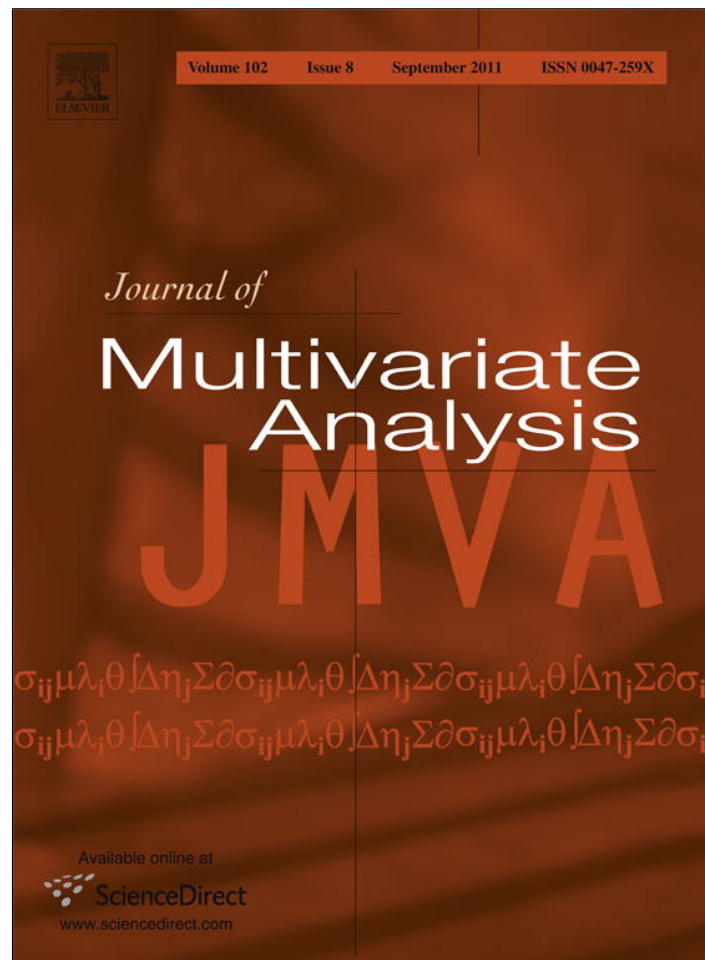


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## Journal of Multivariate Analysis

journal homepage: [www.elsevier.com/locate/jmva](http://www.elsevier.com/locate/jmva)

# A new class of minimum power divergence estimators with applications to cancer surveillance

Nirian Martín <sup>a,\*</sup>, Yi Li <sup>b,c</sup><sup>a</sup> Department of Statistics, Carlos III University of Madrid, 28903 Getafe, Madrid, Spain<sup>b</sup> Department of Biostatistics & Computational Biology, Dana Farber Cancer Institute, United States<sup>c</sup> Department of Biostatistics, Harvard School of Public Health, United States

## ARTICLE INFO

## Article history:

Received 10 May 2010

Available online 7 April 2011

## AMS subject classifications:

62H17

62J12

62P10

## Keywords:

Minimum power divergence estimators

Age-adjusted cancer rates

Annual percent change (APC)

Trends

Poisson sampling

## ABSTRACT

The annual percent change (APC) has been adopted as a useful measure for analyzing the changing trends of cancer mortality and incidence rates by the NCI SEER program. Difficulties, however, arise when comparing the sample APCs between two overlapping regions because of induced dependence (e.g., comparing the cancer mortality change rate of California with that of the national level). This paper deals with a new perspective for understanding the sample distribution of the test-statistics for comparing the APCs between overlapping regions. Our proposal allows for computational readiness and easy interpretability. We further propose a more general family of estimators, namely, the so-called minimum power divergence estimators, including the maximum likelihood estimators as a special case. Our simulation experiments support the superiority of the proposed estimator to the conventional maximum likelihood estimator. The proposed method is illustrated by the analysis of the SEER cancer mortality rates observed from 1991 to 2006.

© 2011 Elsevier Inc. All rights reserved.

## 1. Introduction

According to the World Health Statistics 2009, published by the World Health Organization, in 2004, the age-standardized mortality rate in high-income countries attributable to cancer deaths was 164 per 100,000. Cancer constituted the second highest cause of death after cardiovascular disease (its age-standardized mortality rate was equal to 408 per 100,000). For cancer prevention and control programs, such as the Surveillance, Epidemiology and End Results (SEER) in the United States (US), it is very important to rely on statistical tools to capture downward or upward trends of rates associated with each type of cancer and to measure their intensity accurately. These trends in cancer rates are defined within a specific spatial–temporal framework, that is, different geographic regions and time periods are considered.

Let  $r_{ki}$  be the expected value of the cancer rate associated with region  $k$  and the  $i$ th time point in a sequence of ordered  $I_k$  time points  $\{t_{ki}\}_{i=1}^{I_k}$ . We shall assume that Region 1 starts with the earliest time. Each point represents an equally spaced period of time, for instance a year, and thus without any loss of generality,  $t_{1i} = i$ ,  $i = 1, \dots, I_1$  (any change in origin or scale with respect to time should not affect a measure of trend). The cancer rates are useful for evaluating either the risk of developing cancer (cancer incidence rates) or dying from cancer (cancer mortality rates) at a specific moment. Statistically, the trend in cancer rates is an average rate of change per year in a given relatively short period of time framework when constant change along time has been assumed. The annual percent change (APC) is a suitable measure for comparing recent

\* Corresponding author.

E-mail address: [nirian.martin@uc3m.es](mailto:nirian.martin@uc3m.es) (N. Martín).

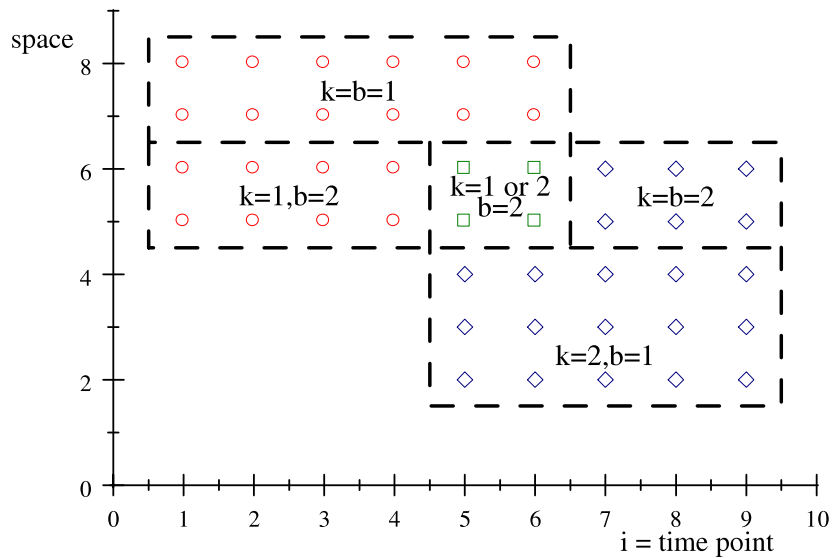


Fig. 1. Two overlapping regions not sharing the same period of time.

trends associated with age-adjusted expected cancer rates:

$$r_{ki} = \sum_{j=1}^J \omega_j r_{kji}, \tag{1}$$

where  $J$  is the number of age groups,  $\{\omega_j\}_{j=1}^J$  is the age distribution of the Standard Population ( $\sum_{j=1}^J \omega_j = 1, \omega_j > 0, j = 1, \dots, J$ ) and  $\{r_{kji}\}_{j=1}^J$  is the set of expected rates associated with the  $k$ th region ( $k = 1, 2$ ) at the time point  $t_{ki}$  ( $i = 1, \dots, I_k$ ), or the  $i$ th year, in each of the age groups ( $j = 1, \dots, J$ ). For example, the SEER Program applies as standard the US population of year 2000 with  $J = 19$  age groups  $[0, 1), [1, 5), [5, 10), [10, 15), \dots, [80, 85), [85, *)$ . More technical details can be found in [4] and [14]. The APC removes differences in scale by considering the proportion  $(r_{k,i+1} - r_{k,i})/r_{k,i} = r_{k,i+1}/r_{k,i} - 1$  under constant change assumption of the expected rates. Proportionality constant  $\theta_k = r_{k2}/r_{k1} = \dots = r_{kI_k}/r_{k,I_k-1}$  constitutes the basis for defining  $APC_k = 100(\theta_k - 1)$  as a percentage associated with the expected rates  $\{r_{kji}\}_{j=1}^J$  of the  $k$ th region. Since the models that deal with the APCs consider the logarithm of age-adjusted cancer rates, the previous formula is usually replaced by

$$APC_k = 100 (\exp(\beta_{1k}) - 1), \tag{2}$$

and we would like to make statistical inferences on parameter  $\beta_{1k}$  (see [3] and [15] for more details about the APC).

The data that are collected for modeling the APC associated with region  $k$ , are:

- $d_{kji}$ , the number of deaths (or incidences) in the  $k$ th region,  $j$ th age group, at the time point  $t_{ki}$ ;
- $n_{kji}$  the population at risk in the  $k$ th region,  $j$ th age group, at the time point  $t_{ki}$ ;

so that the r.v.s that generate  $d_{kji}, D_{kji}$ , are considered to be mutually independent. In a sampling framework we can define the empirical age-adjusted cancer rates as  $R_{ki} = \sum_{j=1}^J \omega_j R_{kji} = \sum_{j=1}^J \omega_j \frac{D_{kji}}{n_{kji}}$ , whose expected value is (1). Even though the assumption of “independence” associated with  $D_{kji}$  simplify the process of making statistical inference, it is in practice common to find situations in which the two APCs to be compared,  $APC_1$  and  $APC_2$ , share some data because there is an overlap between the two regions. For example, in [13] county-level data on 22 selected cancer sites during 1996–2005 are analyzed, so that the APC of each county is compared with the APC of Oregon state. It is not possible to assume independence between the data of counties (local level) and their state (global level). Moreover, the APC comparison between overlapping regions is more complicated when the APCs are not for the same period of time. For instance in the aforementioned study that appeared in [13], while Oregon APC was obtained for a period of time ending in 2005, the US APC was calculated for a period of time ending in 2004 because the US data of year 2005 were not available. Fig. 1 represents the most complicated overlapping case for two regions, where  $\{1, 6\} \times \{5, 8\}$  is the set of points of the first region,  $\{5, 9\} \times \{2, 6\}$  is the set the points of the second region,  $\{5, 6\} \times \{5, 6\}$  is the set of points of the overlapping region (boxed points). Each of the two regions have a portion of space and period of time not contained in the other one (circular points for region 1 and diamond points for region 2).

This paper is structured as follows. In Section 2 different models that establish the relationship between  $r_{ki}$  and  $\beta_{1k}$  are reviewed and the two basic tools for making statistical inferences are presented, the estimators and test-statistics for equal APCs. Specifically, the Age-stratified Poisson Regression model, introduced for the first time in [7], is highlighted as the model that arises as an alternative to improve the previous ones. Based on Power divergence measures, in Section 3 a

family of estimators that generalize the maximum likelihood estimators (MLEs) are considered for the Age-stratified Poisson Regression model. In addition, a new point of view for computing the covariance between the MLEs of  $\beta_{1k}$  is introduced inside the framework of this family of estimators and this is the key for substantially improving the Z-test statistic for testing the equality of APCs for the Age-stratified Poisson Regression model. In addition, such a methodology provides explicit and interpretable expressions of the covariance between the estimators of  $\beta_{1k}$ . We evaluate the performance of the new proposed methodology in Section 4 through a simulation study and we also consider an application example to Breast and Thyroid cancer data from California (CA) and the US population, extracted from the SEER\*STAT software of the SEER Program. Finally in Section 5 some concluding remarks are given.

## 2. Models associated with the annual percent change (APC)

When non-overlapping regions are taken into account, there are basically two models which allow us to estimate the APC starting from slightly different assumptions, the Age-adjusted Cancer Rate Regression model and Age-stratified Poisson Regression model. The main difference between them is based on the probability distribution of  $D_{kji}$ , number of deaths in the  $k$ th region,  $j$ th age group, at the time point  $t_{ki}$ : while the Age-adjusted Cancer Rate Regression model assumes normality for  $\log R_{ki}$  with  $D_{kji}$  having the same mean and variance, the Age-stratified Poisson Regression model assumes directly a Poisson random variable (r.v.) for  $D_{kji}$ . The Age-adjusted Cancer Rate Regression model establishes  $\log R_{ki} = \beta_{0k} + \beta_{1k}t_{ki} + \epsilon_{ki}$ , where  $\epsilon_{ki} \stackrel{\text{ind}}{\sim} \mathcal{N}(0, \sigma_{ki}^2)$  with  $\sigma_{ki}^2 = \sum_{j=1}^J \omega_j^2 r_{kji} / n_{kji} = \sum_{j=1}^J \omega_j^2 m_{kji} / n_{kji}^2$  under

$$E[D_{kji}] = \text{Var}[D_{kji}] = n_{kji}r_{kji} \equiv m_{kji}, \tag{3}$$

i.e.  $\log R_{ki} \stackrel{\text{ind}}{\sim} \mathcal{N}(\log r_{ki}, \sigma_{ki}^2)$  with

$$r_{ki} = \exp(\beta_{0k}) \exp(\beta_{1k}t_{ki}). \tag{4}$$

According to the Age-stratified Poisson Regression model [7],  $D_{kji} \stackrel{\text{ind}}{\sim} \mathcal{P}(n_{kji}r_{kji})$  and for  $r_{kji}$  it holds

$$\log r_{kji} = \beta_{0kj} + \beta_{1k}t_{ki} \quad \text{or} \quad \log \frac{m_{kji}}{n_{kji}} = \beta_{0kj} + \beta_{1k}t_{ki}. \tag{5}$$

Observe that the parametrization of both models is essentially the same because the expected age-adjusted rate  $r_{ki}$  in terms of (5) is equal to (4), where

$$\exp(\beta_{0k}) = \sum_{j=1}^J \omega_j \exp(\beta_{0kj}), \tag{6}$$

and thus for both models it holds that

$$\theta_k = \left( \frac{r_{klk}}{r_{k1}} \right)^{\frac{1}{t_{klk} - t_{k1}}} = \exp(\beta_{1k}). \tag{7}$$

The original estimators associated with the Age-adjusted Cancer Rate Regression model and Age-stratified Poisson Regression model are the Weighted Least Square estimators (WLSE) and Maximum Likelihood estimators (MLE) respectively.

