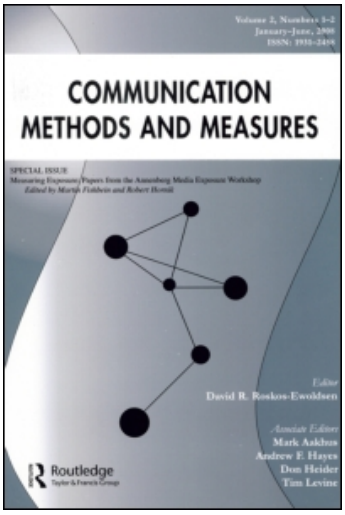

Communication Methods and Measures
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653633

The Asymmetry of Predictive and Descriptive Capabilities in Quantitative Communication Research: Implications for Hypothesis Development and Testing

Daniel J. O'Keefe[a]
[a] Northwestern University,

Online publication date: 14 June 2011

To cite this Article O'Keefe, Daniel J.(2011) 'The Asymmetry of Predictive and Descriptive Capabilities in Quantitative Communication Research: Implications for Hypothesis Development and Testing', Communication Methods and Measures, 5: 2, 113 — 125
To link to this Article: DOI: 10.1080/19312458.2011.568375
URL: http://dx.doi.org/10.1080/19312458.2011.568375

Communication Methods and Measures, *5*(2), 113–125, 2011

ISSN: 1931-2458 print / 1931-2466 online
DOI: 10.1080/19312458.2011.568375

Routledge
Taylor & Francis Group

The Asymmetry of Predictive and Descriptive Capabilities in Quantitative Communication Research: Implications for Hypothesis Development and Testing

Daniel J. O'Keefe
Northwestern University

The bulk of hypotheses in quantitative communication research are directional (e.g., the correlation is positive, the treatment mean is larger than the control mean). For testing such hypotheses, null hypothesis significance testing (NHST) and the use of effect sizes and confidence intervals (ES+CI) are functionally equivalent. ES+CI provides more precise descriptions of research results (effect sizes, confidence intervals) than does NHST, but that descriptive capability exceeds researchers' current predictive capabilities. Developing more refined predictive capabilities will require making good use of the additional information provided by ES+CI—and careful thinking about how such refined hypotheses might be tested.

Null hypothesis significance testing (NHST) has come in for some rough handling in recent years, and rightly so. Certain misunderstandings of NHST have been sufficiently widespread as to be worrisome. For example, the belief that the significance level achieved is an indication of the powerfulness of the effect, or the belief that a nonsignificant result confirms the null hypothesis, or the belief that statistical significance is an indicator of the replicability of results are all properly criticized. The emerging consensus has been that NHST ought to be replaced by an emphasis on effect sizes and confidence intervals (ES+CI), rather than simply whether the null hypothesis is or is not rejected. (For general discussions of such matters, see Harlow, Mulaik, & Steiger, 1997. For specific attention to the context

Correspondence should be addressed to Daniel J. O'Keefe, Department of Communication Studies, Frances Searle Building, Northwestern University, 2240 Campus Drive, Evanston, IL 60208-3545. E-mail: d-okeefe@northwestern.edu

of communication research, see Levine, Weber, Hullett, Park, & Lindsey, 2008; Levine, Weber, Park, & Hullett, 2008.)

A shift from NHST to ES+CI in quantitative communication research raises some complex issues concerning how research hypotheses are expressed and tested. Briefly, the argument of this paper is as follows: NHST and ES+CI provide functionally identical procedures for testing the sorts of hypotheses common in communication research. ES+CI does provide more information than NHST, but this underscores the current gap between our crude predictive capabilities and our refined descriptive capabilities. As ES+CI permits the development of more refined predictions, assessment of those predictions will require some careful thinking about how hypotheses are to be confirmed or disconfirmed. Thus the underlying purpose of this paper is to invite reflection on how communication research hypotheses are formulated and confronted with data. Such reflection requires consideration both of the nature of common hypotheses about communication and the way in which statistical data are brought to bear on such hypotheses.

TESTING DIRECTIONAL HYPOTHESES USING NHST AND ES+CI

NHST and ES+CI represent two different approaches to the treatment of statistical data. But where the question is one of assessing the kinds of hypotheses common in communication research (or in social-scientific research generally), NHST and ES+CI provide what amount to functionally identical procedures. To see this clearly requires understanding just what substantive hypotheses are tested with NHST—a matter unfortunately often obscured in discussions of NHST.

