Published in final edited form as:

Commun Stat Simul Comput. 2016 ; 45(7): 2338–2361. doi:10.1080/03610918.2014.901359.

Evaluating the performance of simultaneous stepwise confidence intervals for the difference between two Poisson rates

Brianna C. Bright and

Graduate Student, Department of Statistics, University of Nebraska-Lincoln, 340 Hardin Hall-North, Lincoln, NE 68583-0963, Phone: (402) 472-7231, Fax: (402) 472-5179

Julia N. Soulakova

Associate Professor, Department of Statistics, University of Nebraska-Lincoln, 340 Hardin Hall-North, Lincoln, NE 68583-0963, Phone: (402) 472-7231, Fax: (402) 472-5179

Brianna C. Bright: Brianna.Bright@gmail.com; Julia N. Soulakova: jsoulakova2@unl.edu

Abstract

We consider the problem of simultaneously estimating Poisson rate differences via applications of the Hsu and Berger stepwise confidence interval method (termed HBM), where comparisons to a common reference group are performed. We discuss continuity-corrected confidence intervals and investigate the HBM performance with a moment-based confidence interval, and uncorrected and corrected for continuity Wald and Pooled confidence intervals. Using simulations, we compare nine individual confidence intervals in terms of coverage probability and the HBM with nine intervals in terms of family-wise error rate and overall and local power. The simulations show that these statistical properties depend highly on parameter settings.

Keywords

multiple testing; simultaneous confidence intervals; Poisson distribution

1. Introduction

When reporting findings from research studies, investigators often emphasize the significance of these findings in terms of p -values and corresponding hypothesis testing decisions. However, confidence intervals can offer important additional information in many studies. For example, if there is no consensus on the value of a clinically relevant threshold, which is the case for many therapeutic areas (Hothorn, 2007), findings from non-inferiority studies involving the threshold are more informative when reported in terms of confidence intervals. Technical guidelines for the registration of pharmaceuticals for human use affirm that confidence intervals are as important as the use of statistical tests for estimating the relationship of dose and response and assert that confidence intervals should be provided when reporting estimates of treatment effects (ICH E9 Expert Working Group, 1999).

Similar recommendations have been made by the International Committee of Medical Journal Editors (2010) for the submission of manuscripts to biomedical journals, suggesting that findings are presented with appropriate indicators of measurement error or uncertainty, such as confidence intervals. Moreover, some study objectives, e.g., demonstrating bioequivalence of two drug compounds, are assessed solely using confidence intervals (U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, 2001).

Statistical methodology for interval estimation of a parameter and the difference between two parameters has been developed and widely discussed with respect to normally distributed (Niwitpong, 2010; and all references therein) and binary (Newcombe, 1998; and all references therein) outcomes. In addition, interval estimation for the Poisson rate, i.e., mean per unit of time, has been addressed (Pearson & Hartley, 1966; Hall, 1982; Schwertman & Martinez, 1994; Brown, Cai, & Dasgupta, 2001; Barker, 2002; Byrne & Kabaila, 2005; Johnson, Kemp, & Kotz, 2005; Swift, 2009; Khamkong, 2012; Patil & Kuklarni, 2012). However, interval estimation for the difference of two Poisson rates has received considerably less attention in the literature, despite the fact that the Poisson distribution is commonly encountered in practice, e.g., transplantation studies (Rochon, Aprile, & Cardella, 1992) and consumer research (Pelosi & Sandifer, 2003; Sharpie, De Veaux, & Velleman, 2010).

In this paper we discuss interval-estimating approaches that can be used to draw simultaneous inferences regarding the differences between two Poisson rates, when several experimental groups are compared to a common reference group in a fixed-sequence fashion. In particular, we discuss applications of the Hsu and Berger (1999) method (termed HBM) with respect to Poisson distributed outcomes, which have not yet been addressed in the literature. We show that this method can be used in practice for identifying the groups that are superior (in terms of the rate) to the reference group, while providing a confidence set for the differences between the rates.

The paper is outlined as follows. The problem formulation and a review of the HBM approach are given in Sections 2.1 and 2.2. A review of individual confidence interval methodology for the difference of two Poisson rates, including several methods for incorporating a continuity correction, is given in Sections 2.3 and 2.4. Two examples of the numeric application of the discussed methods are given in Section 3. Design and results of simulation studies conducted to investigate the performance of methods are given in Sections 4. Finally, a discussion of the advantages and disadvantages of the discussed methods is provided in Section 5.

2. Background

2.1. Problem Formulation

Consider a study with $N + 1$ groups, where the experimental groups are numbered as $1, \dots, N$, and the 0-th group corresponds to the reference group. Suppose that there are n_i subjects in the i -th group, $i = 0, \dots, N$, and the (i, j) -th subject, $i = 0, \dots, N$ and $j = 1, \dots, n_i$ is followed for some time period t_{ij} . Let X_{ij} denote the response for the (i, j) -th subject, i.e., the

number of occurrences of the event of interest in the time period t_{ij} . Furthermore, let λ_i be the true rate of occurrences of the event of interest per unit of time for group i , $i = 0, \dots, N$, such that the rates are non-decreasing, i.e.,

$$\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_N. \quad (1)$$

We assume that $X_{ij} \sim \text{Poisson}(\lambda_i t_{ij})$ and that the X_{ij} 's are mutually independent; thus the total number of occurrences across all subjects in the i -th group, $X_i = \sum_{j=1}^{n_i} X_{ij}$, follows $\text{Poisson}(\lambda_i t_i)$, where $t_i = \sum_{j=1}^{n_i} t_{ij}$ (Casella & Berger, 2002). Note that if all subjects are observed for the same amount of time, which is treated as one unit, then $t_i = n_i$ and $X_i \sim \text{Poisson}(\lambda_i n_i)$; this is another commonly used scenario that we term the one-unit scenario (Weissfeld, St. Laurent, & Moulto, 1991; Schwertman & Martinez, 1994; Stamey & Hamilton, 2006; Krishnamoorthy & Lee, 2012).

Another related scenario, called here the incidence rate scenario, was considered by Liu, Wang, Liu, and Snaveley (2006) with respect to a two-group design. In the incidence rate scenario t_{ij} represents the exposure time for the (i, j) -th subject, where if a subject experiences multiple events then t_{ij} is the time to the first event, and if the subject does not experience the event then t_{ij} is censored at the last follow-up time. In addition, X_i denotes the number of subjects who experienced the event in the i -th group and t_i denotes the total exposure time for the subjects in the i -th group. Then the ratio X_i/t_i provides the incidence rate for the i -th group.

Suppose that the study goal is to obtain the confidence set for the differences $\lambda_i - \lambda_0$, $i = 1, \dots, N$, and then based on this set identify the experimental groups that have significantly larger rates relative to the reference group.

The Bonferroni method (Dunn, 1959, 1961) (termed BONM) is, probably, the most straightforward method for assessing the study goal, and its application to the problem of estimating pair-wise differences of Poisson rates has been discussed in the literature (Scheaffer, 1980). Applications of the BONM approach to the considered settings are illustrated in Example 1. However, in many settings, the HBM can also be used. Similar to the BONM, it does not involve the joint distribution of the associated statistics and thus can be applied as long as individual confidence intervals are available; in addition the HBM may also be more powerful (Hsu & Berger, 1999).

2.2. Hsu and Berger (1999) Stepwise Confidence Intervals

In this section we discuss applications of the HBM to the settings discussed in Section 2.1. This procedure is compatible with the following α -level fixed-sequence hypothesis testing strategy (Hsu & Berger, 1999; Rosenbaum, 2008). Consider the hypotheses

$$H_{0i}: \lambda_i - \lambda_0 = 0 \quad \text{and} \quad H_{ai}: \lambda_i - \lambda_0 > 0, i = 1, \dots, N. \quad (2)$$

Then under the monotonicity assumption (1) these hypotheses are tested in a sequence starting with the N -th hypothesis, each at level α , provided that the previously tested hypothesis is rejected. The procedure stops if a hypothesis is accepted or all hypotheses have been tested and rejected.

