49

Converting rejections into positive stimuli

Donald B. Rubin
Department of Statistics
Harvard University, Cambridge, MA

> *"It's not that I'm so smart, it's just that I stay with problems longer."*
> – Albert Einstein

At first glance this Einstein quotation may seem to have little to do with my title, but those readers who know something of Einstein's early life will recall that these years were not full of recognized scientific successes, but he kept working on his problems. And that is certainly related to why I chose the quote, but there is more to it. I have been fortunate to have had many journal publications, but less than one percent were accepted at first submission — far more were immediately rejected, followed closely by those that were rejected accompanied with the suggestion that it would not be wise to resubmit. However, I cannot think of an instance where this nasty treatment of my magnificent (self-assessed) work (sometimes joint) did not lead to a markedly improved publication, somewhere. In fact, I think that the drafts that have been repeatedly rejected by many different journals possibly represent my best contributions! Certainly the repeated rejections, combined with my trying to address various comments, led to better exposition and sometimes better problem formulation as well.

So here, in an attempt to inspire younger researchers to stay the course, I'll relay some of my stories on the topic, of course using some of my own publications as examples. I'll give only a short summary of each example, hopefully just enough for the reader to get the basic idea of the work (or possibly even read it, or as my wonderful PhD advisor, Bill Cochran, used to say, "I'd prefer if you read it and understood it, but if not, please read it; failing that, just cite it!"). For potential interest, I'll insert the approximate number of `Google Scholar` cites as of August 1, 2013. These counts may be of interest because the relationship between the number of citations and my memory of the paper's ease of acceptance appears to me to be zero (excluding the EM outlier). So young writers, if you think you have a good idea that reviewers do not appreciate, you're not alone, and quite possibly on to a very good idea, especially if the reviewers come across as real experts in their reports, but appear to have off-target comments.

49.1 My first attempt

> "A non iterative algorithm for least squares estimation of missing values in any analysis of variance design." *Journal of the Royal Statistical Society, Series C*, vol. 21 (1972), pp. 136–141. [Number of citations: 58]

This was my first sole-authored submission, and of course, I thought it was very clever, combining simple matrix manipulations with simple computations to generalize an old "Rothamsted" (to use Cochran's word) method to fill in missing data in an experimental design with their least squares estimates — a standard objective in those days (see the target article or Little and Rubin (2002, Chapter 2), for the reason for this objective). When I submitted this, I was still a PhD student, and when I received the report and saw "tentative reject," I was not a happy camper. Cochran calmed me down, and gave me some advice that he learned as a wee Scottish lad on the links: Keep your eye on the ball! Meaning, the objective when writing is to communicate with your readers, and the reviewers are making useful suggestions for improved communication. He went on to say:

> "The Editor is not your enemy — at this point in time, he has no idea who you even are! The Editor sent your draft to people who are more experienced than you, and they are reading it without pay to help you and the journal."

I was calm and the paper was accepted, a revision or two later. I was only fully calm, however, until the next "tentative reject" letter a few months later.

49.2 I'm learning

> "Matching to remove bias in observational studies." *Biometrics*, vol. 29 (1973), pp. 159–183. Printer's correction note in vol. 30 (1974), p. 728. [Number of citations: 392]

> "The use of matched sampling and regression adjustment to remove bias in observational studies." *Biometrics*, vol. 29 (1973), pp. 184–203. [Number of citations: 321]

This pair of submissions was based on my PhD thesis written under Bill's direction — back-to-back submissions, meaning both were submitted at the same time, with the somewhat "aggressive" suggestion to publish them back-to-back if they were acceptable. Both were on matched sampling, which at the time was really an unstudied topic in formal statistics. The only publication that was close was the wonderful classic Cochran (1968). Once again, a tentative rejection, but this time with all sorts of misunderstandings, criticisms and suggestions, that would take voluminous amounts of time to implement, and because at the time I was faculty in the department, I was a busy boy! I again told Bill how furious I was about these reviews, and Bill once again calmed me down and told me to remember what he had said earlier, and moreover, I should realize that these reviewers had spent even more time trying to help me, and that's why their comments were so long. Of course, Bill was correct, and both papers were greatly improved by my addressing the comments — not necessarily accepting the suggestions but addressing them. This lesson is important — if a reviewer complains about something and makes a suggestion as to how things should be changed, you as the author, needn't accept the reviewer's suggestion, but you should fix that thing to avoid the criticism. I was beginning to learn how to communicate, which is the entire point of writing journal articles or books.

