a recommendation to incorporate measures such as those referred to above coming from geriatric assessment teams. Of course achieving that end can be a long and winding road because before any measures can be incorporated into a database, those providing the data have to learn to use the instruments properly and upon becoming proficient in their use, actually employ them and subsequently record the data derived from them consistently and accurately. Much easier said than done.

What Groups are being Compared?

Since the n’s or specific size of the groups being compared were not reported, I had difficulty figuring out what the basis of the comparison consisted of. There was no question about how many fell into the seniors housing and preventive home care category as those numbers were defined up front. What was, however, not reported was how many were in the group of those not receiving service. I doubt the target of the comparison was simply the total number of elderly falling within the definition of the terms of the study (i.e. the population of 52,981 in Sask. over the age of 75 on July 1, 1991 between 1989-97) less those in the respective intervention groups.

If that’s what it was, I’d have to think about what that comparison meant. In that case, my initial reaction is that this would be somewhat like comparing those who are elderly and don’t need care with those who are deemed to warrant it (by virtue of being considered more vulnerable or frail). This would seem to me is as meaningful as comparing those elderly in hospitals with those elderly not in them and finding that the institutionalized group doesn’t live as long or is more likely to end up in nursing homes (notwithstanding that statistical controls are implemented to try to equate the groups for health risk, other services and so on). In any event whatever the comparative group’s size, it was not self evident and knowing that size helps to know who they might constitute and thereby make better sense out of the comparisons rendered.

Determining the Relationship between the Independent and Dependent Variable

There is no question that the way to go towards obtaining more definitive rather than just suggestive results about the service variable of concern would be to conduct this type of study as a classical experiment. In this instance that would entail those deemed eligible for consideration for preventive home care (or seniors housing) would be assigned to one of these treatment condition or nothing at all (i.e. put on a waiting list or left to life as
usual) on some random basis. If it were a comparison of two of the above options, this would essentially entail the flip of a coin determining who gets what service. A simultaneous comparison of two of these services with a control condition would be even better, again based on random assignment of those regarded as eligible. A gold medal could be awarded to the service that wins the competition.

Regional comparisons, while not experiments, would also prove useful here, especially if one region was consistently able to produce more favorable results than the others. A successful program could be examined in terms of trying to identify what components of it might differ from its less successful counterparts elsewhere so as to suggest how to turn the situation around in the others not so fortunate. Once again a prize could go to the winner of the regional derby.

Ethical Imperatives

There would be important ethical issues arising were the experimental route taken. The matter of assigning someone to a treatment condition such as preventive home care when results such as those derived from the present study suggest it may be harmful can raise concerns. At the same time depriving someone of this type of care when it may be deemed to serve them well is also an issue. The catch is we have no ‘a priori’ way of knowing which is which until we conduct the experiment. If it turns out that the interventions provide better results than the non intervention or control condition, then those in the latter groups can be granted the first shot at the intervention of choice, if they are still in a state to take advantage of the opportunity. If it turns out that those in the control group show an advantage over the others, I am not sure we know how to undo the detrimental effects of a treatment here other than to terminate it. Damned if you do; damned if you don’t. ‘To be or not to be’, one won’t know for sure which is the case until the experiment gets done.

Above all don’t be inclined to take too seriously those who, after results are released, claim the study was a waste of time that they could have told you how it would turn out. Even a complete idiot is likely to call a coin toss right half of the time.

Do these issues sound a lot like those faced in clinical trials? The answer is absolutely. The point here is that the standard for assessing the safety and effectiveness of a social intervention should be no less stringent than that used for making the decision about the acceptance of a new medication. One would not likely get away with using a correlational design and the substitution of statistical controls for experimental ones in a drug trial, so why should we settle for anything less when it comes to social programming. There is definitely as much at stake both for better and for worse, including matters of life and death, in one of these instances as the other.

Let me illustrate the ethical issue here by citing the example cited a few years back in the Globe and Mail of the clinical trials involving the use of warfarin (in short, rat poison) for the treatment of arrythmia in a group deemed at high risk of experiencing a stroke done by Graham Turpie at McMaster. This study was conducted as a classic experiment with
random assignment of subjects to either of two treatment conditions: the rat poison or a placebo. The upshot of this study was that the researchers felt compelled to prematurely terminate the study, for reasons, not as you might expect, that the warfarin group experienced higher mortality or morbidity (M + M) but because to the contrary -- they fared so much better than those in the control condition making it unethical to continue the study. Furthermore, these results were not just a fluke at McMasters but agreed with similar outcomes obtained elsewhere. Once again this illustrates how misguided the common sense of some may be.

Admittedly, there are research ethics review boards that would never have allowed studies of this nature to get off the ground, but fortunately there are more enlightened others that would permit it to proceed as in the case of the McMasters study cited above.

No doubt similar ethical concerns could arise with respect to any experimental design concerning preventive home care. One way of addressing such concerns would be to raise the "remember warfarin" example. The other argument is that notwithstanding the results of the HSURC study, conventional practices regarding referral to home care will likely prevail even though it has been deemed to be harmful by this study and an experiment is the most definitive way to go to ensure that we are not continuing to promote or utilize unsafe programs.

**The Meaning of the Independent Variable**

I have come to realize that an experimental approach, while methodologically superior to the survey of health records approach utilized, would still leave important questions unresolved, even if the results of the experiment turned out to be consistent with the survey of records approach that preceded it. The reason for this unresolved uncertainty rests with trying to figure out what are the essential ingredients in the home care that may have been responsible for such an outcome. Home care is by no means a singular homogenous entity that’s readily replicable as, for example, the independent variable(s) in the case of a drug trial. Home care is more accurately construed as an amalgam or treatment cocktail consisting of a diverse set of ingredients including support services, most notable among which may well be homemaking, personal care, meals and referral but certainly not limited to these elements. Add to this the further possible confounding that these elements may also be of differing types, quality and intensities for different clients at different times.