The hypothesis testing for comparing the equality of trends of two regions,  $\mathcal{H}_0 : \text{APC}_1 = \text{APC}_2$ , is according to (2), equivalent to  $\mathcal{H}_0 : \beta_{11} - \beta_{12} = 0$ . Hence, the Z-test statistic for both models can be defined as

$$Z = \frac{\widehat{\beta}_{11} - \widehat{\beta}_{12}}{\sqrt{\widehat{\text{Var}}(\widehat{\beta}_{11} - \widehat{\beta}_{12})}}, \tag{8}$$

where  $\widehat{\beta}_{1k}$ ,  $k = 1, 2$  are the estimators of  $\beta_{1k}$  associated with each region,  $\widehat{\text{Var}}(\widehat{\beta}_{11} - \widehat{\beta}_{12})$  is the estimator of the variance of  $\widehat{\beta}_{11} - \widehat{\beta}_{12}$ ,  $\text{Var}(\widehat{\beta}_{11} - \widehat{\beta}_{12})$ . The expression of the variance is  $\text{Var}(\widehat{\beta}_{11} - \widehat{\beta}_{12}) = \sigma_{11}^2 + \sigma_{12}^2$ , with  $\sigma_{1k}^2 \equiv \text{Var}(\widehat{\beta}_{1k})$ ,  $k = 1, 2$ , for non-overlapping regions. When overlapping regions are taken into account, the methodology for obtaining the estimators as well as Z-test statistic (8) remain valid, but the given expression for  $\text{Var}(\widehat{\beta}_{11} - \widehat{\beta}_{12})$  is no longer valid. When the overlapping regions do not share the same period of time ( $t_{11} \neq t_{21}$  or  $I_1 \neq I_2$ ), we must consider a new reference point for index  $i$ , denoted by  $\bar{I}$ , such that  $t_{i\bar{I}}$  represents the time point within  $\{t_{2i}\}_{i=1}^{I_1}$  where the time series associated with the second region is about to start, i.e. we have  $\{t_{2i}\}_{i=1}^{I_2}$  such that  $t_{21} = t_{i\bar{I}} + 1$ . In particular, if  $t_{1i} = i$ ,  $i = 1, \dots, I_1$ , then  $t_{2i} = \bar{I} + i$ ,  $i = 1, \dots, I_2$ . Observe that  $\{t_{1i}\}_{i=\bar{I}+1}^{I_1}$ , or equivalently  $\{t_{2i}\}_{i=1}^{I_1-\bar{I}}$ , is the time series associated with the overlapping region ( $t_{1i} = t_{2,i-\bar{I}}$ ,  $i = \bar{I} + 1, \dots, I_1$ ). In Fig. 1  $I_1 = 6, I_2 = 5, \bar{I} = 4$  and thus we can distinguish three subregions  $\{5, 6\} \times \{1, \dots, 4\}$ ,  $\{5, 6\} \times \{5, 6\}$  and  $\{5, 6\} \times \{7, \dots, 9\}$ . Without any loss of generality each random variable  $D_{kji}$  can be decomposed into two summands

$$D_{kji} = D_{kji}^{(1)} + D_{kji}^{(2)} \tag{9}$$

where  $D_{kji}^{(1)}$ ,  $i \in \{1, \dots, I_k\}$ , is the number of deaths (or incidences) in the  $k$ th region,  $j$ th age group, at the time point  $t_{ki}$  for the subregion where there is no overlap in space;  $D_{kji}^{(2)}$ ,  $i \in \{1, \dots, I_k\}$ , is the number of deaths (or incidences) in the  $k$ th region,  $j$ th age group, at the time point  $t_{ki}$  for the subregion where there is overlap in space. Similarly,  $n_{kji} = n_{kji}^{(1)} + n_{kji}^{(2)}$  and  $m_{kji}(\boldsymbol{\beta}_k) = m_{kji}^{(1)}(\boldsymbol{\beta}_k) + m_{kji}^{(2)}(\boldsymbol{\beta}_k)$ . Observe that when  $i \in \{\bar{I} + 1, \dots, I_1\}$ , r.v.s  $D_{1ji}^{(2)}$  and  $D_{2j,i-\bar{I}}^{(2)}$  are associated with the same overlapping subregion. Revisiting the example illustrated in Fig. 1, it should be remarked that in the  $y$ -axis (space) there are more points than those that represent one realization of all r.v.s  $D_{kji}^{(b)}$  in each time point, but grouping the points belonging to the same vertical line inside the portion marked in dash we are referring to one realization of them (for instance, for  $t_{11} = 1$  we have two groups of points associated with  $D_{1j1}^{(1)}$ ,  $D_{1j1}^{(2)}$  respectively, while for  $t_{1j5} = t_{2j1} = 5$  we have three groups of points associated with  $D_{1j5}^{(1)}$ ,  $D_{1j5}^{(2)}$  or  $D_{2j1}^{(2)}$ ,  $D_{2j1}^{(1)}$ . Grouping points symbolize different extension in regions. In Fig. 1 there are 20 realizations of all r.v.s  $D_{kji}^{(b)}$  in total, 12 for region 1, 10 for region 2 and 2 r.v.s are shared for both regions.

It is important to understand r.v.s  $D_{kji}^{(b)}$ ,  $b \in \{1, 2\}$  as “homogeneous contributors” with respect to  $D_{kji}$ , i.e.  $D_{kji}^{(b)} \sim \mathcal{P}(m_{kji}^{(b)})$  such that (10) holds, and hence  $\{m_{1ji}^{(2)}(\boldsymbol{\beta}_1)\}_{i=\bar{I}+1}^{I_1}$  and  $\{m_{2ji}^{(2)}(\boldsymbol{\beta}_2)\}_{i=1}^{I_1-\bar{I}}$  are only equal when  $\beta_{11} = \beta_{12}$  (or equivalently, when  $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ ). Now we can say thoroughly that under  $\beta_{11} = \beta_{12}$ , the reason why  $\text{Cov}(\widehat{\beta}_{11}, \widehat{\beta}_{12}) = 0$  is not true inside  $\text{Var}(\widehat{\beta}_{11} - \widehat{\beta}_{12}) = \text{Var}(\widehat{\beta}_{11}) + \text{Var}(\widehat{\beta}_{12}) - 2\text{Cov}(\widehat{\beta}_{11}, \widehat{\beta}_{12})$  for overlapping regions is that  $\{D_{1ji}\}_{i=1, \dots, I_1; j=1, \dots, J}$  and  $\{D_{2ji}\}_{i=1, \dots, I_2; j=1, \dots, J}$  are not independent, because both regions share the same the set of r.v.s  $\{D_{1ji}^{(2)}\}_{i=\bar{I}+1, \dots, I_1; j=1, \dots, J}$  with  $D_{1ji}^{(2)} = D_{2j,i-\bar{I}}^{(2)}$ .

**Assumption 1.**  $D_{kji}^{(b)} \stackrel{\text{ind}}{\sim} \mathcal{P}(m_{kji}^{(b)})$ ,  $b \in \{1, 2\}$ , where for  $n_{kji}^{(b)} > 0$  the following holds

$$m_{kji}^{(b)} = \frac{n_{kji}^{(b)}}{n_{kji}} m_{kji}, \quad b \in \{1, 2\}. \tag{10}$$

We accept the case where  $n_{kji}^{(b)} = 0$ , for some  $b \in \{1, 2\}$ , so that  $D_{kji}^{(b)} = 0$  a.s. (degenerate r.v.) because  $m_{kji}^{(b)} = 0$ .

Regarding the basic models considered in the papers dealing with overlapping regions, the Age-stratified Poisson regression model can be considered as the most realistic one, actually they have been constructed by successive improvements on the previous models so that initially normality assumptions were taken as approximations of underlying Poisson r.v.s. In the first paper concerned about trend comparisons across overlapping regions [7], it is remarked that “... the derivation of  $\text{Cov}(\widehat{\beta}_{11}, \widehat{\beta}_{12}), \dots$ , is nontrivial as it requires a careful consideration of the overlapping of two regions”. The assumption considered by them (which is based on Pickle and White [11]) for the overlapping subregion is similar to the assumption considered herein in the sense that the overlapping subregion follows the same distribution considered for the whole region. A similar criterion was followed in [8,7].

### 3. Minimum power divergence estimators for an age-stratified Poisson regression model with overlapping

Let  $m_s$  be the expected value of the r.v. of deaths (or incidences)  $D_s$  associated with the  $s$ th cell of a contingency table with  $M_k \equiv JI_k$  cells ( $s = 1, \dots, M_k$ ). In this section, we consider model (5) in matrix notation so that the triple indices are unified in a single one by following a lexicographic order. Hence, the vector of cell means  $\mathbf{m}_k(\boldsymbol{\beta}_k) = (m_1(\boldsymbol{\beta}_k), \dots, m_{M_k}(\boldsymbol{\beta}_k))^T = (m_{k11}(\boldsymbol{\beta}_k), \dots, m_{kJI_k}(\boldsymbol{\beta}_k))^T$  of the multidimensional r.v. of deaths (or incidences)  $\mathbf{D}_k = (D_1, \dots, D_{M_k})^T = (D_{k11}, \dots, D_{kJI_k})^T$ , is related to the vector of parameters  $\boldsymbol{\beta}_k = (\beta_{0k1}, \dots, \beta_{0kj}, \beta_{1k})^T \in \Theta_k = \mathbb{R}^{J+1}$  according to

$$\log(\text{Diag}^{-1}(\mathbf{n}_k) \mathbf{m}_k(\boldsymbol{\beta}_k)) = \mathbf{X}_k \boldsymbol{\beta}_k \quad \text{or} \quad \mathbf{m}_k(\boldsymbol{\beta}_k) = \text{Diag}(\mathbf{n}_k) \exp(\mathbf{X}_k \boldsymbol{\beta}_k), \tag{11}$$

where  $\text{Diag}(\mathbf{n}_k)$  is a diagonal matrix of individuals at risk  $\mathbf{n}_k = (n_1, \dots, n_{M_k})^T = (n_{k11}, \dots, n_{kJI_k})^T$  ( $n_s > 0, s = 1, \dots, M_k$ ) and

$$\mathbf{X}_k = \begin{pmatrix} \mathbf{1}_{I_k} & & \mathbf{t}_k \\ & \ddots & \vdots \\ & & \mathbf{1}_{I_k} & \mathbf{t}_k \end{pmatrix}_{JI_k \times (J+1)} = (\mathbf{I}_J \otimes \mathbf{1}_{I_k}, \mathbf{1}_J \otimes \mathbf{t}_k),$$

with  $\mathbf{t}_k \equiv (t_{k1}, \dots, t_{kJ})^T$ , a full rank  $M_k \times (J + 1)$  design matrix. Based on the likelihood function of a Poisson sample  $\mathbf{D}_k$  the kernel of the log-likelihood function is given by

$$\ell_{\boldsymbol{\beta}_k}(\mathbf{D}_k) = \sum_{s=1}^{M_k} D_s \log m_s(\boldsymbol{\beta}_k) - \sum_{s=1}^{M_k} m_s(\boldsymbol{\beta}_k),$$

and thus the MLE of  $\beta_k$  is

$$\hat{\beta}_k = \arg \max_{\beta_k \in \Theta_k} \ell_{\beta_k}(\mathbf{D}_k).$$

It is well known that there is a very close relationship between the likelihood theory and the Kullback–Leibler divergence measure [6]. Focused on a multinomial contingency table it is intuitively understandable that a good estimator of the probabilities of the cells should be such that the discrepancy with respect to the empirical distribution or relative frequencies is small enough. The oldest discrepancy or distance measure we know is the Kullback divergence measure, actually the estimator which is built from the Kullback divergence measure is the MLE. By considering the unknown parameters of a Poisson contingency table, the expected values, rather than probabilities and the observed frequencies rather than relative frequencies, we are going to show how is it possible to carry out statistical inference for Poisson models through power divergence measures. According to the Kullback divergence measure, the discrepancy or distance between the Poisson sample  $\mathbf{D}_k$  and its vector of means  $\mathbf{m}_k(\beta_k)$  is given by

$$d_{\text{Kull}}(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = \sum_{s=1}^{M_k} \left( D_s \log \frac{D_s}{m_s(\beta_k)} - D_s + m_s(\beta_k) \right). \tag{12}$$

Observe that  $d_{\text{Kull}}(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = -\ell_{\beta_k}(\mathbf{D}_k) + C_k$ , where  $C_k$  does not depend on parameter  $\beta_k$ . Such a relationship allows us to define the MLE of  $\beta_k$  as minimum Kullback divergence estimator

$$\hat{\beta}_k = \arg \min_{\beta_k \in \Theta_k} d_{\text{Kull}}(\mathbf{D}_k, \mathbf{m}_k(\beta_k)),$$

and the MLE of  $\mathbf{m}_k(\beta_k)$  functionally as  $\mathbf{m}_k(\hat{\beta}_k)$  due to the invariance property of the MLEs. The power divergence measures are a family of measures defined as

$$d_\lambda(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = \frac{1}{\lambda(1+\lambda)} \sum_{s=1}^{M_k} \left( \frac{D_s^{\lambda+1}}{m_s^\lambda(\beta_k)} - D_s(1+\lambda) + \lambda m_s(\beta_k) \right), \quad \lambda \notin \{0, -1\} \tag{13}$$

such that from each possible value for subscript  $\lambda \in \mathbb{R} - \{0, -1\}$  a different way to quantify the discrepancy between  $\mathbf{D}_k$  and  $\mathbf{m}_k(\beta_k)$  arises. In case of  $\lambda \in \{0, -1\}$ , we define  $d_\lambda(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = \lim_{\ell \rightarrow \lambda} d_\ell(\mathbf{D}_k, \mathbf{m}_k(\beta_k))$ , and in this manner the Kullback divergence appears as special case of power divergence measures when  $\lambda = 0$ ,  $d_0(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = d_{\text{Kull}}(\mathbf{D}_k, \mathbf{m}_k(\beta_k))$  and on the other hand case  $\lambda = -1$  is obtained by changing the order of the arguments for the Kullback divergence measure,  $d_{-1}(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = d_{\text{Kull}}(\mathbf{m}_k(\beta_k), \mathbf{D}_k)$ .

The estimator of  $\beta_k$  obtained on the basis of (13) is the so-called minimum power divergence estimator (MPDE) and it is defined for each value of  $\lambda \in \mathbb{R}$  as

$$\hat{\beta}_{k,\lambda} = \arg \min_{\beta_k \in \Theta_k} d_\lambda(\mathbf{D}_k, \mathbf{m}_k(\beta_k)), \tag{14}$$

and the MPDE of  $\mathbf{m}_k(\beta_k)$  functionally as  $\mathbf{m}_k(\hat{\beta}_{k,\lambda})$  due to the invariance property of the MPDEs. Apart from the MLE ( $\hat{\beta}_k$  or  $\hat{\beta}_{k,0}$ ) there are other estimators that are members of this family of estimators: minimum modified chi-square estimator,  $\hat{\beta}_{k,-2}$ ; minimum modified likelihood estimator,  $\hat{\beta}_{k,-1}$ ; Cressie–Read estimator,  $\hat{\beta}_{k,2/3}$ ; minimum chi-square estimator,  $\hat{\beta}_{k,1}$ . These estimators were introduced and analyzed for multinomial sampling by Cressie and Read [12], but for Poisson sampling were applied for the first time in [10]. The so-called minimum  $\phi$ -divergence estimators are a wider class of estimator that contains MPDEs as special case (see [9] and [5] for more details) and this statistical problem could be easily extended for these estimators.

Taking into account that the asymptotic distribution of all MPDEs tend to be “theoretically” the same, including the MLE, we are going to propose an alternative method for estimating  $\text{Var}(\hat{\beta}_{11} - \hat{\beta}_{12}) = \text{Var}(\hat{\beta}_{11,0} - \hat{\beta}_{12,0})$  that covers a new element for overlapping regions,  $\text{Cov}(\hat{\beta}_{11}, \hat{\beta}_{12}) = \text{Cov}(\hat{\beta}_{11,0}, \hat{\beta}_{12,0})$ . We postulate that for not very large data sets, the MLEs,  $\hat{\beta}_{11,0} - \hat{\beta}_{12,0}$ , might be likely improved by the estimation associated with  $\lambda = 1$ ,  $\hat{\beta}_{11,1} - \hat{\beta}_{12,1}$ , when overlapping regions are considered.

In order to obtain the MPDE of (2),  $\widehat{\text{APC}}_{k,\lambda} = 100(\exp(\hat{\beta}_{1k,\lambda}) - 1)$ , we need to compute the estimator of the parameter of interest by following the next result.