49.3 My first JASA submission

> "Characterizing the estimation of parameters in incomplete data problems." *Journal of the American Statistical Association*, vol. 69 (1974), pp. 467–474. [Number of citations: 177]

This article concerns factoring likelihoods with missing data, which presented generalizations and extensions of prior work done by Anderson (1957) and Lord (1955) concerning the estimation of parameters with special patterns of missing data. Here, the editorial situation was interesting because, when I submitted the draft in 1970, the *JASA* Theory and Methods Editor was Brad Efron, whom I had met a couple of years earlier when he visited Harvard, and the Associate Editor was Paul Holland, my good friend and colleague at Harvard. So, I thought, finally, I will get a fast and snappy acceptance, maybe even right away!

No way! Paul must have (I thought) selected the most confused mathematical statisticians in the world — these reviewers didn't grasp any of the insights in my wonderous submission! And they complained about all sorts of irrelevant things. There is no doubt that if it hadn't been for Paul and Brad, it would have taken years more to get it into *JASA*, or would have followed the path of Rubin (1976) described below, or far worse. They were both helpful in explaining that the reviewers were not idiots, and actually they had some decent suggestions, properly interpreted — moreover, they actually liked the paper — which was very difficult to discern from the reports written for my eyes. Another set of lessons were apparent. First, read between the lines of a report: Editors do not want to over commit for fear the author won't pay attention to the suggestions. Second, reinterpret editorial and reviewer's suggestions in ways that you believe improve the submission. Third, thank them in your reply for suggestions that improved the paper — they did spend time writing reports, so acknowledge it. Fourth, it does help to have friends in positions of power!

49.4 Get it published!

> "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology*, vol. 66 (1974), pp. 688–701. [Number of citations: 3084]

This paper is the one that started my publishing trail to use the potential outcomes notation to define formally causal effects in all situations, not just in randomized experiments as in Neyman (1923). Actually Neyman said he never made that generalization because he never thought of it, and anyway, doing so would be too speculative; see Rubin (2010) for the story on this. Everyone dealing with non-randomized studies for causal effects was using the observed value notation with one outcome (the observed value of the outcome) and one indicator variable for treatments until this paper. So in fact, Rubin (1974a) was the initiating reason for the phrase "Rubin Causal Model" — RCM, coined in Holland (1986).

I wrote this in some form when I was still at Harvard, more as notes for an introductory statistics course for psychologists. Someone suggested that I spruce it up a bit and submit it for publication. I did, but then couldn't get it published anywhere! Every place that I submitted the piece, rejected it. Sometimes the reason was that "every baby statistics student knows this" (I agreed that they should, but then show me where it is written!); sometimes the reason was "it's completely wrong"! And, in fact, I just received (July 2013) an email stating that "the Rubin definition of 'causality' is not appealing to many eminent statisticians." Sometimes the comments were even insulting, especially so because I was submitting statistical work from my position at Educational Testing Service (ETS) rather than a respected university statistics department. I asked around ETS and someone suggested the place, the *Journal of Educational Statistics*, where it ended up — I think that the acceptance was because of some high level intervention from someone who did like the paper but, more importantly, wanted to get me off his back — I honestly do not remember whom to thank.

There are several lessons here. First, it demonstrates that if a publication is good and good people find out about it (again, it helps to know good people), it will get read and cited. So if you are having this kind of problem with something that you are convinced is decent, get it published somewhere, and start citing it yourself in your own publications that are less contentious, and nag your friends to do so! Second, if you are repeatedly told by some reviewers that everyone knows what you are saying, but without specific references, and other reviewers are saying what you are writing is completely wrong but without decent reasons, you are probably on to something. This view is reinforced by the next example. And it reiterates the point that it does help to connect with influential and wise people.