That having been said about this service variable, the same applies, and even moreso, to any control type comparison group which shouldn’t be regarded as an entity in suspended animation in the interim. There will no doubt be a great deal going on with this group, much more than meets the eye, covering the waterfront of official and unofficial service to them, involving a host of good, bad and indifferent effects, possibly even potentiating or neutralizing one another.

Nevertheless an experiment is still the most powerful means at our disposal of determining whether differences between groups occurred as a result of the intervention
provided to them. One must keep in mind, however, that even if such differences are detected, it will still leave up in the air the answers to the questions of why and what component of the intervention (or combinations of them) was responsible for this outcome (as was the case with the Blenkner study referred to above).

I am not intimating that the answers to these questions will forever remain unable to be resolved. To answer them will, however, require even tighter, more complex research designs than those used to date, involving factorial (experimental) designs controlling for the various components of the service provided and control group conditions. We still have a long way to go here in terms of further research that needs to be done before many of these issues get resolved.

Please don’t simply choose to replicate the current research approach. Even if that approach results in similar findings, all it will indicate to a critic is that if you use the same inadequate design, you will no doubt come to the same questionable conclusions.

Allow me to note, somewhat facetiously, that while I am aware that another way to go here might be to employ a research design based on elderly identical twin comparisons. I have summarily ruled out that approach here, even if you could find enough research subjects dealt with under either condition, since it would be opening up a whole other can of worms.

There is only one example that comes to mind of research on a major public health issue where non experimental approaches were the only practical and ethical way to go and that example involves the impact of seat belt legislation on M + M. You simply can't randomly assign citizens to one of these conditions or the other. You have no choice but to rely on data from comparable groups -- in this case (M + M) rates before and after implementation and enforcement of the relevant legislation or comparisons of these rates in comparable jurisdictions during similar time periods, one of which had the legislation in place and the other not (as was the case in the comparisons made way back when, between Alberta which didn't have the legislation and Saskatchewan which did). Those non experimental results were still quite compelling.

That having been noted, don't make what may seem as a logical leap from seat belt legislation research to preventative home care. The nature of the independent variables differ dramatically in either case. In the seat belt situation, it is fairly straightforward to determine whether someone was using one or not at the time of an accident and to relate the M + M statistics to that condition. For reasons already outlined above, the same certainty of what was responsible for a health outcome can't be made with respect to an amorphous variable such as home care, where it's unclear who gets it and why or what it is that is the active ingredient vs the filler or buffer agent. In short, determining the impact of preventive home care cannot be handled in the same way as assessing the effect of seat belt usage for reasons outlined above.
The Risks Associated with Arm Chair Derived Solutions

The study ends with inclusion of a section that tries to make sense of the results by consultation with health program planners and administrators referred to as the Study Working Group, expert panel and others. The important point to note here is that the explanations arising are just guesses by those supposedly in the know about what may be going on. Moreover, these hunches, informed by experience as they might be, might be wrong, in the same way that the original premise concerning the efficacy of home care might not have been borne out.

The main purposes of these resource persons should not be to explain the results but to set the stage for the further research required to enhance that understanding. In short the explanations derived by these types of consultations should not to be regarded as the end of the road but instead should be framed as questions that can be formulated as hypotheses that in turn can be examined by suitable research designs that either uphold or refute them.

Remember it was the believing that we knew what was best for the elderly by discussing it around a table that got us into this apparent problem in the first place (of believing that home care worked well). Let empiricism, not think tanks or armchair speculation, lead the way to enhancement of our understanding of this domain in the future. We’re simply not as smart as we think we are.

There was one conclusion I found highly contentious and did not follow logically from the results obtained, even if you accept that the study was methodologically sound as you can hope to have it. In particular I am referring to the conclusion ‘if preventive home care is of any benefit, it is likely among those seniors more at risk of death and loss of independence (p8)’. I expected that this might be followed through to the recommendation section to argue that accordingly preventive home care should only be made available to this high-risk group. Curiously, no such recommendation was forthcoming.

Back once again to Billy Shakespeare, I would agree that ‘there is something rotten in the state of Denmark’ but I wouldn’t go so far as to suggest that Denmark is rotten. To mix metaphors I fear that the study recommendations may be a case of throwing out the baby with the bath water. To the point, the study can’t shed much empirical light on what aspect of home care (all of it or just some specific part) might be responsible for the harmful effects observed. It’s one thing to detect the smell of something rotten; it is another matter to believe you know where it’s coming from, which is the point on which I and the study authors disagree.

My own interpretation of the results in this case is that the primarily risky part of home care appeared to be that associated with higher levels of nursing care, not the homemaking, personal care or meal provision part of it. Specifically the findings were that the higher the hours of nursing received, the higher the mortality and loss of independence (which was reported to be more than 3 times the rate of the comparison
group). Needless to say I can appreciate that this represents a more than somewhat politically sensitive finding that one might be tempted to skirt around.

Once again we have the problem with a correlational study, of sorting out which is cause and which might be effect or whether this might just represent a spurious relationship between level of nursing care and outcome. As far as cause-effect relationships are concerned, I believe that there are more tenable explanations for this apparent paradox than having to invoke the ‘Thyphoid Mary’ one. My own explanation once again echoes the one arising out of the Blenkner study, that more nursing care leads to a greater likelihood of assessment of someone as requiring more care. Providing more intensive care could, in turn, translate into greater loss of independence which correspondingly undermine one’s ability to live on one’s own or simply the will to live. This situation could then lead, not surprisingly, ultimately to higher mortality and loss of independence.

