The need for objective testing

[ROBERT J. GELLER](#)

[Wyss's letter](#) in week 5 of this debate claims that I made a 'significant error' in my [week 3 article](#). I explain here why this claim should not be accepted, placing my rebuttal in a more general context.

Testable algorithms required

All readers have undoubtedly studied Newton's law of gravitational attraction:

$$F = Gm_1m_2 / r^2$$

It took a great scientist to discover this law, but any competent scientist can use it. Because this law is quantitative, it can easily be checked against observed data, thus permitting it to be verified or disproved. This precision is what led to the discovery that it is only approximately correct, as significant discrepancies were found for cases where relativistic effects are important.

In contrast, prediction proponents provide retrospective 'case studies' that are represented as showing the existence of some precursory phenomenon before some particular earthquake. Unlike Newton's law, there is no formula or algorithm that can be objectively tested using other data. This shortcoming is crucial.

Prediction proponents should, but do not, provide a 'predictor in the box' software package, with all parameters fixed. We could then input any data of the type considered by the method being tested (for example seismicity, geoelectrical, geodetic, *etc.*), using either real-time data or recorded data for regions other than that used to develop the algorithm. The software package would then generate predictions that could be tested against an intelligent null hypothesis such as the automatic alarm strategy¹.

Wyss's criticism rebutted

In his [week 1 article](#), Wyss said that "some [earthquakes] are preceded by increased moment release during the years before them, and some are preceded by seismic quiescence."

In my [week 3 article](#) I commented as follows. "Wyss's contribution to this debate cites both 'increased moment release' (more small earthquakes than usual) and 'seismic quiescence' (fewer small earthquakes than usual) as precursors. Thus it appears that any variability whatsoever in background seismicity can be claimed, after the fact, to have been a precursor. To determine whether these variations in seismicity levels are random fluctuations or real physical phenomena, objective testing of unambiguously stated hypotheses is required."

[Wyss](#) claims that this is "a significant error", because "increased moment release is mostly due to large (M7) and medium magnitude (M6) earthquakes, not small ones." Here he has defined "large", "medium," and "small" in a specific way. The "significant error", if any, that I made was perhaps to be insufficiently precise in defining the size of earthquakes. Similar criticisms could be levelled at most of the articles in this debate, which are written in a quasi-journalistic style to make them easily accessible to readers.

Needed: more statistics, less rhetoric

Individual case studies, rather than objective tests of a hypothesis, are generally presented in support of the 'increased moment release' model. One

recent study² attempts to make progress towards objective testing, but this study's null hypothesis does not appear realistic (see [ref. 3](#)), and the statistical testing does not appear to account for the effect of retroactively adjustable parameters (see [ref. 4](#)).

Another recent study⁵, which appears to have been carefully conducted, found that a random (Poisson) model outperformed the increased moment release model (sometimes also known as 'time-to-failure' analysis) by a factor of four for the particular dataset that was analyzed. (Note that Wyss cited [ref. 2](#) in his Week 1 article, but did not cite [ref. 5](#) in any of his three articles (week 1, 3, 5) to date.)

In my [week 3 article](#) I said that many prediction proponents appeared to have fallen into 'The Gambler's Fallacy' of drawing conclusions from a retrospectively chosen unrepresentative subset of a much larger dataset. The hypotheses of 'seismic quiescence' and 'increased moment release' both appear to be examples of this pitfall. If their proponents disagree with me, they should attempt to refute me, not by quibbling over the definitions of 'small', 'medium', and 'large', but rather by objective statistical testing.

Robert J. Geller

Department of Earth and Planetary Physics,
Graduate School of Science,
Tokyo University,
Bunkyo, Tokyo 113-0033,
Japan.
bob@global.geoph.s.u-tokyo.ac.jp

References

1. Kagan, Y. VAN earthquake predictions: an attempt at statistical evaluation. *Geophys. Res. Lett.* **23**, 1315-1318 (1996).
2. Bowman, D.D., Ouillon, G., Sammis, C.G., Sornette, A. & Sornette, D. An observational test of the critical earthquake concept. *J. Geophys. Res.* **103**, 24,359-24,372 (1998).
3. Stark, P.B. Earthquake prediction: the null hypothesis. *Geophys. J. Int.* **131**, 495-499 (1997).
4. Mulargia, F. Retrospective validation of the time association of precursors. *Geophys. J. Int.* **131**, 500-504 (1997).
5. Gross, S. & Rundle, J. A systematic test of time-to-failure analysis. *Geophys. J. Int.* **133**, 57-64 (1998).