**Proposition 2.** *The MPDE of  $\beta_{1k}$ ,  $\hat{\beta}_{1k,\lambda}$ , is the solution of the nonlinear equation*

$$f(\hat{\beta}_{1k,\lambda}) = \sum_{i=1}^{I_k} t_{ki} \mathcal{Y}_{ki} = 0,$$

with

$$\begin{aligned} \gamma_{ki} &= \sum_{j=1}^J m_{kji}(\hat{\beta}_\lambda) (\varphi_{kji} - 1), \\ m_{kji}(\hat{\beta}_\lambda) &= n_{kji} \exp(\hat{\beta}_{0kj,\lambda}) \exp(\hat{\beta}_{1ki,\lambda} t_{ki}) \quad \text{and} \quad \varphi_{kji} = \left( \frac{D_{kji}}{m_{kji}(\hat{\beta}_\lambda)} \right)^{\lambda+1}, \\ \exp(\hat{\beta}_{0kj,\lambda}) &= \left( \sum_{s=1}^{I_k} p_{kjs} \psi_{kjs}^{\lambda+1} \right)^{\frac{1}{\lambda+1}}, \quad j = 1, \dots, J, \\ p_{kjs} &= \frac{n_{kjs} \exp(\hat{\beta}_{1k,\lambda} t_{ks})}{\sum_{h=1}^{I_k} n_{kjh} \exp(\hat{\beta}_{1k,\lambda} t_{kh})} \quad \text{and} \quad \psi_{kjs} = \frac{D_{kjs}}{n_{kjs} \exp(\hat{\beta}_{1k,\lambda} t_{ks})}. \end{aligned}$$

Our aim is to show that  $\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}$  is asymptotically normal and to obtain an explicit expression of the denominator of the Z-test statistic (8) with MPDEs

$$Z_\lambda = \frac{\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})}}, \tag{15}$$

when the random vectors of observed frequencies of both regions,  $\mathbf{D}_1$  and  $\mathbf{D}_2$ , share some components (those belonging to the overlapping subregion). Since (15) is approximately standard normal for  $\min\{N_1, N_2\}$  large enough, we can test  $\mathcal{H}_0 : \text{APC}_1 = \text{APC}_2 (\beta_{11} = \beta_{12})$  vs.  $\mathcal{H}_1 : \text{APC}_1 \neq \text{APC}_2 (\beta_{11} \neq \beta_{12})$ , so that if the value of  $|Z_\lambda|$  is greater than the quantile  $z_{1-\frac{\alpha}{2}}$  (i.e.,  $\Pr(Z_\lambda < z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ ),  $\mathcal{H}_0$  is rejected with significance level  $\alpha$ .

The following result is the key result for estimating the variances and covariance of the estimators of interest,  $\hat{\beta}_{1k,\lambda}$ ,  $k = 1, 2$ . It allows us to establish a linear relationship between the parameter of interest and the observed frequencies under Poisson sampling when the expected total mean  $N_k$  in each region ( $k = 1, 2$ ) is large enough and the way that  $N_k$  increases is given in Assumption 3.

**Assumption 3.**  $m_{kji}^*(\beta_k^0) = m_{kji}(\beta_k^0)/N_k$  remains constant as  $N_k$  increases, that is,  $m_{kji}(\beta_k^0)$  increases at the same rate as  $N_k$ .

**Theorem 4.** The MPDE of  $\beta_{1k}$ ,  $\hat{\beta}_{1k,\lambda}$ ,  $k = 1, 2$ , can be expressed as

$$\hat{\beta}_{1k,\lambda} - \beta_{1k}^0 = \sigma_{1k}^2 \tilde{\mathbf{t}}_k^T(\beta_k^0) \mathbf{X}_k^T (\mathbf{D}_k - \mathbf{m}_k(\beta_k^0)) + o\left(\left\| \frac{\mathbf{D}_k - \mathbf{m}_k(\beta_k^0)}{N_k} \right\|\right),$$

where superscript 0 is denoting the true and unknown value of a parameter, o is denoting a little o function for a stochastic sequence (see Chapter 14 in [2]) and

$$\sigma_{1k}^2 = (\tilde{\mathbf{t}}_k^T(\beta_k^0) \mathbf{X}_k^T \text{Diag}(\mathbf{m}_k(\beta_k^0)) \mathbf{X}_k \tilde{\mathbf{t}}_k(\beta_k^0))^{-1} = \left( \sum_{j=1}^J \sum_{i=1}^{I_k} m_{kji}(\beta_k^0) (t_{ki} - \tilde{t}_{kj}(\beta_k^0))^2 \right)^{-1}, \tag{16}$$

$$\tilde{\mathbf{t}}_k^T(\beta_k^0) = (-\tilde{t}_{k1}(\beta_k^0) \quad \dots \quad -\tilde{t}_{kj}(\beta_k^0) \quad 1),$$

$$\tilde{t}_{kj}(\beta_k^0) = \frac{\sum_{i=1}^{I_k} m_{kji}(\beta_k^0) t_{ki}}{\sum_{i=1}^{I_k} m_{kji}(\beta_k^0)}. \tag{17}$$

**Theorem 5.** The MPDE of  $\beta_{1k}$ ,  $\hat{\beta}_{1k,\lambda}$ ,  $k = 1, 2$ , is asymptotically Normal, unbiased and with variance equal to (16).

Note that Theorem 5 would be more formally enunciated in terms of  $\sqrt{N_k}(\hat{\beta}_{1k,\lambda} - \beta_{1k}^0)$ , because  $\sigma_{1k}^2$  is not constant as  $N_k$  increases. We have avoided that in order to focus directly on the estimator of interest. Due to Assumption 3 and  $\tilde{t}_{kj}(\beta_k^0) = \sum_{i=1}^{I_k} m_{kji}^*(\beta_k^0) t_{ki}$ , what is constant is

$$\text{Var}(\sqrt{N_k} \hat{\beta}_{1k,\lambda}) = N_k \sigma_{1k}^2 = \left( \sum_{j=1}^J \sum_{i=1}^{I_k} m_{kji}^*(\beta_k^0) (t_{ki} - \tilde{t}_{kj}(\beta_k^0))^2 \right)^{-1}.$$

Let  $N$  be the total expected value of the region constructed by joining regions 1 and 2. Note that  $N \leq N_1 + N_2$ , being only equal with non-overlapping regions. In order to establish the way that  $N$  increases with respect to  $N_k$ , we shall consider throughout the next assumption.

**Assumption 6.**  $N_k^* = \frac{N_k}{N}$  ( $k = 1, 2$ ) is constant as  $N$  increases, that is  $N$  increases at the same rate as  $N_k$ .

Note that for overlapping regions,  $N_1^* + N_2^* > 1$  holds and under the hypothesis that  $\beta_{11}^0 = \beta_{12}^0$ , we have a common true parameter vector  $\beta^0 \equiv \beta_k^0$  ( $k = 1, 2$ ). Hence, under the hypothesis that  $\beta_{11}^0 = \beta_{12}^0$ , since  $N_1^* + N_2^* = 1 + \sum_{j=1}^J \sum_{i=1}^{I_1-\bar{i}} m_{2kj}^{(2)}(\beta^0)/N$  is constant, the overlapping death fraction,  $\sum_{j=1}^J \sum_{i=1}^{I_1-\bar{i}} m_{2kj}^{(2)}(\beta^0)/N$ , is also constant as  $N$  increases.

**Theorem 7.** Under the hypothesis that  $\beta_{11}^0 = \beta_{12}^0$ , the MPDE of  $\beta_{11} - \beta_{12}$ ,  $\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}$ , is decomposed as

$$\begin{aligned} \hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda} &= X_1 + X_2 + X_3 + Y, \\ X_1 &= \sigma_{11}^2 \tilde{\mathbf{t}}_1^T(\beta^0) \mathbf{X}_1^T (\mathbf{D}_1^{(1)} - \mathbf{m}_1^{(1)}(\beta^0)), \\ X_2 &= -\sigma_{12}^2 \tilde{\mathbf{t}}_2^T(\beta^0) \mathbf{X}_2^T (\mathbf{D}_2^{(1)} - \mathbf{m}_2^{(1)}(\beta^0)), \\ X_3 &= (\sigma_{11}^2 \tilde{\mathbf{t}}_1^T(\beta^0) \bar{\mathbf{X}}_1^T - \sigma_{12}^2 \tilde{\mathbf{t}}_2^T(\beta^0) \bar{\mathbf{X}}_2^T) (\bar{\mathbf{D}}^{(2)} - \bar{\mathbf{m}}^{(2)}(\beta^0)), \\ Y &= o\left(\left\|\frac{\mathbf{D}_1 - \mathbf{m}_1(\beta^0)}{N_1}\right\|\right) + o\left(\left\|\frac{\mathbf{D}_2 - \mathbf{m}_2(\beta^0)}{N_2}\right\|\right), \end{aligned} \tag{18}$$

where  $\bar{\mathbf{X}}_k$  is an amplified  $J(I_1 + I_2) \times (J + 1)$  matrix of  $\mathbf{X}_k$ ,

$$\begin{aligned} \bar{\mathbf{X}}_k &= \begin{pmatrix} \bar{\mathbf{1}}_k & & \bar{\mathbf{t}}_k \\ & \ddots & \vdots \\ & & \bar{\mathbf{1}}_k & \bar{\mathbf{t}}_k \end{pmatrix}_{J(I_1+I_2) \times (J+1)} = (\mathbf{I}_J \otimes \bar{\mathbf{1}}_k, \mathbf{1}_J \otimes \bar{\mathbf{t}}_k), \\ \bar{\mathbf{1}}_1^T &= (\mathbf{1}_{I_1}^T, \mathbf{0}_{I_1+I_2-I_1}^T) \quad \text{and} \quad \bar{\mathbf{1}}_2^T = (\mathbf{0}_I^T, \mathbf{1}_{I_2}^T), \\ \bar{\mathbf{t}}_1^T &= (\mathbf{t}_{I_1}^T, \mathbf{0}_{I_1+I_2-I_1}^T) \quad \text{and} \quad \bar{\mathbf{t}}_2^T = (\mathbf{0}_I^T, \mathbf{t}_{I_2}^T), \end{aligned}$$

and  $\bar{\mathbf{D}}^{(2)} = (\bar{D}_{111}, \dots, \bar{D}_{1J, I_1+I_2})^T$ ,  $\bar{\mathbf{m}}^{(2)}(\beta^0) = (\bar{m}_{111}^{(2)}(\beta^0), \dots, \bar{m}_{1J, I_1+I_2}^{(2)}(\beta^0))^T$  are the vectors obtained joining  $\mathbf{D}_k^{(2)}$  for  $k = 1, 2$  and  $\mathbf{m}_k^{(2)}(\beta^0)$  for  $k = 1, 2$  respectively, i.e.

$$\begin{aligned} \bar{\mathbf{D}}^{(2)} &= ((D_{111}, \dots, D_{1J\bar{i}}), (\mathbf{D}_2^{(2)})^T)^T, \quad \mathbf{D}_2^{(2)} = (D_{211}, \dots, D_{2J\bar{i}})^T, \\ \bar{\mathbf{m}}^{(2)}(\beta^0) &= ((m_{111}^{(2)}(\beta^0), \dots, m_{1J\bar{i}}^{(2)}(\beta^0)), \mathbf{m}_2^{(2)}(\beta^0))^T, \quad \mathbf{m}_2^{(2)}(\beta^0) = (m_{211}^{(2)}(\beta^0), \dots, m_{2J\bar{i}}^{(2)}(\beta^0))^T. \end{aligned}$$

**Theorem 8.** Under the hypothesis that  $\beta_{11}^0 = \beta_{12}^0$ , the asymptotic distribution of  $\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}$  is central Normal with

$$\text{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}) = \sigma_{11}^2 + \sigma_{12}^2 - 2\sigma_{11}^2\sigma_{12}^2\xi_{12}$$

where  $\sigma_{1k}^2$  is equal to

$$\sigma_{1k}^2 = \left( \sum_{j=1}^J \sum_{i=1}^{I_k} m_{kji}(\beta^0) (t_{ki} - \tilde{t}_{kj}(\beta^0))^2 \right)^{-1} = \left( \sum_{j=1}^J \sum_{i=1}^{I_k} m_{kji}(\beta^0) t_{ki}^2 - \sum_{j=1}^J m_{kj\bullet} \tilde{t}_{kj}^2(\beta^0) \right)^{-1}, \tag{19}$$

with  $m_{kj\bullet} = \sum_{i=1}^{I_k} m_{kji}(\beta^0)$ ,  $\tilde{t}_{kj}(\beta^0)$  is (17) and

$$\begin{aligned} \xi_{12} &= \sum_{j=1}^J \sum_{i=1}^{I_1-\bar{i}} \frac{n_{2ji}^{(2)}}{n_{2ji}} m_{2ji}(\beta^0) (t_{2i} - \tilde{t}_{1j}(\beta^0))(t_{2i} - \tilde{t}_{2j}(\beta^0)) \\ &= \sum_{j=1}^J \sum_{i=1}^{I_1-\bar{i}} \frac{n_{2ji}^{(2)}}{n_{2ji}} m_{2ji}(\beta^0) (t_{2i}^2 + \tilde{t}_{1j}(\beta^0)\tilde{t}_{2j}(\beta^0)) - \sum_{j=1}^J \sum_{i=1}^{I_1-\bar{i}} \frac{n_{2ji}^{(2)}}{n_{2ji}} m_{2ji}(\beta^0) t_{2i} (\tilde{t}_{1j}(\beta^0) + \tilde{t}_{2j}(\beta^0)). \end{aligned} \tag{20}$$

That is, the covariance between  $\hat{\beta}_{11,\lambda}$  and  $\hat{\beta}_{12,\lambda}$  is given by

$$\sigma_{1,12} = \text{Cov}(\hat{\beta}_{11,\lambda}, \hat{\beta}_{12,\lambda}) = \sigma_{11}^2\sigma_{12}^2\xi_{12}, \tag{21}$$

and the correlation by  $\rho_{1,12} = \text{Cor}(\hat{\beta}_{11,\lambda}, \hat{\beta}_{12,\lambda}) = \sigma_{11}\sigma_{12}\xi_{12}$ .



For the expression in the denominator of (15), we need to obtain the MPDEs of  $\sigma_{1k}^2$ ,  $k = 1, 2$  and  $\xi_{12}$ ,  $\hat{\sigma}_{1k,\lambda}^2$ ,  $k = 1, 2$  and  $\hat{\xi}_{12,\lambda}$  respectively. A way to proceed is based on replacing  $\beta^0$  by the most efficient MPDE

$$\hat{\beta}_\lambda^0 \equiv \begin{cases} \hat{\beta}_{1,\lambda}^0, & \text{if } N_1 \geq N_2 \\ \hat{\beta}_{2,\lambda}^0, & \text{if } N_1 < N_2. \end{cases}$$

An important advantage of this new methodology is that the expression of the denominator of (15) is explicit, easy to compute and can be interpreted easily. The term (20) determines the sign of (21). The structure of (20) is similar to the covariance proposed in the model of Li et al. [8] for WLSEs or as well as for the estimators in the model of Li and Tiwari [7]. We can see that if there is no time point shared by the two regions, i.e.  $\bar{I} \geq I_1$ , then  $\hat{\sigma}_{1,12,\lambda} = 0$  and  $\widehat{\text{Var}}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}) = \hat{\sigma}_{11,\lambda}^2 + \hat{\sigma}_{12,\lambda}^2$ ; if there is no space overlap, then it holds  $m_{2ji}^{(2)}(\hat{\beta}_\lambda^0) = 0$  for all  $i$  and  $j$  belonging to the overlapping subregion and hence  $\hat{\sigma}_{1,12,\lambda} = 0$  and  $\widehat{\text{Var}}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}) = \hat{\sigma}_{11,\lambda}^2 + \hat{\sigma}_{12,\lambda}^2$ . On the other hand, when the two regions to be compared share at least one time point and there is space overlap,  $\widehat{\text{Var}}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}) = \hat{\sigma}_{11,\lambda}^2 + \hat{\sigma}_{12,\lambda}^2 - 2\hat{\sigma}_{1,12,\lambda}$  holds, with  $\hat{\sigma}_{1,12,\lambda} \neq 0$ . Moreover, when the period of time not shared by the two regions is large (small), the covariance tends to be negative (positive) because the average values,  $\tilde{t}_{1j}(\hat{\beta}_{1,\lambda}^0)$  and  $\tilde{t}_{2j}(\hat{\beta}_{2,\lambda}^0)$ , are more separated from (closer to) the time points associated with the overlapping subregion. We shall later analyze this behavior through a simulation study, and we shall now investigate how is the structure of  $\xi_{12}$  when the two regions to be compared share the whole period of time.