That such a vicious circle might occur in the first place (if it does at all) may have to do with the way the nursing profession is educated. A basic principle of nursing care, as I understand it, appears to be that nurses are taught that it’s better to be safe than sorry or that when in doubt, move a patient towards or refer them on to the more intensive level. Better to believe there is a disorder when there is none than the other way around. (The same principle probably applies to the other care givers as well but not to the same extent because they lack the authority to make these referrals.) In this instance this may well ends up producing a self-fulfilling prophecy, at least in comparison to those not receiving this type of attention.

Please don’t regard my own efforts at explanation here as more than somewhat hypocritical in light of my being so critical about those pretending to know rather than to show what’s right. I don’t purport to be closer to the angels or smarter than others here. My own hunches are not intended to serve as anything more than the basis for further research to clarify this issue.

One of the ways of testing out a hypothesis deriving from my own tentative explanation would be to examine why the health costs of those associated with home care are higher than for those not receiving this care. My guess is that this disaggregation of the data would reveal that institutional care (whether hospital care or in a nursing home) is what is responsible for the difference in cost.

**The Costs Associated with Finding Answers**

Doing this type of research doesn’t come cheap. It takes time and the access to the data sets costs money. How much was paid in this case, we haven’t been told. No doubt it’s a major expenditure, the costs associated with which would likely be prohibitive for any other than the independently wealthy (of which there are few among us), or those with major external funding for this purpose (of which there aren’t likely many of those with excess capacity to take this on around either). A related question arises in the instance of the cost to someone who may wish to subject the data set used here to a secondary analysis, a reworking of which could lead to different conclusions.
The point I’m getting at is that this data set, and others like it should be made available gratis to any researcher deemed qualified to handle it, as opposed to having to pay for such access. (This recommendation is made on the assumption that the data has been suitably rendered anonymous to meet ethical research requirements and commercial exploitation of it has also been ruled out). The potential fruits of these endeavors are of a nature that they hold promise of offering incalculable dividends for the public good, let alone major tax savings if they lead to programming efficiencies being realized. It goes without saying that the current research being reviewed is a prime example. I’ll go so far as to argue that there is so much of public benefit at stake here that our researchers should actually be enticed to exploit these excellent provincial health information databases by being offered monetary incentives to do so by those responsible for controlling access to them, rather than being discouraged from this end by having to pay for the product.

Another important element has to do with recognizing the critical value of engaging in ongoing evaluative research on these types of programs. Those exercises are essential in determining how well they fare; however, rarely are they done at all and at best only intermittently. While the costs associated with carrying out these evaluations can be substantial, they hold promise of identifying programs that fail to meet their intent and possibly efficiencies that could be implemented. So while they don’t come cheap, we also cannot afford to fail to engage in them, because of the payoff they hold promise to provide. It goes without saying that controversy may arise as a result, especially if vested interests are threatened, but so what, as long as it serves the best interests of those we intend to serve.

Conducting evaluative research should be seen as being as useful an exercise as conducting regular audits about how money for a program is spent to ensure it has been spent properly and wisely. That this type of financial audit of a program might yield consistently favorable results over the years is not a reason not to conduct them in the future. Neither does the fact that the audit for one program might have come out all right mean that one might not be required of a similar program elsewhere. Engaging in regular audits of how money for a social program has been spent is the standard accepted way we conduct our business in this respect. Similarly, determining the bang for a buck associated with a program or comparison of program alternatives to dealing with a particular problem should be given no less attention, even if the cost of doing so might detract from other programming. More attention has to be given to our determining how well we are doing than simply going on doing what we’ve done, without knowing objectively how well that’s working out.

Program Cost Comparisons

This is one part of the study I couldn’t take very seriously since the costs calculated for each of the programs were so narrowly defined. These cost comparisons were essentially those related to health care and as such showed only that seniors’ housing is cheaper than home care in terms of health budgets. This finding gives no indication that the latter is less of a drain on the public purse or on expenditures by others, so we could basically be making a decision about whether the money to support these elderly was coming out of
one of our pockets or the other. Twisted the other way around and taking the same approach as in this study, only this time with the agency responsible for seniors housing conducting the study, they might show that home care is cheaper than seniors housing for a given individual because the bills in this case are paid either by the person concerned, their families or Health (via institutionalization). A kind way of putting it would be to call this type of cost analysis conducted incomplete. ‘The unkindest cut’ would be to call it just plain silly. I am assuming the intent here was not simply to play to health care administrators who might be delighted to see this type of analysis as an avenue for offloading costs associated with care to other sectors even if those other sectors are simply other parts of government.

No Variance to Analyze in terms of Cost Comparisons

The summary copy of the report I reviewed only presented means (average costs) for comparative purposes when just as informative would have been the corresponding standard deviations. It is the latter that would be quite useful to examine in order to ‘eyeball’ how much overlap might have existed between groups being compared. While there were numerous indications of statistical measures of centrality, not a single measure of variability was anywhere to be seen.

The Art and Science of Data Analysis

I want to make an important distinction between data analysis and interpretations derived from them. In the case of quantitative approaches the data analytic techniques have a mathematical basis. One could refer to this part of the data analysis process as its objective component.

A somewhat separate consideration has to do with the determination of the most suitable statistical test of a research hypothesis and subsequent appropriate interpretation of the results. Unfortunately this aspect is mitigated by knowledge of the confidence that can be assigned to the quality of the data, research design, the selection of the most suitable statistical analysis and it’s inherent limitations, violation of assumptions underlying it and so on. This makes the interpretations of results more a matter of an art than science. In short while statistical calculations involve objective and consistent processes, the selection of the most appropriate test and interpretation of what it means is often an art. Being more art than science, it implies that different researchers will not always see eye to eye in this respect and often beg to differ, as is the case here.