49.5 Find reviewers who understand

> "Inference and missing data." *Biometrika*, vol. 63 (1976), pp. 581–592 (with discussion and reply). [Number of citations: 4185]

This article is extremely well known because it established the basic terminology for missing data situations, which is now so standard that this paper often isn't cited for originating the ideas, although often the definitions are summarized somewhat incorrectly. As Molenberghs (2007) wrote: "... it is fair to say that the advent of missing data methodology as a genuine field within statistics, with its proper terminology, taxonomy, notation and body of results, was initiated by Rubin's (1976) landmark paper." But was this a bear to get published! It was rejected, I think twice, from both sides of *JASA*; also from *JRSS B* and I believe *JRSS A*. I then decided to make it more "mathy," and I put in all this measure theory "window dressing" (a.s., a.e., both with respect to different measures because I was doing Bayesian, repeated sampling and likelihood inference). Then it got rejected twice from *The Annals of Statistics*, where I thought I had a chance because I knew the Editor — knowing important people doesn't always help. But when I told him my woes after the second and final rejection from *The Annals*, and I asked his advice on where I should send it next, he suggested "Yiddish Weekly" — what a great guy!

But I did not give up even though all the comments I received were very negative; but to me, these comments were also very confused and very wrong. So I tried *Biometrika* — home run! David Cox liked it very much, and he gave it to his PhD student, Rod Little, to read and to contribute a formal comment. All those prior rejections created, not only a wonderful publication, but lead to two wonderful friendships. The only real comment David had as the Editor was to eliminate all that measure theory noise, not because it was wrong but rather because it just added clutter to important ideas. Two important messages: First, persevere if you think that you have something important to say, especially if the current reviewers seem not up to speed. Second, try to find a sympathetic audience, and do not give up.

49.6 Sometimes it's easy, even with errors

> "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society, Series B*, vol. 39 (1977), pp. 1–38 (joint work with A.P. Dempster and N. Laird, published with discussion and reply). [Number of citations: 34,453]

Those early years at ETS were tough with respect to getting articles accepted, and I think it *is* tougher submitting from less academically prestigious places. But publishing things became a bit easier as I matured. For example, the EM paper was accepted right away, with even invited discussion. It was to be a read paper in London in 1976, the trip where I met Rod Little and David Cox in person — the latter mentioned that he really wasn't fond of the title of the already accepted Rubin (1976) because something that's missing can't be "given" — the Latin meaning of data. And this rapid acceptance for the EM paper was despite having one of its proofs wrong — misapplication of the triangle inequality! Wu (1983) corrected this error, which was not critical to the fundamental ideas in the paper about the generality of the missing data perspective. In statistics, ideas trump mathematics — see Little's (2013) Fisher lecture for more support for this position. In this case, a rapid acceptance allowed an error to be published and corrected by someone else. If this can be avoided it should be, even if it means withdrawing an accepted paper; three examples of this follow.

49.7 It sometimes pays to withdraw the paper!

It sometimes pays to withdraw a paper. It can be good, it can be important, and even crucial at times, as the following examples show.

49.7.1 It's good to withdraw to complete an idea

> "Parameter expansion to accelerate EM: The PX-EM algorithm." *Biometrika*, vol. 85 (1998), pp. 755–770 (joint work with C.H. Liu and Y.N. Wu). [Number of citations: 243]

This submission was done jointly with two exceptionally talented former PhD students of mine, Chuanhai Liu and Ying Nian Wu. It was a technically very sound article, which introduced the PX-EM algorithm, an extension of EM. If correctly implemented, it always converged in fewer steps than EM — nice. But after the submission was accepted by an old friend, Mike Titterington at *Biometrika*, there was an intuitive connection that I knew had to be there, but that we had not included formally; this was the connection between PX-EM and ANCOVA, which generally creates more efficient estimated treatment effects by estimating a parameter whose value is known to be zero (e.g., the difference in the expected means of covariates in the treatment and control groups in a completely randomized experiment is zero, but ANCOVA estimates it by the difference in sample means). That's what PX-EM does — it introduces a parameter whose value is known, but estimates that known value at each iteration, and uses the difference between the estimate and the known value to obtain a larger increase in the actual likelihood. But we hadn't done the formal math; so we withdrew the accepted paper to work on that.