The HBM can be outlined as follows. Let L_i be a $100(1 - \alpha)\%$ lower one-sided individual confidence bound for the difference $\lambda_i - \lambda_0$, $i = 1, \dots, N$.

Step 1: If $L_N < 0$, then assert $\lambda_N - \lambda_0 < 0$ and go to Step 2. If $L_N \geq 0$, then assert $\lambda_N - \lambda_0 \geq L_N$ and stop.

Step 2: If $L_{N-1} > 0$, then assert $\lambda_{N-1} - \lambda_0 > 0$ and go to Step 3. If $L_{N-1} \leq 0$, then assert $\lambda_{N-1} - \lambda_0 \geq L_{N-1}$ and stop.

...

Step N : If $L_1 > 0$, then assert $\lambda_1 - \lambda_0 > 0$ and go to step $N + 1$. If $L_1 \leq 0$, then assert $\lambda_1 - \lambda_0 \geq L_1$ and stop.

Step $N + 1$: Assert that $\lambda_i - \lambda_0 > 0$, for all $i = 1, \dots, N$ and stop.

Suppose M ($1 \leq M \leq N$) is the step in which the procedure stops. Then the resulting stepwise confidence intervals have a lower bound of zero for the first $M - 1$ compared differences, L_{N-M+1} for the M -th compared difference $\lambda_{N-M+1} - \lambda_0$, and $-\infty$ for the other differences.

While, originally, the HBM was presented in the framework of efficacy trials targeted at determining the minimum effective dose (MED) of a drug in the case of normally distributed outcomes (Hsu & Berger, 1999), applications with respect to other problems and distributional settings have been discussed in the literature (Tao, Shi, Guo, & Gao, 2002; Bretz, Hothorn, & Hsu, 2003; Chen, 2008; Peng, Lee, Davis, & Wang, 2008). In particular, Chen (2008) considered the problem of identifying the MED in the case of binary outcomes. The author investigated the HBM based on the individual Wald and Peskun (1993) confidence intervals in terms of power and Type I error rate with respect to nine types of dose-response curves. It was confirmed that the HBM could offer substantial advantages in terms of power over the Holm (1979) step-down procedure. Chen (2008) also showed that the HBM with the (uncorrected for continuity) Wald interval can result in an inflated overall Type I error rate.

2.3. Individual Confidence Intervals for the Difference of Two Poisson Rates

As is mentioned in Section 2.2., in order for the HBM stepwise confidence intervals to form a $100(1 - \alpha)\%$ simultaneous confidence set, the individual confidence intervals should have a $100(1 - \alpha)\%$ confidence level. In this section we review three methods for constructing the asymptotic two-sided and one-sided individual confidence interval for the difference between two Poisson rates: two well-known methods, i.e., Wald and Pooled, and one recently proposed method, i.e., the Moment-based method. We use the following representation for all confidence intervals for the difference $\lambda_i - \lambda_0$ (i is fixed, $1 \leq i \leq N$): the Wald, Pooled, and Moment-based $100(1 - \alpha)\%$ two-sided confidence intervals are,

respectively, denoted by (W_L, W_U) , (P_L, P_U) , and (M_L, M_U) , whereas the corresponding $100(1 - \alpha)\%$ one-sided analogs are, respectively, denoted by $(W, +\infty)$, $(P, +\infty)$, and $(M, +\infty)$.

Let x_0 and x_i be the observed total number of occurrences in the reference and the i -th experimental group, respectively, i.e., the observed values of X_0 and X_i , respectively. Construct $\hat{\lambda}_0 = x_0/t_0$, $\hat{\lambda}_i = x_i/t_i$, and $\hat{d}_i = \hat{\lambda}_i - \hat{\lambda}_0$. Let $Z_{1-\alpha}$ denote the $(1 - \alpha)$ th quantile of the standard normal distribution. Then the standard Wald bounds (Weissfeld et al., 1991; Schwertman & Martinez, 1994; Stamey & Hamilton, 2006; Liu, Wang, Liu, & Snavey, 2006; Krishnamoorthy & Lee, 2012) are given by

$$W_L = \hat{d}_i - z_{1-\frac{\alpha}{2}}(SE), W_U = \hat{d}_i + z_{1-\frac{\alpha}{2}}(SE), W = \hat{d}_i - z_{1-\alpha}(SE), \quad (3)$$

$$\text{where } SE = \sqrt{\frac{\hat{\lambda}_i}{t_i} + \frac{\hat{\lambda}_0}{t_0}}.$$

Liu et al. (2006) discussed applications of the two-sided Wald interval (and several others) to the incidence rate scenario. The authors conducted a simulation study, where the time to event followed the exponential distribution with parameter λ_i for the experimental treatment (or λ_0 for the reference treatment), and the maximum follow-up time was 1 unit. The simulation configurations included rates between 0 and 5 for each group, and a sample size of 100 per group. It was shown that when the rate is small (less than 1) for both groups, then the coverage probability of the Wald interval could be less than nominal, and when at least one rate is relatively large (1 to 5), then the Wald method has appropriate coverage probability. A similar conclusion was drawn by other authors (Weissfeld et al., 1991; Stamey & Hamilton, 2006), who compared the performance of the Wald interval with the performance of several Bayesian methods and an analogous Wald method that is based on the Student's t distribution instead of the standard normal distribution (Stamey & Hamilton, 2006); the one-unit scenario was also considered in these settings (Weissfeld et al., 1991; Stamey & Hamilton, 2006).

Next we discuss the Pooled method for constructing the asymptotic two-sided and one-sided individual confidence interval for the difference between two Poisson rates. The Pooled method incorporates the following bounds:

$$P_L = \hat{d}_i - z_{1-\frac{\alpha}{2}}(SE), P_U = \hat{d}_i + z_{1-\frac{\alpha}{2}}(SE), P = \hat{d}_i - z_{1-\alpha}(SE),$$

$$\text{where } SE = \sqrt{\hat{\lambda} \left(\frac{1}{t_i} + \frac{1}{t_0} \right)}$$

and $\hat{\lambda} = (x_i + x_0)/(t_i + t_0)$. When each subject can experience at most one event, the Pooled interval is mathematically equivalent to the ‘two-by-two table’ interval given by Miettinen (1975) and described by Liu et al. (2006) with respect to the incidence rate problem. Using simulations, briefly described above, Liu et al. (2006) concluded that the ‘two-by-two table’ interval has poor performance: it provides appropriate coverage only when rates are similar. The Wald interval either outperforms or performs similarly to the ‘two-by-two table’ interval in all scenarios reported by Liu et al. (2006).

The third method that we review is the Moment-based interval that was recently proposed by Krishnamoorthy and Lee (2012). The Moment-based interval has the following bounds:

$$M_L = \hat{d}_i + \frac{z_{1-\frac{\alpha}{2}}^2}{2} \left(\frac{1}{t_i} - \frac{1}{t_0} \right) - Z_{1-\frac{\alpha}{2}}(SE_1), \quad M_U = \hat{d}_i + \frac{z_{1-\frac{\alpha}{2}}^2}{2} \left(\frac{1}{t_i} - \frac{1}{t_0} \right) + Z_{1-\frac{\alpha}{2}}(SE_1),$$

$$M = \hat{d}_i + \frac{Z_{1-\alpha}^2}{2} \left(\frac{1}{t_i} - \frac{1}{t_0} \right) - Z_{1-\alpha}(SE_2),$$

$$\text{where } SE_1 = \sqrt{\frac{\hat{\lambda}_i + \hat{\lambda}_0}{t_i + t_0} + \frac{z_{1-\frac{\alpha}{2}}^2}{4} \left(\frac{1}{t_i} - \frac{1}{t_0} \right)^2} \text{ and } SE_2 = \sqrt{\frac{\hat{\lambda}_i + \hat{\lambda}_0}{t_i + t_0} + \frac{z_{1-\alpha}^2}{4} \left(\frac{1}{t_i} - \frac{1}{t_0} \right)^2}.$$

As is mentioned by Krishnamoorthy and Lee (2012), under the balanced one-unit scenario the Moment-based confidence interval simplifies to the corresponding standard Wald interval, given in (3).