Uses and Abuses of Statistical Tests of Significance

Before I embark upon this section there is a clarification required about what actually constituted the number of observations on which the analysis for this study were based. The Summary Report indicates that number to be 26,490; however, the ‘Methods’ portion published separately identifies the population of study as actually twice as large as 52,981. The reason for the discrepancy is that this population was randomly subdivided into two parts with one of these halves used for model development and the other for
model testing. In other words the values derived from the first go round of analysis were applied to the second group to see how well they held up to cross validation.

The approach used above is fairly standard fare but does raise some important questions. My suggestion would be that rather than dividing the 53K population randomly in half as was done, I would have thought more useful would have been to subdivide it in half on the basis of time. For example, my advice would have been to use the first four years of data for model development and the data for the remaining years for testing of how well it held up to this type of replication over time. Were it to hold up under these circumstances, it could have provided useful information about the confidence we could place in the model for understanding what was going on at other points in time (outside of the time frame on which the initial model was based).

As it stands all I can say is that the results of a random division of the population in two would result in predictable consistent results showing limited discrepancy between model development and testing of it. In fact you could conduct an unlimited number of such random splits, the distribution of which would yield the standard error of these estimates. None of these results would offer any practical significance. In short – whether differences were found or not as a result – so what?

One would not expect anywhere near as consistent results in the split I have suggested which would give a much better idea of the value of the model outside of the groups on which it had been tested.

The second point that needs to be made is the failure of the study to precisely define the population of interest to which their results can be generalized. Specifically is the population meant to imply elderly in general, or more narrowly to the 52,981 identified in their database? My suspicion is that it is the more narrowly defined latter group. This is an exceedingly important matter that needs to be clarified up front because of its bearing on the group to which the results apply: namely do they apply just to the group under study or some broader population of elderly both here and elsewhere.

If their population of interest is much broader than the 53K group, then we have a problem with violation of the assumption of randomness of selection since that group cannot be construed as randomly selected from the broader population of the elderly. Without random selection from the population of interest the meaning of inferential statistics derived becomes blurred and accordingly proves useless.

One further point to note here is that a randomly selected sub sample drawn from a very large super sample can be deemed representative of that super sample. Where, however, the latter has not been similarly drawn from a larger population, this only makes the random sub sample representative of the super sample, not of the larger population. In short, a random sub sample drawn from a biased sample does not provide unbiased estimates of the population from which the latter was selected, only unbiased estimates of that sub sample.
With random sampling we encounter unsystematic error which then allows us to use inferential statistics to serve as an estimate of sampling error via the significance level indicated. In non random sampling the sampling error is systematic in which case inferential statistics cannot provide any unbiased estimate of the probability the results obtained occurred by chance.

Although references to tests of the statistical significance of various results only appear as an occasional footnote, I want to emphasize my criticism that these tests simply don’t apply here. The fact is the cases under study cannot be construed as any random sample of a population of elderly in general since they actually constitute the entire population of those 75 and over within Saskatchewan during the 1991-98 time interval. Accordingly since the data set contains information on an entire population, no inferential measures are needed and any differences, even minor ones, are real differences. Some of the smaller differences observed in this case may be considered trivial ones but making that type of determination is a judgment about what appears to be meaningful rather than anything informed by a test of statistical significance. In short, inferential statistics involving tests of significance apply only to samples; descriptive statistics suffice in describing parameters of populations, which is what we have in this study.

The only way tests of statistical significance could be justified in this study is based on the assumption that the target group for analysis could be regarded as a random sample of this population of elderly over other points in time or location. That type of sampling clearly didn’t happen here. Rather than risk that assumption of the representativeness of the group under study being false, I’d leave that question up to empirical verification to see whether characteristics of the target group of elderly on the variables of interest remain invariant over time and place or not. The study also makes it very clear that the characteristics of the elderly under care constitute a moving target changing rather dramatically in age composition over time. Similarly since the observations for the study in this case only come from Saskatchewan, I would not risk assuming that these constitute random samples of observations of others, coming from elsewhere, especially those that might be served by different programs. All this rests on the tenability of the assumption that seen one (data set of elderly) you’ve seen ‘em all. That may be true and then again, maybe not.

Once again I would leave it to comparative studies to determine whether the results obtained here apply in other jurisdictions and/or at other points in time, even in our province—not on inferential leaps from the present one supposedly guided by tests of significance. Accordingly feel free to ignore these tests of significance reported here and above all don’t make the mistake of interpreting them to imply which results reflect real or important differences.

I am well aware that the study’s approach to using tests of significance, even in questionable instances like this one, represents conventional practice, that research will be written up in this way to make it appear sophisticated and proper, whether doing it that way really makes much sense or not. I trust that in time these conventions will be replaced by more parsimonious and sensible ones. Looking at the results in this case
from a descriptive statistics point of view makes for easier and simpler reading by removing a superfluous, obfuscating and pretentious level of analysis. What you see is what you’ve got and what we’ve got simply applies to Saskatchewan elderly at a specific point in time (1989-97). As I see it statistical significance here ends up being a case of ‘much ado about nothing’.

There’s no question that certain of the findings of the foregoing will be upheld by virtue of further study of this population here, at other points in time, as well as elsewhere. The point is that at the present time we don’t know which of these findings will survive more meaningful attempts at cross validation. The answer to such questions will not be derived from pondering the statistical significance of the present findings but by conducting replication/extension studies with these other sectors and points in time.