An important source of confusion is a common objection to NHST, namely, that the null hypothesis is already known to almost certainly be false, so testing is not needed in order to reject it.[1] For example: "The null hypothesis, taken literally (and that's the only way you can take it in formal hypothesis testing), is *always* false in the real world. . . . If it is false, even to a tiny degree, it must be the case that a large enough sample will produce a significant result and lead to its rejection. So if the null hypothesis is always false, what's the big deal about rejecting it?" (Cohen, 1990, p. 1308). Or: "The population mean difference may be trivially small but will always be positive or negative. As a consequence we should not set forth a null hypothesis because to do so is unrealistic and misleading" (Jones & Tukey, 2000, p. 412). Or: NHST involves "the testing of a null hypothesis that

[1]Here, "null hypothesis" is used in its familiar sense, to refer to what is sometimes called a "nil hypothesis," that is, a hypothesis of zero effect (no difference, no relationship). There is another broader meaning in which the "null hypothesis" is that hypothesis whose rejection (nullification) constitutes evidence for some particular alternative hypothesis.

could not really be true to begin with" (Loftus, 1996, p. 162) and so "testing the null hypothesis . . . cannot provide new information. All it can do is indicate whether there is enough statistical power to detect whatever differences among the population means must be there to begin with" (p. 163). Or: "The null hypothesis tends to be highly improbable and therefore an inaccurate initial assumption for our tests" (Hullett, 2007, p. 275). Or: "to reject the null hypothesis as false does not tell an investigator anything that was not known already" (Fidler & Loftus, 2009, p. 28). Or: "If the null hypothesis is false anyway, then disproving it is both unimpressive and uninformative" (Levine, Weber, Hullett, et al., 2008, p. 176).

This criticism of NHST is misplaced. It is not a good reason to abandon NHST. Even if one accepts the premise that the null hypothesis is a priori likely to be false, it is a mistake to conclude that NHST must perforce be uninformative. The mistake derives from a misapprehension concerning *why* the null hypothesis might be of interest. The reason the null hypothesis is commonly of interest is not its literal value. Rather, the null hypothesis is characteristically of interest because it marks an important *boundary*.

For example, consider a test of the significance of a sample correlation, where the null hypothesis is that the population correlation is zero (.00000000...). Even though we may justifiably presume that the population correlation is not in fact exactly zero, we may *not* justifiably presume that the population correlation is (say) positive. That is, even though we can appropriately assume that the population correlation is nonzero, we do not know whether the population correlation is positive or negative.

A statistically significant result means that the population correlation is unlikely to be zero; that is, a statistically significant result excludes the null hypothesis from the range of plausible population values. But, more important, a statistically significant result also excludes one *direction* of population effect. By definition, a statistically significant result rules out not only zero as a plausible population value, but also all values on the other side of zero from the obtained sample result: "Rejection of a given null hypothesis implies the rejection not only of the particular null value in question, but also of all of the values in the end of the distribution that is opposite to the end in which the observed value resides" (Cortina & Dunlap, 1997, p. 168).

Conversely, if the sample result is nonsignificant (so that we cannot reject the null hypothesis that the population correlation is .00), then we cannot tell whether the population effect is positive or negative. *That* is why assessing the plausibility of the null hypothesis is important—not because of the literal possibility of a zero population correlation but because of the importance of being able to confidently identify the direction of effect in the population. Being able to identify the direction of the population effect necessarily implies being able to exclude zero as a plausible population effect—and NHST provides precisely a test of whether zero is a plausible population effect (i.e., a test of whether the null hypothesis is plausible).

So although we may be justified in believing from the outset that the population correlation is nonzero—that is, we know the null hypothesis is literally false—we cannot know in advance whether it is false because the population correlation is positive or because the population correlation is negative. Any sample result that includes the null hypothesis as plausible (i.e., any nonsignificant sample result) necessarily includes both positive and negative population correlations as plausible.