Both Chuanhai and Yingnian were fine with that decision. My memory is that we basically destroyed part of a Christmas holiday getting the idea down correctly. We were now ready to resubmit, and it was not surprising that it was re-accepted overnight, I think. Another lesson: Try to make each publication as clean as possible — you and your coauthors will have to live with the published result forever, or until someone cleans it up!

49.7.2 It's important to withdraw to avoid having a marginal application

> "Principal stratification for causal inference with extended partial compliance: Application to Efron–Feldman data." *Journal of the American Statistical Association*, vol. 103 (2008), pp. 101–111 (joint work with H. Jin). [Number of citations: 65]

This paper re-analyzed a data set from an article (Efron and Feldman, 1991) on noncompliance, but I think that Hui Jin and I approached it more appropriately using principal stratification (Frangakis and Rubin, 2002). I had a decade to ponder the issues, the benefit of two great economics coauthors in the interim (Angrist et al., 1996), a wonderful PhD student (Constantine Frangakis) to help formulate a general framework, and a great PhD student to work on the example. The submission was accepted fairly quickly, but as it was about to go to the Copy Editors, I was having my doubts about the last section, which I really liked in principle, but the actual application didn't make complete scientific sense, based on my experience consulting on various pharmaceutical projects. So I wanted to withdraw and to ask my coauthor, who had done all the extensive computing very skillfully, to do all sorts of new computing. Her initial reaction was something like: Had I lost my mind? Withdraw a paper already accepted in *JASA*?! But wasn't the objective of writing and rewriting to get the paper accepted? But after listening to the reasons, she went along with my temporary insanity, and she executed the final analyses that made scientific sense with great skill and care. Of course the paper was re-accepted. And it won the Mitchell prize at the Joint Statistical Meetings in 2009 for the best Bayesian paper.

The message here is partly a repeat of the above one regarding publishing the best version that you can, but it is more relevant to junior authors anxious for publications. I surely know how difficult it can be, certainly in the early years, to build a CV and get promoted; but that's short term. Eventually real quality will triumph, and don't publish anything that you think may haunt you in the future, even if it's accepted in a top journal. As Pixar's Jay Shuster put it: "Pain is temporary, 'suck' is forever." By the way, Hui Jin now has an excellent position at the International Monetary Fund.

49.7.3 It's really important to withdraw to fix it up

> "Multiple imputation by ordered monotone blocks with application to the Anthrax Vaccine Research Program," *Journal of Computational and Graphical Statistics*; 2013; in press (joint work with F. Li, M. Baccini, F. Mealli, E.R. Zell, and C.E. Frangakis)

This publication hasn't yet appeared, at least at the time of my writing this, but it emphasizes the same point, with a slightly different twist because of the multiplicity of coauthors of varying seniority. This paper grew out of a massive joint effort by many people, each doing different things on a major project. I played the role of the MI-guru and organizer, and the others were absorbed with various computing, writing, and data analytic roles. Writing a document with five major actors was complicated and relatively disorganized — the latter issue, my fault. But then all of a sudden, the paper was written, submitted, and remarkably the first revision was accepted! I now had to read the entire thing, which had been "written" by a committee of six, only two of which were native English speakers! Although some of the writing was good, there were parts that were confusing and other parts that appeared to be contradictory. Moreover, on closer examination, there were parts where it appeared that mistakes had been made, mistakes that would take vast amounts of time to correct fully, but that only affected a small and orthogonal part of the paper. These problems were really only evident to someone who had an overview of the entire project (e.g., me), not reviewers of the submission.

I emailed my coauthors (some of whom were across the Atlantic) that I wanted to withdraw and rewrite. Initially, there seemed to be some shock (but wasn't the purpose of writing to get things published?), but they agreed — the more senior authors essentially immediately, and more junior ones after a bit of contemplation. The Editor who was handling this paper (Richard Levine) made the whole process as painless as possible. The revision took months to complete; and it was re-accepted over a weekend. And I'm proud of the result. Same message, in some sense, but wise Editors want to publish good things just as much as authors want to publish in top flight journals.

