Biometrika Advance Access published November 25, 2007

Biometrika (2007), *pp*. 1–4 doi: 10.1093/biomet/asm080

Printed in Great Britain

A note on repeated *p*-values for group sequential designs

By MARTIN POSCH

Section of Medical Statistics, Medical University of Vienna, A-1090 Vienna, Austria
martin.posch@meduniwien.ac.at

GERNOT WASSMER

Department of Medical Statistics, Informatics and Epidemiology, University of Cologne, D-50937 Köln, Germany
gernot.wassmer@uni-koeln.de

AND WERNER BRANNATH

Section of Medical Statistics, Medical University of Vienna, A-1090 Vienna, Austria
werner.brannath@meduniwien.ac.at

Summary

One-sided confidence intervals and overall p-values for group-sequential designs are typically based on a sample space ordering which determines both the overall p-value and the corresponding confidence bound. Accordingly, the strength of evidence against the null hypothesis is consistently measured by both quantities such that the order of the p-values of two distinct sample points is consistent with the order of the respective confidence bounds. An exception is the commonly used repeated p- values and repeated confidence intervals. We show that they are not ordering-consistent in the above sense and propose an alternative repeated p-value which is ordering-consistent and has the monitoring property of the classical repeated p-value in being valid even when deviating from the prefixed stopping rule.

Some key words: Group-sequential design; Repeated confidence interval; Repeated p-value; Sample space ordering.

1. Introduction

Both one-sided confidence intervals and p-values map the sample space to the real line and thus define an ordering in the sample space. For single-stage tests of one-parameter families the orderings induced by p-values and confidence intervals typically coincide. For group-sequential tests there is no single natural ordering: there are many ways of defining an order between sample points at different stages. Orderings that have received attention are the stagewise ordering, the maximum-likelihood ordering, the likelihood ratio ordering and the score test ordering; see Jennison & Turnbull (2000) for a review. Typically, the same sample space ordering is used for the construction of p-values and confidence intervals. This guarantees that, for two different trials following the same group-sequential design, the trial with the larger lower confidence bound is also the trial with the smaller overall p-value.

Exceptions are the repeated confidence interval (Jennison & Turnbull, 1984) and the repeated p-value (Jennison & Turnbull, 2000) when defined via a common family of group-sequential boundaries. We show that they lead to different sample space orderings and are therefore not ordering-consistent. This implies that, for two sample points in the same group-sequential design, the sample point with the larger lower confidence bound may have the larger overall p-value.

We propose an overall p-value that is ordering-consistent to the repeated confidence interval. It is a repeated p-value in the sense that it satisfies the monitoring property and is valid at each stage of the group-sequential trial.

2. Repeated one-sided confidence intervals and p-values and their ordering inconsistency

2·1. *Group-sequential designs*

Consider a one-sided group-sequential test with K stages for testing $H_0 : \mu = 0$ against $H_1 : \mu > 0$. Let $Z_1, \ldots, Z_K$ denote a standardized test statistic available at stage $k = 1, \ldots, K$, such that $Z_k \sim N\{\mu(I_k)^{1/2}, 1\}$, where I_k is the information for μ at stage k, and $(Z_1, \ldots, Z_K)$ are multivariate normal with $\text{cov}(Z_{k_1}, Z_{k_2}) = (I_{k_1}/I_{k_2})^{1/2}$ for $k_1 \leqslant k_2$. Consider a family of group-sequential boundaries, such as those from the Δ-class of boundaries proposed by Wang & Tsiatis (1987). The corresponding critical values are

$$c_k(\alpha) = c_\alpha \left(\frac{I_k}{I_1}\right)^{\Delta - 0\cdot5}, \quad k = 1, \ldots, K, \tag{1}$$

where Δ defines the shape of the boundaries, and the constant c_α, is chosen such that

$$P_0\left(\bigcup_{k=1}^{K} \{Z_k \geqslant c_k(\alpha)\}\right) = \alpha, \tag{2}$$

where P_0 denotes the probability under the null hypothesis. This gives the O'Brien and Fleming boundaries for $\Delta = 0$ and the Pocock boundaries for $\Delta = 0\cdot5$.