Too much replication going on as it is, you say? I’d reply, quite the contrary, not enough. In any event let’s not confuse replication as a process with replication as an outcome. If conducting the same study over and over again yields identical results or at least similar ones, in hindsight it may seem like a waste of time, especially considering the reluctance of mainstream journals to publish these types of research findings. On the other hand, these previous results may not bear up to replication across groups or time, which then becomes quite newsworthy. The point is that we have no a priori way of knowing which way it will turn out until we endeavor to repeat the study in question under similar or other circumstances.

**Management and Meaning of Uncertainty**

**When Tests of Significance apply**

As indicated above I have serious reservations about whether the tests of significance reported above are appropriate in this study. At the same time a determination may be made by others that they are. This then leads to the question of understanding what they mean, even if not here, at least in other cases where they may arise and be appropriately applied.

I want to reiterate and expand on my earlier point that a lot more scholars and professionals read about inferential stats and even use them than necessarily understand what they mean. If you are confident that you have them nailed down pretty well, then just skip this section. If not, or you just wouldn’t mind a refresher on some of the fundamentals conveyed in a non-technical and simple way, please read on.

Being able to see what appropriate statistical inference entails from a number of different angles may ultimately help to crystallize what they are all about, or not about as well. Using this type of triangulation is how I wound up learning what they meant and hopefully this type of piece will help you in the same way too. My own biggest regret is that I never encountered a lynch pin like the one I am attempting to compose here that could have helped me out in my own struggles along the line of this ‘stats for dummies’ effort before you here.
I also want to caution that while this might help one who has an unsure footing in the statistical realm to get a better sense of balance and confidence in this domain, this is not intended as a substitute for a more formal technical involvement with the subject. It should offer a useful foot in the door of the subject but not to be seen as any attempt at a last word on the topic.

Types of Statistical Errors

I will begin by giving a simplified outline of the various elements of uncertainty concerning decision making in research and then applying such results to practice situations. The most familiar of these relates to inferential statistics. Where there has been random sampling and as noted above, even where there hasn’t been, there is frequent reference in research that the results have achieved statistical significance at some benchmark level such as .05, .01 or more stringent yet. This is known as the probability of making a type 1 error which is one of concluding that there is a something happening, when the effect observed is really due to sampling error or variability (5 times out of 100, 1 out of 100 and so on). Expressed otherwise, draw 100 samples of a given size from the same population and at the .05 level, 5 of them will meet or exceed that level of significance. In short the significant result is just a fluke, nothing is really going on. It’s all a case of sampling variability or sampling error. The medical clinical analogy to a type 1 error would be the probability of a false positive diagnostic result. The criminal analogue would be of finding an innocent person guilty.

Please note the following sections will refer to various parallels between statistical decision making and decision making in other areas, particularly medicine. These can best be understood by examining the tabular representations of selected examples and corresponding explanations as contained in Appendix 1. If you get confused about the examples presented and how they were intended to be represented, please consult this appendix for clarification. In fact my draft for this section required me to prepare such tables to avoid getting confused myself about making the appropriate links. In the end I realized that if these tables would help me to explain the connections, it should also serve the reader in understanding this section.

Where do these critical values of .05, .01 and so on come from? They originate in the same murky depths where a whole host of other variables which in nature are actually continuous are manufactured into categorical ones. Include here arbitrary categorizations such as speeding limits, .08 count for legal intoxication, definition of passing grades, age of majority and so on and so forth. Who sets these, what’s the rationale behind the setting of the specific borders or how far back do any of these go? Who knows the answers. The best single reply I can come up with is that this is the way it is, so just grin and bear it because there hardly seems to be any point in arguing about them or challenging many of these definitions. The fact is that practices, once entrenched, can become conventions that often take on a life of their own and ultimately as a result become hard to unseat, not unlike the challenge constituted by home care. All I can say is how paradoxical it is that there still remains so much unscientific business in our current scientific practices.
Many journals will be reluctant to accept an article submitted for consideration that doesn’t report significant results. This can produce a peculiar bias about what actually gets published. For example, if one did any particular study 100 times (done by the same person or different ones) where there was actually no difference between groups being compared in the population from which they were drawn, at the .05 level, about 5 of these studies would achieve statistical significance. The crunch comes in that these significant findings are the one’s likely to catch the attention of the journal editors and moreover might possibly might even be regarded as evidence of replication of one another, while the remaining 95 languish in some trash heap for fugitive studies and never even get submitted.

There is another more direct way of increasing the chance of obtaining significant results if that is the be all and the end all of the research exercise and that is wherever possible to base the study in question on large samples, the bigger the better. Since it is unlikely in real life that the null hypothesis is ever true in any situation, all that is required is to have large enough samples to rule out that the differences observed could have occurred by chance through sampling from the same population. One caution that must be kept in mind is that the statistically significant findings that have been observed in such cases may be so miniscule that they are of no meaningful or practical significance. Bigger samples usually mean more statistically significant results but not necessarily a corresponding increase in the discovery of interesting useful findings. As indicated earlier it is usually theory, not statistical significance that would guide determination of which findings also met a criterion of practical relevance.

The second type of statistical error and one which is less familiar, is, not surprisingly, referred to as a type 2 error. In this case the question becomes one of the probability of incorrectly rejecting an effect when one actually exits (namely of accepting the null hypothesis when it is false). In other words one concludes the differences evidenced are due to sampling error when they are actually due to differences in the populations from which the samples were drawn. Succeeding in avoiding this type of error depends on a number of factors including the sample size (the larger the better), the size of the expected effect (the bigger it is, the easier it is to detect), the level of the type 1 error (the more stringent it is, the more difficult to detect an effect) and finally the precision of the measured variable (the lower the reliability, the more noise involved with it, the harder to detect a real effect which can otherwise be obscured by the measurement error). The medical counterpart here would be the probability of obtaining a false negative diagnostic finding or in the criminal justice instance of allowing a guilty party to go free.