**Corollary 9.** When  $\bar{I} = 0$  and  $I_1 = I_2$ , under the hypothesis that  $\beta_{11}^0 = \beta_{12}^0$

$$\xi_{12} = \frac{1}{\sigma_{1(2)}^2} + \sum_{j=1}^J \frac{m_{1j\bullet}^{(1)} m_{2j\bullet}^{(1)}}{m_{1j\bullet} m_{2j\bullet}} m_{j\bullet}^{(2)} (\tilde{t}_{1j}^{(1)}(\beta^0) - \tilde{t}_{1j}^{(2)}(\beta^0)) (\tilde{t}_{2j}^{(1)}(\beta^0) - \tilde{t}_{2j}^{(2)}(\beta^0)), \tag{22}$$

with

$$\frac{1}{\hat{\sigma}_{1(2)}^2} = \sum_{j=1}^J \sum_{i=1}^{I_2} m_{2ji}^{(2)}(\beta^0) (t_{2i} - \tilde{t}_{2j}^{(2)}(\beta^0))^2,$$

$$\tilde{t}_{kj}^{(b)}(\beta^0) = \frac{\sum_{i=1}^{I_k} m_{kji}^{(b)}(\beta^0) t_{ki}}{\sum_{i=1}^{I_k} m_{kji}^{(b)}(\beta^0)},$$

$$m_{kj\bullet}^{(b)} = \sum_{i=1}^{I_k} m_{kji}^{(b)}(\beta^0), \quad m_{kj\bullet} = m_{kj\bullet}^{(1)} + m_{kj\bullet}^{(2)},$$

$$m_{j\bullet}^{(2)} = \sum_{i=1}^{I_2} m_{2ji}^{(2)}(\beta^0) = \sum_{i=1}^{I_1} m_{1ji}^{(2)}(\beta^0).$$

$\sigma_{1(2)}^2$  represents the variance of  $\hat{\beta}_{12,\lambda}$  focused on the overlapping subregion. In particular, if region 2 is completely contained in region 1,  $\xi_{12} = 1/\sigma_{1(2)}^2 = 1/\sigma_{12}^2$ ,  $m_{2j\bullet}^{(1)} = 0$  for all  $j = 1, \dots, J$ , and hence

$$\text{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}) = \sigma_{12}^2 - \sigma_{11}^2. \tag{23}$$

#### 4. Simulation studies and analysis of SEER mortality data

When dealing with asymptotic results, it is interesting to analyze the performance of the theoretical results in an empirical framework. Specifically, for Poisson sampling what is important to calibrate is the way that the total expected value of deaths (or incidences)  $N_k$  affects the precision of the results. Other characteristics such as the percentage of overlapping regions “in space” or “in time”, as well as the suitable choice of  $\lambda$  values are also worth to be analyzed. As a preliminary study, before focusing on  $N_k$ , we have considered thyroid cancer mortality (rare cancer) in three regions, Western (W) US population (composed of Arizona, New Mexico and Texas), South Western (SW) US population (composed of Arizona, California and Nevada) and West Coast (WC) US population (composed of California, Oregon and Washington). APC comparison of W vs. SW (Arizona is shared) on one hand and SW vs. WC (California is shared) on the other hand are considered. We have taken different scenarios with different time periods, 1998–2007 for SW in all scenarios and 1986–1995, 1989–1998, 1992–2001, 1995–2004 and (1998–2007) for the other region (W or WC) in each of scenarios  $A'$ ,  $B'$ ,  $C'$ ,  $D'$  and  $E'$  respectively. In Table 1 the percentage of expected deaths in the regions to be compared with respect to

**Table 1**  
Overlapping percentages for W vs. SW and SW vs. WC in five scenarios.

Space \ Time	sc A'	sc B'	sc C'	sc D'	sc E'
W vs. SW	18.96%; 13.03%	12.66%; 9.12%	6.94%; 5.24%	1.66%; 1.32%	0%; 0%
SW vs. WC	81.80%; 78.39%	59.09%; 54.06%	34.75%; 30.30%	8.93%; 7.40%	0%; 0%

the shared part (the percentages of overlapping) are shown, when  $\beta_{11} = \beta_{12} = -0.005$  ( $APC_1 = APC_2 \simeq -0.5$ ) for W vs. SW, and  $\beta_{13} = \beta_{14} = 0.02$  for SW vs. WC ( $APC_3 = APC_4 \simeq 2.02$ ). Observe that in the same scenario but different couple of comparisons, the change in overlapping percentage is due to the space overlapping (the overlapping percentages are greater for SW vs. WC, actually the shared part is a large state, California). In addition, we have chosen some values of  $\lambda$ ,  $\lambda \in \{-0.5, 0, \frac{2}{3}, 1, 1.5\}$ , in order to compare the performance of minimum power divergence estimators. In Table 3 these results are shown for W vs. SW. From scenario B' to E' (i.e. when the overlapping percentage is increasing), the covariance is increasing, starts with negative values at B' (1 time point is shared), decreases at E' (4 time points are shared), later positive values but small are reached at F' (7 time points are shared) and finally at E' (10 time points are shared) ends with positive and high values. It seems that more or less the sign of the covariance changes in the middle of time points considered for each of the regions. In scenario A' the theoretical covariance is zero, actually the two regions do not share observations. By asterisk we have marked the variances and significance levels obtained by simulation which are greater than its corresponding theoretical values, in order to visualize them as the worst cases. From the results it is concluded the minimum power divergence estimators with  $\lambda = 1$ , that is the minimum chi-squared estimators provide empirically efficient estimators and their Z-test statistics have good performance with respect to the theoretical significance level in the sense that tend to be much smaller. We have omitted the results for SW vs. WC because we have seen that the space overlapping by itself do not affect much the covariances of  $\hat{\beta}_{k1,\lambda}$ . That is, there were no remarkable difference among the covariances in case of choosing SW vs. WC rather than W vs. SW, because the sign of the covariances starts at the same scenario and it is just the value of the covariance what marks the difference between both of them. The behavior of minimum power divergence estimators is very similar too. Hence, in the simulation study that follows we are going to focus only on fixed overlapping percentages and one of them is going to be 100% and the focus of interest are going to be the MLEs and the MCSEs.

For studying the precision of the results when  $N_k$  changes, we have considered three proportionality constants  $\kappa \in \{1, \frac{1}{100}, \frac{1}{300}\}$  associated with  $N_k$  in each of the following scenarios for Regions 1 and 2, with  $\beta_{1k} \in \{0.02, 0.005, 0, -0.005\}$  being equal for both ( $k = 1, 2$ ) as it is required for the null hypothesis, i.e.  $APC_1 = APC_2 \simeq 2.02$ ,  $APC_1 = APC_2 \simeq 0.50$ ,  $APC_1 = APC_2 \simeq 0$ ,  $APC_1 = APC_2 \simeq -0.50$ :

- Scenario A: Low level overlapping regions,  $I_1 = 6, I_2 = 11, I_1 - \bar{I} = 3$ .
- Scenario B: Medium level overlapping regions,  $I_1 = 10, I_2 = 11, I_1 - \bar{I} = 7$ .
- Scenario C: High level overlapping regions,  $I_1 = 8, I_2 = 8, I_1 - \bar{I} = 8$ .

The values of  $n_{kji}$  have been obtained from real data sets for female:

- Scenario A: Region 1 = United States (US) during 1993–1998, Region 2 = California (CA) 1996–2006.
- Scenario B: Region 1 = US during 1993–2002, Region 2 = CA during 1996–2006.
- Scenario C: Region 1 = US during 1999–2006, Region 2 = CA during 1999–2006.

From the same data sets we have taken  $\beta_{0kj} = \log(\kappa D_{kj1}/n_{kj1}) - \beta_{k1}t_{k1}$ , focused on the Breast cancer for the first year of the time interval ( $i = 1$ ). All these data were obtained from the SEER database and hence we are taking into account  $J = 19$  age groups. Once the previous parameters have been established we can compute in a theoretical framework the individual variances of estimators  $\hat{\beta}_{k1,\lambda}, \sigma_{1k}^2$ , covariance  $\sigma_{1,12}$  and  $\text{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}) = \sigma_{11}^2 + \sigma_{12}^2 - 2\sigma_{1,12}$ . We can also compute the theoretical value of  $\eta_k \equiv N_k/(J I_k)$ , the average expected value per cell, which is useful to see if the value of  $N_k$  is large enough, these values are in Table 2.

Since both regions share a common space, we have generated firstly its death counts by simulation and thanks to the Poisson distribution's reproductive property under summation, we have generated thereafter the death counts for each region by adding the complementary Poisson observations. In Tables 4, 6 and 8 are summarized the theoretical results as well as those obtained by simulation for the MLEs and in Tables 5, 7 and 9 for the MCSEs. The variances and covariances appear multiplied by  $10^9$  in all the tables. We have added tilde notation for those parameter that have been calculated by simulation with  $R = 22000$  replications:

$$\tilde{\sigma}_{1k,\lambda}^2 = \frac{1}{R} \sum_{r=1}^R (\hat{\beta}_{1k,\lambda}(r) - \tilde{E}[\hat{\beta}_{1k,\lambda}])^2, \quad \tilde{E}[\hat{\beta}_{1k,\lambda}] = \frac{1}{R} \sum_{r=1}^R \hat{\beta}_{1k,\lambda}(r),$$

$$\tilde{\sigma}_{1,12,\lambda} = \frac{1}{R} \sum_{r=1}^R (\hat{\beta}_{11,\lambda}(r) - \tilde{E}[\hat{\beta}_{11,\lambda}])(\hat{\beta}_{12,\lambda}(r) - \tilde{E}[\hat{\beta}_{12,\lambda}]).$$

It is important to remark that such a large quantity of replications have been chosen in order to reach a reliable precision in the simulation study (e.g., it was encountered that  $R = 10000$  was not large enough). The last column is referred to the

**Table 2**  
Average total expected means of deaths per cell.

$\kappa$	$\beta_{1k}$	Scenario A		Scenario B		Scenario C	
		$\eta_1$	$\eta_2$	$\eta_1$	$\eta_2$	$\eta_1$	$\eta_2$
1	0.020	2538.24	331.42	2741.10	331.42	2493.85	265.98
1	0.005	2441.69	292.43	2552.96	292.43	2360.81	251.71
1	0.000	2410.67	280.62	2494.19	280.619	2318.67	247.19
1	-0.0050	2380.23	269.35	2437.28	269.35	2277.59	242.79
$\frac{1}{100}$	0.020	25.38	3.31	27.41	3.31	24.94	2.66
$\frac{1}{100}$	0.005	24.42	2.92	25.53	2.92	23.61	2.52
$\frac{1}{100}$	0.000	24.11	2.81	24.94	2.81	23.19	2.47
$\frac{1}{100}$	-0.0050	23.80	2.69	24.37	2.69	22.77	2.43
$\frac{1}{300}$	0.020	8.46	1.10	9.14	1.10	8.31	0.89
$\frac{1}{300}$	0.005	8.14	0.97	8.51	0.97	7.87	0.84
$\frac{1}{300}$	0.000	8.03	0.93	8.31	0.93	7.73	0.82
$\frac{1}{300}$	-0.0050	7.93	0.90	8.12	0.90	7.59	0.81

**Table 3**  
Minimum power divergence estimators with  $\lambda \in \{-0.5, 0, \frac{2}{3}, 1, 1.5\}$  for scenarios A', B', C', D' and E'.

sc	$\lambda$	$\sigma_{11}^2$	$\tilde{\sigma}_{11,\lambda}^2$	$\sigma_{12}^2$	$\tilde{\sigma}_{12,\lambda}^2$	$\sigma_{1,12}$	$\tilde{\sigma}_{1,12,\lambda}$	$Var(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{\alpha}_\lambda$
A'	-0.5	106106.94	117206.91	88722.57	96004.29	0.00	-190.20	194829.51	*213591.59	*0.056
A'	0	106106.94	106482.52	88722.57	88399.60	0.00	-131.23	194829.51	*195144.57	0.050
A'	$\frac{2}{3}$	106106.94	100968.49	88722.57	84561.80	0.00	-64.51	194829.51	185659.31	0.047
A'	1	106106.94	99842.89	88722.57	83793.94	0.00	-34.77	194829.51	183706.37	0.047
A'	1.5	106106.94	99346.15	88722.57	83510.99	0.00	2.27	194829.51	182852.60	0.049
B'	-0.5	106106.94	117293.27	83850.45	92311.97	-4020.16	-3833.28	197997.72	*217271.80	*0.058
B'	0	106106.94	106707.01	83850.45	85398.66	-4020.16	-3490.04	197997.72	*199085.75	0.051
B'	$\frac{2}{3}$	106106.94	101342.71	83850.45	81753.09	-4020.16	-3229.76	197997.72	189555.32	0.049
B'	1	106106.94	100261.70	83850.45	80985.96	-4020.16	-3142.66	197997.72	187532.98	0.049
B'	1.5	106106.94	99807.30	83850.45	80649.82	-4020.16	-3047.64	197997.72	186552.39	*0.052
C'	-0.5	106106.94	116056.24	79295.40	84620.24	-6035.64	-5099.81	197473.63	*210876.09	*0.055
C'	0	106106.94	105572.08	79295.40	78400.39	-6035.64	-4630.90	197473.63	193234.25	0.048
C'	$\frac{2}{3}$	106106.94	100178.76	79295.40	75138.00	-6035.64	-4302.56	197473.63	183921.87	0.046
C'	1	106106.94	99090.96	79295.40	74470.54	-6035.64	-4199.62	197473.63	181960.74	0.046
C'	1.5	106106.94	98646.02	79295.40	74214.18	-6035.64	-4094.68	197473.63	181049.56	0.049
D'	-0.5	106106.94	115548.66	74971.59	81107.54	2294.32	2271.85	176489.89	*192112.50	*0.057
D'	0	106106.94	104872.37	74971.59	75820.32	2294.32	2148.49	176489.89	176395.71	0.050
D'	$\frac{2}{3}$	106106.94	99400.99	74971.59	72923.02	2294.32	2060.12	176489.89	168203.77	0.048
D'	1	106106.94	98300.76	74971.59	72306.86	2294.32	2034.16	176489.89	166539.31	0.050
D'	1.5	106106.94	97854.16	74971.59	72044.77	2294.32	2011.33	176489.89	165876.27	*0.052
E'	-0.5	106106.94	115740.28	70747.13	75885.14	15621.44	17152.32	145611.20	*157320.78	*0.055
E'	0	106106.94	105114.37	70747.13	71094.62	15621.44	16123.83	145611.20	143961.33	0.048
E'	$\frac{2}{3}$	106106.94	99710.57	70747.13	68383.44	15621.44	15273.05	145611.20	137547.92	0.047
E'	1	106106.94	98636.63	70747.13	67789.30	15621.44	14953.01	145611.20	136519.90	0.047
E'	1.5	106106.94	98219.30	70747.13	67513.89	15621.44	14557.46	145611.20	136618.26	0.049

exact significance level associated with the Z-test obtained by simulation when the nominal significance level is given by  $\alpha = 0.05$ ,

$$\tilde{\alpha}_\lambda = \frac{1}{R} \sum_{r=1}^R I(|Z_\lambda(r)| > z_{0.975}),$$

where I() is an indicator function and  $z_{0.975} \simeq 1.96$  the quantile of order 0.975 for the standard normal distribution.

It can be seen as expected, that in Scenario 3 the covariance is positive in all the cases, while in Scenario 1 the covariance is negative. It is clear that the precision for  $Var(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$  as well as for  $\tilde{\alpha}_\lambda$  gets better as  $\kappa$  increases. While for large data sets ( $\kappa = 1$ ) there is no best choice regarding  $\lambda$ , for small data sets ( $\kappa = 1/300$ ) the choice in favor of  $\lambda = 1$  is clear because estimators  $\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}$  are more efficient, in fact  $Var(\hat{\beta}_{11,1} - \hat{\beta}_{12,1}) < Var(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}) < Var(\hat{\beta}_{11,0} - \hat{\beta}_{12,0})$ , and the exact significance levels or estimated type I error is less than for  $\lambda = 0$  in all the cases ( $\tilde{\alpha}_1 \leq \tilde{\alpha}_0$ ). Since perhaps type II error could be better for MLEs, the power functions for both estimators have been studied. In particular, for  $\kappa = 1/300$  it was observed the same behavior as appears in Fig. 2: in equidistant differences regarding  $\beta = \beta_{11} - \beta_{12}$ , when  $\beta_{11}^0$  is fixed, if error II is better for MLEs when  $\beta > 0$  ( $\beta < 0$ ) then error II is better for MCSEs when  $\beta < 0$  ( $\beta > 0$ ). Hence, in overall

**Table 4**

Scenario A: maximum likelihood estimators ( $\lambda = 0$ ).