2·2. *The repeated confidence interval*

The $(1 - \alpha)100\%$ repeated lower confidence bound at stage k, LB_k, is given by

$$\text{LB}_k = \frac{z_k - c_k(\alpha)}{(I_k)^{1/2}}, \tag{3}$$

(Jennison & Turnbull, 1984) where z_k denotes the observed test statistic at interim analysis k, and $c_k(\alpha)$ are the stopping boundaries of the group-sequential test.

Repeated confidence bounds are reported by standard software packages for group-sequential trials (Addplan, 2005; East, 2005). They owe their popularity to their monitoring property: they are valid at each stage of the trial, even if one does not adhere to the stopping rule since $P_\mu(\cup_{j \leqslant K}\{\mu \geqslant \text{LB}_j\}) = \alpha$.

2·3. *The repeated p-value*

Jennison & Turnbull (2000, p. 202) propose repeated p-values which share the monitoring property of the repeated confidence intervals and are valid at each interim analysis. They consider a specific family of critical boundaries $c_k(\alpha')$ and the corresponding repeated confidence bounds $\text{LB}_k(\alpha')$. The repeated p-value at stage k is defined as the largest level α' such that the value 0 is included in the confidence interval. Since the inclusion of 0 in the confidence interval at level α' is equivalent to $z_k < c_k(\alpha')$, the repeated p-value is given by

$$p_k^{\text{r}} = \sup\{\alpha' : z_k < c_k(\alpha')\}. \tag{4}$$

If $c_k(\alpha')$ is continuously decreasing in α' then the stage-k repeated p-value equals the level α', where $z_k = c_k(\alpha')$. Note that the repeated p-value shares the monitoring property of the repeated confidence bound since $P_0(\cup_{j \leqslant K}\{p_j^{\text{r}} \leqslant \alpha\}) = P_0(\cup_{j \leqslant K}\{Z_j \geqslant c_j(\alpha)\}) = \alpha$.

The repeated p-value is consistent with the repeated confidence interval in the following sense: whenever the repeated p-value is less than α, the confidence interval excludes all parameter values of the null hypothesis. However, as shown in the following section, the repeated p-value in general induces a different ordering in the sample space than the repeated confidence interval.

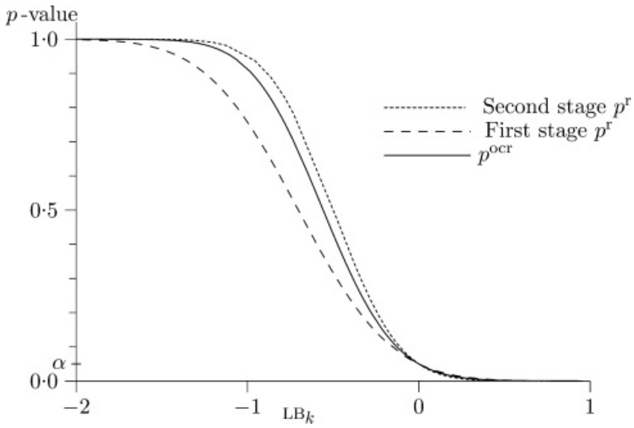

Fig. 1. Repeated p-values as functions of the lower bound LB_k of the repeated confidence bound in a two-stage design with constant critical values, $\alpha = 0{\cdot}05$, $I_2 = 2\,I_1$. The curves correspond to first stage p^r, dashed, second stage p^r, dotted, and p^{ocr}, solid.

2·4. *The ordering inconsistency of the repeated confidence interval and the repeated p-value*

Consider two trials following the same group-sequential design. Let k_1, k_2, LB_{1,k_1}, LB_{2,k_2}, p_{1,k_1} and p_{2,k_2} denote the stopping times, confidence bounds and p-values of the two trials. The p-value is said to be ordering-consistent with the confidence bound if $p_{1,k_1} \leqslant p_{2,k_2}$ implies that $\text{LB}_{1,k_1} \geqslant \text{LB}_{2,k_2}$ for all stopping times and sample points.