One underlying source of confusion here may be an implicit supposition that in NHST one's choice of conclusions is binary (namely, retain or reject the null hypothesis). A clearer picture is that three possible conclusions are available: (1) the direction of effect is positive (X and Y are positively correlated, mean A is larger than mean B, etc.), (2) the direction of effect is negative (X and Y are negatively correlated, mean A is smaller than mean B, etc.), or (3) the direction of effect cannot be determined (for discussion of this point, see Harris, 1997a, 1997b; Tukey, 1991). If the possible conclusions to be drawn from significance tests had been clearly understood as these three options (rather than the two options of retaining or rejecting the null hypothesis), then it might have been more clear that the null hypothesis's being a priori false is not necessarily a reason for avoiding NHST. (Appreciating this point might also have made it less likely that nonsignificant results would have been misinterpreted as confirming the null hypothesis.)

The parallel with effect-sizes-and-confidence-intervals should be apparent: When effect sizes and confidence intervals are computed, researchers will often be interested in whether zero falls in the confidence interval. For example, if the 95% confidence interval around a sample correlation value includes zero, that result is of interest—not because the researcher is interested in the possibility that the population correlation might actually be precisely .00000000... (after all, that is almost certainly false) but rather because if zero falls in the confidence interval, then both positive and negative population correlations are plausible.

So the reason the null hypothesis is important in NHST is the same reason that, in ES+CI, it is important whether the confidence interval includes a zero effect: not because the zero value itself is plausible but because the zero value marks an important boundary between two crucially different substantive alternatives—and hence being able to rule out zero as a plausible possibility necessarily means being able to rule out one of those two substantive possibilities. So even though the null hypothesis (that the effect = .00000000...) may always be false, this does not make a test of the null hypothesis unimpressive or uninformative. Any assessment that permits conclusions about the direction of effect is almost certainly informative, and may be impressive as well.

Against that backdrop, then, consider the current state of theorizing about communication phenomena (or, for that matter, nearly any social-scientific phenomena). Researchers are simply not in a good position to make specific quantitative predictions such as "the population correlation is .37." Instead, predictions

(hypotheses) specify only the *direction* of effect, such as, "the population correlation is positive." (Any doubts about the frequency of directional hypotheses, and the dearth of specific quantitative predictions, can be removed by the briefest glance at primary research reports in communication.)

Perhaps the implication is obvious. Given that predictions about communication phenomena commonly concern only the direction, not the magnitude, of effects, the appropriate statistical procedures are ones suitable to assessing the direction (not the specific magnitude) of effects. Perhaps at some point in the future researchers can be concerned about whether to predict that the correlation is .25 or .40. But currently, researchers are generally prepared to offer only a directional prediction—and for the purposes of assessing such predictions (testing such hypotheses), NHST and ES+CI are equally well-suited.

To put this point another way: With respect to this one specific purpose—the purpose of testing the sorts of hypotheses common in communication research—ES+CI is no better than NHST. Communication research hypotheses are overwhelmingly directional, and for testing such hypotheses, NHST and ES+CI are functionally equivalent. A given data set provides equally good evidence for conclusions about the direction of effect no matter whether the data are analyzed using NHST procedures or ES+CI procedures. For addressing such questions, NHST and ES+CI are "two ways to get the same answer" (Natrella, 1960, p. 20).

To be clear, the claim is not that ES+CI is not an improvement over NHST. The claim is that testing directional hypotheses—the most common sort of hypotheses in communication research—is equally well accomplished by NHST procedures (which tell a researcher whether the direction of effect can be confidently identified, and if so what that direction is) and by ES+CI procedures (which tell a researcher whether the direction of effect can be confidently identified, and if so what that direction is).

ES+CI does, however, offer an advance over NHST for purposes other than the testing of directional hypotheses. Specifically, ES+CI procedures provide valuable information that is obscured by, or unavailable in, common forms of reporting NHST analyses. This difference between NHST and ES+CI is the focus of the next section.

LOOKING AHEAD: USING ES+CI TO FORMULATE RANGE HYPOTHESES

The Current Gap Between Descriptive and Predictive Capabilities

The point of the preceding section might be captured this way: NHST procedures should not be understood as aimed at (literally) assessing the null hypothesis—is the population correlation zero? Is there zero difference between the treatment

condition and the control on the outcome variable? Rather, common NHST procedures can be seen as aimed at assessing the *direction* of effect—is the population correlation positive or negative? Does the population effect favor the treatment condition or the control? A nonsignificant result means that the direction of effect is uncertain; a significant result means that the direction of effect can be confidently identified. However, simply reporting whether a result is statistically significant (as one might do in traditional NHST applications) provides less information than is provided by ES+CI analyses. Reporting effect sizes and confidence intervals is now widely, and appropriately, recognized as valuable.