$\kappa$	$\beta_{1k}$	$\sigma_{11}^2$	$\tilde{\sigma}_{11,\lambda}^2$	$\sigma_{12}^2$	$\tilde{\sigma}_{12,\lambda}^2$	$\sigma_{1,12}$	$\tilde{\sigma}_{1,12,\lambda}$	$Var(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{\alpha}_\lambda$
1	0.020	1188.50	1196.88	1468.14	1475.38	-152.70	-153.71	2962.03	*2979.67	0.049
1	0.005	1233.49	1245.97	1653.48	1644.16	-166.41	-159.88	3219.78	3209.89	0.050
1	0.000	1248.91	1237.81	1720.50	1707.01	-171.20	-156.44	3311.81	3257.70	0.049
1	-0.005	1264.55	1276.82	1790.33	1801.21	-176.11	-185.35	3407.10	*3448.73	*0.052
$\frac{1}{100}$	0.020	118849.86	120545.66	146813.66	146902.40	-15269.95	-16186.46	296203.41	*299820.97	*0.052
$\frac{1}{100}$	0.005	123349.03	123414.45	165348.33	167479.53	-16640.50	-16374.79	321978.37	*323643.55	*0.052
$\frac{1}{100}$	0.000	124891.15	124618.15	172050.41	173875.81	-17119.82	-17946.96	331181.19	*334387.88	0.050
$\frac{1}{100}$	-0.005	126455.01	125135.69	179033.49	181356.96	-17610.72	-15914.04	340709.94	338320.73	0.047
$\frac{1}{300}$	0.020	356549.59	359581.84	440440.97	451288.03	-45809.84	-53204.15	888610.23	*917278.18	*0.052
$\frac{1}{300}$	0.005	370047.09	373291.90	496045.00	503332.51	-49921.51	-51558.77	965935.10	*979741.96	0.050
$\frac{1}{300}$	0.000	374673.44	375119.77	516151.22	532280.30	-51359.46	-50448.54	993543.56	*1008297.13	*0.051
$\frac{1}{300}$	-0.005	379365.02	380562.71	537100.47	562780.79	-52832.16	-58143.46	1022129.82	*1059630.42	*0.054

**Table 5**

Scenario A: minimum chi-square estimators ( $\lambda = 1$ ).

$\kappa$	$\beta_{1k}$	$\sigma_{11}^2$	$\tilde{\sigma}_{11,\lambda}^2$	$\sigma_{12}^2$	$\tilde{\sigma}_{12,\lambda}^2$	$\sigma_{1,12}$	$\tilde{\sigma}_{1,12,\lambda}$	$Var(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{\alpha}_\lambda$
1	0.020	1188.50	1196.61	1468.14	1474.56	-152.70	-153.88	2962.03	*2978.92	0.049
1	0.005	1233.49	1245.63	1653.48	1642.03	-166.41	-158.86	3219.78	3205.38	0.049
1	0.000	1248.91	1237.43	1720.50	1704.17	-171.20	-156.10	3311.81	3253.80	0.049
1	-0.005	1264.55	1276.42	1790.33	1797.77	-176.11	-185.22	3407.10	*3444.64	*0.051
$\frac{1}{100}$	0.020	118849.86	118678.59	146813.66	131711.15	-15269.95	-14717.95	296203.41	279825.64	*0.051
$\frac{1}{100}$	0.005	123349.03	121351.67	165348.33	148155.30	-16640.50	-14873.11	321978.37	299253.18	0.049
$\frac{1}{100}$	0.000	124891.15	122229.69	172050.41	152728.68	-17119.82	-16259.50	331181.19	307477.36	0.048
$\frac{1}{100}$	-0.005	126455.01	122628.20	179033.49	158913.98	-17610.72	-14215.39	340709.94	309972.96	0.045
$\frac{1}{300}$	0.020	356549.59	342888.09	440440.97	340354.35	-45809.84	-42220.99	888610.23	767684.41	0.050
$\frac{1}{300}$	0.005	370047.09	354377.57	496045.00	369058.29	-49921.51	-41173.07	965935.10	805782.01	0.045
$\frac{1}{300}$	0.000	374673.44	356408.32	516151.22	388239.25	-51359.46	-38265.08	993543.56	821177.73	0.045
$\frac{1}{300}$	-0.005	379365.02	360799.63	537100.47	403732.21	-52832.16	-47176.61	1022129.82	858885.06	0.045

**Table 6**

Scenario B: maximum likelihood estimators ( $\lambda = 0$ ).

$\kappa$	$\beta_{1k}$	$\sigma_{11}^2$	$\tilde{\sigma}_{11,\lambda}^2$	$\sigma_{12}^2$	$\tilde{\sigma}_{12,\lambda}^2$	$\sigma_{1,12}$	$\tilde{\sigma}_{1,12,\lambda}$	$Var(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{\alpha}_\lambda$
1	0.020	234.90	234.40	1468.14	1461.71	12.72	6.74	1677.59	*1682.63	0.050
1	0.005	251.10	252.79	1653.48	1648.47	13.91	10.97	1876.77	*1879.32	*0.052
1	0.000	256.77	255.02	1720.50	1713.16	14.35	7.81	1948.56	*1952.57	0.050
1	-0.005	262.57	261.96	1790.33	1792.15	14.83	17.78	2023.24	2018.56	0.049
$\frac{1}{100}$	0.020	23489.78	23328.11	146813.66	147774.27	1272.17	181.93	167759.10	*170738.52	*0.053
$\frac{1}{100}$	0.005	25109.90	24424.06	165348.33	147273.99	1390.58	1546.90	187677.08	168604.26	0.049
$\frac{1}{100}$	0.000	25676.71	25666.21	172050.41	171995.21	1435.45	822.83	194856.21	*196015.76	*0.052
$\frac{1}{100}$	-0.005	26257.50	26172.50	179033.49	179024.69	1483.35	708.28	202324.30	*203780.65	*0.051
$\frac{1}{300}$	0.020	70469.35	71112.09	440440.97	442433.57	3816.51	2392.77	503277.31	*508760.12	*0.052
$\frac{1}{300}$	0.005	75329.71	74737.59	496045.00	510147.59	4171.74	3181.74	563031.24	*578521.71	*0.053
$\frac{1}{300}$	0.000	77030.13	76849.11	516151.22	521168.35	4306.36	2781.06	584568.62	*592455.34	0.050
$\frac{1}{300}$	-0.005	78772.49	79582.80	537100.47	545463.20	4450.04	5288.71	606972.89	*614468.57	0.050

terms we recommend using MCSE rather than MLEs for small data sets. This is the case of the study illustrated for instance in [13] where there are a lot of cases such that the value of  $\hat{\eta}_k = \sum_{j=1}^J \sum_{i=1}^k d_{kji} / (Jl_k)$  is quite low (moreover, several cases such that  $\hat{\eta}_k < 12/19$  appear without giving any estimation “due to instability of small numbers”).

We have applied our proposed methodology to compare with real data the APC in the age-adjusted mortality rates of WC, WS and W (described at the beginning of this section) for different periods of time, 1969–1983, 1977–1991 and 1990–1999 respectively, with both estimators and for Thyroid cancer (rare cancer). The third one differs from the rest in the sense that it considers a shorter period of time for its study. The rates are expressed per 100000 individuals at risk. In Fig. 3 the fitted models are plotted and from them it seems at first sight that there is a decreasing trend for Thyroid cancer in WC and SW, and null or decreasing trend in W. The specific values for estimates and test-statistics  $Z_\lambda$ , for  $\lambda = 0, 1$ , are summarized in

**Table 7**

Scenario B: minimum chi-square estimators ( $\lambda = 1$ ).

$\kappa$	$\beta_{1k}$	$\sigma_{11}^2$	$\tilde{\sigma}_{11,\lambda}^2$	$\sigma_{12}^2$	$\tilde{\sigma}_{12,\lambda}^2$	$\sigma_{1,12}$	$\tilde{\sigma}_{1,12,\lambda}$	$Var(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\widetilde{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{\alpha}_\lambda$
1	0.020	234.90	234.36	1468.14	1459.11	12.72	6.73	1677.59	*1680.02	0.050
1	0.005	251.10	252.87	1653.48	1646.88	13.91	10.79	1876.77	*1878.16	*0.052
1	0.000	256.77	255.00	1720.50	1710.41	14.35	7.53	1948.56	*1950.35	0.050
1	-0.005	262.57	261.91	1790.33	1790.65	14.83	17.89	2023.24	2016.78	0.049
$\frac{1}{100}$	0.020	23489.78	23039.44	146813.66	132848.97	1272.17	204.75	167759.10	155478.91	*0.056
$\frac{1}{100}$	0.005	25109.90	24424.06	165348.33	147273.99	1390.58	1546.90	187677.08	168604.26	0.049
$\frac{1}{100}$	0.000	25676.71	25227.72	172050.41	152123.09	1435.45	950.07	194856.21	175450.67	0.049
$\frac{1}{100}$	-0.005	26257.50	25713.71	179033.49	157016.58	1483.35	608.74	202324.30	181512.81	0.050
$\frac{1}{300}$	0.020	70469.35	68417.63	440440.97	333558.97	3816.51	2545.19	503277.31	396886.23	*0.055
$\frac{1}{300}$	0.005	75329.71	71630.87	496045.00	375196.57	4171.74	2568.93	563031.24	441689.58	0.049
$\frac{1}{300}$	0.000	77030.13	73435.63	516151.22	380384.11	4306.36	1980.50	584568.62	449858.76	0.046
$\frac{1}{300}$	-0.005	78772.49	75952.38	537100.47	394349.52	4450.04	3665.47	606972.89	462970.97	0.046

**Table 8**

Scenario C: maximum likelihood estimators ( $\lambda = 0$ ).

$\kappa$	$\beta_{1k}$	$\sigma_{11}^2$	$\tilde{\sigma}_{11,\lambda}^2$	$\sigma_{12}^2$	$\tilde{\sigma}_{12,\lambda}^2$	$\sigma_{1,12}$	$\tilde{\sigma}_{1,12,\lambda}$	$Var(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\widetilde{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{\alpha}_\lambda$
1	0.020	505.19	502.38	4753.38	4766.57	505.19	515.35	4248.20	4238.26	0.050
1	0.005	532.12	529.53	5006.55	4962.78	532.12	527.77	4474.43	4436.77	0.049
1	0.000	541.45	543.59	5094.21	5129.48	541.45	549.19	4552.76	*4574.69	0.050
1	-0.005	550.96	550.15	5183.56	5202.96	550.96	563.62	4632.60	4625.87	0.051
$\frac{1}{100}$	0.020	50518.62	50823.72	475338.31	480772.05	50518.62	52893.29	424819.68	*425809.19	0.050
$\frac{1}{100}$	0.005	53212.46	53963.47	500654.98	500398.48	53212.46	53825.39	447442.52	446711.16	0.049
$\frac{1}{100}$	0.000	54145.29	53655.97	509420.86	511073.61	54145.29	55232.68	455275.57	454264.23	0.050
$\frac{1}{100}$	-0.005	55096.19	55610.24	518356.01	521012.61	55096.19	56166.72	463259.82	*464289.42	0.050
$\frac{1}{300}$	0.020	151555.86	149118.50	1426014.92	1461021.71	151555.86	152950.11	1274459.05	*1304240.00	0.051
$\frac{1}{300}$	0.005	159637.38	161042.69	1501964.94	1529631.00	159637.38	160328.08	1342327.55	*1370017.53	0.049
$\frac{1}{300}$	0.000	162435.88	162828.12	1528262.58	1534795.18	162435.88	165625.27	1365826.70	*1366372.76	0.047
$\frac{1}{300}$	-0.005	165288.56	165289.23	1555068.02	1599312.35	165288.56	168363.58	1389779.46	*1427874.42	0.050

**Table 9**

Scenario C: minimum chi-square estimators ( $\lambda = 1$ ).

$\kappa$	$\beta_{1k}$	$\sigma_{11}^2$	$\tilde{\sigma}_{11,\lambda}^2$	$\sigma_{12}^2$	$\tilde{\sigma}_{12,\lambda}^2$	$\sigma_{1,12}$	$\tilde{\sigma}_{1,12,\lambda}$	$Var(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\widetilde{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda})$	$\tilde{\alpha}_\lambda$
1	0.020	505.19	502.28	4753.38	4756.74	505.19	514.11	4248.20	4230.79	*0.051
1	0.005	532.12	529.39	5006.55	4956.27	532.12	527.17	4474.43	4431.32	0.049
1	0.000	541.45	543.50	5094.21	5120.89	541.45	549.07	4552.76	*4566.24	0.050
1	-0.005	550.96	550.04	5183.56	5194.95	550.96	563.79	4632.60	4617.41	*0.051
$\frac{1}{100}$	0.020	50518.62	49937.40	475338.31	417941.93	50518.62	47230.21	424819.68	373418.90	0.050
$\frac{1}{100}$	0.005	53212.46	53092.20	500654.98	434030.70	53212.46	48517.38	447442.52	390088.14	0.048
$\frac{1}{100}$	0.000	54145.29	52785.97	509420.86	441697.10	54145.29	49680.01	455275.57	395123.06	0.046
$\frac{1}{100}$	-0.005	55096.19	54621.08	518356.01	449926.19	55096.19	50651.05	463259.82	403245.16	0.047
$\frac{1}{300}$	0.020	151555.86	141857.76	1426014.92	1037025.40	151555.86	119830.05	1274459.05	939223.06	0.048
$\frac{1}{300}$	0.005	159637.38	153101.35	1501964.94	1075673.38	159637.38	123845.16	1342327.55	981084.41	0.046
$\frac{1}{300}$	0.000	162435.88	154380.49	1528262.58	1074400.76	162435.88	128138.62	1365826.70	972504.01	0.043
$\frac{1}{300}$	-0.005	165288.56	57146.31	1555068.02	1110194.55	165288.56	131114.59	1389779.46	1005111.67	0.044

**Table 10.** Apart from the appropriate test-statistic, we have included naive test-statistics  $\tilde{\lambda}_\lambda$ , for  $\lambda = 0, 1$  that are obtained by applying the methodology for non-overlapping regions. For Thyroid cancer there is no evidence for rejecting the hypothesis of equal APCs for WS and W but it is not clear WC and WS. Looking at the confidence intervals for each region, observe that for WC and WS the test-statistic has more power to discriminate differences than for WS and W, because the variability is less (the period of time considered for W is shorter). The hypothesis of equal APCs is rejected with 0.05 significance level for WC and WS when using the naive test, and cannot be rejected when using the proper test-statistic for overlapping regions (anyway, its  $p$ -value is close to 0.05). When dealing with common cancer types the same value of APC differences on the sample would probably lead to reject the null hypothesis.

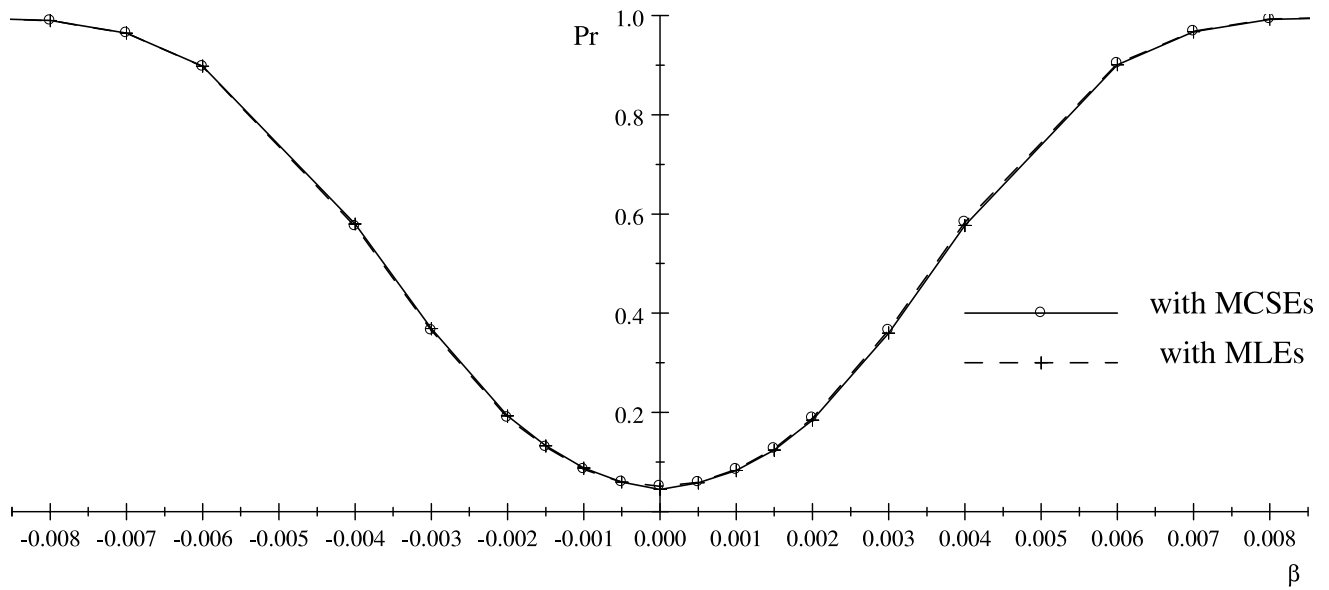


Fig. 2. Power function in terms of  $\beta = \beta_{11} - \beta_{12}$  when  $\beta_{11}^0 = 0$ , for Scenario A and  $\kappa = 1/300$ .