To see that the repeated confidence bound and the repeated p- value (4) may not be ordering-consistent consider, for example, a Pocock-type two-stage group-sequential design with constant critical boundaries $c_k(\alpha) = c_\alpha$. Each value u of the repeated p-value is attained for two different sample points: one from the first stage where $z_1 = c_u$ and the other from the second stage where $z_2 = c_u$. While these sample points correspond to identical p-values, they correspond in general to different repeated confidence bounds, namely $\text{LB}_1 = (c_u - c_\alpha)/(I_1)^{1/2}$ and $\text{LB}_2 = (c_u - c_\alpha)/(I_2)^{1/2}$, which differ if $c_u \neq c_\alpha$ since $I_2 > I_1$. Figure 1 shows for each possible p-value the corresponding lower confidence bounds, one from the sample point of the first stage, dashed line, and the other from the sample point of the second stage, dotted line. Note that the dashed and dotted lines in Fig. 1 cross when the p-value is equal to α. This implies that, if the confidence bound is kept fixed, the ordering of the corresponding first- and second-stage p-values switch. Furthermore it can be seen that an ordering inconsistency may occur if $p_1^r < p_2^r < \alpha$ or $p_1^r > p_2^r > \alpha$.

3. The ordering-consistent repeated p-values

We derive a repeated p-value for each stage k that is ordering-consistent with the repeated confidence interval and has the monitoring property. Let $F(x)$ denote the cumulative distribution function of $\max_{j \leqslant K} \text{LB}_j$, under the null, of a trial in which we always pass to stage K, even if a stopping boundary is crossed earlier. Then an ordering-consistent repeated p- value at stage k is given by

$$p_k^{ocr} = 1 - F(\text{LB}_k), \tag{5}$$

where LB_k denotes the observed stage-k confidence bound. The order consistency of p_k^{ocr} follows immediately from the monotonicity of F. Note that the stage-k p-value is equal to the probability that in an independent trial at any stage a larger confidence bound than LB_k is observed. To see that p_k^{ocr} is a conservative p-value that satisfies the monitoring property, we show that it coincides with the classical repeated p-value (4) for a special family of stopping boundaries. By (5) the ordering-consistent p-value can be written as

$$p_k^{\text{ocr}} = P_0\left(\bigcup_{j=1}^{K}[\{Z_j - c_j(\alpha)\}/(I_j)^{1/2} \geqslant \text{LB}_k]\right)$$
$$= P_0\left(\bigcup_{j=1}^{K}\{Z_j \geqslant c_j(\alpha) + (I_j)^{1/2}\,\text{LB}_k\}\right), \tag{6}$$

where LB_k is the observed repeated confidence bound and Z_j, $j \leqslant K$, are the z-scores from an independent replication of the group-sequential trial under the null hypothesis. Thus, the ordering-consistent p-value coincides with the classical repeated p-value for the specific family of stopping boundaries given by

$$c_k^{\text{ocr}}(\alpha') = c_k(\alpha) + (I_k)^{1/2}\, c_{\alpha'},$$

where for each α' the constant $c_{\alpha'}$ is chosen such that the group-sequential test with boundaries $c_k^{\text{ocr}}(\alpha')$ has level α'.

The family c_k^{ocr} does not follow any of the commonly used shapes. In particular, it does not belong to the Δ-class family even if the level α boundaries $c_k(\alpha)$ do. Note that c_k^{ocr} is the only family of stopping boundaries that leads to an ordering-consistent p-value. This follows by arguments similar to those in §2·4. The solid line in Fig. 1 shows the ordering-consistent repeated p-value as a function of the confidence bound when starting with Pocock-type boundaries $c_k(\alpha) = c_\alpha$ at level α. Irrespective of the stage considered the lower confidence bound determines the p-value uniquely.

In principle, one could also consider the classical repeated p- value (4) for some common family of rejection boundaries $c_k(\alpha')$, such as the Δ-class, and construct a modified repeated confidence interval that is ordering-consistent with this p- value. This can be achieved by defining dual tests chosen from this family.

ACKNOWLEDGEMENT

We wish to thank Peter Bauer and Franz König for many helpful suggestions.

REFERENCES

ADDPLAN (2005). *Adaptive Designs—Plans and Analyses. Release 3. Software Documentation.* Cologne: Addplan GmbH. http://www.addplan.com.

EAST (2005). *Software for Advanced Clinical Trial Design, Simulation and Monitoring. Version 4*. Cambridge, MA: Cytel Inc. http://www.cytel.com.

JENNISON, C. & TURNBULL, B. W. (1984). Repeated confidence intervals for group sequential clinical trials. *Contr. Clin. Trials* **5**, 33–45.

JENNISON, C. & TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton, FL: Chapman and Hall/CRC.

WANG, S. K. & TSIATIS, A. A. (1987). Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics* **43**, 193–9.

[*Received November* 2006, *Revised May* 2007]