CONFLICT OF INTEREST
The author states no conflict of interest.

REFERENCE
Bryld LE, Hindsbergen C, Kyvik KO, Agner T, Menne T (2004) Genetic factors in nickel allergy evaluated in a population-based female twin sample. *J Invest Dermatol* 123:1025–9

Chen JJ, Liang YH, Zhou FS, Yang S, Wang J, Wang PG, *et al.* (2006) The gene for a rare autosomal dominant form of pompholyx maps to chromosome 18q22.1–18q22.3. *J Invest Dermatol* 126:300–4

Diepgen TL (2003) Occupational skin disease data in Europe. *Int Arch Occup Environ Health* 76:331–8

Flohr C, Johansson SG, Wahlgren CF, Williams HC (2004) How atopic is atopic dermatitis? *J Allergy Clin Immunol* 114:150–8

Irvine AD, McLean WHI (2006) Breaking the (un)sound barrier: filaggrin is a major gene for atopic dermatitis. *J Invest Dermatol* 126:1200–2

Lerbaek A, Kyvik KO, Mortensen J, Bryld LE, Menné T, Agner T (2007) Heritability of hand eczema is not explained by comorbidity with atopic dermatitis. *J Invest Dermatol* 127:1632–40

McLean WHI, Hull PR (2007) Breach delivery: increased solute uptake points to a defective barrier in atopic dermatitis. *J Invest Dermatol* 127:8–10

Meding B, Jarvholm B (2004) Incidence of hand eczema-a population-based retrospective study. *J Invest Dermatol* 122:873–7

Smit HA, van Rijssen A, Vandenbroucke J, Coenraads PJ (1994) Individual susceptibility and the incidence of hand dermatitis in a cohort of apprentice hairdressers and nurses. *Scand J Work Environ Health* 20:113–21

Svensson A (1988) Hand eczema: an evaluation of the frequency of atopic background and the difference in clinical pattern between patients with and without atopic dermatitis. *Acta Derm Venereol* 68:509–13

See related article on pg 1641

Casting Light on Evidence

Jonathan L. Rees[1]

Clinicians are growing inured to the fact that rarely do the results of a single randomized controlled trial (RCT) decisively change clinical practice. Nor should they. In a new RCT, Kirke and colleagues have compared the therapeutic effects of different types of UVB radiation on clearance of psoriasis. Their study is unlikely to be the last word on this topic.

Journal of Investigative Dermatology (2007) **127**, 1570–1571. doi:10.1038/sj.jid.5700775

Not long ago, in an age of optimism (and statistical naiveté), it seemed that you could read the results of a randomized controlled trial (RCT) at the weekend and change your prescribing habits accordingly on Monday morning. It was as though the RCT was a truth machine: you just fed in the patients' details, added a little randomization, mechanically applied some statistical test, and out popped a God's-eye view of how the Universe worked. A veritable epistemological engine. Things do not seem quite so simple anymore. Indeed, in a recent article, Ioannidis (2005) argued cogently, if a little mischievously, that the results of most clinical studies, including RCTs, were mistaken. With this in mind, what are we to make of the report by Kirke *et al.* (2007, this issue) that there is no difference in efficacy between narrowband phototherapy and (selective) broadband UVB lamps for the treatment of psoriasis? Should those recently purchased narrowband cabinets be put to one side?

The use of UVB phototherapy to treat psoriasis has a long history, but it is probably only since the introduction of psoralen plus UVA (PUVA) that interest has focused on how this older modality of therapy could be made more efficacious. Subsequent clinical application has owed more to incremental improvement than to any "eureka" moments. However, if there was an elegant experiment, rather than the brute force of clinical trials, it was that reported by Parrish and Jaenicke in this journal in 1981 (Parrish and Jaenicke, 1981). In this study, Parrish and Jaenicke exposed psoriatic plaques to monochromatic sources and showed that wavelengths less than 290 nm were erythemogenic but had little efficacy. The implication was that it would be better to use lamps of longer wavelengths with the UVC filtered out. Although the exact relative efficacy of different longer UVB wavelengths was unclear, the invention of the 311 nm (Philips TL-01) lamp soon led to their widespread use, at least in Europe. In a pattern common to the way many medical innovations are adopted, an apparent physiological rationale, coupled with a few initial small-scale reports and enthusiasm by early adopters, led to widespread changes in clinical practice. Which is where the exemplary study by Kirke *et al.* (2007) comes into play. Is there anything special about TL-01, or is it just that UVB lamps contaminated with UVC are less efficacious?

Kirke and colleagues (2007) screened 192 subjects in a single university hospital center, of whom 124 were eligible and 24 declined to participate. The remaining 100 were randomized, 50 in each group, to receive either TL-01 or (selective) UVB with UV6 fluorescent lamps. (UV6 bulbs have negligible output in the UVC range; see Kirke *et al.* (2007) for a comparison of the spectra of the two tube types.) The primary outcome was the number of treatments to clearance, and the proportion of patients clearing was also reported. The median number of exposures was 28.4 for TL-01 and 30.4 for UV6. With TL-01, 56% of patients cleared, whereas with UV6, 40% cleared. Neither of these comparisons was significant at the conventional statistical level. So, in the light of the general claims by Ioannidis (2005), how do we interpret the data from this study as evidence? What does it mean?

