Explanatory & Pragmatic Research

Joel M Miller, PhD
jmm@eidactics.com

Eidactics • eidactics.com

First published in 1967, and re-published in 2009, despite its importance for biomedical research, the compelling thesis of Schwartz & Lellouch remains little known and largely ignored.

Clinical researchers are often pressed to denigrate their works as "pilot studies", perhaps practically useful, but not up to the standards of true scientific research.

Schwartz and Lellouch argue, on methodologic, mathematical, ethical, and political grounds, that this widely and tenaciously held view is false and destructive.

Explanatory studies, say S&L, have wrongly come to be considered the "gold standard" of clinical research, perhaps because drug companies like that they can demonstrate superiority of a new drug, even if by a margin that is practically unimportant.

Whereas the high variability of patients and treatments in pragmatic studies certainly limits the conclusions that can be drawn, they have a critical advantage: greater variability means that where statistically significant differences are found they are

- associated with larger absolute effects, and are
- generalizable to broader patient and treatment populations.

The statistically significant findings in pragmatic studies are therefore more likely than those in strictly-controlled, homogeneous-group, explanatory studies to have broad, real-world significance.

Journal of Clinical Epidemiology 62 (2009) 499–505

ORIGINAL ARTICLE

Explanatory and Pragmatic Attitudes in Therapeutical Trials

Daniel Schwartz, Joseph Lellouch

Unité de Recherches Statistiques, Institut National de la Santé et de la Recherche Medicale, 94 Villejuif, France

Accepted 30 January 2009

Schwartz & Lellouch (1967) observed that the notion of "comparison of treatments" referred to at least two fundamentally different projects:

- Explanatory studies, which test hypotheses about underlying processes, and are designed to find differences, however small.
  - Experimental & control groups are as uniform as possible. Subject numbers may be large.
  - Treatments are strictly controlled.
  - Studies are conducted under laboratory conditions.
  - Results are assessed by conventional tests of statistical significance
- Pragmatic studies intend to help make real-world decisions, to determine which of two treatments we should prefer.
  - Subjects are typical patients – those to which the treatment would be applied – with no further selection for uniformity: "prospective observational clinical series".
  - Treatments are flexible, customized to best benefit individual patients. Auxiliary treatments are not excluded.
  - Study is conducted under normal settings, eg, in the flow of a physician's usual practice.
  - Results are assessed with tests of statistical significance, but also considering patient interests and costs.

Strabismus: Injection Treatment vs Incisional Surgery

- It is generally agreed that early correction of congenital strabismus can have both visual and non-visual advantages. Incisional surgery, however, results in extraocular scarring making any subsequent surgeries more difficult, requires protracted anesthesia that may be hazardous in itself, and is costly. Reasonably enough, surgeons prefer to delay surgical correction.
- But, what is there were an injection treatment for strabismus which had none of the above disadvantages, and which could therefore be comfortably used at an earlier age than incisional surgery?
- How should we compare injection treatment with incisional surgery?

Equalized vs Optimized Treatment Groups

- **Explanatory Design** – we <u>equalize</u> treatment groups so that, ideally, they differ only in the treatment under test.
  - Yields the clearest information concerning mechanisms.
  - Generally, one or the other treatment is applied non-optimally, so such tests may be little help making practical, clinical decisions.
- **Pragmatic Design** – we <u>optimize</u> treatments, appropriate to the nature of the treatments themselves.
  - We learn less about underlying mechanisms of action because treatment groups differ.

Outcome	Explanatory Design (both groups treated at 3 yo)	Pragmatic Design (surgery at 3 yo, injection at 2.5 yo)
Research Questions	Both outcomes are informative	**Surg>Inj:** mature system may have been better stabilized **Inj>Surg:** immature system may have been more malleable.
Practical Application	**Inj>Surg:** prefer injection **Surg>Inj:** we learn nothing because injection timing was non-optimal	Both outcomes are useful

Withdrawals

Suppose a set of patients with infantile esotropia, deemed suitable for either surgical or injection treatment, are randomly placed in one or the other treatment group. What happens when some injection patients need multiple injections or withdraw to have surgery?