However, the difference between the information provided by ES+CI and by NHST underscores the current gap between our ability to *describe* research results (where we can give exact effect sizes, with confidence intervals) and our ability to *predict* research results (where, generally, we can sensibly make only directional predictions). To elaborate, given some particular empirical result—say, an experimental comparison of a media-literacy treatment condition and a control treatment, where the outcome of interest is the ability to detect advertising ploys—the result can be described with considerable precision. The effect size (the observed difference between the experimental and control conditions on the outcome variable) can be reported to any number of decimal places, and the confidence interval can be given an equally precise description. But the researchers' prediction will almost certainly have been purely directional (e.g., that the ability to detect advertising ploys would be superior in the treatment condition as compared to the control condition). In short, at present, our ability to *describe* empirical results outstrips our ability to plausibly *predict* empirical results.

Narrowing the Gap: Using ES+CI to Refine Predictions

Given this difference between predictive and descriptive capabilities, one might wonder: if researchers have only directional hypotheses, why should one care about the additional information provided by ES+CI? In fact, there are many reasons for wanting that additional information. For example, effect magnitude information permits one to undertake sensible cost-benefit analyses (because one will know something of the size of effects, not merely their direction). Or, as another example, such information enables the use of equivalence testing, which (in canonical form) asks whether two conditions (e.g., two treatments) differ enough to matter as opposed to being functionally equivalent (for a general discussion, see Wellek, 2002; for discussion of applications in communication, see Levine, Weber, Park, & Hullett, 2008, pp. 199–201).

In the present context, however, the benefit of greatest relevance is in some ways the simplest: Knowing the CI narrows the range of plausible population values. Where NHST procedures indicate only that the population correlation

is positive, confidence-interval information shows that the range of plausible population values is, for example, from .14 to .38.

This, in turn, may underwrite more refined hypotheses in the future—specifically, hypotheses that specify a predicted *range* of effects (e.g., "we expect the population correlation to be between .10 and .40"). For the moment, researchers may need to be content with cruder directional hypotheses, but as evidence accumulates about effect sizes and their associated confidence intervals, researchers may be in a position to offer plausible hypotheses that narrow the range of expected effects.

A range hypothesis might take a variety of forms. "The population correlation is between .10 and .40" ($.10 \leq r_{pop} \leq .40$) identifies a relatively narrow range. A hypothesis such as "the population correlation is at least .15" ($r_{pop} \geq .15$)—or, more generally, "the population correlation is not small" where some criterion is specified for "not small"—identifies a broader range. And, of course, directional hypotheses are range hypotheses with a very broad range; for example, hypothesizing that the population correlation is positive is the equivalent of the hypothesis that $.00 < r_{pop} \leq 1.00$.

Developing more refined range predictions, however, will require not merely that effect sizes and confidence intervals be reported but also that researchers use such information to refine their predictive capabilities. The challenge of this task should not be underestimated. Even researchers who are required to report CIs often ignore CIs in discussing results (Fidler, Thomason, Cumming, Finch, & Leeman, 2004). Thus, reporting effect sizes and confidence intervals is a necessary (and valuable) step, but researchers need to begin to pay attention to that information.

LOOKING AHEAD: ASSESSING RANGE HYPOTHESES

Range Predictions and Confidence Intervals

As more refined range predictions begin to be offered, the natural question to be faced is how such predictions ought to be assessed given some obtained sample result. It would not be proper to assess a range prediction simply by seeing whether the sample value lies within the specified range. Suppose, for instance, that the prediction is "the population correlation is between .20 and .30" and that the sample correlation is .24 ($N = 100$). The observed sample r falls within the specified range, but the 95% CI around that sample value includes values outside the predicted range (the 95% CI has limits of .05 and .42). That is, these data are consistent with a belief that the population correlation is (for example) .10. Because the range of plausible population values includes values outside the predicted range, the prediction cannot be said to be confirmed by the data, even though the sample value falls within the predicted range.

