HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2016 March 30.

Published in final edited form as:

Stat Med. 2015 March 30; 34(7): 1094–1095. doi:10.1002/sim.6430.

Comments on 'A critical look at prospective surveillance using a scan statistic' by 'T. Correa, M. Costa, and R. Assunção

Martin Kulldorff[a,*] and **Ken Kleinman**[a]

[a]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA 02215, United States.

We appreciate the opportunity that this paper and the journal editors have provided us to clarify some important issues.

There is consensus that one should adjust for repeated analyses when conducting prospective disease surveillance. Correa et al. [1] are concerned about p-values adjusted for prior analyses in a prospective scan surveillance setting, an advanced option in the SaTScan software [2]. The key claim, stated in the last paragraph, is that *'For both types of adjustments, … for each n [=analysis], the probability of type I error is much smaller than the significance level, since the p-values [are] in the interval 0.9,1.0. This lack of control of the type I error probability and of the average run length lead us to strongly oppose the use of the scan statistic in the prospective context with the SaTScan™ signal rules'*. The complaint refers to p-values adjusted for either all prior analyses or the prior 100 analyses (situations c and d in the paper), but not to the default or usual signal rules in SaTScan software (situation b).

Correa et al.'s claim is based on a misunderstanding.

When adjusting for repeated analysis of the data, under the null hypothesis we should have a 0.05 probability of rejecting the null *across all analyses*. Situations c/d will only give an adjusted p<0.05 when rejecting the null at the overall alpha 0.05 level for the many sequential analyses. This means that unadjusted p-values must be much smaller than 0.05 in order to generate a signal (=alarm) rejecting the null, and that the adjusted p-values are greater than 0.9 for most single analyses of the data (Figure 4). This is exactly the way it should be. Rather than showing *"a lack of control of the type I error"*, it demonstrates that the desired adjustment for repeated analyses of the data is working. It is not the adjusted p-value from a *single* analysis, but the *minimum* of the adjusted p-values over *repeated* analyses, that should be approximately uniform [0, 1].

In particular, suppose we adjust for an overall alpha level of 0.05 across a sequence of 100 analyses. The null probability of a signal during those 100 analyses will then be 0.05. Over 300 such analyses it will be somewhat less than $1-(1-0.05)^3=0.14$, since there is some dependency between the three sets of 100 analyses. In the authors' simulations, the

*Correspondence to: Martin Kulldorff, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 133 Brookline Avenue, Boston, MA 02215, United States. Martin_Kulldorff@hms.harvard.edu.

proportion was in fact 0.13 (section 3.2.3). The extensive simulations conducted by Correa et al. hence confirm that the adjusted p-values (situation d) properly account for repeated sequential analyses.

P-values adjusted for all prior analyses (situation c) should be avoided for long sequences of data [3], as rejection becomes increasingly difficult with time (Figure 3). P-values adjusted for a fixed number of prior analyses are always appropriate (situation d), but can be confusing. This is why we continue to recommend that users use recurrence intervals (RI) based on standard unadjusted p-values (situation b) [4, 5]. This is the default signal rule in SaTScan that we and others normally use to account for the repeated analyses in prospective space-time scan disease surveillance [6, 7].

A RI of (say) 365 analyses means that the expected number of signals of equal or greater strength is one during an arbitrary 365-analysis period. Hence, a public health department can use this to control the expected number of false signals they have to deal with, e.g., by setting a signal threshold of RI > 365.

The RI is calculated for each analysis of the data, and then compared to a pre-specified signal threshold. The run length (RL) is the number of analyses it takes until some statistic exceeds a signal threshold. There is no reason why the RI at the time of a signal and the RL should be correlated, and the lack of correlation in Figure 7d is expected. This is also true for other prospective methods such as the CUSUM. With a signal threshold of RI > 365 days, one might think that the average run length (ARL) would also be 365 days, but that is not true, although there is obviously a correlation between the threshold used and the ARL.

In summary, the authors' offer no objection to situation b, the SaTScan default, despite stating that they *"strongly oppose the use of the scan statistic with the usual signal rules in the prospective context"*, and their concerns about situations c/d are based on misunderstandings. The prospective space-time scan statistic is a valuable disease surveillance tool used by public health departments around the world, and we will continue to recommend using it with the recurrence interval based signal rules available in SaTScan.

Acknowledgements

This commentary was funded by the National Cancer Institute, grant #R01CA165057.

References

1. Correa T, Assunção R, Costa M. A critical look at prospective surveillance using a scan statistic. Statistics in Medicine. 2015 *X:xxx-xxx*.
2. Kulldorff, M. Information Management Services, Inc.. SaTScan™: Software for the spatial and space-time scan statistics. [www.satscan.org].
3. Kulldorff M. Prospective time-periodic geographical disease surveillance using a scan statistic. Journal of the Royal Statistical Society. 2001; A164:61–72.
4. Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. American Journal of Epidemiology. 2004; 159:217–224. [PubMed: 14742279]
5. Kleinman K. Generalized Linear Models and Generalized Linear Mixed Models for Small?Area Surveillance. Spatial and Syndromic Surveillance for Public Health. 2005:77–94.

6. Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Medicine. 2005; 2:216–224.

7. Viñas M, Tuduri E, Galar A, Yih WK, Pichel M, Stelling J, Brengi S, Della Gaspera A, van der Ploeg C, Bruno S, Rogé A, Caffer M, Kulldorff M, Galas M. Laboratory-based prospective surveillance for community outbreaks of Shigella spp. in Argentina. PLoS Neglected Tropical Diseases. 2013; 7:e2521. [PubMed: 24349586]