49.8 Conclusion

I have been incredibly fortunate to have access to sage advice from wonderful mentors, obviously including advice about how to react to rejected submissions. It may not always be true, and I do know of some gross examples, but in the vast majority of cases, Editors and reviewers are giving up their time to try to help authors, and, I believe, are often especially generous and helpful to younger or inexperienced authors. Do not read into rejection letters personal attacks, which are extremely rare. The reviewers may not be right, but only in rare situations, which I believe occur with submissions from more senior authors, who are "doing battle" with the current Editors, is there any personal animus. As Cochran pointed out to me about 1970, they probably don't even know anything about you, especially if you're young. So my advice is: Quality trumps quantity, and stick with good ideas even when you have to do battle with the Editors and reviewers — they are not perfect judges but they are, almost uniformly, on your side.

Whatever wisdom is offered by this "fireside chat" on dealing with rejections of journal submissions, owes a huge debt to the advice of my mentors and very respected folks along my path. So with the permission of the Editors of this volume, I will follow with a description of my incredible good fortune to meet such folks. As one of the wisest folks in our field (his name is hidden among the authors of the additional references) once said to me: If you ask successful people for their advice on how to be successful, their answers are, "Be more like me." I agree, but with the addition: "And meet wonderful people." This statement creates a natural transition to the second part of my contribution to this 50th anniversary volume, on the importance of listening to wise mentors and sage colleagues. I actually wrote the second part before the first part, but on rereading it, I feared that it suffered from two problems; one, it sounded too self-congratulatory and almost elitist. The Editors disagreed and thought it actually could be a helpful chapter for some younger readers, perhaps because it does illustrate how good fortune plays such an important role, and I certainly have had that with respect to the wonderful influences I've had in my life. The advice: Take advantage of such good fortune!

References

Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 94:444–472.

Anderson, T.W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, 52:200–203.

Cochran, W.G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, 24:295–313.

Dempster, A.P., Laird, N., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38.

Efron, B. and Feldman, D. (1991). Compliance as an explanatory variable in clinical trials. *Journal of the American Statistical Association*, 86:9–17.

Frangakis, C.E. and Rubin, D.B. (2002). Principal stratification in causal inference. *Biometrics*, 58:21–29.

Holland, P.M. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81:945–960.

Jin, H. and Rubin, D.B. (2008). Principal stratification for causal inference with extended partial compliance: Application to Efron–Feldman data. *Journal of the American Statistical Association*, 103:101–111.

Li, F., Baccini, M., Mealli, F., Zell, E.R., Frangakis, C.E., and Rubin, D.B. (2013). Multiple imputation by ordered monotone blocks with application to the Anthrax Vaccine Research Program. *Journal of Computational and Graphical Statistics*, in press.

Little, R.J.A. (2013). In praise of simplicity not mathematistry! Ten simple powerful ideas for the statistical scientist. *Journal of the American Statistical Association*, 108:359–369.

Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd edition. Wiley, New York.

Liu, C.H., Rubin, D.B., and Wu, Y.N. (1998). Parameter expansion to accelerate EM: The PX-EM Algorithm. *Biometrika*, 85:755–770.

Lord, F.M. (1955). Estimation of parameters from incomplete data. *Journal of the American Statistical Association*, 50:870–876.

Molenberghs, G. (2007). What to do with missing data? *Journal of the Royal Statistical Society, Series A*, 170:861–863.

Neyman, J. (1923). On the application of probability theory to agricultural experiments: Essay on principles. Section 9. *Roczniki Nauk Rolniczych*, 10:1–51. [English translation of the original Polish article available in *Statistical Science*, 5:465–472.]

Rubin, D.B. (1972). A non iterative algorithm for least squares estimation of missing values in any analysis of variance design. *Journal of the Royal Statistical Society*, *Series C*, 21:136–141.

Rubin, D.B. (1973a). Matching to remove bias in observational studies. *Biometrics*, 29:159–183. [Printer's correction note 30, p. 728.]

Rubin, D.B. (1973b). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, 29:184–203.

Rubin, D.B. (1974a). Characterizing the estimation of parameters in incomplete data problems. *The Journal of the American Statistical Association*, 69:467–474.

Rubin, D.B. (1974b). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701.

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63:581–592.

Rubin, D.B. (2010). Reflections stimulated by the comments of Shadish (2009) and West & Thoemmes (2009). *Psychological Methods*, 15:38–46.

Wu, C.F.J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11:95–103.