



Translating Financial Research Into Action

MEMBERSHIP NOTE #1:

**How to Read Research Papers
With a Discerning Eye:
Take the Best and Leave the Rest**

by François Gadenne, September 2018

ABSTRACT

If fake news (opinions presented as facts) affects our understanding of reality, how does imperfect research (errors presented as facts) affect your business? How do you know that your best practices are good enough? Answering the question requires an evaluation process.

Informal evaluation processes rely on trust. Such evaluation processes range from trusting a brand, an author, a school of thought, or the editors of a peer-reviewed journal.

Financial professionals working under a standard of competence as well as a standard of conduct cannot rely on the power of trust alone. Good intentions enacted without sufficient care, skill, and caution will lead to negative outcomes for both clients and advisors.

Professionals need a formal research screening process. This Note from the Curve Triangle & Rectangle InstituteSM (CTRISM) documents a two-step screening process that you can put to work right away.

The first step identifies the type of questions that a research paper can legitimately ask and answer.

The second step helps you form an opinion about the validity of the answers to the legitimate questions.

This paper starts with three important reasons why this topic is important for your business, it continues by outlining Step 1 and Step 2, and concludes with a summary of CTRI's mission.

WHO IS THE IDEAL READER FOR THIS NOTE?

The ideal readers for this Note are retirement-focused advisors and the executives of the companies that support them with professional services, investment products, insurance products, and software solutions.

This Note starts with two important reasons why you need a formal process to identify and evaluate the financial research papers that support the best practices you adopt.

RETIREMENT PLANNING STARTS AS A DECISION PROBLEM THAT RETIREMENT MANAGEMENT TURNS INTO A MEASUREMENT PROBLEM

Advisors are bombarded with “research” on a daily basis, including marketing materials, white papers, press articles, practitioner papers, academic papers, etc.

Financial professionals working under a standard of competence as well as a standard of conduct cannot rely on the power of trust alone. Good intentions enacted without sufficient care, skill, and caution will lead to negative outcomes for both clients and advisors.

A retirement plan is a long-term, adaptive, and protective structure. Clients want to protect their wealth and their ability to support a suitable standard of living in retirement. Thus, a plan’s value resides in the domain of moral choice (client and advisor values, goals, etc.) before the domain of economics (client and advisor asset growth, income flow, etc.)

Retirement planning starts as a decision problem based on such moral and economic inputs and continues as a measurement problem based on outcomes: first launch the rocket in a specific direction, then manage course corrections to the destination.

The decision problem is a clinical issue that qualifies your diagnosis of ambiguous values, goals, and expected payoffs. The measurement problem is a statistical issue that quantifies your degree of doubt and optimization of the payoffs.

Advisors need strong conceptual foundations to justify both retirement planning and retirement management best practices. Do you distinguish between retirement planning and retirement management best practices? What evaluation process do you use as you adopt best practices in your business?

LONGEVITY IS A PLANNING RISK FOR CLIENTS AND A PRACTICE RISK FOR ADVISORS

Advisors need to build their retirement planning and management work on solid foundations that will pass the test of time. Retirement plans are not monthly, quarterly, or annual services. A retirement plan may last more than 30 years.

Longevity risk is not only a planning risk for clients, it is also a practice risk for retirement advisors who often work with clients to their time of death and beyond, advising the surviving spouse and/or the next generation.

Both investment and retirement advisors have business models that are akin to selling a put option, receiving a premium in exchange for being held responsible on the downside.

Unlike investment advisors, however, retirement advisors are selling a longer-duration put option. The longer the duration, the greater the compounding of initial planning errors, and the higher the practice risk.

“SPINACH” AS A CODE WORD FOR RESEARCH ERROR PRESENTED AS FACTS

Do we know the right things for the right reasons?

Did your mother make you eat spinach because it was high in iron? Did you make your children eat spinach for the same reason? Well, high iron in spinach is not a valid research answer, let alone a valid parental practice.