Krishnamoorthy and Lee (2012) compared the coverage probability for the Moment-based interval to several other methods (Li, Tang, Poon, & Tang, 2011) in the one-unit scenario. It was shown that the Moment-based interval offers substantial improvement in terms of the coverage probability over the Wald interval in some instances, e.g., unbalanced designs with moderate sample sizes (ranging from 10 to 60 per group) and small rate values (ranging from 0.2 to 1.2). In addition, in the balanced settings with moderate rates (at least 2), the Moment-based interval, which is equivalent to the Wald interval in this setting, was shown to perform satisfactorily.

2.4. Continuity-Corrected Individual Confidence Intervals for the Difference between Two Poisson Rates

Continuity-corrected methods for interval estimation were widely discussed with respect to binomial outcomes (Yates, 1934; Adler, 1951; Pirie & Hamdan, 1972; Schouten, Molenaar, van Strike, & Boomsma, 1980; Upton, 1982; Hauck & Anderson, 1986; Haviland, 1990; Martin Andres & Tapia Garcia, 2004; Martin Andres, Tapia Garcia, & Del Moral Avila, 2005; Soualakova & Bright, 2013). The most commonly used approaches incorporate the Yates continuity correction (Yates, 1934; Adler, 1951; Upton, 1982; Hauck & Anderson, 1986; Haviland, 1990), denoted by c_Y , Schouten-Molenaar-van Strik-Boomsma correction

(Schouten et al., 1980), denoted by c_S , and Hauck-Anderson correction (Hauck & Anderson, 1986), denoted by c_{HA} . In our settings, these corrections are given by

$$c_Y = \frac{1}{2t_0} + \frac{1}{2t_i}, c_S = \frac{1}{2\max\{t_0, t_i\}}, \text{ and } c_{HA} = \frac{1}{2\min\{t_0, t_i\}}.$$

Note that for $c_Y > c_{HA} > c_S > 0$ for $t_0 \neq t_i$; in the case of equal time intervals (i.e., $t_0 = t_i = t$), and $c_Y = 1/t$ and $c_S = c_{HA} = 1/(2t)$. In addition, the correction based on methods by Pirie and Hamdan (1972), Martin Andres and Tapia Garcia (2004), and Martin Andres et al. (2005) was addressed for binomial settings (Soualakova & Bright, 2013). The latter correction is very close to zero for large time intervals (or large sample sizes in the one-unit scenario) and thus is not discussed in this paper.

Although, many authors (Schwertman & Martinez, 1994; Barker, 2002; Byrne & Kabaila, 2005; Khamkong, 2012; Patil & Kuklarni, 2012) have studied the performance of the continuity-corrected Wald interval for a single Poisson rate, to the best of our knowledge, there has been only one article, by Schwertman and Martinez (1994), that has addressed continuity corrections for confidence intervals around the difference of two Poisson rates. Schwertman and Martinez (1994) discussed eleven methods for constructing approximate confidence intervals for the difference of two Poisson rates in the one-unit scenario. These intervals are based on the Poisson approximation to the binomial distribution. The authors assessed the coverage probability for the standard Wald interval, as well as the Wald interval with the Yates and Hauck-Anderson continuity corrections, and a continuity correction of 0.5 in balanced cases. The standard Wald interval was shown to perform well in terms of probability coverage in the balanced setting and several unbalanced settings, where the sample sizes were not largely different. In addition, the continuity-corrected Wald intervals were shown to be superior in some unbalanced settings, where the sample sizes were quite different.

The continuity-corrected Wald and Pooled confidence interval methods for the difference of two Poisson rates have a similar form to the ones proposed for binomial outcomes (Schwertman & Martinez, 1994). The Wald and Pooled $100(1 - \alpha)\%$ two-sided continuity-corrected confidence intervals are, respectively, given (and denoted) by $(W_L - c, W_U + c)$ and $(P_L - c, P_U + c)$, whereas the corresponding $100(1 - \alpha)\%$ one-sided continuity-corrected analogs are, respectively, given by $(W - c, +\infty)$, and $(P - c, +\infty)$, where c is c_Y , c_S , or c_{HA} ; the case with $c = c_0 = 0$ corresponds to the standard (uncorrected) intervals.

For simplicity, the continuity-corrected Wald methods are abbreviated as Wald- c_0 , Wald- c_Y , Wald- c_S , and Wald- c_{HA} . Similar abbreviations are used for the continuity-corrected Pooled methods.

3. Examples

We present two examples: Example 1 illustrates the possible differences in conclusions based on the HBM and BONM approaches, and corresponds to an observational study with the one-unit scenario. Example 2 illustrates the possible differences in conclusions among

HBM approaches incorporating different individual confidence intervals, and corresponds to an observational study with the general scenario, i.e., where the number of occurrences per time period is considered. In both examples we assume that all assumptions stated in Section 2.1 hold. Issues related to possible overdispersion (Consul & Famoye, 1992) and other violations to the assumptions are not discussed; the examples are presented purely for illustrative purposes.

Example 1. Brener et al. (2002) described the results of a reliability study conducted to assess the data quality of the 1999 Youth Risk Behavior Survey. A subsample of high school students responded to two questionnaires administered about two weeks apart. Among a number of collected items regarding youth health risk behaviors, information on the number of sex partners within the last three months was gathered. This reliability study did not incorporate any adjustments for the complex design that are commonly used in the analyses of regular Youth Risk Behavior Surveillance System data.

Suppose our goal is to seek evidence as to whether the mean number of sex partners (for the past three months) is significantly higher for 10th, 11th, and 12th graders when compared to 9th graders. We assume that the monotonicity assumption (1) holds, i.e., the mean number of sex partners is non-decreasing as grade increases. In addition to hypothesis testing decisions we are interested in the simultaneous 95% one-sided lower confidence bounds for the differences between the means.

Based on the baseline (1st wave) data, the calculated total number (over all students in the particular grade) of sex partners is 432 for 1,289 9th graders, 521 for 1,375 10th graders, 570 for 944 11th graders, and 523 for 680 12th graders. Thus, the observed means for the 9th, 10th, 11th, and 12th grades are, respectively, given by 0.3351, 0.3758, 0.6038, and 0.7691.

Table 1 depicts the results corresponding to the individual and BONM one-sided lower confidence bounds. As the table illustrates, all HBM interval estimating methods lead to the same conclusion: there are significant differences between the 10th, 11th, and 12th grade means when compared to the 9th grade mean. Similarly, all BONM interval estimating approaches lead to the same conclusion, but the conclusion is different from the above one: there are significant differences only between the 11th and 12th grade means when compared to the 9th grade mean. We also observe that in this example, when one uses the same type of overall approach (HBM or BONM), for the Wald and Pooled intervals, the Yates correction results in the widest intervals and standard interval estimating methods result in the narrowest intervals. Also, overall, the Pooled intervals are narrower than the Moment-based intervals, which are narrower than the Wald intervals.

Example 2. Boice, Preston, Davis, and Monson (1991) explored the relationship between repeated exposure to X-ray fluoroscopy and incidence of breast cancer in women. In the study, the hospital records of female patients who have been treated for pulmonary tuberculosis between 1925 and 1954 in Massachusetts were examined. For those patients that received lung collapse treatment using repeated X-ray fluoroscopy, the records were used to determine the radiation dose absorbed by the breast. The authors categorized the dose levels in centigray (cGy) into four groups corresponding to 1 – 99 cGy, 100 – 199 cGy,

200 – 299 cGy, and at least 300 cGy. The unexposed group served as the reference group and consisted of women who did not receive lung collapse therapy, i.e., the radiation absorbed dose 0 is cGy. We assume that the monotonicity assumption (1) holds, i.e., the breast cancer rate is non-decreasing in dose level group. Table 2 presents a subset of the data given by Boice et al. (1991).

