Editorials

Perspectives on evaluating community programmes

One year ago we published a report by Svanström describing the evaluation of the Lidköping programme. The paper, as initially submitted, posed many problems for the editors and our reviewers. Much of the difficulty arose from the statistical procedures used to assess the results. As is so often the case, their interpretation was made more difficult and uncertain because the design of the evaluation was, itself, complicated. Despite this, we were keen to publish the Lidköping paper because of the widespread and growing interest in community based injury prevention programmes. In brief, it was important to know if this bandwagon was moving in the right direction; if it should be halted; or if it needed to modify its course.

In this issue, Langley and Alsop takes the assessment process one step further. They were concerned about the interpretation of the results in the Lidköping paper and decided to reanalyse the data using a different approach. Although some readers may view any dispute over what is the best analytic approach as esoteric, the statistical mysteries can be easily unravelled.

In the first paper, Svanström et al chose to compare injury rates using a regression procedure that assumes the events are evenly dispersed across the areas being compared and thus have a roughly normal distribution. In contrast, Langley and Alsop believe that relatively rare events are more likely to be unevenly distributed and that a Poisson approach is more appropriate. Perhaps surprisingly, the result of their secondary analysis largely supported the original conclusion that the intervention may have been effective. Before agreeing to publish their critique, however, we thought it should be subjected to the same rigorous review process as any other, and, in particular, we wanted assurances that the statistical concerns raised by the ‘critics’ were themselves valid.

Thus readers will find a third viewpoint on the topic from two eminent statisticians, Hanley and Choi, who were asked to comment on the methods used in both Svanström’s and Langley’s papers. Their comments were illuminating and it seemed wise to publish them (p 133). In essence, their conclusion is that both statistical approaches are acceptable but that each falls somewhat short of the mark.

The fact that the conclusion from Langley and Alsop’s reanalysis is similar to Svanström’s may be interpreted by some as the final word on the effectiveness of community interventions. But many questions remain and there are several lessons to be learned from this controversy. First, we should agonize less about statistical niceties and concentrate more on study designs. For reasons that, in retrospect, are understandable, the Lidköping evaluation relied on a comparison between the target community and the four bordering counties as well as the rest of the county in which Lidköping is situated. This design clouds the issue and may load the dice against finding a sharp difference between the communities being compared. It is possible that in view of their proximity, the ‘control’ communities received many of the messages and other components of the intervention.

Second, although choosing the most appropriate statistical procedure is undoubtedly important, solid findings are usually robust and unlikely to differ greatly regardless of how they are analyzed. Third, every new approach to injury prevention must be evaluated rigorously and a randomized trial is always preferable. Fourth, the evaluation process should be conceived at the time when the intervention is planned, not as an afterthought. Fifth, it is always preferable that the evaluation be carried out by independent parties, not those responsible for the intervention.

Against these, perhaps self evident reminders, it is important to acknowledge the special problems community interventions present for evaluators. For example, evaluators and those responsible for these programmes are frequently expected to demonstrate effects in unrealistically short time frames. Sample size considerations are especially challenging. It can be argued that each community is a unit, equivalent to a single case, so that no matter how large its population the effective N equals one. A third problem is knowing what to measure: even if children adopt a targeted behaviour (for example, helmet wearing) and an injury specific reduction follows, it cannot be assumed that the intervention has equally affected other behaviours. Ideally, targeted and non-targeted injuries must be separated in the analyses. Finally, it is essential that every element of the intervention be thoroughly documented. Imagine the consternation if an evaluation conclusively showed that a programme was effective but no one else could replicate it because what was done was not recorded in sufficient detail.