To be fair, it started in 1871 with a research error:

The Prudence of Eating Spinach for Iron*

- Did you make your children eat spinach because people in a position of authority told you it was high in iron?
- The genealogy of the Spinach/Iron Contradictory Prescriptions:
 - Professor von Wolff (1871)
 - Contaminated vessels and the charcoal used to heat them
 - Generations of textbooks and researchers replicate the error, including “Handbuch der Emahrungslehre” by von Nordee and Saloman (1920)

However, the correct answer, the correction of the initial error, has been known since the XIXth Century:

The Prudence of Not Eating Spinach for Iron*

- Did you stop making your children eat spinach because you were told Professor von Wolff misplaced the decimal point?
 - Professor Arnold E. Bender (1972, 1977)
- The Spinach Myth turns into the decimal “Good Story”
 - Professor Terry Hamblin (1981)
 - Popularizes Bender’s explanation, including a mention of Popeye
- But ... correct values, the “Real Story”, were known for a long time...
 - Professor von Bunge (1892, 1902)
 - Professor Sherman (1907)
- Prudent Spinach Advice: What you do and why you do it

**Source: Dr. Mike Sutton (Criminologist). Department of Social Sciences. Nottingham Trent University*

We even get it wrong in cartoons. Popeye did not eat spinach for its iron content, he ate it for its vitamin A: “Spinach is full of vitamin A, an’ that’s what makes hoomans strong an’ helty.”

As if this was not amazing enough, the nutritional value of iron from Spinach is even lower than what is properly measured because of the presence of oxalic acid that blocks iron absorption. A 13.5 oz can of spinach may have 6.6 mg of iron but oxalic acid limits its availability to 1 mg.

This is why we use “spinach” as a CTRI code-word for research error promoted as fact. How much ‘spinach’ do we have in our business practices?

Planning based on goals, measuring based on outcomes, and mitigating risks such as longevity require valid answers to important questions. Specific research papers will answer some questions but not other questions.

WHAT ARE THE LEGITIMATE QUESTIONS THAT A RESEARCH PAPER CAN ANSWER?

Researchers use statistics to develop the answers to their questions. Similarly, we can evaluate papers, and thus the legitimacy of their research questions, based on where they fit on a ladder of statistical processes. Judea Pearl calls this ladder of statistical processes and related questions the Ladder of Causation.

The ladder has three rungs:

- Descriptive statistical processes
- Prescriptive statistical processes
- Predictive statistical processes

When we look at a research paper, we first ask: What type of questions does this paper seek to answer? Are they “What if I See?” questions, “What if I Do?” questions, or “Why?” questions?

Papers that ask “What if I See?” questions are observational, data-driven research processes looking for associations between variables using statistics

and probabilities. Such papers answer descriptive questions about population averages:

- Would seeing variable X change my belief in variable Y?
- What can a sample from a population tell us about the population's average?
- What can a symptom tell us about a problem?
- What population-level information can we summarize from observing the data?

Papers that ask “What if I Do?” questions are interventional, experiment-driven research processes looking for differences between tested populations using Randomized Controlled Trials (RCTs).

Such papers answer prescriptive questions about population averages:

- What happens to Y if we do X?
- What are the estimated effects (safety, effectiveness, efficiency) if we take this course of action?
- What population-level interpretations can we make from the data above and beyond summarizing the data in order to make a prescriptive recommendation?

Papers that ask “Why?” questions are causal research processes based on the development of counterfactuals and thought-experiments using structural causal models (SCMs) to develop predictive statements for specific events in the data set. Such papers answer predictive questions about specific individuals:

- What would have happened to Y if X had not happened?
- Was it X that caused Y?
- Would the client's outcomes have improved if we had not made this decision?

A great danger in research is trying to claim something that it cannot support, such as an observational research paper trying to recommend a prescriptive action or a causal explanation.

For instance, if observational data from a study of tens of thousands of retirees shows that income and marriage have a high positive correlation, would this

result make it a justified and protective best practice to provide a positive answer when un-married clients ask if their income will increase if they get married?

Additionally, would such a result justify public statements by the research's authors about the potential fiduciary risk to advisors who do not recommend marriage as a retirement planning tool in the best interest of the client?

In the absence of looking at (i) the type of research involved, (ii) the list of variables and related assumptions, and (iii) the choice of analytical model, there are common sense situations where we can rely on intuition to guide our thinking. However, when we move past simple problems to where intuition can no longer guide the analysis, we need to formalize and automate our ability to read research in order to sort the over-reaching chaff from the justified and protective wheat.

If you do not have such formalized automation to support your reading of financial research, you will benefit from using this first, do-it-yourself step of determining the types of questions the research can legitimately answer. Developing this habit of mind will give you a first-order screen to identify questions that you can, before you look at the answers, expect to bring safely into your practice.

HOW CAN WE FORM AN OPINION ABOUT THE VALIDITY OF ANSWERS TO LEGITIMATE QUESTIONS?

Now that we have a process to determine the legitimate questions in the research, we need a second process to determine the validity of the answers.

Research questions are based on models of reality because reality is too complicated to question directly. The validity of the answers from the model comes from the degree of fitness between the model and reality, akin to the process followed by a tailor fitting flat, two-dimensional cloth panels to a curved, three-dimensional body. At what point do you know that

this suit really looks good on you instead of hearing an unjustified, conflicted opinion?

Note that we have now turned our focus from understanding the legitimacy of the questions that can be answered from the research, to understanding the specific data and how the data were generated in order to validate the degree of fitness of the answers.

This part of the process brings assumptions to the surface, to create transparency, and thus an estimate of validity:

- What are the assumptions behind the variables?
- What are the assumptions behind the model?
- Did the research test for the distortions that may derive from the statistical machinery?

Assumptions are often amputations of reality because they are so restrictive as to be un-realistic. How much validity would you give to your tailor's work if a key feature of their data set and tailoring process was the elimination of waist sizes smaller than a 34 and larger than 38? That sub-population average may still look close enough to the overall population average. On the other hand, how many of us would find the results to be suits that do not fit us at all? What do you know about the assumptions behind your foundational research papers?

ASSUMPTIONS ABOUT RESEARCH VARIABLES SET THE STAGE FOR COMPARABILITY

Variables for statistical analysis, and the events that quantify their values, can be dependent or independent of one another. Research seeks to discover dependent relationships between input variables and output variables. Valid input/output relationships justify our choices of advisory best practices.

One of the key assumptions, necessary for the proper functioning of statistical, mathematic machinery, is the independence of input variables. This is important because a lack of independence among input variables would create spurious connections

and correlations that would affect the results in hidden and mistaken ways.

We now seek to document data generation assumptions such as:

- What is the full list of input and output variables?
- How were the values of the input variables collected?

Can you list the variables used in the specific research paper that you are reading? Look at the list, does it seem reasonable to attach the assumption of independence to all input variables? How does this list of input variables compare to other studies you are reading?

There are other, probability-based assumptions about input variables that you should know in order to qualify the validity and thus the applicability of the answers in your practice. Some of these assumptions may be hard to find explicitly in the papers and illustrate the need for the automated documentation processes and artificial intelligence platforms. This is a key focus of CTRI.

Examples of such probability-based assumptions include:

- Continuous change in the values of variables, as opposed to discontinuities over the range of the values
- Identically distributed values of the variables as opposed to variables operating under different shapes of probability distributions
- Normal distribution of the values of the variables as contrasted with distributions with non-normal mean, variance, skew, kurtosis, etc. such as Power Curves (e.g. 80/20 rule)

Data collection processes range from biased samples (such as using the students in the researcher's class because they are easily accessed and directed) to fully randomized samples with control groups. Clearly, a research paper based on the former will not have the same intrinsic validity as the latter.

The section you just read extended our understanding of the data from the statistical values of the variables, to the extra-statistical assumptions behind the variables and the collection processes used to generate the values of the variables. We will now turn our attention to the assumptions behind the statistical models.

ASSUMPTIONS ABOUT THE RESEARCH MODEL SETS THE STAGE FOR STATISTICAL PARADOXES

Key questions, to help document the assumptions behind the statistical model that processes the data related to the variables, include:

- What are the nodes and connections in the model and can you create a graph of it?
- What are the structural equations representing the nodes and connections in the model? (Note: this is the mathematization of the graph.)
- How was the model implemented and calculated (e.g. Excel model, statistical package, custom app, etc.?)

This level of reading research papers moves from making lists in writing, to extracting mathematical equations and algorithmic code. This will take more time than advisors can put into the evaluation of each and every paper. This also illustrates the need for the automated documentation processes and artificial intelligence platforms, a key focus of CTRI.

Advisors can see for themselves that research papers based on Excel models that process a small number of artificially created examples should have much less validity (if any) than papers based on well documented structural equations processing large amounts of randomized samples with control groups.

In the previous section, we continued our exploration of extra-statistical features of the data by including the extra-statistical features of the methods that can be used to model and interpret the findings. We are now turning our attention to analytical limitations that are inherent to the use of statistical machinery.

THE LIMITS OF STATISTICAL MACHINERY CREATE PARADOXES

Financial research deals with lists of variables, data sets, and probability distributions that are too large for a direct examination of each and all possible combination of the values. Statistics is a body of knowledge that makes it possible to summarize, albeit with some loss of information, the relevant features of the values and probabilities.

These simplifying representations of statistical associations include concepts that are familiar to advisors, including measures of expected values, variances, co-variances, and regressions. However, the loss of information that derives from using statistical measures can be significant and will sometimes create paradoxes.

Statistics operate at the level of population averages and emphasize that correlation is not causation. However, our minds interpret data, implicitly or explicitly, based on causal assumptions. This creates space for paradoxes that cannot be resolved with statistical methods.