The authors computed the ratio of observed versus expected cases of breast cancer for the combined exposed group to be 1.29 and the ratio for the unexposed group to be 0.86, and concluded that the combined exposed group exhibited excess risk of developing breast cancer while the unexposed group did not. It was also found that the risk of breast cancer increased significantly with increasing radiation dose.

We use the data summary to seek evidence that the rate of breast cancer is significantly higher in the radiation exposed groups when compared to the unexposed group and to obtain simultaneous 97.5% one-sided lower confidence bounds for the differences in rates. Table 3 presents the resulting HBM lower bounds. Based on these results we draw the following conclusions. The HBM with the Wald- c_Y and Wald- c_{HA} intervals result in no significant differences between the rates, thus all radiation dose groups are concluded to be equivalent in terms of the breast cancer rate. The Wald- c_0 and Wald- c_S intervals result in only one significantly higher rate (than the 0-dose rate), which corresponds to the highest radiation dose. The other methods all lead to significantly higher rates for the three highest radiation doses and non-significant difference between the rates corresponding to doses of 1 – 99 cGy and 0 cGy.

4. Simulation studies

Simulation studies are conducted to compare the performance of the nine discussed individual continuity-corrected and uncorrected two-sided confidence intervals for the difference between two Poisson rates in terms of their coverage probability, and secondly to investigate the performance of the corresponding HBMs (as a hypothesis testing approach) in terms of family-wise error rate and power.

4.1. Performance of Individual Confidence Intervals

First, for each simulation configuration $(\lambda_0, t_0, \lambda_1, t_1)$ specified below, the data were generated from two independent Poisson distributions, $Poisson(\lambda_0 t_0)$ and $Poisson(\lambda_1 t_1)$. Then the two-sided 95% confidence interval methods were applied to these data and the corresponding confidence bounds were noted. The above steps were repeated 10,000 times, leading to a maximum possible simulation error of 0.005. The percentage of replications where the difference of rates, $\lambda_1 - \lambda_0$, was correctly captured between the lower and upper bounds was computed for each approach. This percentage provided an estimate of the true coverage probability. A method was said to result in under-coverage in the considered setting if the observed percentage was below 94.67%, where $94.6\% = 95.0\% - 1.645 \sqrt{95.0\% * 5.0\% / 10,000}$.

Two sets of simulation configurations were considered. The first set included the relatively small rates $\lambda_0 = 0.002$ and $\lambda_1 = 0.003, 0.004, 0.020, 0.200$, and 1.000; while the second set

included the relatively large rates $\lambda_0 = 1.0$ and $\lambda_1 = 2.0, 5.0$, and 10.0 . For each pair (λ_0, λ_1) several cases of equal and unequal time intervals were considered. Equal time interval settings (t_0, t_1) included $t_0 = t_1 = t$, where $t = 500, 1,000, 10,000$, and $50,000$. Unequal time interval settings (t_0, t_1) included settings with $t_0 < t_1$ such as $(200, 500)$, $(500, 1,000)$, $(6,000, 10,000)$, and $(30,000, 50,000)$, and settings with $t_0 > t_1$ such as $(500, 200)$, $(1,000, 500)$, $(10,000, 6,000)$, and $(50,000, 30,000)$. Thus, we considered 36 settings with equal time intervals, 36 settings with $t_0 < t_1$, and 36 settings with $t_0 > t_1$.

Tables 4 and 5 present the results for selected configurations with relatively small rates, as well as the summary results across all corresponding settings for equal- and unequal-interval settings, respectively. Note that in the case of equal time intervals, the uncorrected Wald, uncorrected Pooled, and Moment-based intervals reduce to the same formula, thus the continuity-corrected intervals do not differ based on Wald, Pooled, or Moment-based approaches. As is illustrated in Table 4, for equal-time interval settings, the methods with the Yates correction do not result in under-coverage in any of the considered settings. The other methods result in under-coverage in some of these settings, where the uncorrected approaches (including the Moment-based method) perform the worst in terms of the probability coverage.

In settings with $t_0 < t_1$, as is illustrated in Table 5, for large t_0 and t_1 values, all methods perform satisfactorily. However, only Wald- c_Y , Wald- c_{HA} , Pooled- c_Y , and Pooled- c_{HA} intervals do not result in under-coverage in any settings. In settings with $t_0 > t_1$, the Moment-based confidence interval is the only approach that does not result in under-coverage. The uncorrected Pooled confidence interval consistently results in under-coverage and the corrected Pooled intervals also perform unsatisfactorily in the majority of settings. The Wald confidence intervals exhibit under-coverage only for smaller t_0 and t_1 values, where the highest proportion of under-coverage is observed for the uncorrected method and the smallest proportion is observed for the Yates-corrected method.

Finally, we briefly discuss results corresponding to settings with relatively large rates; the exact simulation results are not displayed. In equal time interval settings and settings with $t_0 < t_1$, none of the methods resulted in under-coverage. Although the coverage probabilities in the equal time interval setting were very similar across methods, for settings with $t_0 < t_1$ the Pooled intervals exhibited higher coverage probability than the other intervals. In settings with $t_0 > t_1$, the Wald and Moment-based confidence intervals perform the best with no under-coverage; the Pooled intervals perform unsatisfactory and are the worst among the considered methods.

4.2. Performance of the HBM (as a Hypothesis Testing Approach) with Uncorrected and Continuity-corrected Confidence Intervals

The second simulation study was conducted to investigate the family-wise error rate (FWER) and power of the HBM as a hypothesis testing method in the case of $N = 2$. The null hypotheses $H_{01}: \lambda_1 - \lambda_0$ and $H_{02}: \lambda_2 - \lambda_0 = 0$ are tested in a fixed-sequence manner starting with H_{02} , and the corresponding alternatives are given by $H_{a1}: \lambda_1 - \lambda_0 > 0$ and $H_{a2}: \lambda_2 - \lambda_0 > 0$. The simulation study was performed as follows. First, for each design configuration specified below, the data were generated from three independent Poisson

distributions, $Poisson(\lambda_i t_i; i = 0, 1, 2)$. Then, the individual one-sided 97.5% lower confidence bounds for $\lambda_j - \lambda_0, j = 1, 2$, were constructed and corresponding decisions in terms of accepting or rejecting the hypotheses (in the fixed-sequence manner) were noted. The above steps were repeated 10,000 times, leading to a maximum possible simulation error of 0.005. The percentage of replications where at least one true null hypothesis was rejected provided the estimated FWER. An estimated FWER exceeding 2.8%, where $2.8\% = 2.5\% + 1.645 \sqrt{97.5\% * 2.5\% / 10,000}$, indicates that the FWER is inflated above the nominal level of 2.5%. The proportion of replications where all false and only false null hypotheses were rejected provided the estimated overall power of the fixed-sequence approach. The proportion of replications where a false null was rejected (independently of the decision regarding the other hypothesis) provided the estimated local power for the hypothesis.

4.2.1. Family-wise error rate—The following cases were considered: H_{01} is true and H_{02} is false (case 1), H_{01} is false and H_{02} is true (case 2), and H_{01} is true and H_{02} is true (case 3). Note that in case 2 the monotonicity assumption (1) does not hold; this case is considered in order to illustrate the performance of the fixed-sequence method (in terms of FWER) in settings where the rates do not satisfy (1). Two sets of Poisson rate values were considered. The first set included relatively small rates between 0.002 and 0.2 (these configurations are shown in Table 6), while the second set included larger rates between 1 and 10 in a similar pattern. In case 1 four configurations of rates $(\lambda_0, \lambda_1, \lambda_2)$ were considered, in case 2 three configurations of rates were considered, and in case 3 four configurations of rates were considered. For each of these configurations two settings of (t_0, t_1, t_2) were considered, i.e. (15,000, 30,000, 50,000) and (50,000, 30,000, 15,000), leading to settings in total.