With respect to type 1 and 2 errors, once again I reiterate that the bigger the sample size, the better but only if the cases have been sampled according to the conditions of randomness and independence of selection. Also because the errors associated with any sample size can be precisely estimated (which one can’t do with a biased sample), the usual rhetorical reply to the question by a researcher of how large a sample size is needed is responded to by the question from the statistician - ‘well that depends on how precise you want to be?’ or ‘How much do you have to spend?’ In short type 1 error gets
reduced by increasing the sample size. As this size increases beyond 30 cases it becomes a case of diminishing returns so that the reduction in sampling error of going from, for example, 20 cases to 30 is much greater than it is going from 100 to 300, though the larger number still provides the better estimate. The same applies to type 2 errors although other factors outlined above also have a bearing here.

For any given effect size, the larger the sample, once again presuming that it has been randomly and independently selected, the less likely any observed results are attributable to a sampling anomaly. The point that must be kept in mind here is that while the larger the sample size, the lower the corresponding level of the possibility of a type 1 or 2 errors, these types of errors can never be eliminated completely unless of course the sample size is as large as the population which is never the case. For a case point, look no further than the recent presidential election in the United States which was so close no poll could have ever called if correctly.

An upshot of this type of dithering leads to the kind of comment once attributed to Harry Truman who indicated one of his most prized advisors, if he could ever find one, would be a one armed economist. In other words someone who couldn’t say, "on the one hand it could turn out like this, on the other...." In analysing research it would be nice to be able to say they indicate one way or the other, however, the realities of science usually don’t allow such clear cut choices. The best outcome of any inferential study is to leave us in a state of sometimes reducible, but nevertheless unavoidable, uncertainty.

If the group under study is not a randomly selected sample, meaning that there may be systematic error involved, bigger does not necessarily mean better. Bigger in this case may mean more of unknown bias effecting the results. Moreover increasing the size of the study group will not ultimately compensate for the lack of randomness until possibly the size of the group under study gets almost as large as the population of interest at which point you don’t have to worry about sampling any more anyway. In short bigger does not necessarily mean better unless random and independent sampling are involved.

From both a practical as well as theoretical point of view a small random sample of a handful of cases is generally going to provide a more reliable indicator of a population parameter than a convenience sample conducted on hundreds or more. Sure the study based on the larger numbers may look to the uneducated public to be more impressive, but the smaller one is likely to yield more accurate results or at least yield results with a margin of error that can be accurately assessed.

According to the conventions evident in the research literature, the predominance of the reference to type 1 errors (p>.05, .01 and so on) suggests that the most important errors are those of deciding there is an effect where there really isn’t (a false positive finding so to speak), with much lesser and limited attention ever given to type 2 errors, of erroneously deciding that nothing is going on when there really is (the possibility of a false negative).
Extending these statistical considerations to the medical analogy of clinical decision making, a similar situation arises in cases of uncertainty, when one decides that it is preferable to risk the error of concluding the presence of a disease that may not be present (to risk acting on a possible false positive finding), than it is to err by failing to detect a condition that is really there (of failing to act as a result of a false negative finding). In short the latter is generally seen as the more serious sin. Similarly better for the smoke detector to produce a false alarm and people to run away scared as a result, than for the detector to miss the real thing. At least in clinical medicine both types of errors are considered whereas in research published, reference is frequently only made to type 1. In short much published research suggests that the only error worth worrying about and keeping from rearing its ugly head no more than 5 times out of 100 or often less is that of deciding there is an effect present when it is really not there. Rarely is consideration given to type 2 errors or the capacity of the statistical procedure utilized to detect an effect if it is really present in the population from which the samples have been taken.

A ridiculous extreme way of eliminating either of these errors, nicely illustrates the trade-offs that are involved if you want to minimize one of them but to do so at the expense of the other. Want to avoid acting on false positives or false alarms, believe nothing, simply fall asleep at the switch and disregard any such signals or findings or yank the battery out of the detector. Want to eliminate false negatives in relation to truth determination, believe everything and act on it too. In real life, of course, there are usually more subtle shifts of risks between choosing cynicism or naivete.

From a practical point of view deciding which error is worse depends entirely on the consequences associated with acting, when one needn’t and failing to, when one should. Is it better to be safe than sorry or is it better to avoid being sorry for playing it too safe? The answer depends on analysis of the matrix of possible outcomes (positive and negative) and the probabilities associated with each of them if you’re wrong and willingness to accept those risks in relationship to the benefits that could be derived. Ethically, the ultimate decision obviously has to be rendered by the one effected by the judgment. Frequently this does not become any easy task and even when all goes well, it sometimes becomes a case of type 3 errors instead of good judgment.

What is a type 3 error? I can’t relocate the reference where I first read about it 30 years ago but it’s not a big deal since it was probably raised by the original author more in jest than to be taken seriously. This one entails coming to the right conclusion for the wrong reasons. It could very well apply in the case of the study under review here. I am at a loss for a medical example here other than being just plain lucky, not smart, in making a proper diagnostic decision for a particular patient.

Statistically speaking, while complete certainty can never be realized, the probability of the two types of error (i.e. type 1 and 2) can be progressively minimized simply by increasing sample size (remember not just making these larger but also via random selection). The corresponding way of reducing error in the analogous situations referred to above is to attempt to maximize the amount of reliable and valid information that can be brought to bear on a decision. At the same time one must recognize that sometimes
there may be a clock ticking away in the background conspiring against giving one the luxury of having an unlimited time frame to arrive at a decision so that if the decision isn’t made in time, there may be no decision to make.