[1]*Department of Dermatology, University of Edinburgh, Edinburgh, United Kingdom*

Correspondence: Prof. Jonathan L. Rees, Department of Dermatology, University of Edinburgh, Room 4.018, First Floor, Lauriston Building, Lauriston Place, Edinburgh EH3 9HA, United Kingdom. E-mail: jonathan.rees@ed.ac.uk

First, the failure to find a difference does not mean one doesn't exist. Perhaps if the study had been larger the authors would have found a difference. Perhaps. On the other hand, the authors cite the usual power calculations, and, as with most contemporary clinical trials, the data have been reported in a particular frequentist way: that is, you first assume no difference, and then argue that if an extreme event beyond a particular critical value occurs, your initial hypothesis is to be rejected. If you fail to achieve this critical value,

> Is there anything special about TL-01, or is it just that UVB lamps contaminated with UVC are less efficacious?

you say there are no good grounds for rejecting your initial hypothesis. We go from hypothesis to data, not data to hypothesis (Goodman, 1999a; Goodman, 1999b; Royall, 1997). Do the results mean that the two treatments are equivalent (as in an equivalence trial)? The answer is no; the trial was not designed to answer this question. But if we step outside a strict Neyman–Pearson frequentist interpretation of the results, what do we think? Are the data still compatible with TL-01 being more efficacious than UV6? Answer: yes. Are the data compatible with UV6 being more efficacious? Answer: yes. Indeed, the answer to these seeming paradoxes is that you can only interpret the trial as evidence in favor of one particular hypothesis over another in the light of what you thought before the trial was carried out. The trial results should shift your prior opinion, but whether the conclusion is decisive depends on how extreme your viewpoint was before the study. If you were fairly indifferent to which UVB modality was more efficacious, the study provides little reason to change this viewpoint. However, if, for whatever reason, you were convinced that TL-01 was more efficacious, then the study will not convert you — but it should lessen your certainty. It is for these reasons that trials are not truth machines and the ghost of judgment cannot be exorcised. As Ioannidis (2005) and others have concluded (Vandenbroucke, 1997), one reason trials are not as informative as we would like is that often we go into them with a very weak prior hypothesis of what is going on. In these circumstances, imbuing rare events with biological significance may be misleading. Nor should we blame the statistician.

But there are other issues, possibly even more important ones, than arguments about how to interpret statistical tests. As has been commented on before, many RCTs are extremely poor surveys (Kravitz *et al.*, 2004; Longford and Nelder, 1999). That is to say, it is unclear how representative the patients studied in a trial are of the larger target population. Unlike in many studies reported in dermatology in which very few patients from a particular center are enrolled, in the study by Kirke *et al.* (2007) a very high proportion of the patients from their center were recruited. It would seem likely to me that the patients studied are very similar to the ones I see in my phototherapy clinic. By contrast, in many multicenter studies one suspects that the patients chosen to take part are very atypical. Unfortunately, such considerations, because they are hard (but not impossible) to formalize, tend to get neglected in many reviews. An RCT on a small nonrandom sample may be more misleading as evidence than a larger observational study that enrolls most subjects within a defined geographical population.

Another issue, much discussed by Kirke *et al.* (2007), relates to safety. Again, in paying too much attention to statistical testing, we may be accused, like the proverbial drunk, of searching for our lost keys under the light because that is where we can see. Numbers have a habit of trumping judgment even when they shouldn't. Kirke *et al.* (2007) correctly state that we have no robust data on the safety of TL-01 (nor, for that matter, UV6). Trials in the particular context of psoriasis often cannot measure what some of us consider the biggest hazard — cancer, although in this case our anxieties are somewhat lessened, as the treatments are topical. Because there are no direct data germane to this issue, the authors reasonably use extant data on the relation between wavelength and skin cancer to argue that TL-01 lamps are likely to be more hazardous than UV6 tubes. Although I cannot dispute their figures, I do not find this a decisive reason to favor UV6, but, on the other hand, nor do I believe that we will ever get clean long-term data on the hazards of UV in psoriasis without the confounding effects of personal sun exposure or systemic medication. If there is a big difference, we might see it; otherwise, we will probably remain ignorant.

So if today is Saturday, what should I learn for Monday? First, RCTs are very much operational research — health technology assessment — rather than science. They are often noisy and cumbersome, even when they are as good as the one under discussion. Second, I hope somebody goes back and revisits the earlier work of Parrish and Jaenicke (1981). As for therapy, I will carry on as before: I am not convinced that there is much special about TL-01 beyond the absence of UVC. Sadly, we now need some more trials to prove me wrong.

CONFLICT OF INTEREST
The author states no conflict of interest.

REFERENCES

Goodman SN (1999a) Toward evidence-based medical statistics. 1. The *P* value fallacy. *Ann Intern Med* 130:995–1004

Goodman SN (1999b) Toward evidence-based medical statistics. 2. The Bayes factor. *Ann Intern Med* 130:1005–13

Ioannidis JP (2005) Why most published research findings are false. *PLoS Med* 2:e124

Kirke SM, Lowder S, Lloyd JJ, Diffey BL, Matthews JNS, Farr PM (2007) A randomized comparison of selective broadband UVB and narrowband UVB in the treatment of psoriasis. *J Invest Dermatol* 127:1641–46

Kravitz RL, Duan N, Braslow J (2004) Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q* 82:661–87

Longford NT, Nelder JA (1999) Statistics versus statistical science in the regulatory process. *Stat Med* 18:2311–20

Parrish JA, Jaenicke KF (1981) Action spectrum for phototherapy of psoriasis. *J Invest Dermatol* 76:359–62

Royall R (1997) *Statistical Evidence: A Likelihood Paradigm*. Chapman & Hall/CRC: Boca Raton, 208pp

Vandenbroucke JP (1997) Homoeopathy trials: going nowhere. *Lancet* 350:824