Table 10

Thyroid cancer mortality trends comparison among WC, SW and W during 1969–1983, 1977–1991 and 1990–1999 respectively: maximum likelihood estimators and minimum chi-square estimators.

Region	$k$	$\lambda$	$\hat{\beta}_{1k,\lambda}$	$\hat{\beta}_{0k,\lambda}$	$\hat{\sigma}_{1k,\lambda}^2$	$\hat{\sigma}_{1,k,k+1,\lambda}^2$	$\widehat{APC}_{k,\lambda}$	$CI_{APC_{k,\lambda}}(95\%)$
WC	1	0	-0.0267	-0.3680	$2.923 \times 10^{-5}$	$-77.292 \times 10^{-5}$	-2.639	(-3.665, -1.601)
	1	1	-0.0268	-0.3241	$2.785 \times 10^{-5}$	$-73.429 \times 10^{-5}$	-2.646	(-3.648, -1.635)
SW	2	0	-0.0107	-0.5404	$3.044 \times 10^{-5}$	$-32.915 \times 10^{-5}$	-1.064	(-2.128, 0.011)
	2	1	-0.0106	-0.4943	$2.888 \times 10^{-5}$	$-3.1074 \times 10^{-5}$	-1.053	(-2.089, -0.005)
W	3	0	0.0003	-0.7939	$13.064 \times 10^{-5}$		0.031	(-2.184, 2.297)
	3	1	-0.0012	-0.7084	$12.421 \times 10^{-5}$		-0.117	(-2.275, 2.088)

Z-test statistics for WC vs. SW:  $Z_{12,0} = -1.85, \tilde{Z}_{12,0} = -2.08; Z_{12,1} = -1.92, \tilde{Z}_{12,1} = -2.16$ .

Z-test statistics for SW vs. W:  $Z_{23,0} = -0.85, \tilde{Z}_{23,0} = -0.87; Z_{23,1} = -0.75, \tilde{Z}_{23,1} = -0.76$ .

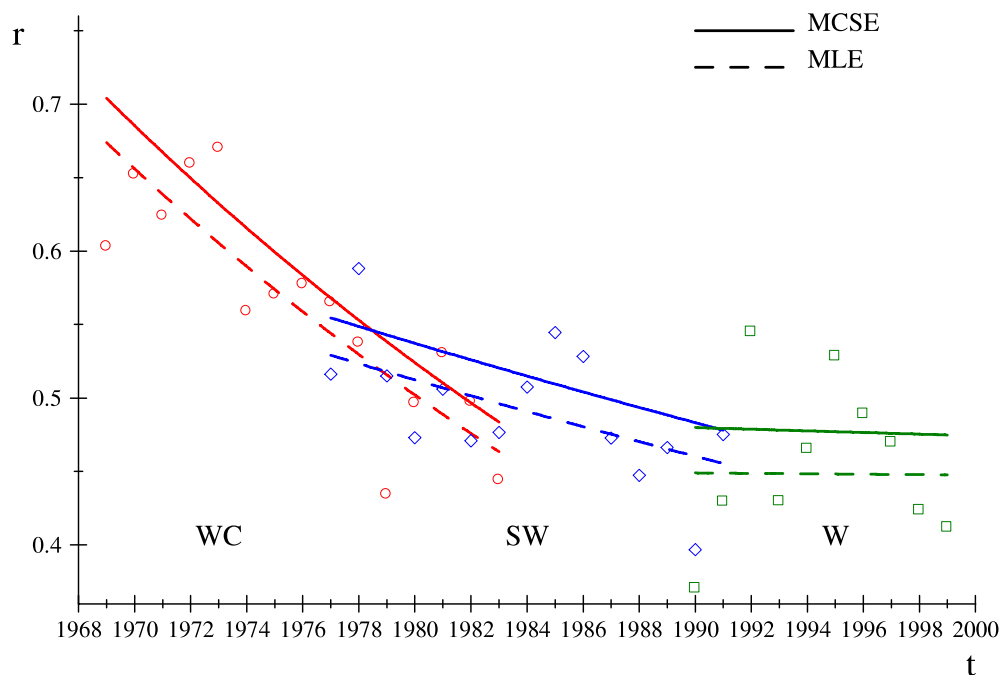


Fig. 3. MCSE and MLE for Thyroid cancer mortality trends in WC, SW and W.

**5. Concluding remarks**

In this work, we have dealt with an important problem of comparing the changing trends of cancer mortality/incidence rates between two overlapping regions. Our new proposal allows us to correctly account for the correlation induced by the overlapping regions when drawing statistical inference. The better finite sample performance of the minimum chi-square estimators, in comparison with the maximum likelihood estimators, suggests the practical utility of the proposed methods especially when comparing the APCs of rare cancers. Not only do our results verify the claim of Berkson [1] that the efficiency of the maximum likelihood estimator is questionable for the finite sample size situations, they also encompass the Poisson models, for which the power divergence based theoretical results (in particular for the minimum chi-square estimators) have remained elusive. In this paper, we have mainly focused on comparing two regions. Extending the methods to accommodate more than two regions simultaneously is certainly worthy of future investigations.

**Acknowledgments**

The authors thank the associate editor and two referees for their valuable comments and suggestions. This work was carried out during the stay of the first author as Visiting Scientist at Harvard University and Dana Farber Cancer Institute, supported by the Real Colegio Complutense and grant MTM2009-10072. Li's work was partially supported by R01CA95747 and 1P01CA134294.

**Appendix**

**Proof of Theorem 4.** Let  $\Delta_{M_k}$  be the set with all possible  $M_k$ -dimensional probability vectors and  $C^{M_k} = (0, 1) \times \dots \times (0, 1)$ . The way in which  $N$  increases is so that  $\text{Diag}^{-1}(\mathbf{n}_k) \mathbf{m}_k(\boldsymbol{\beta}_k)$  does not change, hence  $m_s(\boldsymbol{\beta}_k)$  and  $n_s$  increase at the same time ( $s = 1, \dots, M_k$ ). This means that as  $N_k$  increases, parameter  $\boldsymbol{\beta}_k$  does not suffer any change and neither does the normalized mean vector of deaths,  $\mathbf{m}_k^*(\boldsymbol{\beta}) = \frac{1}{N_k} \mathbf{m}_k(\boldsymbol{\beta}_k)$ . Note that  $\mathbf{m}_k^*(\boldsymbol{\beta}_k) \in \Delta_{M_k} \subset C^{M_k}$ . Let  $V \subset \mathbb{R}^{J+1}$  be a neighborhood of  $\boldsymbol{\beta}_k^0$  and a function

$$\mathbf{F}_{N_k}^{(\lambda)} = (F_1^{(\lambda)}, \dots, F_{J+1}^{(\lambda)}) : C^{M_k} \longrightarrow \mathbb{R}^{J+1},$$

so that

$$F_i^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k) = \frac{\partial d_\lambda(N_k \mathbf{m}_k^*, \mathbf{m}_k(\boldsymbol{\beta}_k))}{\partial \theta_{ki}}, \quad i = 1, \dots, J + 1,$$

with  $\boldsymbol{\beta}_k = (\beta_{0k1}, \dots, \beta_{0kj}, \beta_{1k})^T = (\theta_{k1}, \dots, \theta_{kj}, \theta_{k,J+1})^T \in V$  and  $\mathbf{m}_k^* = (m_1^*, \dots, m_{M_k}^*)^T \in \Delta_{M_k} \subset C^{M_k}$ . More thoroughly, considering  $\mathbf{X}_k = (x_{si})_{s=1, \dots, M_k; i=1, \dots, J+1}$  and  $d_\lambda(\mathbf{D}_k, \mathbf{m}_k(\boldsymbol{\beta}_k)) = \sum_{s=1}^{M_k} m_s(\boldsymbol{\beta}_k) \phi_\lambda(\frac{D_s}{m_s(\boldsymbol{\beta})})$ , where

$$\phi_\lambda(x) = \begin{cases} \frac{x^{\lambda+1} - x - \lambda(x - 1)}{\lambda(\lambda + 1)}, & \lambda(\lambda + 1) \neq 0, \\ \lim_{\alpha \rightarrow \lambda} \phi_\alpha(x), & \lambda(\lambda + 1) = 0, \end{cases}$$

it holds

$$F_i^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k) = \sum_{s=1}^{M_k} m_s(\boldsymbol{\beta}) x_{si} \left( \phi_\lambda \left( \frac{N m_s^*}{m_s(\boldsymbol{\beta}_k)} \right) - \frac{N_k m_s^*}{m_s(\boldsymbol{\beta}_k)} \phi'_\lambda \left( \frac{N m_s^*}{m_s(\boldsymbol{\beta}_k)} \right) \right).$$

It can be seen that replacing  $\mathbf{m}_k^*$  by  $\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)$ ,  $\boldsymbol{\beta}_k$  by  $\boldsymbol{\beta}_k^0$ , it holds  $F_i^{(\lambda)}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0), \boldsymbol{\beta}_k^0) = 0$ , for all  $i = 1, \dots, M_k$ . We shall now establish that Jacobian matrix

$$\frac{\partial \mathbf{F}_{N_k}^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial \boldsymbol{\beta}_k} = \left( \frac{\partial F_i^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial \theta_{kj}} \right)_{i,j=1, \dots, M_k+J+1}$$

is nonsingular when  $(\mathbf{m}_k^*, \boldsymbol{\beta}_k) = (\mathbf{m}_k^*(\boldsymbol{\beta}_k^0), \boldsymbol{\beta}_k^0)$ . For  $i, j = 1, \dots, J + 1$

$$\begin{aligned} \frac{\partial F_i^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial \theta_{kj}} &= \frac{\partial}{\partial \theta_j} \frac{\partial d_\lambda(N \mathbf{m}_k^*, \mathbf{m}_k(\boldsymbol{\beta}_k))}{\partial \theta_{ki}} \\ &= \frac{\partial}{\partial \theta_{kj}} \left( \sum_{s=1}^{M_k} m_s(\boldsymbol{\beta}_k) x_{si} \left( \phi_\lambda \left( \frac{N_k m_s^*}{m_s(\boldsymbol{\beta}_k)} \right) - \frac{N_k m_s^*}{m_s(\boldsymbol{\beta}_k)} \phi'_\lambda \left( \frac{N_k m_s^*}{m_s(\boldsymbol{\beta}_k)} \right) \right) \right) \\ &= \sum_{s=1}^{M_k} m_s(\boldsymbol{\beta}_k) x_{si} x_{sj} \left( \phi_\lambda \left( \frac{N_k m_s^*}{m_s(\boldsymbol{\beta})} \right) - \frac{N_k m_s^*}{m_s(\boldsymbol{\beta})} \phi'_\lambda \left( \frac{N_k m_s^*}{m_s(\boldsymbol{\beta})} \right) \right) + \sum_{s=1}^{M_k} N_k m_s^* x_{si} x_{sj} \frac{N_k m_s^*}{m_s(\boldsymbol{\beta})} \phi''_\lambda \left( \frac{N_k m_s^*}{m_s(\boldsymbol{\beta})} \right), \end{aligned}$$

and because  $\phi_\lambda(1) = \phi'_\lambda(1) = 0$ , and  $\phi''_\lambda(1) = 1$  for all  $\lambda$ ,

$$\left. \frac{\partial F_i^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial \theta_{kj}} \right|_{(\mathbf{m}_k^*, \boldsymbol{\beta}_k) = (\mathbf{m}_k^*(\boldsymbol{\beta}_k^0), \boldsymbol{\beta}_k^0)} = N_k \sum_{s=1}^{M_k} m_s^*(\boldsymbol{\beta}_k^0) x_{si} x_{sj}.$$

Hence,

$$\left( \frac{\partial \mathbf{F}_{N_k}^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial \boldsymbol{\beta}_k} \right)^{-1} \Big|_{(\mathbf{m}_k^*, \boldsymbol{\beta}_k) = (\mathbf{m}_k^*(\boldsymbol{\beta}_k^0), \boldsymbol{\beta}_k^0)} = N_k \mathbf{X}_k^T \text{Diag}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)) \mathbf{X}_k.$$

Applying the Implicit Function Theorem there exist:

- a neighborhood  $U_k$  of  $(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0), \boldsymbol{\beta}_k^0)$  in  $C^{M_k} \times \mathbb{R}^{J+1}$  such that  $\partial \mathbf{F}^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k) / \partial \boldsymbol{\beta}_k$  is nonsingular for every  $(\mathbf{m}_k^*, \boldsymbol{\beta}_k) \in U_k$ ;
- an open set  $A_k \subset C^{M_k}$  that contains  $\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)$ ;
- and a unique, continuously differentiable function  $\tilde{\boldsymbol{\beta}}_k^{(\lambda)} : A_k \rightarrow \mathbb{R}^{J+1}$  such that  $\tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)) = \boldsymbol{\beta}_k^0$  and

$$\{(\mathbf{m}_k^*, \boldsymbol{\beta}_k) \in U_k : \mathbf{F}_{N_k}^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k) = \mathbf{0}\} = \{(\mathbf{m}_k^*, \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*)) : \mathbf{m}_k^* \in A_k\}.$$

Since

$$\min_{\mathbf{m}_k^* \in A_k} d_\lambda(\mathbf{m}_k(\boldsymbol{\beta}_k^0), \mathbf{m}_k(\tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*))) = \min_{\boldsymbol{\beta}_k \in \Theta_k} d_\lambda(\mathbf{m}_k(\boldsymbol{\beta}_k^0), \mathbf{m}_k(\boldsymbol{\beta}_k)),$$

it holds

$$\tilde{\boldsymbol{\beta}}_k^{(\lambda)} \left( \arg \min_{\mathbf{m}_k^* \in A_k} d_\lambda(\mathbf{m}_k(\boldsymbol{\beta}_k^0), \mathbf{m}_k(\tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*))) \right) = \arg \min_{\boldsymbol{\beta}_k \in \Theta_k} d_\lambda(\mathbf{m}_k(\boldsymbol{\beta}_k^0), \mathbf{m}_k(\boldsymbol{\beta}_k)),$$

that is

$$\tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)) = \arg \min_{\boldsymbol{\beta}_k \in \Theta_k} d_\lambda(N_k \mathbf{m}_k^*(\boldsymbol{\beta}_k^0), \mathbf{m}_k(\boldsymbol{\beta}_k)). \tag{24}$$

Furthermore, from the properties of power divergence measures and because  $\tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)) = \boldsymbol{\beta}_k^0$ , we have

$$0 = d_\lambda(\mathbf{m}_k(\boldsymbol{\beta}_k^0), \mathbf{m}_k(\tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)))) < d_\lambda(\mathbf{m}_k(\boldsymbol{\beta}_k^0), \mathbf{m}_k(\boldsymbol{\beta}_k)), \quad \forall \mathbf{m}_k(\boldsymbol{\beta}_k) \neq \mathbf{m}_k(\boldsymbol{\beta}_k^0).$$