As Table 6 illustrates, in settings with relatively small rates and $t_0 = 15,000, t_1 = 30,000$, and $t_2 = 50,000$ only the HBM with the Pooled or Moment-based intervals perform consistently well. In settings with $t_0 = 50,000, t_1 = 30,000$, and $t_2 = 15,000$, all HBM approaches perform satisfactory except for the HBM with Pooled- c_0 and Moment-based intervals; the latter approaches inflate the FWER in one case. In settings with larger rates, all confidence intervals resulted in a FWER of less than 2.8%, with FWER values ranging from 2.3% to 2.7%; the specific results are not presented.

4.2.2. Overall and local power—Simulations to evaluate the overall power were performed for a number of rate settings (some of these configurations are depicted in Table 7, the other configurations corresponded to larger rates where $\lambda_0 = 1$ and λ_1 and λ_2 were chosen to be between 1 and 10) with $t_0 = 15,000, t_1 = 30,000$, and $t_2 = 50,000$, and the same rate settings with $t_0 = 50,000, t_1 = 30,000$, and $t_2 = 15,000$. Non-monotone settings were also considered.

First, the results for settings with $t_0 = 15,000, t_1 = 30,000$, and $t_2 = 50,000$ are discussed. The results corresponding to the first set of rate configurations are depicted in Table 7. As is illustrated, in the setting of $\lambda_0 = 0.0002, \lambda_1 = 0.002$, and $\lambda_2 = 0.004$, i.e., H_{02} is the only false hypothesis with the true difference of 0.002, all methods perform with overall power of at least 95.0%. However, the methods perform poorly in the other three settings: the overall

power is below the 80% level for any method. The HBM with Wald confidence intervals achieve higher (overall and local) power than do the HBM with Pooled or Moment-based intervals, and uncorrected methods outperform the corresponding continuity-corrected ones. Among continuity-corrected methods, methods with the c_S correction correspond to the highest power, followed by c_{HA} , and then by c_Y . Thus, the HBM with the Wald- c_0 interval has the highest power while the HBM with the Pooled- c_Y interval has the lowest power.

Next, the results for settings with $t_0 = 50,000$, $t_1 = 30,000$, and $t_2 = 15,000$ are discussed. The results corresponding to the first set of rate configurations are illustrated in Table 8. Again, all methods perform with high power in the setting of $\lambda_0 = 0.002$, $\lambda_1 = 0.002$, and $\lambda_2 = 0.004$, where the overall power is at least 94.9% for all methods, while the methods perform poorly in the other three settings, with overall power below the 80% level for any method. The HBMs with Pooled and Moment-based intervals achieve higher power than do the HBMs with Wald intervals, where the HBM with the Wald- c_Y interval achieves the lowest power. In addition, uncorrected methods outperform the corresponding continuity-corrected ones.

We also note that while in several settings the HBM with Moment-based and Pooled- c_0 intervals resulted in the same power values, see Tables 7 and 8, the actual computed lower bounds were different.

In additional settings with larger rate values, all intervals perform very well in terms of the overall and local power (at least 97.5%). Thus, the exact results are not presented.

5. Discussion

Based on the results of our simulation study conducted with respect to probability coverage of individual two-sided confidence intervals, we can draw the following conclusions when relatively small Poisson rates are considered. First, in the equal-time interval settings (balanced one-unit scenarios), the methods with the Yates correction perform adequately well, while the uncorrected approaches perform unsatisfactory. Second, in settings with $t_0 < t_1$ (the experimental group sample size is larger than the reference group sample size in the one-unit case), the Wald and Pooled intervals with Yates and Hauck-Anderson corrections perform consistently well; the other methods have appropriate probability coverage only for relatively large values of t_0 and t_1 . Finally, in settings with $t_0 > t_1$ (the experimental group sample size is smaller than the reference group sample size in the one-unit scenario), the Moment-based confidence interval is the only approach that does not result in under-coverage. In contrast, simulations conducted with relatively large Poisson rates resulted in appropriate coverage for all confidence interval methodology, except in the setting of $t_0 > t_1$, where the Pooled intervals consistently resulted in under-coverage.

The second simulation study was conducted to evaluate the power and FWER of the HBM. Based on the observed results, we can conclude that the performance of the HBM as a hypothesis testing approach highly depends on what method is used for constructing the individual confidence intervals. Although, we did not consider any settings with equal-time intervals, we anticipate that methods with the Yates correction will perform the best in terms

of control of the FWER, while uncorrected methods will perform the best in terms of power but may inflate the FWER. When relatively small Poisson rates are considered in settings where $t_0 < t_1 < t_2$ and the time intervals are relatively large, only the HBM with Pooled or Moment-based intervals control the FWER. Among these methods, uncorrected Pooled and Moment-based intervals result in the highest overall power of the HBM; these two methods perform similarly, where the average overall power in the setting of relatively small Poisson rates is observed as 60.3% for the Moment-based and 59.8% for the uncorrected Pooled methods. When relatively small Poisson rates are considered in settings with $t_0 > t_1 > t_2$ and the time intervals are relatively large, all methods tend to control the FWER except for the HBM with the uncorrected Pooled and Moment-based intervals (inflation was observed in one case). Out of methods that do appear to control the FWER, the HBM with the Pooled- c_S outperforms the other methods, where it has the average overall power 70.9%, closely followed by the HBM with the Pooled- c_{HA} (69.9%) and Pooled- c_Y (69.2%). For relatively large Poisson rate configurations, all methods control the FWER and perform with high power.

We also explored the potential benefits of the Pirie-Hamdan (Pirie & Hamdan, 1972) and Martin Andres-Tapia Garcia (Marin Andres & Tapia Garcia, 2004; Martin Andres, Tapia Garcia, & Del Moral Avilia, 2005) corrections, but in all considered settings the correction was very close to 0, thus the resulting approach performed similarly to the uncorrected methods and therefore, the results are not presented here.

Although it is mentioned that monotonicity of parameters is not mandatory for the HBM (Hsu & Berger, 1999; Peng et al., 2008), the applications of the fixed-sequence testing and the HBM may be not logical in general cases. In these cases other simultaneous confidence band methods such as the Bonferroni, Scheffe's, Tukey's, or Dunnett's method (Hsu, 1996) should be considered. Also, note that when the number of comparisons is large, the HBM is expected to have substantial advantages in power over Bonferroni, Holm, and other methods; however, the investigation of this anticipated result is outside the scope of this paper.

Acknowledgements

Research of Julia N. Soulakova reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number R03CA165831. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors are also thankful to Nancy Brener (Survey Operations and Dissemination Team Leader, Division of Adolescent and School Health Centers for Disease Control and Prevention) for providing us with the 1999 Youth Risk Behavior Survey data.