**Consequences of Violation of Assumptions Underlying Inferential Statistics**

Normally one would expect a discussion of the assumptions underlying inferential statistics to precede any explanation of them. In that sense my approach here may seem to be somewhat ass backwards. There is, however, ‘method in my madness’ in that there is usually such scant attention paid to this aspect that it frequently doesn’t even seem to be given consideration, except possibly after the fact, so I will follow that same beaten path myself.

Although every statistic is associated with it’s own specific assumptions which prescribe where it can be used appropriately and under what conditions and where not, one assumption all inferential statistics share in common is that of randomness and independence of selection of observations for study from a specifically defined population or universe of study.

Just as a refresher, randomness of selection means that every member in the population of interest has an equal opportunity for being selected for study. Independence refers to the fact that selection of any unit of study, person for example, is independent of every other. In other words securing one case does not mean that one can subsequently restrict one’s interest to others that might be conveniently located nearby.

As mentioned earlier adherence to this assumption is integral and crucial to the appropriate use of inferential statistics. Why then is it so often ignored with impunity, for example by basing large sets of studies on what are essentially non randomly selected convenience samples? More specifically I am referring to studies which are exclusively based on convenience samples like those involving introductory psychology students (when the population of interest has no intent to restrict itself to this group), volunteers, focus groups and so on.

The answer is that those using such convenience groupings of research subjects are essentially banking on the fact that this type of selectivity bias probably doesn’t matter or at least hope and pray that that’s the case. Essentially the assumption made by these researchers is that regardless of the fact that strict adherence to the principles of sampling has not occurred, it doesn’t really matter and it’s safe to assume that ‘seen (the results of) one(of these groups), you’ve seen ‘em all’.

You can count the authors of the home care study here too. And now for what may probably be seem as an unexpected concession by me. They may well be right that it doesn’t matter. The catch is that the only way one will know for sure is by replication with more appropriately selected samples or broader populations of interest.
I suspect it all boils down to what might be at stake if generalizations made or implied are wrong. There’s certainly evidence in marketing research that the results based on convenience samples may not augur well for what goes on in the larger population. Need I cite any more examples here than that of ‘new coke’ about which the taste testers group provided rave reviews. Well even so, their customers didn’t swallow it, resulting in one of the greatest marketing fiascoes in history. This is not an isolated example either.

You can even use cross species examples where the safety of thalidomide had supposedly been demonstrated by showing no adverse effects on ‘Norvegicus Albinus’ such as the original usage for thalidomide but…. Many more examples could be presented but won’t be because I believe the point has been made.

So the question boils down to, given the fact that while sometime you can get away with non random sampling and other times not, what should one do with the home care study at hand? As I’ve mentioned earlier, my idea of the way to go is not by way of any contemplation of one’s naval nor by searches of the literature for precedents or theory to guide one but via efforts at replication of the findings obtained this last go round with other sectors in time and place using better experimental research designs.

**The research design issue**

As mentioned above the tighter the research design, the further we move from results that are merely suggestive to ones that are more definitive. In operational terms this entails moving from correlational research designs to quasi-experimental ones (involving statistical controls over confounding variables) and ultimately to true experiments, the virtual optimal defense in our armament against confounding effects. The important point to note here is that while the progression towards true experiments increasingly rules out competing explanations, it can never rule all of these out completely. In short, while improvement of our study designs will increase our confidence that we really know what’s going on, the increased confidence never ends up in certainty.

Certainty is often more of an ideal than an achievable reality. This is especially true in the case of assessing preventive home care which is associated with so many imponderables and uncertainties about both of what it entails and what it effects, that we will never likely know what’s going on for sure. I am not saying this to suggest that efforts to nail it down are exercises in futility but that no matter how well we do the research job of attempting to understand it, there will remain an element of not knowing for certain whether it really works or hurts or what about it might cause that effect.

Here I’d also like to make the important distinction between knowing what might apply in general (as might be revealed by a research study) as opposed to what may be best for a specific individual at a particular point in time since most research on interventions demonstrate overlapping variance. This latter comment is not to be interpreted that I am advocating disregard of research results, even from those derived from optimal designs, in place of idiographic approaches (i.e. arguing that a case at hand is unique and no other research results apply to it) but that decision making in this regard involves a complex
matrix of possibilities and unknowns that can result in decisions about interventions for individuals that are incorrect and being able to estimate that fallibility.

In short it is as important that we recognize and acknowledge what we know, as well as it is to be aware of what we don’t. All of the above have a bearing on how we chose to act on the results of this study as well as others that impact on the health and well being of our population. The highest form of intellectual evolution for the practitioner is not just knowing a lot but also knowing what you don’t know and having the honesty and humility to admit it. The highest virtue for the researcher is to seek to narrow that uncertainty.

Where do we go from here?

Fortunately the recommendations provided by the report were much more reasonable than the digested versions of the report that appeared in the media. At the same time this is also unfortunate since it is unlikely that many of the front line staff, elderly or those in their families concerned about their well being would likely read the actual report. Accordingly I regard it as highly imperative that a public education component be developed by HSURC that advises those involved in this area, either as service providers or recipients or those such as family members assisting the latter, on how to proceed during this interim period of uncertainty until more research clarification is forthcoming. This is necessary in order to address any confusion or panic that might arise among those directly involved in this type of care trying to decide where to turn to now. Unfortunately, I have no idea about what HSURC’s educational component to this study entails.

While it may not help the confusion to state that more questions may have been raised by this study than have been answered at this time, it would be an honest reply, at least until further study results are available. In the meantime I would hope that people will not feel compelled to flee home care as though it were the plague no more than it would be appropriate for them to merely regard it a matter of life as usual in the meantime.