Presumably, a range prediction should be assessed by seeing whether, given the sample data, the range of plausible population values falls entirely within the predicted range—which can be determined by examining the 95% CI around the sample value. If the 95% CI falls entirely inside the predicted range, then the prediction presumably may be taken as confirmed (i.e., not rejected); if any part of the CI falls outside the range, however, the hypothesis is rejected.

Approached in this way, the ability of sample data to shed light on a range hypothesis depends jointly on two considerations (among others): the width of the predicted range and the width of the confidence interval. First: The narrower the predicted range, the more difficult it will be (ceteris paribus) to obtain confirming evidence. Expressed the other way around, as the width of the predicted range increases, the prediction becomes progressively easier to confirm—and correspondingly less informative. Everything else being equal, this procedure makes a prediction such as "The population correlation is between .20 and .80" easier to confirm than a prediction such as "The population correlation is between .20 and .30."

Second, the wider the confidence interval, the more difficult it will be (ceteris paribus) to obtain evidence confirming a predicted range. To be certain of being able to confirm a given range prediction, the CI must be narrower than the predicted range. (If the CI is guaranteed to be wider than the predicted range, then it will be impossible for the CI to fit within the predicted range.) Practically speaking, narrowing the width of the CI requires increasing the number of cases. For instance, for a sample r of .24 to have its 95% CI fall entirely within the range from .20 to .30 would require something in the neighborhood of $N = 2200$.

Directional Predictions (and Statistical Power) Revisited

A directional prediction is a range prediction (e.g., $.00 < r_{pop} \leq 1.00$). When directional predictions are understood in this way, it can easily be seen that NHST provides a procedure precisely parallel to the confidence-interval-based procedure just described for testing range predictions generally. For example, for a directional hypothesis that the population correlation is positive, if the sample correlation is positive and significantly different from zero, then the relevant 95% CI falls entirely within the predicted range (and so excludes zero). If the sample correlation is not significantly different from zero, then the relevant 95% CI includes zero—and includes both positive and negative values (i.e., values outside the specified range), thus leading to the conclusion that the evidence is insufficient to confirm the prediction. That is, the evidentiary basis by which NHST assesses directional hypotheses is precisely the same as that of the ES+CI procedure suggested here for range hypotheses, namely, the answer to the question "is the range of plausible population values entirely contained within the hypothesized range?"

Seeing this clearly permits one to appreciate the continuing importance of statistical power, even if NHST is replaced by ES+CI procedures. One might think

that abandoning NHST will naturally mean that questions of statistical power ("what are the chances of obtaining a statistically significant result?") will become irrelevant. After all, if questions of statistical significance are misplaced, then surely it will no longer be relevant to be concerned about one's chances of obtaining a statistically significant result: "If significance testing is no longer used, then the concept of statistical power has no place and is not meaningful" (Schmidt, 1996, p. 124).

Approached from the perspective of ES+CI, a statistical power figure is simply the answer to the question "How likely is it that these data will permit one to identify the population *direction* of effect?" (i.e., "How likely is it that these data will yield a confidence interval that excludes zero?"). Where statistical power is low, one has little chance of being able to confidently say what the direction of effect is. Given that researchers characteristically are hoping simply to get the direction of effect right (never mind exactly what the effect size is), statistical power is crucial.[2]

Where narrower range predictions are concerned (e.g., ".20 $\leq r_{\text{pop}} \leq$ .40,"), questions about statistical power can be reformulated as a matter of the width of the confidence interval. Instead of asking, "How likely is it that these data will permit identification of the population direction of effect?" (the NHST statistical power question), researchers will ask, "How likely is it that these data will yield a confidence interval sufficiently narrow to fit within the predicted range?" Even in the absence of such narrower range predictions, of course, narrower confidence intervals are desirable, because they constrict the range of plausible population values. But, as researchers come to offer narrower range predictions, familiar statistical-power questions will surface in new (but parallel) forms. For example, just as under NHST procedures researchers can engage in sample-size planning so as to ensure sufficient power to detect the expected population effect size, so under ES+CI procedures researchers can engage in similar planning so as to ensure a sufficiently narrow CI to confirm the expected population range.