References

- Adler F. Yates' correction and the statisticians. *Journal of the American Statistical Association*. 1951; 46(256):490–501.
- Barker L. A comparison of nine confidence intervals for a Poisson parameter when the expected number of events is 5. *The American Statistician*. 2002; 56(2):85–89.
- Boice JD, Preston D, Davis FG, Monson RR. Frequent chest x-ray fluoroscopy and breast cancer incidence among tuberculosis patients in Massachusetts. *Radiation Research*. 1991; 125:215–222.
- Brener ND, Kann L, McManus T, Kinchen SA, Sunberg EC, Ross JG. Reliability of the 1999 Youth Risk Behavior Survey Questionnaire. *Journal of Adolescent Health*. 2002; 31:336–342. [PubMed: 12359379]

- Bretz F, Hothorn LA, Hsu JC. Identifying effective and/or safe doses by stepwise confidence intervals for ratios. *Statistics in Medicine*. 2003; 22:847–858. [PubMed: 12627404]
- Brown LD, Cai T, Dasgupta A. Interval estimation for a binomial proportion (with discussion). *Statistical Science*. 2001; 16:101–133.
- Byrne J, Kabaila P. Comparison of Poisson confidence intervals. *Communications in Statistics – Theory and Methods*. 2005; 34:545–556.
- Casella, G.; Berger, RL. *Statistical inference*. 2nd ed.. Pacific Grove, CA: Duxbury; 2002.
- Chen JT. Inference on the minimum effective dose using binary data. *Communications in Statistics – Theory and Methods*. 2008; 37:2124–2135.
- Consul PC, Famoye F. Generalized Poisson Regression Model. *Communications in Statistics – Theory and Methods*. 1992; 21(1):89–109.
- Dunn OJ. Confidence intervals for the means of dependent, normally distributed variables. *Journal of the American Statistical Association*. 1959; 54(287):613–621.
- Dunn OJ. Multiple comparisons among means. *Journal of the American Statistical Association*. 1961; 56(293):52–64.
- Hall P. Improving the normal approximation when constructing one-sided confidence intervals for binomial or Poisson parameters. *Biometrika*. 1982; 69(3):647–652.
- Hauck WW, Anderson S. A comparison of large-sample confidence interval methods for the difference of two binomial probabilities. *The American Statistician*. 1986; 40(4):318–322.
- Haviland MG. Yates's correction for continuity and the analysis of 2×2 contingency tables. *Statistics in Medicine*. 1990; 9:363–367. [PubMed: 2362976]
- Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 1979; 6:65–70.
- Hothorn LA. How to deal with multiple treatment or dose groups in randomized clinical trials. *Fundamental & Clinical Pharmacology*. 2007; 21:137–154. [PubMed: 17391286]
- Hsu, JC. *Multiple Comparisons, Theory and Methods*. New York, NY: Chapman & Hall; 1996.
- Hsu JC, Berger RL. Stepwise confidence intervals without multiplicity adjustment for dose-response and toxicity studies. *Journal of the American Statistical Association*. 1999; 94:468–482.
- ICH E9 Expert Working Group. Statistical principles for clinical trials: ICH harmonized tripartite guideline. *Statistics in Medicine*. 1999; 18:1905–1942. [PubMed: 10532877]
- International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication. 2010. Retrieved from http://www.icmje.org/urm_full.pdf
- Johnson, N.; Kemp, A.; Kotz, S. *Univariate discrete distributions*. 3rd ed.. Hoboken, NJ: Wiley-Interscience; 2005.
- Khamkong M. Approximate confidence interval for the mean of Poisson distribution. *Open Journal of Statistics*. 2012; 2:204–207.
- Krishnamoorthy K, Lee M. New approximate confidence intervals for the difference between two Poisson means and comparison. *Journal of Statistical Computation and Simulation*. 2012
- Li HQ, Tang ML, Poon WY, Tang NS. Confidence intervals for difference between two Poisson rates. *Communications in Statistics – Simulation and Computation*. 2011; 40:1478–1493.
- Liu GF, Wang J, Liu K, Snaveley DB. Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials. *Statistics in Medicine*. 2006; 25:1275–1286. [PubMed: 16138360]
- Martin Andres A, Tapia Garcia JM. Optimal unconditional asymptotic test in 2x2 multinomial trials. *Communications in Statistics - Simulation and Computation*. 2004; 33(1):83–97.
- Martin Andres A, Tapia Garcia JM, Del Moral Avila MJ. Unconditional inferences on the difference of two proportions in cross-sectional studies. *Biometrical Journal*. 2005; 47(2):177–187. [PubMed: 16389914]
- Miettinen, O. *Principles of epidemiologic research*. Department of Epidemiology and Biostatistics, School of Public Health, Harvard University; 1975. Unpublished Course Text.
- Miettinen O, Nurminen M. Comparative analysis of two rates. *Statistics in Medicine*. 1985; 4:213–226. [PubMed: 4023479]

- Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*. 1998; 17:873–890. [PubMed: 9595617]
- Niwitpong S. Confidence interval for the difference of two normal population means with a known ratio of variances. *Applied Mathematical Sciences*. 2010; 4(8):347–359.
- Patil VV, Kuklarni HV. Comparison of confidence intervals for the Poisson mean: some new aspects. *REVSTAT – Statistical Journal*. 2012; 10(2):211–227.
- Pearson, ES.; Hartley, HO., editors. *Biometrika tables for statisticians*. 3rd ed.. Vol. Vol. 1. Cambridge: Cambridge University Press; 1966.
- Pelosi, MK.; Sandifer, TM. *Elementary statistics*. New York, NY: Wiley; 2003.
- Peng J, Lee CC, Davis KA, Wang W. Stepwise confidence intervals for monotone dose-response studies. *Biometrics*. 2008; 64:877–885. [PubMed: 18177461]
- Peskun P. A new confidence interval method based on the normal approximation for the difference of two binomial probabilities. *Journal of the American Statistical Association*. 1993; 88:656–661.
- Pirie WR, Hamdan MA. Some revised continuity corrections for discrete distributions. *Biometrics*. 1972; 28:693–701.
- Rochon J, Aprile MA, Cardella CJ. Analyzing the number of “rejection episodes” in renal transplant studies. *Transplantation*. 1992; 53:1236–1242. [PubMed: 1604478]
- Rosenbaum PR. Testing hypotheses in order. *Biometrika*. 2008; 95(1):248–252.
- Scheaffer RL. Multiple comparisons for Poisson rates. *Journal of Quality Technology*. 1980; 12(2):94–97.
- Schouten HJA, Molenaar IW, van Strike R, Boomsma A. Comparing two independent binomial proportions by a modified chi square test. *Biometrical Journal*. 22(3):241–248.
- Schwertman N, Martinez R. Approximate confidence intervals for the difference in two Poisson parameters. *Journal of Statistical Computation and Simulation*. 1994; 50:235–247.
- Schwertman N, Martinez R. Approximate Poisson confidence limits. *Communications in Statistics – Theory and Methods*. 1994; 23(5):1507–1529.
- Sharpie, NR.; De Veaux, RD.; Velleman, PF. *Business statistics*. Boston, MA: Addison Wesley; 2010.
- Soulakova JN, Bright BC. Applications of asymptotic confidence intervals with continuity corrections for asymmetric comparisons in non-inferiority trials. *Pharmaceutical Statistics*. 2013; 12:147–155. [PubMed: 23554217]
- Stamey J, Hamilton C. A note on confidence intervals for a linear function of Poisson rates. *Communications in Statistics – Simulation and Computation*. 2006; 35:849–856.
- Swift MB. Comparison of confidence intervals for a Poisson mean – further considerations. *Communications in Statistics – Theory and Methods*. 2009; 38:748–759.
- Tao J, Shi NZ, Guo J, Gao W. Stepwise procedures for the identification of minimum effective dose with unknown variances. *Statistics & Probability Letters*. 2002; 57:121–131.
- Upton GJG. A comparison of alternative tests for the 2×2 comparative trial. *Journal of the Royal Statistical Society (Series A)*. 1982; 145(1):86–105.
- U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Guidance for Industry, Statistical Approaches to Establishing Bioequivalence. 2001. Retrieved from <http://www.fda.gov/downloads/Drugs/Guidances/ucm070244.pdf>
- Weissfeld L, St. Laurent R, Moulto L. Confidence intervals for the comparison of two poisson distributions. *Communications in Statistics – Theory and Methods*. 1991; 20(10):3071–3081.
- Yates F. Contingency tables involving small numbers and the test. *Journal of Royal Statistical Society (Supplement)*. 1934; 1:217–234.