By applying the chain rule for obtaining derivatives on  $\mathbf{F}_k^{(\lambda)}(\mathbf{m}_k^*, \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0))) = \mathbf{0}$  with respect to  $\mathbf{m}_k^* \in A_k$ , we have

$$\left. \frac{\partial \mathbf{F}_{N_k}^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial \mathbf{m}_k^*} \right|_{\boldsymbol{\beta}_k = \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*)} + \left. \frac{\partial \mathbf{F}_N^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial \boldsymbol{\beta}_k} \right|_{\boldsymbol{\beta}_k = \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*)} \frac{\partial \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*)}{\partial \mathbf{m}_k^*} = \mathbf{0},$$

so that for  $\mathbf{m}_k^* = \mathbf{m}_k^*(\boldsymbol{\beta}_k^0)$

$$\left. \frac{\partial \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*)}{\partial \mathbf{m}_k^*} \right|_{\mathbf{m}_k^* = \mathbf{m}_k^*(\boldsymbol{\beta}_k^0)} = - \left( \frac{\partial \mathbf{F}_N^{(\lambda)}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0), \boldsymbol{\theta})}{\partial \boldsymbol{\beta}_k} \right)^{-1} \left. \frac{\partial \mathbf{F}^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial \mathbf{m}_k^*} \right|_{(\mathbf{m}_k^*, \boldsymbol{\beta}_k) = (\mathbf{m}_k^*(\boldsymbol{\beta}_k^0), \boldsymbol{\beta}_k^0)}.$$

The last expression is part of the Taylor expansion of  $\tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*)$  around  $\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)$

$$\tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*) = \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)) + \left. \frac{\partial \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*)}{\partial \mathbf{m}_k^*} \right|_{\mathbf{m}_k^* = \mathbf{m}_k^*(\boldsymbol{\beta}_k^0)} (\mathbf{m}_k^* - \mathbf{m}_k^*(\boldsymbol{\beta}_k^0)) + o(\|\mathbf{m}_k^* - \mathbf{m}_k^*(\boldsymbol{\beta}_k^0)\|).$$

Taking derivatives on  $F_i^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)$  with respect to  $m_j^*$

$$\begin{aligned} \frac{\partial F_i^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial m_j^*} &= \frac{\partial}{\partial m_j^*} \frac{\partial d_\lambda(N_k \mathbf{m}_k^*, \mathbf{m}_k(\boldsymbol{\beta}_k))}{\partial \theta_{ki}} \\ &= \frac{\partial}{\partial m_j^*} \sum_{s=1}^{M_k} m_s(\boldsymbol{\beta}_k) x_{si} \left( \phi_\lambda \left( \frac{N m_s^*}{m_s(\boldsymbol{\beta}_k)} \right) - \frac{N m_s^*}{m_s(\boldsymbol{\beta}_k)} \phi'_\lambda \left( \frac{N m_s^*}{m_s(\boldsymbol{\beta}_k)} \right) \right), \\ &= -N_k \frac{N_k m_j^*}{m_j(\boldsymbol{\beta}_k)} x_{ji} \phi'' \left( \frac{N_k m_j^*}{m_j(\boldsymbol{\beta}_k)} \right), \end{aligned}$$



that is

$$\left. \frac{\partial F_i^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial m_j^*} \right|_{(\boldsymbol{\beta}_k, \mathbf{m}_k^*) = (\boldsymbol{\beta}_k^0, \mathbf{m}_k^*(\boldsymbol{\beta}_k^0))} = -N_k x_{ji},$$

and hence

$$\left. \frac{\partial \mathbf{F}_{N_k}^{(\lambda)}(\mathbf{m}_k^*, \boldsymbol{\beta}_k)}{\partial \mathbf{m}_k^*} \right|_{(\mathbf{m}_k^*, \boldsymbol{\beta}_k) = (\mathbf{m}_k^*(\boldsymbol{\beta}_k^0), \boldsymbol{\beta}_k^0)} = \left( \left. \frac{\partial F_i^{(\lambda)}(\mathbf{m}^*, \boldsymbol{\theta})}{\partial m_j^*} \right|_{(\mathbf{m}^*, \boldsymbol{\theta}) = (\mathbf{m}^*(\boldsymbol{\beta}_0), \boldsymbol{\theta}_0)} \right)_{i=1, \dots, B; j=1, \dots, M_k} = -N_k \mathbf{X}_k^T,$$

and

$$\tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*) = \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)) + (\mathbf{X}_k^T \text{Diag}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)) \mathbf{X}_k)^{-1} \mathbf{X}_k^T (\mathbf{m}_k^* - \mathbf{m}_k^*(\boldsymbol{\beta}_k^0)) + o(\|\mathbf{m}_k^* - \mathbf{m}_k^*(\boldsymbol{\beta}_k^0)\|). \tag{25}$$

It is well known that for Poisson sampling  $\frac{D_k}{N_k}$  converges almost surely (a.s.) to  $\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)$  as  $N_k$  increases, which means that  $\frac{D_k}{N_k} \in A_k$  a.s. for  $N_k$  large enough and thus according to the Implicit Function Theorem  $(\frac{D_k}{N_k}, \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\frac{D_k}{N_k})) \in U$  a.s. for  $N_k$  large enough. We can conclude from (24)

$$\tilde{\boldsymbol{\beta}}_k^{(\lambda)}\left(\frac{D_k}{N_k}\right) = \arg \min_{\boldsymbol{\beta}_k \in \Theta_k} d_\lambda\left(N_k \frac{D_k}{N_k}, \mathbf{m}_k(\boldsymbol{\beta}_k)\right) = \arg \min_{\boldsymbol{\beta}_k \in \Theta_k} d_\lambda(D_k, \mathbf{m}_k(\boldsymbol{\beta}_k)),$$

which means that  $\widehat{\boldsymbol{\beta}}_{k,\lambda} = \tilde{\boldsymbol{\beta}}_k^{(\lambda)}(\frac{D_k}{N_k})$ , and hence from (25)

$$\widehat{\boldsymbol{\beta}}_{k,\lambda} - \boldsymbol{\beta}_k^0 = (\mathbf{X}_k^T \text{Diag}(\mathbf{m}_k(\boldsymbol{\beta}_k^0)) \mathbf{X}_k)^{-1} \mathbf{X}_k^T (D_k - \mathbf{m}_k(\boldsymbol{\beta}_k^0)) + o\left(\left\| \frac{D_k - \mathbf{m}_k(\boldsymbol{\beta}_k^0)}{N_k} \right\|\right).$$

Taking into account that  $\widehat{\boldsymbol{\beta}}_{1k,\lambda} - \boldsymbol{\beta}_{1k}^0 = \mathbf{e}_{j+1}^T (\widehat{\boldsymbol{\beta}}_{k,\lambda} - \boldsymbol{\beta}_k^0)$ , where  $\mathbf{e}_{j+1}^T = (0, \dots, 0, 1)$ , we are going to show that  $\mathbf{e}_{j+1}^T (\mathbf{X}_k^T \text{Diag}(\mathbf{m}_k(\boldsymbol{\beta}_k^0)) \mathbf{X}_k)^{-1} = \sigma_{1k}^2 \tilde{\mathbf{t}}_k^T(\boldsymbol{\beta}_k^0)$ . For that purpose we consider the design matrix partitioned according to  $\mathbf{X}_k = (\mathbf{U}, \mathbf{v})$ , where  $\mathbf{U} = \mathbf{I}_j \otimes \mathbf{1}_{l_k}$ ,  $\mathbf{v} = \mathbf{1}_j \otimes \mathbf{t}_k$ , so that for

$$\begin{aligned} (\mathbf{X}_k^T \text{Diag}(\mathbf{m}_k(\boldsymbol{\beta}_k^0)) \mathbf{X}_k)^{-1} &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}, \\ \begin{cases} \mathbf{A}_{11} = \mathbf{U}^T \text{Diag}(\mathbf{m}_k(\boldsymbol{\beta}_k^0)) \mathbf{U} = \text{Diag}(\{N_{kj}^j\}_{j=1}^J), \\ \mathbf{A}_{12} = \mathbf{U}^T \text{Diag}(\mathbf{m}_k(\boldsymbol{\beta}_k^0)) \mathbf{v} = \left( \sum_{i=1}^{l_k} m_{k1i}(\boldsymbol{\beta}_k^0) t_{ki}, \dots, \sum_{i=1}^{l_k} m_{kj_i}(\boldsymbol{\beta}_k^0) t_{ki} \right)^T = \mathbf{A}_{21}^T, \\ \mathbf{A}_{22} = \mathbf{v}^T \text{Diag}(\mathbf{m}_k(\boldsymbol{\beta}_k^0)) \mathbf{v} = \sum_{j=1}^J \sum_{i=1}^{l_k} m_{kj_i}(\boldsymbol{\beta}_k^0) t_{ki}^2, \end{cases} \end{aligned}$$

we can use formula

$$\begin{cases} \mathbf{B}_{11} = \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{B}_{22} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \\ \mathbf{B}_{21} = \mathbf{B}_{12}^T = -\mathbf{B}_{22} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \\ \mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12})^{-1}. \end{cases} \tag{26}$$

It follows that

$$\begin{aligned} \mathbf{e}_{j+1}^T (\mathbf{X}_k^T \text{Diag}(\mathbf{m}_k(\boldsymbol{\beta}_k^0)) \mathbf{X}_k)^{-1} &= (\mathbf{B}_{21} \quad \mathbf{B}_{22}) = (-\mathbf{B}_{22} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \quad \mathbf{B}_{22}) \\ &= \mathbf{B}_{22} (-\mathbf{A}_{21} \mathbf{A}_{11}^{-1} \quad \mathbf{1}), \end{aligned}$$

where

$$\begin{aligned} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} &= \left( \sum_{i=1}^{l_k} m_{k1i}(\boldsymbol{\beta}_k^0) t_{ki}, \dots, \sum_{i=1}^{l_k} m_{kj_i}(\boldsymbol{\beta}_k^0) t_{ki} \right) \text{Diag}(\{N_{kj}^{-1}\}_{j=1}^J) \\ &= \left( N_{k1}^{-1} \sum_{i=1}^{l_k} m_{k1i}(\boldsymbol{\beta}_k^0) t_{ki}, \dots, N_{kj}^{-1} \sum_{i=1}^{l_k} m_{kj_i}(\boldsymbol{\beta}_k^0) t_{ki} \right) = (\tilde{t}_{k1}(\boldsymbol{\beta}_k^0), \dots, \tilde{t}_{kj}(\boldsymbol{\beta}_k^0)) \end{aligned}$$

and

$$\begin{aligned} \mathbf{B}_{22} &= \left( \sum_{j=1}^J \sum_{i=1}^{I_k} m_{kji}(\boldsymbol{\beta}_k^0) t_{ki}^2 - \sum_{j=1}^J \left( \sum_{i=1}^{I_k} m_{kji}(\boldsymbol{\beta}_k^0) \right) \tilde{t}_{kj}^2(\boldsymbol{\beta}_k^0) \right)^{-1} \\ &= \left( \sum_{j=1}^J \sum_{i=1}^{I_k} m_{kji}(\boldsymbol{\beta}_k^0) t_{ki}^2 - \sum_{j=1}^J \left( \sum_{i=1}^{I_k} m_{kji}(\boldsymbol{\beta}_k^0) \right) \tilde{t}_{kj}^2(\boldsymbol{\beta}_k^0) \pm \sum_{j=1}^J \left( \sum_{i=1}^{I_k} m_{kji}(\boldsymbol{\beta}_k^0) t_{kj} \right) \tilde{t}_{kj}(\boldsymbol{\beta}_k^0) \right)^{-1} \\ &= \left( \sum_{j=1}^J \sum_{i=1}^{I_k} m_{kji}(\boldsymbol{\beta}_k^0) (t_{ki} - \tilde{t}_{kj}(\boldsymbol{\beta}_k^0))^2 \right)^{-1}, \end{aligned}$$

because  $\sum_{j=1}^J \left( \sum_{i=1}^{I_k} m_{kji}(\boldsymbol{\beta}_k^0) \right) \tilde{t}_{kj}^2(\boldsymbol{\beta}_k^0) = \sum_{j=1}^J \left( \sum_{i=1}^{I_k} m_{kji}(\boldsymbol{\beta}_k^0) t_{kj} \right) \tilde{t}_{kj}(\boldsymbol{\beta}_k^0)$ .

**Proof of Theorem 5.** Reformulating Theorem 4 we obtain

$$\sqrt{N_k} (\hat{\beta}_{1k,\lambda} - \beta_{1k}^0) = \mathbf{a}_k^T \sqrt{N_k} (\mathbf{D}_k - \mathbf{m}_k(\boldsymbol{\beta}_k^0)) + o \left( \left\| \sqrt{N_k} \left( \frac{\mathbf{D}_k}{N_k} - \mathbf{m}_k^*(\boldsymbol{\beta}_k^0) \right) \right\| \right),$$

with  $\mathbf{a}_k^T \equiv \sigma_{1k}^2 \tilde{\mathbf{t}}_k^T(\boldsymbol{\beta}_k^0) \mathbf{X}_k^T$ . We would like to calculate the asymptotic distribution as a linear function of

$$\sqrt{N_k} \left( \frac{\mathbf{D}_k}{N_k} - \mathbf{m}_k^*(\boldsymbol{\beta}_k^0) \right) \xrightarrow[N_k \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}, \text{Diag}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0))).$$

Since

$$\begin{aligned} \text{Var} \left( \mathbf{a}_k^T \sqrt{N_k} (\mathbf{D}_k - \mathbf{m}_k(\boldsymbol{\beta}_k^0)) \right) &= \mathbf{a}_k^T \text{Var} \left( N_k \sqrt{N_k} \left( \frac{\mathbf{D}_k}{N_k} - \mathbf{m}_k^*(\boldsymbol{\beta}_k^0) \right) \right) \mathbf{a}_k \\ &= N_k^2 \mathbf{a}_k^T \text{Diag}(\mathbf{m}_k^*(\boldsymbol{\beta}_k^0)) \mathbf{a}_k = N_k \sigma_{1k}^2, \end{aligned}$$

it holds

$$\mathbf{a}_k^T \sqrt{N_k} (\mathbf{D}_k - \mathbf{m}_k(\boldsymbol{\beta}_k^0)) \xrightarrow[N_k \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, N_k \sigma_{1k}^2). \tag{27}$$

Taking into account that  $o \left( \left\| \sqrt{N_k} \left( \frac{\mathbf{D}_k}{N_k} - \mathbf{m}_k^*(\boldsymbol{\beta}_k^0) \right) \right\| \right) = o(\text{Op}(1)) = o_p(1)$ , according to the Slutsky's Theorem, the asymptotic distribution of  $\sqrt{N_k} (\hat{\beta}_{1k,\lambda} - \beta_{1k}^0)$  must coincide with the asymptotic distribution of (27).

**Proof of Theorem 7.** From Theorem 4 subtracting  $\hat{\beta}_{12,\lambda} - \beta_{12}^0$  to  $\hat{\beta}_{11,\lambda} - \beta_{11}^0$  we get

$$\begin{aligned} (\hat{\beta}_{11,\lambda} - \beta_{11}^0) - (\hat{\beta}_{12,\lambda} - \beta_{12}^0) &= \sigma_{11}^2 \tilde{\mathbf{t}}_1^T(\boldsymbol{\beta}_1^0) \mathbf{X}_1^T \left( (\mathbf{D}_1^{(1)} - \mathbf{m}_1^{(1)}(\boldsymbol{\beta}_1^0)) - (\mathbf{D}_1^{(2)} - \mathbf{m}_1^{(2)}(\boldsymbol{\beta}_1^0)) \right) \\ &\quad - \sigma_{12}^2 \tilde{\mathbf{t}}_2^T(\boldsymbol{\beta}_2^0) \mathbf{X}_2^T \left( (\mathbf{D}_2^{(1)} - \mathbf{m}_2^{(1)}(\boldsymbol{\beta}_1^0)) - (\mathbf{D}_2^{(2)} - \mathbf{m}_2^{(2)}(\boldsymbol{\beta}_1^0)) \right) + o \left( \left\| \frac{\mathbf{D}_1 - \mathbf{m}_1(\boldsymbol{\beta}_1^0)}{N_1} \right\| \right) - o \left( \left\| \frac{\mathbf{D}_2 - \mathbf{m}_2(\boldsymbol{\beta}_2^0)}{N_2} \right\| \right). \end{aligned}$$

Observe that  $\mathbf{X}_k^T \mathbf{D}_k^{(2)} = \bar{\mathbf{X}}_k^T \bar{\mathbf{D}}^{(2)}$ ,  $k = 1, 2$ , and under  $\beta_{11}^0 = \beta_{12}^0$  it holds  $\mathbf{X}_k^T \mathbf{m}_k^{(2)}(\boldsymbol{\beta}_k^0) = \bar{\mathbf{X}}_k^T \bar{\mathbf{m}}^{(2)}(\boldsymbol{\beta}^0)$ ,  $k = 1, 2$ . In addition,  $o(\cdot)$  function is not affected by the negative sign and under  $\beta_{11}^0 = \beta_{12}^0$  it holds  $\boldsymbol{\beta}_1^0 = \boldsymbol{\beta}_2^0$  and thus we obtain (18).