For instance, advisors need to know about Simpson's paradox. In this statistical paradox, the statistical associations that can be calculated in sub-populations, e.g. genders, contradict the findings at the overall population level.

Simpson's example (1951) presents a drug test on a general population, thus both male and female patients, where those who took the drug recovered at a lower rate than those who did not take the drug. However, when the data is analyzed by gender, more men who took the drug recovered than men who did not, and more women who took the drug recovered, than women who did not. Thus, the findings show a practice (a drug in this case) that appears to help men as well as women but does not help the general population. As you can see, this is paradoxical, not in the expected order of things.

This is important to know, when reading research, because the sub-populations could be types of clients such as high net-worth, affluent, mass-market, over-funded, constrained, under-funded, etc. Findings and recommendations that you would bring into your practice based on results that apply to the overall population may reverse when applied to the specific type of clients that you serve.

How many research papers have you read that did not test for Simpson's paradox? How many current "best practices" at the population level could be labeled "not so good" or worse for your specific target market?

CONCLUSION

In closing, let's summarize the two-step, framework that you can use to establish your level of confidence in the legitimacy of the questions and the validity of the answers, that you read in financial research papers.

The first step is to understand the type of questions that a research paper can legitimately ask and answer. Are they "What if I See?" questions, "What if I Do?" questions, or "Why?" questions?

The second step brings the research paper's assumptions to the surface and thus an estimate of validity of the answers to the legitimate questions:

These two steps will help you identify and qualify the legitimate findings, as opposed to the over-reaching findings, that can be claimed by a research paper. This process should help you improve your ability to translate research into retirement planning and retirement management best practices. CTRISM automation, described in the appendix, will take you to the next level of performance and protection.

SOURCES

- 2018, *The Book of Why*, Judea Pearl and Dana McKenzie, Basic Books
- 2017, *The Shapes of Retirement Planning: Are You a Curve, a Triangle, or a Rectangle?*, Patrick Collins and François Gadenne, *Investments & Wealth Monitor*, July/August 2017
- 2016, *Causal Inference in Statistics*, Judea Pearl, Madelyn Glymour, Nicholas P. Jewell, Wiley
- 2014, *Willful Ignorance, The Mismeasure of Uncertainty*, Herbert Weisberg, Wiley
- 2012, *Models. Behaving. Badly.* Emanuel Derman, Simon and Schuster

APPENDIX

WHAT IS CTRI?

Launched in January 2018, The Curve, Triangle, & Rectangle InstituteSM (CTRISM) is a 501(c)6, not-for-profit, membership association focused on the translation of financial research into action.

THE PROBLEM

Longevity risk is not only a planning risk for clients, it is also a practice risk for retirement advisors who often work with clients to their time of death and beyond. Both investment and retirement advisors have business models that are akin to selling a put option, receiving a premium in exchange for being held responsible on the downside.

Unlike investment advisors, however, retirement advisors and their solution providers are selling a longer-duration put option. The longer the duration, the greater the industry's risk. Advisors and providers use research, academic and otherwise, to justify their business practices. How do you know that your retirement best practices are good enough?

UNIQUE ADVANTAGE

CTRI continues and extends the 12 years of research from the Retirement Income Industry Association[®] (RIIA[®]) whose brand and professional designation (the RMA[®]) were acquired by the Investments & Wealth Institute[®] (formerly IMCA[®]) in 2017.

RIIA's research was summarized in the paper, "The Shapes of Retirement: Are you a Curve, a Triangle, or a Rectangle?" published in the July/August 2017 issue of IMCA's journal, the *Investments and Wealth Monitor*.

VALUE PROPOSITION

CTRI turns research into action by curating important research papers, extracting features from the papers in a meta-database, structuring the features in the form of a Wiki for review by members, authors, and peers, and completing the process with machine reviews performed by Artificial Intelligence learners. The AI learners, a Recommender system, a Prescriber system, and a Predictive system seek to discover deeper connections across industry silos.

This combination of academic research and fintech technologies provides CTRI members with the following benefits:

- Controlled experimentation with disruptive technologies
- Access to cross-silo thinking and solutions
- Managed innovation risk by keeping the necessary failures on a small scale before the vetted, winning solutions can be scaled up internally

THE SOLUTION

Planning based on goals, managing performance based on outcomes, and mitigating risks such as longevity require valid answers to important questions. Specific research papers will answer some questions but will not answer other questions. What are the legitimate questions that a specific research paper can answer? What are the valid answers to these questions?

CTRI's combination of descriptive, prescriptive, and predictive analytics translates foundational research into action by providing members with a higher level of industry-wide sources and insights for product development, advisory process solutions, and client engagement models.