Table 1

Individual(top entry) and Bonferroni (bottom entry) simultaneous 95% one-sided lower bounds for Example 1

	Comparisons		
	12 th Grade versus 9 th Grade	11 th grade versus 9 th grade	10 th Grade versus 9 th Grade
Wald- c_0	0.3726	0.2193	0.0057
	0.3546	0.2048	-0.0055
Wald- c_Y	0.3715	0.2184	0.0049
	0.3535	0.2039	-0.0062
Wald- c_S	0.3722	0.2189	0.0053
	0.3542	0.2045	-0.0058
Wald- c_{HA}	0.3719	0.2188	0.0053
	0.3539	0.2043	-0.0059
Pooled- c_0	0.3797	0.2215	0.0056
	0.3637	0.2076	-0.0056
Pooled- c_Y	0.3786	0.2206	0.0049
	0.3626	0.2067	-0.0063
Pooled- c_S	0.3793	0.2211	0.0053
	0.3633	0.2072	-0.0059
Pooled- c_{HA}	0.3789	0.2209	0.0052
	0.3630	0.2071	-0.0060
Moment-based	0.3736	0.2197	0.0056
	0.3562	0.2055	-0.0056

Table 2
Example 2 data summary for each radiation absorbed group (data taken from Boice et al.(1991))

	Radiation Dose to Breast(cGy)				
	0	1–99	100–199	200–299	300+
Total Number of Breast Cancer Cases	87	75	44	14	9
Total Time at Risk (years)	48919	33724	15453	3757	1675
Annual Rate of Breast Cancer Cases	0.00177	0.00222	0.00285	0.00373	0.00537

Table 3

Hsu and Berger (1999) simultaneous 97.5% one-sided lower confidence bounds

	Radiation Absorbed Dose Group (cGy) Comparison			
	300+ vs. 0	200–299 vs. 0	100–199 vs. 0	1–99 vs. 0
Wald- c_0	0	−0.00004		
Wald- c_Y	−0.00024			
Wald- c_S	0	−0.00005		
Wald- c_{HA}	−0.00023			
Pooled- c_0	0	0	0	−0.00017
Pooled- c_Y	0	0	0	−0.00019
Pooled- c_S	0	0	0	−0.00018
Pooled- c_{HA}	0	0	0	−0.00018
Moment-based	0	0	0	−0.00016

Table 4

Selected results from and summary of simulated coverage probability(%) of the individual two-sided 95% confidence intervals for equal-interval settings

	c_0	c_Y	$c_S(c_{HA})$
$\lambda_0 = 0.002, \lambda_1 = 0.003,$			
$t_0 = t_1 = 50,000$	95.0	95.6	95.3
$t_0 = t_1 = 10,000$	95.0	96.5	95.9
$t_0 = t_1 = 1,000$	92.7	98.4	96.9
$t_0 = t_1 = 500$	89.4	99.9	91.1
$\lambda_0 = 0.002, \lambda_1 = 0.004,$			
$t_0 = t_1 = 50,000$	95.0	95.7	95.5
$t_0 = t_1 = 10,000$	95.0	96.2	95.8
$t_0 = t_1 = 1,000$	94.9	97.6	97.0
$t_0 = t_1 = 500$	86.2	94.8	93.8
$\lambda_0 = 0.002, \lambda_1 = 0.020,$			
$t_0 = t_1 = 50,000$	95.4	95.7	95.6
$t_0 = t_1 = 10,000$	94.8	95.6	95.3
$t_0 = t_1 = 1,000$	94.2	96.4	95.5
$t_0 = t_1 = 500$	93.2	96.3	95.2
$\lambda_0 = 0.002, \lambda_1 = 0.200,$			
$t_0 = t_1 = 50,000$	95.0	95.1	95.2
$t_0 = t_1 = 10,000$	95.1	95.3	95.2
$t_0 = t_1 = 1,000$	94.5	95.3	95.0
$t_0 = t_1 = 500$	94.6	95.8	95.7
$\lambda_0 = 0.002, \lambda_1 = 1.000,$			
$t_0 = t_1 = 50,000$	95.2	95.2	95.2
$t_0 = t_1 = 10,000$	95.0	95.1	95.0
$t_0 = t_1 = 1,000$	95.1	95.5	95.2
$t_0 = t_1 = 500$	94.9	95.3	95.2
Minimum coverage probability	86.2	94.8	91.1
Maximum coverage probability	95.4	99.9	97.0
Average coverage probability	94.0	96.1	95.2
Proportion of under-coverage	0.35	0.00	0.10

Note: Values below 94.6% are highlighted.

Table 5

Selected results from and summary of simulated coverage probability (%) of the individual two-sided 95% confidence intervals for unequal-interval settings, $\lambda_0 = 0.002$

				Wald				Pooled				Moment-based
				c_0	c_Y	c_S	c_{HA}	c_0	c_Y	c_S	c_{HA}	
Λ_1	t_0	t_1										
0.003	30,000	50,000	94.9	95.7	95.2	95.4	96.0	96.8	97.1	96.4	96.7	94.8
	6,000	10,000	94.6	96.3	95.3	95.7	95.9	96.4	96.7	96.6	96.6	94.7
	500	1,000	94.9	99.9	96.9	97.6	93.6	99.4	96.6	96.6	96.6	93.6
	200	500	83.9	100.0	84.8	99.9	83.2	99.8	83.2	99.8	99.8	77.5
1.000	30,000	50,000	95.0	95.1	95.0	95.0	98.7	98.8	98.8	98.7	98.7	95.0
	6,000	10,000	94.8	94.9	94.8	94.8	98.7	98.8	98.7	98.7	98.7	94.8
	500	1,000	95.0	95.4	95.1	95.3	99.5	99.6	99.5	99.5	99.5	95.0
	200	500	94.6	95.8	95.2	95.7	99.8	99.8	99.8	99.8	99.8	94.5
Summary for settings with $t_1 > t_0$												
Minimum coverage probability				83.9	94.9	84.8	94.8	83.2	96.8	83.2	96.5	77.5
Maximum coverage probability				95.9	100.0	98.3	99.9	99.8	99.9	99.8	99.9	95.2
Average coverage probability				94.1	96.8	94.8	96.3	96.5	98.8	97.0	98.5	92.5
Proportion of under-coverage				0.20	0.00	0.10	0.00	0.15	0.00	0.10	0.00	0.45
0.003	50,000	30,000	95.2	96.0	95.4	95.6	93.9	94.9	94.3	94.5	94.5	95.3
	10,000	6,000	94.9	96.3	95.5	95.7	93.9	95.8	94.6	95.0	95.0	95.0
	1,000	500	76.6	98.5	89.4	93.6	93.4	99.5	95.5	95.6	95.6	98.4
	500	200	74.5	100.0	77.9	100.0	74.6	99.9	77.9	99.0	99.0	99.0
1.000	50,000	30,000	94.9	95.0	95.0	95.0	86.9	87.0	86.9	87.0	87.0	95.0

			Wald			Pooled			Moment-based		
				c_0	c_Y	c_S	c_{HA}	c_0	c_Y	c_S	c_{HA}
Λ_1	t_0	t_1									
	10,000	6,000	95.0	95.1	95.0	95.0	95.1	86.8	87.1	86.8	86.9
	1,000	500	95.0	95.5	95.3	95.3	95.3	83.8	84.7	83.9	84.5
	500	200	94.8	95.5	95.0	95.1	95.1	78.3	80.4	79.6	79.6
Summary for settings with $t_0 > t_1$											
Minimum coverage probability			55.1	93.4	77.9	86.2	74.6	80.4	77.9	79.6	94.7
Maximum coverage probability			95.2	100.0	95.5	100.0	93.9	99.9	96.0	99.7	99.0
Average coverage probability			90.1	95.9	92.5	95.0	86.2	90.8	87.8	90.2	95.7
Proportion of under-coverage			0.40	0.10	0.30	0.15	1.00	0.65	0.85	0.75	0.00

Note: Values below 94.6% are highlighted.