A PROCESS-BASED INTERPRETATION OF NHST

The preceding discussion has been phrased in terms that will be familiar to communication researchers, with the primary inferential task being treated as a matter

[2]This is a simplified treatment, because statistical power depends on a variety of factors, including the putative size of the population effect. So an answer to the question "How likely is it that these data will permit identification of the population direction of effect?" depends in part on the population effect. The larger the population effect (everything else being equal), the more likely it is that a study's confidence interval will exclude zero, that is, the more likely it is that a study will be able to confidently identify the direction of effect.

of generalizing from a sample to a population. It is worth noticing, however, that the same arguments and conclusions are also entirely compatible with a rather different—and arguably more attractive—way of conceptualizing NHST, namely, as a basis for conclusions about *processes* rather than *populations* (e.g., Frick, 1998).

On this alternative view, one thinks of data not as a sample drawn from a population but as "output of a noisy process," output that potentially contains a mixture of signal and noise (Konold & Pollatsek, 2002, p. 264). Questions of inference to some "population" are put aside in favor of questions about the identification of the underlying processes that gave rise to the data, questions about whether a signal is present. The purpose of statistical significance testing, from this perspective, is to assess whether the observed data could be the result of chance: "Rejecting the null hypothesis would allow the conclusion that the difference between groups was not caused by chance fluctuation in the process but instead by some systematic difference in the treatment of the two groups" (Frick, 1998, p. 530).[3]

This process-based approach has been motivated, at least in part, by the felt need to give a more realistic account of the uses of NHST. For example, "a process view better covers the range of statistical situations in which [researchers] are interested" such as those that "have no real population" (Konold & Pollatsek, 2002, p. 265).[4] As another example, as Frick (1998) has pointed out, in experimental social-scientific work "experimenters rarely make any attempt to randomly sample." But that need not be problematic for NHST: "The assumption of random sampling from a population is . . . unnecessary" (p. 527) because "there is another justification for conventional statistical tests that does not rely on the assumption of random sampling" (pp. 529–530). The alternative assumption is that some process produced the data, with one candidate process being random chance (and another candidate process being some nonchance process). Hence, on this approach, "the null hypothesis is phrased in terms of process, so rejecting the null hypothesis leads to a conclusion about process" (p. 530).

One of the attractive aspects of this alternative way of thinking about NHST is that it underscores that significance testing cannot address questions of generalization. Rather than interpreting the results of significance tests as being about some population (to which one seeks to generalize), instead those results are seen as speaking to some process (that gave rise to the data). "With the process-based

[3]Notice that under this process-based interpretation of NHST, the null hypothesis is such as to not necessarily be a priori false. The null hypothesis is that random chance, not a substantive process, gave rise to the observed data. Thusly conceived, the null can indeed be plausibly entertained.

[4]At least where claims about effects of message varieties are concerned—a common kind of question in communication research—there is an argument to be made that message classes should not be conceived of as populations (in the usual sense); see Jackson (1992, pp. 132–136).

interpretation, statistical [significance] testing does not provide generality—it is used to establish the finding" in that particular study (Frick, 1998, p. 531). That is, "rejecting the null hypothesis . . . leads to a conclusion about process, applying to only the subjects in the experiment (e.g., that some difference in the treatment of two groups caused the difference in average scores)" (p. 527).

For present purposes, the relevant point is that the preceding analysis of the relationship of NHST and ES+CI fits quite nicely, mutatis mutandis, within such an alternative framework for understanding NHST. To re-express that analysis briefly: The substantive hypotheses of interest to researchers are commonly ones about underlying processes, but the theoretical understanding of these processes is currently sufficient to yield only directional predictions (that some nonrandom process was at work such that, e.g., the data would yield a positive correlation). With respect to the assessment of such directional hypotheses, NHST and ES+CI are functionally equivalent: Rejection of the null hypothesis permits one to conclude that a nonrandom process was involved, whereas a failure to reject the null hypothesis means that one cannot rule out the possibility that the observed data arose by chance.

However, ES+CI provides additional useful information that is obscured in NHST analyses. This information concerns the size of the observed effect and its plausible range, that is, the strength of the observed signal and the plausible range of values for that signal in the given study. Such information can be valuable for many reasons, such as providing a basis for cost-benefit analyses (balancing the likely size of the effect against the costs of obtaining it) or equivalence testing (e.g., seeing whether a new treatment provides a sufficiently large improvement over existing ones). The most immediate benefit of such information, however, may simply be a deepened understanding of the underlying process being investigated—deepened by virtue of having some sense of the strength of the signal generated by the process in a given study. Such information provides a basis for more refined future hypotheses, ones that specify not simply the direction of effect expected from the hypothesized process but also the rough magnitude of that effect.