**Proof of Theorem 8.** We can consider the following decomposition

$$\sqrt{N} (\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}) = (N \mathbf{a}_1^T) \sqrt{N} \frac{\mathbf{D}_1 - \mathbf{m}_1(\boldsymbol{\beta}^0)}{N} + (N \mathbf{a}_2^T) \sqrt{N} \frac{\mathbf{D}_2 - \mathbf{m}_2(\boldsymbol{\beta}^0)}{N} + \sqrt{N} Y, \tag{28}$$

with

$$\sqrt{N}Y = o\left(\frac{1}{N_1^*} \left\| \frac{\mathbf{D}_1 - \mathbf{m}_1(\boldsymbol{\beta}^0)}{\sqrt{N}} \right\|\right) + o\left(\frac{1}{N_2^*} \left\| \frac{\mathbf{D}_2 - \mathbf{m}_2(\boldsymbol{\beta}^0)}{\sqrt{N}} \right\|\right),$$

rather than (18). Note that from Assumptions 3 and 6  $\mathbf{m}_k(\boldsymbol{\beta}^0)/N = N_k^* \mathbf{m}^*(\boldsymbol{\beta}^0)$  is constant as  $N$  increases and hence  $\sqrt{N}Y = o\left(\left\| \frac{\mathbf{D}_1 - \mathbf{m}_1(\boldsymbol{\beta}^0)}{\sqrt{N}} \right\|\right) + o\left(\left\| \frac{\mathbf{D}_2 - \mathbf{m}_2(\boldsymbol{\beta}^0)}{\sqrt{N}} \right\|\right) = o(O_p(1)) + o(O_p(1)) = o_p(1)$ . We would like to calculate the asymptotic distribution as a linear function of

$$\sqrt{N} \frac{\mathbf{D}_k - \mathbf{m}_k(\boldsymbol{\beta}^0)}{N} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}, \text{Diag}(N_k^* \mathbf{m}^*(\boldsymbol{\beta}^0))).$$

From (28) and by applying Slutsky's theorem we can conclude that the asymptotic distribution of  $\sqrt{N}(\widehat{\beta}_{11,\lambda} - \widehat{\beta}_{12,\lambda})$  is central Normal. In order to calculate the variance we shall follow (18) so that

$$\sqrt{N}(\widehat{\beta}_{11,\lambda} - \widehat{\beta}_{12,\lambda}) = \sqrt{N}X_1 + \sqrt{N}X_2 + \sqrt{N}X_3 + \sqrt{N}Y,$$

with

$$\begin{aligned} \sqrt{N}X_1 &= \mathbf{a}_1^T \sqrt{N}(\mathbf{D}_1^{(1)} - \mathbf{m}_1^{(1)}(\boldsymbol{\beta}^0)), \\ \sqrt{N}X_2 &= \mathbf{a}_2^T \sqrt{N}(\mathbf{D}_2^{(1)} - \mathbf{m}_2^{(1)}(\boldsymbol{\beta}^0)), \\ \sqrt{N}X_3 &= (\bar{\mathbf{a}}_1^T - \bar{\mathbf{a}}_2^T) \sqrt{N}(\bar{\mathbf{D}}^{(2)} - \bar{\mathbf{m}}^{(2)}(\boldsymbol{\beta}^0)), \\ \sqrt{N}Y &= o_p(1) \end{aligned}$$

where  $\bar{\mathbf{a}}_k^T \equiv \sigma_{1k}^2 \tilde{\mathbf{t}}_k^T(\boldsymbol{\beta}^0) \bar{\mathbf{X}}_k^T$ , and  $X_1, X_2$  and  $X_3$  are independent random variables. Since

$$\begin{aligned} \text{Var}(\sqrt{N}X_k) &= \text{Var}(\mathbf{a}_k^T \sqrt{N}(\mathbf{D}_k^{(1)} - \mathbf{m}_k^{(1)}(\boldsymbol{\beta}^0))) \\ &= N \mathbf{a}_k^T \text{Diag}(\mathbf{m}_k^{(1)}(\boldsymbol{\beta}^0)) \mathbf{a}_k, \quad k = 1, 2, \\ \text{Var}(\sqrt{N}X_3) &= \text{Var}((\bar{\mathbf{a}}_1^T - \bar{\mathbf{a}}_2^T) \sqrt{N}(\bar{\mathbf{D}}^{(2)} - \bar{\mathbf{m}}^{(2)}(\boldsymbol{\beta}^0))) \\ &= N(\bar{\mathbf{a}}_1^T - \bar{\mathbf{a}}_2^T) \text{Diag}(\bar{\mathbf{m}}^{(2)}(\boldsymbol{\beta}^0))(\bar{\mathbf{a}}_1 - \bar{\mathbf{a}}_2) \\ &= N(\bar{\mathbf{a}}_1^T \text{Diag}(\bar{\mathbf{m}}^{(2)}(\boldsymbol{\beta}^0)) \bar{\mathbf{a}}_1 + \bar{\mathbf{a}}_2^T \text{Diag}(\bar{\mathbf{m}}^{(2)}(\boldsymbol{\beta}^0)) \bar{\mathbf{a}}_2 - 2\bar{\mathbf{a}}_1^T \text{Diag}(\bar{\mathbf{m}}^{(2)}(\boldsymbol{\beta}^0)) \bar{\mathbf{a}}_2) \\ &= N(\mathbf{a}_1^T \text{Diag}(\mathbf{m}_1^{(2)}(\boldsymbol{\beta}^0)) \mathbf{a}_1 + \mathbf{a}_2^T \text{Diag}(\mathbf{m}_2^{(2)}(\boldsymbol{\beta}^0)) \mathbf{a}_2 - 2\sigma_{11}^2 \sigma_{12}^2 \xi_{12}), \end{aligned}$$

with

$$\begin{aligned} \xi_{12} &= \tilde{\mathbf{t}}_1^T(\boldsymbol{\beta}^0) \bar{\mathbf{X}}_1^T \text{Diag}(\bar{\mathbf{m}}^{(2)}(\boldsymbol{\beta}^0)) \bar{\mathbf{X}}_2 \tilde{\mathbf{t}}_2(\boldsymbol{\beta}^0) \\ &= \sum_{j=1}^J \sum_{i=1}^{I_1 - \bar{I}} m_{2ji}^{(2)}(\boldsymbol{\beta}^0) (t_{2i} - \tilde{t}_{1j}(\boldsymbol{\beta}^0))(t_{2i} - \tilde{t}_{2j}(\boldsymbol{\beta}^0)) \\ &= \sum_{j=1}^J \sum_{i=1}^{I_1 - \bar{I}} \frac{n_{2ji}^{(2)}}{n_{2ji}} m_{2ji}^{(2)}(\boldsymbol{\beta}^0) (t_{2i} - \tilde{t}_{1j}(\boldsymbol{\beta}^0))(t_{2i} - \tilde{t}_{2j}(\boldsymbol{\beta}^0)), \end{aligned}$$

it holds

$$\begin{aligned} \text{Var}(\sqrt{N}(X_1 + X_2 + X_3)) &= N(\mathbf{a}_1^T \text{Diag}(\mathbf{m}_1^{(1)}(\boldsymbol{\beta}^0) + \mathbf{m}_1^{(2)}(\boldsymbol{\beta}^0)) \mathbf{a}_1 \\ &\quad + \mathbf{a}_2^T \text{Diag}(\mathbf{m}_2^{(1)}(\boldsymbol{\beta}^0) + \mathbf{m}_2^{(2)}(\boldsymbol{\beta}^0)) \mathbf{a}_2 - 2\sigma_{11}^2 \sigma_{12}^2 \xi_{12}) \\ &= N(\sigma_{11}^2 + \sigma_{12}^2 - 2\sigma_{11}^2 \sigma_{12}^2 \xi_{12}), \end{aligned}$$

that coincides with the asymptotic variance of  $\sqrt{N}(\widehat{\beta}_{11,\lambda} - \widehat{\beta}_{12,\lambda})$ .

**Proof of Corollary 9.** Since

$$\tilde{t}_{kj}(\boldsymbol{\beta}^0) = \tilde{t}_{kj}^{(2)}(\boldsymbol{\beta}^0) + \frac{m_{kj\bullet}^{(1)}}{m_{kj\bullet}} (\tilde{t}_{kj}^{(1)}(\boldsymbol{\beta}^0) - \tilde{t}_{kj}^{(2)}(\boldsymbol{\beta}^0)), \quad k = 1, 2,$$

formula (20) can be rewritten as

$$\begin{aligned} \xi_{12} &= \sum_{j=1}^J \sum_{i=1}^{I_1-\bar{i}} m_{2ji}^{(2)}(\beta^0) \left( t_{2i} - \tilde{t}_{1j}^{(2)}(\beta^0) - \frac{m_{1j\bullet}^{(1)}}{m_{1j\bullet}} (\tilde{t}_{1j}^{(1)}(\beta^0) - \tilde{t}_{1j}^{(2)}(\beta^0)) \right) \left( t_{2i} - \tilde{t}_{2j}^{(2)}(\beta^0) - \frac{m_{2j\bullet}^{(1)}}{m_{2j\bullet}} (\tilde{t}_{2j}^{(1)}(\beta^0) - \tilde{t}_{2j}^{(2)}(\beta^0)) \right) \\ &= \sum_{j=1}^J \sum_{i=1}^{I_1-\bar{i}} m_{2ji}^{(2)}(\beta^0) (t_{2i} - \tilde{t}_{2j}^{(2)}(\beta^0))^2 + \sum_{j=1}^J \sum_{i=1}^{I_1-\bar{i}} m_{2ji}^{(2)}(\beta^0) \frac{m_{1j\bullet}^{(1)}}{m_{1j\bullet}} \frac{m_{2j\bullet}^{(1)}}{m_{2j\bullet}} (\tilde{t}_{1j}^{(1)}(\beta^0) - \tilde{t}_{1j}^{(2)}(\beta^0)) (\tilde{t}_{2j}^{(1)}(\beta^0) - \tilde{t}_{2j}^{(2)}(\beta^0)) \\ &\quad - \sum_{k=1}^2 \sum_{j=1}^J \sum_{i=1}^{I_1-\bar{i}} m_{2ji}^{(2)}(\beta^0) (t_{2i} - \tilde{t}_{2j}^{(2)}(\beta^0)) \frac{m_{kj\bullet}^{(1)}}{m_{kj\bullet}} (\tilde{t}_{kj}^{(1)}(\beta^0) - \tilde{t}_{kj}^{(2)}(\beta^0)). \end{aligned}$$

The last summand is canceled because

$$\begin{aligned} &\sum_{j=1}^J \sum_{i=1}^{I_1-\bar{i}} m_{2ji}^{(2)}(\beta^0) (t_{2i} - \tilde{t}_{2j}^{(2)}(\beta^0)) \frac{m_{kj\bullet}^{(1)}}{m_{kj\bullet}} (\tilde{t}_{kj}^{(1)}(\beta^0) - \tilde{t}_{kj}^{(2)}(\beta^0)) \\ &= \sum_{j=1}^J \frac{m_{kj\bullet}^{(1)}}{m_{kj\bullet}} (\tilde{t}_{kj}^{(1)}(\beta^0) - \tilde{t}_{kj}^{(2)}(\beta^0)) \sum_{i=1}^{I_1-\bar{i}} m_{2ji}^{(2)}(\beta^0) (t_{2i} - \tilde{t}_{2j}^{(2)}(\beta^0)) \end{aligned}$$

and  $\sum_{i=1}^{I_1-\bar{i}} m_{2ji}^{(2)}(\beta^0) (t_{2i} - \tilde{t}_{2j}^{(2)}(\beta^0)) = 0$ . Hence, it holds (22).

If region 2 is completely contained in region 1,  $\xi_{12} = 1/\sigma_{12}$ , and therefore

$$\text{Var}(\hat{\beta}_{11,\lambda} - \hat{\beta}_{12,\lambda}) = \sigma_{12}^2 + \sigma_{11}^2 - 2\sigma_{12}^2\sigma_{11}^2\xi_{12} = \sigma_{12}^2 + \sigma_{11}^2 - 2\sigma_{11}^2,$$

and it follows (23).

## References

- [1] J. Berkson, Minimum chi-square, not maximum likelihood! *Annals of Statistics* 8 (1980) 457–487.
- [2] Y.M.M. Bishop, S.E. Fienberg, P.W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, 1995.
- [3] M. Fay, R. Tiwari, E. Feuer, Z. Zou, Estimating average annual percent change for disease rates without assuming constant change, *Biometrics* 62 (2006) 847–854.
- [4] M.J. Horner, L.A.G. Ries, M. Krapcho, N. Neyman, R. Aminou, N. Howlander, S.F. Altekruse, E.J. Feuer, L. Huang, A. Mariotto, B.A. Miller, D.R. Lewis, M.P. Eisner, D.G. Stinchcomb, B.K. Edwards (Eds.), *SEER Cancer Statistics Review, 1975–2006*, National Cancer Institute, Bethesda, MD, [http://seer.cancer.gov/csr/1975\\_2006/](http://seer.cancer.gov/csr/1975_2006/).
- [5] P.B. Imrey, Power divergence methods, in: P. Armitage, T. Colton (Eds.), *Encyclopedia of Biostatistics*, John Wiley and Sons, New York, 2005.
- [6] S. Kullback, R.A. Leibler, On information and sufficiency, *The Annals of Mathematical Statistics* 22 (1951) 79–86.
- [7] Y. Li, R.C. Tiwari, Comparing trends in cancer rates across overlapping regions, *Biometrics* 64 (2008) 1280–1286.
- [8] Y. Li, R.C. Tiwari, Z. Zou, An age-stratified model for comparing trends in cancer rates across overlapping regions, *Biometrical Journal* 50 (2008) 608–619.
- [9] L. Pardo, *Statistical Inference Based on Divergence Measures*, in: *Statistics: Textbooks and Monographs*, Chapman & Hall/CRC, New York, 2006.
- [10] L. Pardo, N. Martín, Homogeneity/heterogeneity hypotheses for standardized mortality ratios based on minimum power-divergence estimators, *Biometrical Journal* 51 (2009) 819–836.
- [11] L.W. Pickle, A.A. White, Effects of the choice of age-adjustment method on maps of death rates, *Statistics in Medicine* 14 (1995) 615–627.
- [12] N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society B* 46 (1988) 440–464.
- [13] C. Riddell, J.M. Pliska, *Cancer in Oregon, 2005: Annual Report on Cancer Incidence and Mortality among Oregonians*. Department of Human Services, Oregon Public Health Division, Oregon State Cancer Registry, Portland, Oregon, 2008. <http://egov.oregon.gov/DHS/ph/oscar/arpt2005/ar2005.pdf>.
- [14] R.C. Tiwari, L. Clegg, Z. Zou, Efficient interval estimation for age-adjusted cancer rates, *Statistical Methods in Medical Research* 15 (2006) 547–569.
- [15] K.A. Walters, Y. Li, R.C. Tiwari, Z. Zou, A weighted-least-squares estimation approach to comparing trends in age-adjusted cancer rates across overlapping regions, *Journal of Data Science* 8 (2010) 631–644.