Table 6

Simulated family-wise error rate (%) of the HBM based on the one-sided 97.5% lower confidence intervals (the following abbreviation is used: W for Wald, P for Pooled, and MB for Moment-based)

	$\lambda_0 = \lambda_1 = 0.002, \lambda_0 < \lambda_2$				$\lambda_0 = \lambda_2 = 0.002, \lambda_0 < \lambda_1$				$\lambda_0 = \lambda_1 = \lambda_2$			
	λ_2				λ_2				λ_2			
	0.003	0.004	0.020	0.200	0.003	0.020	0.200	0.200	0.002	0.003	0.020	0.200
$t_0 = 15,000, t_1 = 30,000, t_2 = 50,000$												
W- q_0	3.1	3.3	3.4	2.9	3.9	3.4	3.7	3.8	3.5	3.1	2.7	
W- c_Y	2.5	2.6	2.7	2.1	3.2	2.6	3.0	3.1	2.8	2.9	2.7	
W- c_S	3.0	3.0	3.1	2.6	3.7	3.2	3.6	3.7	3.2	3.0	2.7	
W- c_{HA}	2.7	2.8	2.8	2.3	3.4	2.7	3.2	3.3	2.9	2.9	2.7	
P- q_0	2.3	2.4	2.5	2.0	2.5	1.9	2.4	2.2	2.3	2.6	2.6	
P- c_Y	1.8	1.8	1.8	1.5	2.0	1.5	1.8	1.7	1.7	2.4	2.5	
P- c_S	2.1	2.1	2.3	1.8	2.4	1.8	2.2	2.1	2.1	2.6	2.6	
P- c_{HA}	2.0	1.9	2.0	1.6	2.1	1.6	1.9	1.8	1.9	2.4	2.5	
MB	2.3	2.4	2.5	2.0	2.5	1.9	2.4	2.2	2.3	2.6	2.6	
$t_0 = 50,000, t_1 = 30,000, t_2 = 15,000$												
W- q_0	1.6	2.1	2.0	2.1	1.6	1.6	1.6	1.6	1.6	2.0	2.2	
W- c_Y	1.2	1.8	1.6	1.7	1.3	1.3	1.2	1.3	1.3	1.8	2.2	
W- c_S	1.5	2.0	1.8	2.0	1.5	1.5	1.5	1.5	1.6	1.9	2.2	
W- c_{HA}	1.4	1.9	1.7	1.9	1.4	1.3	1.3	1.3	1.4	1.8	2.2	
P- q_0	2.1	2.5	2.5	2.5	2.9	2.6	2.7	2.8	2.5	2.2	2.3	
P- c_Y	1.6	2.1	2.0	2.1	2.3	2.1	2.2	2.3	2.1	2.1	2.3	
P- c_S	1.9	2.4	2.2	2.3	2.8	2.5	2.6	2.7	2.4	2.2	2.3	
P- c_{HA}	1.8	2.3	2.1	2.2	2.4	2.2	2.3	2.5	2.1	2.1	2.3	
MB	2.1	2.5	2.5	2.5	2.9	2.6	2.7	2.8	2.5	2.2	2.3	

Note: Values greater than 2.8% are highlighted

Simulated overall and local power of the HBM based on the one-sided 97.5% lower confidence interval (setting of $t_0 = 15,000, t_1 = 30,000, t_2 = 50,000$

Table 7

		$\lambda_0 = 0.002,$ $\lambda_1 = 0.002,$ $\lambda_2 = 0.004$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.003,$ $\lambda_2 = 0.004$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.003,$ $\lambda_2 = 0.003$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.004,$ $\lambda_2 = 0.003$
Wald- c_0	Overall Power	0.950	0.540	0.472	0.625
	Local Power H_{01}	-	0.540	0.555	0.969
	Local Power H_{02}	0.983	0.985	0.634	0.627
Wald- c_Y	Overall Power	0.955	0.503	0.429	0.584
	Local Power H_{01}	-	0.504	0.514	0.960
	Local Power H_{02}	0.981	0.980	0.593	0.586
Wald- c_S	Overall Power	0.953	0.526	0.458	0.610
	Local Power H_{01}	-	0.526	0.541	0.965
	Local Power H_{02}	0.983	0.983	0.621	0.611
Wald- c_{HA}	Overall Power	0.954	0.515	0.441	0.595
	Local Power H_{01}	-	0.515	0.527	0.962
	Local Power H_{02}	0.981	0.981	0.604	0.597
Pooled- c_0	Overall Power	0.952	0.492	0.403	0.546
	Local Power H_{01}	-	0.492	0.504	0.957
	Local Power H_{02}	0.975	0.974	0.552	0.548
Pooled- c_Y	Overall Power	0.952	0.450	0.357	0.498
	Local Power H_{01}	-	0.450	0.463	0.947
	Local Power H_{02}	0.969	0.970	0.503	0.500
Pooled- c_S	Overall Power	0.953	0.479	0.386	0.529
	Local Power H_{01}	-	0.480	0.489	0.953
	Local Power H_{02}	0.974	0.974	0.536	0.531
Pooled- c_{HA}	Overall Power	0.952	0.464	0.371	0.513
	Local Power H_{01}	-	0.464	0.476	0.950
	Local Power H_{02}	0.971	0.971	0.519	0.515

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

		$\lambda_0 = 0.002,$ $\lambda_1 = 0.002,$ $\lambda_2 = 0.004$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.003,$ $\lambda_2 = 0.004$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.003,$ $\lambda_2 = 0.003$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.004,$ $\lambda_2 = 0.003$
Moment-based	Overall Power	0.952	0.492	0.421	0.546
	Local Power H_{01}	-	0.492	0.504	0.957
	Local Power H_{02}	0.975	0.974	0.552	0.548

Simulated overall and local power of the HBM based on the one-sided 97.5% lower confidence interval (setting of $t_0 = 50,000$, $t_1 = 30,000$, $t_2 = 15,000$)

Table 8

		$\lambda_0 = 0.002,$ $\lambda_1 = 0.002,$ $\lambda_2 = 0.004$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.003,$ $\lambda_2 = 0.004$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.003,$ $\lambda_2 = 0.003$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.004,$ $\lambda_2 = 0.003$
Wald- c_0	Overall Power	0.952	0.763	0.436	0.530
	Local Power H_{01}	-	0.780	0.762	0.999
	Local Power H_{02}	0.973	0.971	0.528	0.530
Wald- c_Y	Overall Power	0.949	0.736	0.395	0.491
	Local Power H_{01}	-	0.756	0.740	0.998
	Local Power H_{02}	0.966	0.964	0.487	0.491
Wald- c_S	Overall Power	0.952	0.754	0.425	0.521
	Local Power H_{01}	-	0.771	0.754	0.999
	Local Power H_{02}	0.972	0.970	0.519	0.521
Wald- c_{HA}	Overall Power	0.949	0.746	0.406	0.501
	Local Power H_{01}	-	0.766	0.749	0.998
	Local Power H_{02}	0.968	0.965	0.496	0.501
Pooled- c_0	Overall Power	0.956	0.790	0.506	0.608
	Local Power H_{01}	-	0.800	0.784	0.999
	Local Power H_{02}	0.982	0.981	0.605	0.608
Pooled- c_Y	Overall Power	0.957	0.766	0.470	0.576
	Local Power H_{01}	-	0.778	0.761	0.999
	Local Power H_{02}	0.978	0.977	0.572	0.576
Pooled- c_S	Overall Power	0.957	0.781	0.496	0.600
	Local Power H_{01}	-	0.792	0.774	0.999
	Local Power H_{02}	0.981	0.980	0.598	0.600
Pooled- c_{HA}	Overall Power	0.956	0.775	0.481	0.584
	Local Power H_{01}	-	0.787	0.769	0.999
	Local Power H_{02}	0.979	0.978	0.580	0.584

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

		$\lambda_0 = 0.002,$ $\lambda_1 = 0.002,$ $\lambda_2 = 0.004$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.003,$ $\lambda_2 = 0.004$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.003,$ $\lambda_2 = 0.003$	$\lambda_0 = 0.002,$ $\lambda_1 = 0.004,$ $\lambda_2 = 0.003$
Moment-based	Overall Power	0.956	0.790	0.506	0.608
	Local Power H_{01}	-	0.800	0.783	0.999
	Local Power H_{02}	0.982	0.981	0.605	0.608