On this approach, statistical power matters not because it influences one's ability to generalize to a population, but because it influences one's ability to detect a signal amidst noise, that is, one's ability to rule out chance (as opposed to some substantive process) as an explanation for the observed data. Statistical power represents the answer to the question, "How likely is it that these data will permit one to detect the signal created by the hypothesized underlying substantive process?" Statistical power is thus naturally influenced by, inter alia, the strength of the signal assumed to emanate from the underlying process (rather than a "population" effect size). Where the signal of interest is expected to be relatively weak, researchers will want to take care to plan studies such that there will be some good chance of detecting any such signal, which, in practical terms, commonly

means ensuring a sufficiently large number of cases (e.g., participants). But being able to make sensible estimates of the likely signal strength from a given process will require attending closely to the ES+CI information in previous studies for whatever insight it might provide.

In short, the present analysis of the relationship of NHST and ES+CI remains intact if one replaces the common image of NHST as involving inferences about populations with an alternative (and, it must be said, more plausible) conceptualization that stresses conclusions about underlying processes. One can discard notions of "populations" as targets of generalization and still appreciate both the current gap between our descriptive and predictive capabilities and the potential usefulness of our current descriptive resources in refining our understandings of communication phenomena.

CONCLUSION

Even with increasing use of ES+CI procedures, communication researchers will still largely be testing directional hypotheses. The promise of ES+CI is not that it is a better way to *test* those directional hypotheses. Rather, ES+CI offers a more informative way to *describe* research results, and as a consequence it offers a basis for the long-term development of more refined hypotheses. Developing and testing such hypotheses, however, will require greater attention to the descriptive riches afforded by ES+CI, and more careful thinking about how statistical evidence can be brought to bear on empirical claims.

REFERENCES

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304–1312.

Cortina, J. M., & Dunlap, W. P. (1997). On the logic and purpose of significance testing. *Psychological Methods*, *2*, 161–172.

Fidler, F., & Loftus, G. R. (2009). Why figures with error bars should replace p values: Some conceptual arguments and empirical demonstrations. *Zeitschrift Fur Psychologie*, *217*, 27–37.

Fidler, F., Thomason, N., Cumming, G., Finch, S., & Leeman, J. (2004). Editors can lead researchers to confidence intervals, but can't make them think: Statistical reform lessons from medicine. *Psychological Science*, *15*, 119–126.

Frick, R. W. (1998). Interpreting statistical testing: Process and propensity, not population and random sampling. *Behavior Research Methods, Instruments, and Computers*, *30*, 527–535.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Harris, R. J. (1997a). Reforming significance testing via three-valued logic. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 145–174). Mahwah, NJ: Erlbaum.

Harris, R. J. (1997b). Significance tests have their place. *Psychological Science*, *8*, 8–11.

Hullett, C. R. (2007). Concerns about error go beyond cumulating Type I error: A response to Matsunaga. *Communication Methods and Measures*, *1*, 275–279.

Jackson, S. (1992). *Message effects research: Principles of design and analysis*. New York, NY: Guilford Press.

Jones, L. V., & Tukey, J. W. (2000). A sensible formulation of the significance test. *Psychological Methods*, *5*, 411–414.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, *33*, 259–289.

Levine, T. R., Weber, R., Hullett, C., Park, H. S., & Lindsey, L. L. (2008). A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research*, *34*, 171–187.

Levine, T. R., Weber, R., Park, H. S., & Hullett, C. (2008). A communication researchers' guide to null hypothesis significance testing and alternatives. *Human Communication Research*, *34*, 188–209.

Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science*, *5*, 161–171.

Natrella, M. G. (1960). The relation between confidence intervals and tests of significance. *American Statistician*, *14*(1), 20–22, 38.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, *1*, 115–129.

Tukey, J. W. (1991). The philosophy of multiple comparisons. *Statistical Science*, *6*, 100–116.

Wellek, S. (2002). *Testing statistical hypotheses of equivalence*. Boca Raton, FL: Chapman & Hall/CRC.