Reviewer's report

Title: Minimum follow-up time required for the estimation of statistical cure of cancer patients: verification using data from 42 cancer sites in the SEER database

Version: 1 Date: 30 November 2004

Reviewer: Colin Mathers

Reviewer's report:

General

This is an interesting paper which addresses a topic of some importance for cancer epidemiology, namely how long a period is required for followup in order to ascertain survival rates for specific cancers by site. It needs some revision to provide more evidence that the lognormal distribution provides an acceptable fit to the data and to address the question of how long follow up is required if the lognormal curve is an acceptable fit.

---------------------------------------------------------------------------------------------------------------------------

Major compulsory revisions

The concept of threshold year is defined in the appendix and is an essentially arbitrary choice of the point where cumulative predicted risk of death reduces to 2.5%. The authors need to justify this choice. Why would 5% not be equally useful? Is there some specific basis for 2.5%. This should be discussed in the introduction as central to the purpose of this paper.

An alternate choice might be to choose a threshold where risk of cancer death becomes equal to risk of other causes of death. This would vary with age and with cancer site.

The goodness of fit of the lognormal distribution should be illustrated with some graphical examples for several sites. The choice of this distribution should be justified somewhat further. Other models of cancer survival have used the Weibull distribution for survival times. It would be useful to have some discussion on this.

On page 7, it is mentioned that seven sites failed in the test for lognormality. This is a relatively high proportion of sites and some more discussion (and possibility graphical illustration) is needed to assess why this is happening. Does this call into question the lognormal assumption generally, or are there specific reasons for failure for these sites?

If the lognormal distribution provides a good fit to death distribution data, then it follows that observations for a number of years considerably shorter than the threshold year may be adequate to fit the distribution and to predict the survival rate. The authors need to address this fundamental question, or to justify their conclusion that the threshold year specifies the number of years of followup required. It should be feasible with the data used in this paper, to do simulations for x years of followup data, fit lognormal distribution and calculate statistical cure rate. X can then be varied to see what is the lowest x that gives acceptable prediction (with a specified uncertainty range).

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research
interests

**Quality of written English:** Acceptable

**Statistical review:** No

**Declaration of competing interests:**

I declare that I have no competing interests