- **Explanatory Approach:** We seek to compare effects of surgery with effects of an injection (or some predetermined number of injections). Enrollees in the study therefore need to be carefully qualified, perhaps at significant expense, to minimize likelihood of withdrawal. Groups will consequently tend to be highly homogenous. Analysis will need to somehow explain and discount those withdrawals that nevertheless occur, if possible.

- **Pragmatic Approach:** We compare a group given surgery with a group given an injection, possibly repeated, possibly followed by surgery, so that there are no true withdrawals. Injection treatment is defined to "absorb" repeated injections and recourse to surgery. The comparison is basically between the following practical treatment options: [1] only surgery is available, and [2] both injections and surgery are available, be used as appropriate.

Statistical Comparison

Suppose we want to compare BPX injection with BPX + epinephrine.

- **Explanatory Approach:** We might be interested in knowing whether epinephrine conferred any clear advantage. We would then perform a traditional "superiority" test of significance with
  - a small probability of wrongly concluding epinephrine to be helpful (eg, $\alpha = .05$)
  - a high probability of concluding epinephrine to be useful if it really is (large $1 - \beta$)

  Both may require large sample sizes.
- **Pragmatic Approach:** In contrast, if we simply want to use the better injectate, there is no reason to perform any significance tests – we need only choose the injection with the better mean outcome.

Clinical Research Environment

- **Mature Science:** Many diseases now have effective treatments.
- **Research Ethics:** Where effective treatments exist it may be unethical to use placebo controls. Human research may only be possible using active controls, that is, comparing proposed new treatments against established treatments, already known to be effective.
- **Superiority Testing:** On the model of placebo-control trials, active-control trials have sought to demonstrate proposed new treatments to be superior ($p < .05$) to existing treatments.
- **Other Costs & Benefits:** But treatments have clinically relevant features other than effectiveness, such as price, side-effects, and ease of use, which may be difficult to quantify and compare. Thus it is reasonable to require that a proposed new treatment merely be equal in effectiveness to an existing treatment.
- **Non-inferiority Testing:** Unfortunately, it's logically impossible to prove equal effectiveness (ie, to assert the null hypothesis), so non-inferiority testing attempts to demonstrate that proposed new treatments are unlikely to be inferior to existing treatments by more than an "equivalence margin". It is consequently possible for treatments shown to be "non-inferior" to be inferior, minimally effective, or ineffective.
- **Inferiority Creep:** Treatments established by non-inferiority testing, despite being slightly inferior, may subsequently be active controls in non-inferiority tests, resulting in even less effective treatments.

Politics

- **Superiority Testing:** Drug companies like standard $p < .05$ superiority testing because by controlling variability with large, uniform subject groups and rigidly-controlled treatment conditions, they can demonstrate statistically significant advantages for treatments that may have little or no practical advantages over existing treatments.
- **Non-inferiority Testing:** Drug companies must absolutely love having only to demonstrate "non-inferiority"!
- **Pragmatic Attitude:** But, as Schwartz & Lellouch argue, it is surely not wrong to consider the treatment "context", which includes not only effectiveness, but all costs and benefits, although they may be difficult to quantify and compare.
- **Clinicians & Scientific Reviewers** will need to understand and accommodate the complex realities of comparing treatments in a practical, clinical environment.
- **Regulators:** As efficacy becomes recognized as one among many complex, clinically-relevant features of a treatment, bureaucratic regulators will become less able to evaluate new treatments. Their role may be reduced to assuring safety.
- **Animal Testing:** In this complex environment of relative effectiveness and costs, given the danger of inferiority creep, and the ethical problems of human placebo testing, animal testing seems poised to become a critical reference for absolute efficacy.

Ethics

Fundamental research aimed at the verification of a biological hypothesis is best done on a highly-selected, and so, relatively arbitrary population, quite unlike a typical patient population, which is ultimately treated as a means rather than an end.

As such, the use of human subjects must be impermissible except in special cases. Normally, explanatory work must be done on animals, therapeutic trials on human subjects being limited to pragmatic experiments.

[after Schwartz & Lellouche 1967]