It’s a tall order to be sure to endeavor to disseminate research results of this nature to the general public, service providers and consumers of the services concerned in a form that they can understand and act upon appropriately. It has, however, to happen and so far this has not occurred to the extent that I believe is necessary. Failing to engage in such an educational strategy, the research effort might itself ultimately be construed as an example of iatrogenesis, leaving those it was supposed to assist worse off because the decision makers did not interpret properly what direction was indicated by it. Yes, even research can do harm, if not properly communicated, comprehended and acted upon.

Insight, whatever that might be in this case, does not translate into action any more than learning about the health hazards of smoking causes people to butt out. The general population including professionals too, often suffer from "cognitive atherosclerosis,” a hardening of the categories, and it takes a bit more than exposure to a truth to get anyone to change their ways.
Conclusion

There are some rather interesting dynamics associated with this study that could be described as providing a serious challenge to the contemporary mindset about the usefulness of a program intended to assist elderly deemed to be able to derive advantage from this type of service. I doubt there were many who weren’t startled by the findings and include me here. After sleeping on it for awhile I came to realize that all wasn’t as unequivocal as it had initially seemed, resulting in my preparation of this critical review.

The major concern I have about the conclusion and related recommendations is that the group comparisons rendered (home care vs. seniors housing) are based on the assumption of *ceteris paribus* or AOTBE [all other things being equal] among them which I am not convinced is the case. Were these to have been the results of an experiment rather than a correlational study much of my skepticism would have been defused. In an experiment the assumption of group equivalence is addressed by research design (i.e. randomly assigning subjects to treatment condition). Random assignment aims to achieve equivalence of groups by having any factor or combination of them that might influence outcome having equal likelihood of occurring in one of the groups as any other. Consequently any differences observed between groups are over and above these possible confounding effects.

In a correlational design or quasi-experiment, as was used in the study under review, comparability of the groups was attempted to be achieved statistically-an acceptable approach where an experimental manipulation might not be possible either because the variable can’t be subjected to experimental manipulation (e.g. demographics) or hasn’t been (as is the case with the assignment to treatment condition here).

In any event the results of an experiment (once again AOTBE) are generally regarded as more trustworthy than those derived from a statistically controlled retrospective study. This doesn’t diminish the importance of the study that was done but emphasizes it’s primary role, as I see it, of serving a heuristic purpose to set the stage for proceeding to the next level -of determining whether these results are upheld through an experiment.

The other important issue here has to do with the matter of where the ‘burden of proof’ should rest in this case-especially one involving a sacred cow as venerated as home care. If we were dealing with a new program proposal for which a pilot study based on a correlational design suggested was counterproductive or otherwise ineffectual, it might be quite a straightforward matter to pull the plug. Basically it would be incumbent on the protagonists for the program to provide the evidence that it worked.

In the case, however, of an established program such as home care, the burden of proof usually shifts from one of having to prove that it works to showing fairly unequivocally that it doesn’t. Adapting the presumption of innocence principle of criminal law, an established program is generally deemed to be effective until it is shown beyond a reasonable doubt that it isn’t. For an extreme example of how difficult it can become to dislodge entrenched interests, just think of the tobacco industry.
I would contend that with preventive home care most of the evidence against it is primarily circumstantial and as such raises the question of reasonable doubt that it might be performing better than it seems to the authors of the HSURC study in question.

The practical rule that this example establishes, which those responsible for program development should be mindful of, is that it’s always a lot easier to start things up than it is to wind them down. So to mix metaphors look cautiously and carefully before you leap in to get a new one up and running or it may come back to haunt its initiators.
References


**APPENDIX 1**

**Clinical Decision Model**

<table>
<thead>
<tr>
<th>Actual Presence of Illness</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Decision About Illness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>Correct Decision</td>
<td>False Positive</td>
</tr>
<tr>
<td>Absent</td>
<td>False Negative</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

**NOTE:** Medical maxim is better to believe there is a disease when there isn't (false positive) than miss it if it is present.

**Scientific Decision Model**

<table>
<thead>
<tr>
<th>Actual Truth of Proposition</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Decision About Proposition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>Correct Decision</td>
<td>Naivete (acceptance of an untruth)</td>
</tr>
<tr>
<td>False</td>
<td>Cynicism (rejection of a truth)</td>
<td>Correct</td>
</tr>
</tbody>
</table>

**NOTE:** In science the consequences of cynicism are regarded as preferable to those associated with naivete.

**Statistical Decision Model**

<table>
<thead>
<tr>
<th>Reality</th>
<th>Null hypothesis* true/alternate** false</th>
<th>Null hypothesis false/alternate true</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>----Decision---- About Null</strong></td>
<td>Accept null and reject alternative</td>
<td>Correct Decision</td>
</tr>
<tr>
<td></td>
<td>Reject null and accept alternate</td>
<td>Type 1 Error. Conclude something is going on when it isn't.</td>
</tr>
</tbody>
</table>

* Null hypothesis states there is no effect or relationship between or among the variables subjected to a test of significance.
** Alternate hypothesis states that an effect or relationship is evident..
**Criminal Law**

<table>
<thead>
<tr>
<th>Reality</th>
<th>Guilty</th>
<th>Innocent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guilty</td>
<td>Correct decision</td>
<td>An erroneous decision about guilt</td>
</tr>
<tr>
<td>Reject</td>
<td>An erroneous finding of innocence</td>
<td>Correct decision</td>
</tr>
</tbody>
</table>

**NOTE:** Legal convention is that the error of finding someone guilty who is actually innocent is more serious than that of finding someone innocent who's actually guilty. Better to let 100 guilty get off free than to falsely convict an innocent person.