Reviewer's report

**Title:** A critical review of the research literature on Six Sigma, Lean & StuderGroup's Hardwiring Excellence: the need to demonstrate and communicate the effectiveness of transformation strategies in healthcare

**Version:** 1  **Date:** 12 February 2009

**Reviewer:** Paula Michelle Whitty

**Reviewer's report:**

Thank you for asking me to review this interesting article on an important topic for health care systems worldwide, even if the review was limited to the US. The findings are significant for the policy community, so it is essential that the article commands confidence.

While more critical reviewers might disagree, I think a particular strength of the article is that the authors did not limit inclusion of studies on research design criteria. The articles are critiqued, so weaker designs are identified as such in the review. Not only does this make the article more likely to be read by a wider policy community, it also shows up the dearth of formal evaluation of any type. The recommendations for improving the routine evaluation of such interventions in practice are also particularly helpful to the policy community.

However, there are limitations of this article as a literature review which, while not billed as a systematic review, could still nevertheless improve the presentation of its methods (and/or the description of its limitations if this is not possible).

So my comments are limited to areas where I think the authors could significantly improve their article. All of these are 'major compulsory revisions':

**Major compulsory revisions**

1. Title and abstract – neither make it clear that the review was limited to US studies only, and this needs to be corrected.

2. The results and conclusions sections from the abstract could be more representative of their findings. The results as presented here are too even-handed and, to be frank, I didn’t understand the conclusions!

3. Research question – the authors do not explicitly address a key question about all of these interventions in this review, which is the ‘sustainability of effect’ on practice or culture. For example, a process improvement workshop may well produce a short-term change in practice, but does it last? While the decay of effect is mentioned in one study in the results, explicit identification of this issue upfront is warranted.

4. Search strategy – the databases searched were a little limited, even for the US only, as was the basic key word search. US studies could be published in
non-US databases, so their inclusion would have been advisable. An example of relevant databases to search for this topic would be:

Cochrane Effective Practice and Organisation of Care Group (EPOC) specialised register, Cochrane Controlled Trials Register, MEDLINE, EMBASE, PsycLit, Cinahl, ABI Inform, the Science Citation Index, ERIC, the System for Information on the Grey Literature and the Health Management Information Consortium database.

A check on the missing databases to see if they have missed anything obvious would reassure this reviewer. Or, at the very least, acknowledgement of the limitations.

The discussion also needs to acknowledge that this was not a systematic search: some interventions which would have met the criteria to be lean etc, may not have been described as such and therefore missed on a simple keyword search. Further, the authors did not hand search relevant journals to look for studies.

5. Data extraction – acknowledge that the level of detail that would be extracted in a Cochrane EPOC review was not achieved, in particular on the interventions (‘what types of training, education?’ etc), and therefore the ability of the reader to transfer to their own setting would be limited.

6. Results – there is some confusion in the description of biases that were not avoided by the constituent studies. For example, confounding biases cannot be eliminated by statistical analyses: this is a design issue, only fully satisfactorily dealt with by randomisation.

7. Results – the critique of the Interrupted Time Series (ITS) studies could be more in-depth. For example, a single group ITS is not as robust as a controlled ITS. I refer the authors to ‘Craig R. Ramsay, Lloyd Matowe, Roberto Grilli, Jeremy M. Grimshaw, Ruth E. Thomas. Interrupted Time Series designs in Health Technology Assessment: lessons from two systematic reviews of behaviour change strategies. International Journal of Technology Assessment in Health Care, 19:4 (2003), 613–623’ for further guidance.

Hope this is helpful.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests.