



Replicability in agricultural field experiments

William Lockeretz

In reading a report on an agricultural field experiment, a question that one legitimately can (and should) ask is: Under what other conditions would the conclusions still hold? Because agriculture is subject to so many sources of variability, we all know not to expect experimental results to be uniform across highly varied conditions. But we can still hope that they are not very sensitive to minor changes. For example, would the conclusions still hold on a field with the same soil type a few hundred meters away? Would they hold when the management system was changed slightly, let us say by as much as it typically varies among neighboring farmers who all use "standard" production practices? Unless we expect the answer to be "yes" — and certainly if we don't even know the answer — there isn't much point in doing field experiments. Isn't the whole idea behind scientific experimentation that it can teach us something that has broader applicability, something that can be abstracted from the specific circumstances under which the experiment was done?

Unfortunately, customary experimental protocols do not generally deal with this matter. Typically, they do not provide the necessary redundancy, such as more than one location, or more than one version of the same basic management system (apart from variations in the treatment variables

that are the specific focus of study). Except for replication over years — which is customary in publishable field research — replication often means only that a given treatment is represented by multiple plots within the same experimental area. Such an experiment is thus a point measurement, its conclusions holding for sure only at that location, and leaving unaddressed a key question: Where else, if anywhere, do the conclusions hold?

An article in this issue suggests — indirectly and unintentionally — that neglecting this question can be very unfortunate. "Managing interference in a sweet corn-white clover living mulch system," by Albert Fischer and Larry Burrill, has precisely the redundancy that is needed to examine replicability empirically. The result is distressing.

The paper reports two related experiments, both involving a living mulch between rows of sweet corn. (I am referring to the two experiments reported in Tables 3 and 4. Another experiment reported in the same paper, in which planting density was varied systematically (Figs. 1 and 2), is not relevant here.) Although the two experiments are called "separate," there are two treatments that appear in both: row widths of 76 and 38 cm respectively (and correspondingly, stands of 66,000 and 79,000 plants/ha), with mulch. (These correspond to the second and third rows of Tables 3a and 4a in one experiment, and the first and second rows of Tables 3b and 4b in the other.) All other reported management variables (irrigation, fertilization, tillage, etc.) were identical between the two experiments, and the two experimental sites were nearby, on the same soil type.

The two experiments could hardly be more alike. Yet when we compare their results, they can hardly be called alike. The experimental protocols allow us to compare the two experiments with regard to four findings: how crop growth rate (CGR), yield of marketable ears, production of clover dry matter, and leaf area index (LAI) differed between wide and narrow rows in 1985. There is no problem with yield of marketable ears (which, fortunately, is among the most important outcome variables of the study). However, the other three present serious discrepancies.

If the authors, as they might have, had done only the first experiment, their conclusions would have been clear: going from wide to narrow rows raises both CGR and LAI substantially (CGR by 78%, from 7.6 to 13.5; LAI by 100%, from 1.3 to 2.6), and lowers clover dry matter (by 61%, from 1,811 to 711 kg/ha). All these effects were statistically significant (Tables 3a and 4a). But it happens that the other experiment provides us with independent measurements of the same effects, although that was not why the experiment was done. If it had been the only experiment — it, too, is self-contained, and might meaningfully have been done by itself — we would have reached a different conclusion: going from wide to narrow rows increases CGR by only 37% instead of 78% (from 15.2 to 20.9) and LAI by only 24% instead of 100% (from 1.07 to 1.33), and decreases clover dry matter by only 31% instead of 61% (from 1,580 to 1,090), with none of these effects statistically significant (Tables 3b and 4b).

continued on p. 93

The graciousness of Albert Fischer in cooperating with the preparation of this editorial is sincerely appreciated.

William Lockeretz is Technical Editor of *AJAA* and Research Professor, School of Nutrition, Tufts University, Medford, MA 02155.

proaching the Twenty-first Century. Iowa State Univ. Press, Ames. pp. 82-108.

5. Merriam-Webster. 1983. Webster's Ninth New Collegiate Dictionary.

Merriam-Webster Inc., Springfield, Mass.

6. Rzewnicki, P.E., R. Thompson, G.W. Lesoing, R.W. Elmore, C.A. Francis, A.M. Parkhurst, and R.S. Moomaw.

1988. On-farm experiment designs and implications for locating research sites. *Amer. J. Alternative Agric.* 3(4):168-173.



EDITORIAL cont'd from p. 2

What should authors do when faced with such results? The choice that most researchers would probably make (and the one chosen here) would be not to compare results across the two experiments. After all, agriculture is highly sensitive to many sources of variability, and we should not expect that two experiments necessarily would give identical results.

But this answer doesn't deal with the discrepancies; it simply defines them away by declaring the two experiments to be "independent," even though they were identical in all reported conditions. Most agricultural researchers probably would accept this reasoning. To do so is dangerous, however. Not to deal with the apparent discrepancies is to say, in essence, that an experimental result does not necessarily apply *anywhere* else (or that one is not on guard against the possibility of having made a serious mistake). Researchers are accustomed to acknowledging that a specific result might not apply on a different soil type, or under a different production practice. But are we prepared to acknowledge something much more worrisome, namely, that a conclusion reached at one location does not let us confidently say anything about *any* other place whatever?

My interpretation of the discrepancies in the present paper is not that agricultural phenomena are so variable that it is hopeless to try to learn anything from a field experiment. Rather, it is that we must try harder. The discrepancies must have an explanation. However, the explanation lies outside what was taken into account in the paper, because in all reported conditions the two experiments were *identical*. Therefore, either a mistake was made, or there was an important but overlooked difference between the experimental sites.

Clearly, there is a problem here. Yet both experiments — either of which might have been published alone, in which case there would have been no indication of any

problem — followed standard procedures. Maybe this means that standard procedures are not always good enough. Perhaps they should routinely offer more protection against mistakes, through redundancies that provide us with consistency checks (as happened to be available in this paper, although not by intention). Second, to the extent feasible, the experimental design should encompass more sources of variation, both management-related and environmental, beyond those that are the primary focus of the experiment. This will tell us how sensitive the results are to changes in other conditions. Otherwise, we don't really know what is going on; something else than what the experiment was intended to learn about might be more important than the experimental variables. In the present paper, for example, something caused some results for the *same treatment* in different experiments to diverge by as much as *different treatments* in the same experiment. Similarly, in many field experiments, although the treatment variables have a statistically significant effect, they "explain" only a small fraction of the total observed variance. This should be a warning that we may be missing the point entirely. If we have that much plot-to-plot variation just from inhomogeneities *within* the experimental area, even though the entire experimental area was supposed to be uniform (except in the treatment variables), what does that tell us about how much the results might have changed on a different field?

Granted, to include variables that otherwise would have been ignored means we must devote more effort per experimental question, which reduces the number of questions we can investigate. However, doing fewer experiments, but doing them thoroughly and convincingly, may be preferable to doing more experiments where the ignored variations overwhelm the phenomenon being studied. We don't learn much from an experiment where the range of applicability is unknown, or, even

worse, where it might not extend at all beyond the specific circumstances under which the data were collected.

No doubt some *AJAA* readers have thoughts on this question. We invite you to send them to us; if we get an adequate response, we will publish them as a forum. Perhaps we can stimulate more thinking about what we think is a critical but inadequately discussed issue in agricultural research.

Government Pledges to Reduce Pesticides, Promote Sustainable Ag

Only days before the release of the National Research Council's report on pesticide risks in children, the Clinton Administration pledged to reduce pesticide use and promote sustainable agriculture.

A joint statement by Agriculture Secretary Mike Espy, Environmental Protection Agency Administrator Carol Browner, and Food and Drug Administration Commissioner David Kessler, M.D., said, "We will intensify our effort to reduce the use of higher-risk pesticides and to promote integrated pest management, including biological and cultural control systems and other sustainable agriculture practices, under the leadership of the USDA. We will work side by side with American farmers to help test and implement improved and safer methods of pest management already used by many farmers."

Several days later, Browner said, "We'll work with American farmers to promote alternative farming methods that don't use any pesticides at all," and said EPA would monitor more closely pesticide residues on fruits and vegetables, and gather more information on children's eating habits.