25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

A visual examination of Selikoff's 20-year rule using correspondence analysis and the Cressie-Read family of divergence statistics

Asma Alzahrani [a], **Eric J. Beh** [b,c] and **Elizabeth Stojanovski** [a]

[a] *School of Information and Physical Sciences, University of Newcastle, Australia*
[b] *National Institute for Applied Statistics Research Australia, University of Wollongong, Australia*
[c] *Centre for Multi-Dimensional Data Visualisation, Stellenbosch University, South Africa*
Email: asma.alzahrani@uon.edu.au

Abstract: Irving Selikoff's impact on our understanding of the side effects of exposure to asbestos fibres is now very well understood. In his seminal 1981 paper, Selikoff established a "20-year rule" where he concluded that it took about 20 years for workers to exhibit abnormal chest x-rays that signal the presence of asbestosis in at least one of their lungs; asbestosis is the scaring and inflammation of the lung tissue caused by the strong, but flexible, asbestos fibres. This paper presents an analysis of Selikoff's asbestos data using a new method of correspondence analysis for two-way contingency tables that uses the Cressie-Read family of divergence statistics as the general measure of association. While such a family incorporates a wide range of different chi-squared statistics, we shall focus our analysis on Pearson's statistic and the Freeman-Tukey statistic. We also show how one may identify whether each of the categories of the contingency table is a statistically significant contributor to the association structure between the variables. We show that, for Selikoff's data, all years of occupational exposure to asbestos, and all grades of asbestosis that he considered, are statistically significant contributors to the association.

Keywords: *Correspondence analysis, Cressie-Read family of divergence statistics, Pearson's chi-squared statistic, Freeman-Tukey statistic, asbestos fibres*

1. INTRODUCTION

In 1990 The New York Times introduced Irving Selikoff to its readers as "short, he is round, he is old" with "an aura of a savant of the previous century" and that he is "America's most prominent researcher in the field of asbestos disease" (Hooper, 1990). Irving Selikoff was a chest physician who started working at Mount Sanai Hospital in New York from the 1940s until his retirement in 1985. His longevity at the hospital saw much of his work focus on the effects of asbestos and its health risks due to occupational exposure. His landmark 1981 paper published in the *Bulletin of the New York Academy of Medicine* highlighted the dangers of occupational exposure to asbestos fibres, despite much of the work being undertaken during the 1960's (Selikoff, 1981). In his paper, Selikoff describes the early history of such studies and discusses the visit of a gentleman to Dr Montague Murray of Charing Cross Hospital in London in 1898. The gentleman came in with a shortness of breath and died shortly after. According to Dr Murray, the cause of death was linked to the new asbestos factories that were built in the city around 1880. Selikoff goes on to describe similar cases recorded in the US and Europe in the first half of the 20th century until "the pace increased during the 1950s" when more of such cases "began to be seen and questions had to be answered" (p. 948). While at Mount Sanai Hospital in 1963, Selikoff describes that "a small union of insulation workers in the New York metropolitan area" visited the hospital with about half having "abnormal [X-ray] films" (p. 948). These 1117 workers were all exposed to asbestos fibres and upon examination concluded that a link existed between their years of exposure to the fibres and the severity of asbestosis they were diagnosed with; he classified the severity on an ordinal scale from "None" to "Grade 3" (most severe case). Selikoff found that the more severe cases of asbestosis were found in those workers who had been exposed to the fibre for approximately 20 years and so dubbed this his "20-year rule" (p. 948). Selikoff's observations are summarised in Table 1. Interestingly, on page 102 of the Asbestos Workers' Recovery Act, heard before the Subcommittee on Health of the United States Senate, dated September 9, 1985, a statement by Edward J. Carlough, then President of the Sheet Metal Works International Association describes Selikoff's rule as follows: "There is what Dr. Selikoff calls the 20-year rule. The problem doesn't even begin to show up in a membership until they have worked at least 20 years in the industry. Their exposure may have been minimal at some point, but it takes a minimum of that time before tests even show up".

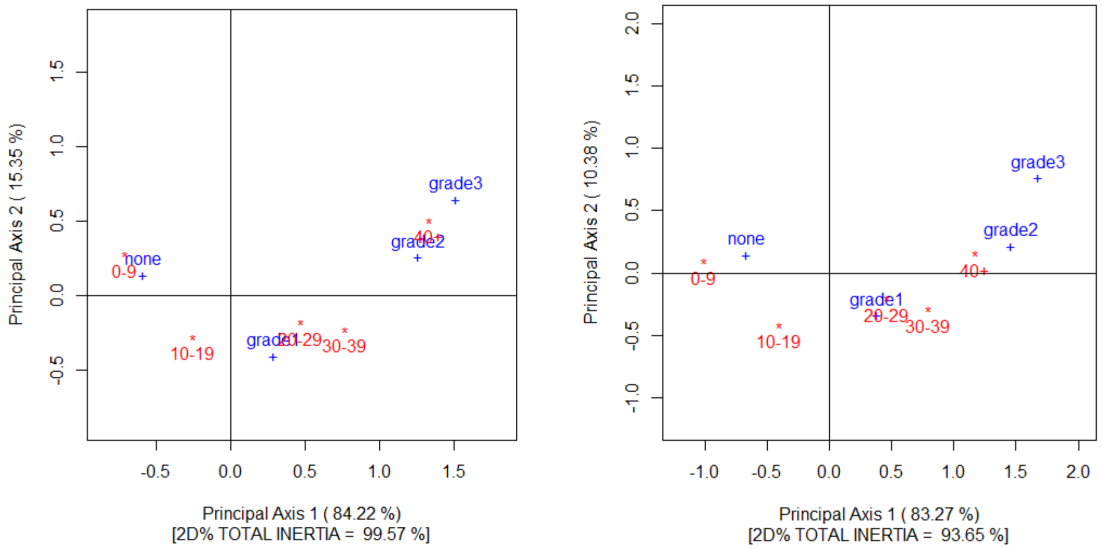
Table 1. Selikoff's data for 1117 New York workers with occupational exposure to asbestos

Occupational exposure (yrs)	Asbestos grade diagnosed				Total
	None	Grade 1	Grade 2	Grade 3	
0–9	310	36	0	0	346
10–19	212	158	9	0	379
20–29	21	102	49	17	194
30–39	25	102	49	18	194
40 +	7	35	51	28	121
Total	575	366	126	50	1117

The 5x4 contingency table of Table 1 has a Pearson chi-squared statistic of $X^2 = 648.81$ with 12 degrees of freedom. Thus, the p-value of this statistic is less than 0.0001 and one may therefore conclude that there is a statistically significant association between a workers' years of exposure to asbestos fibres and their diagnosed severity of asbestosis. Pearson's product moment correlation of Table 1 is 0.69 and, with a p-value < 0.0001, the association is a positive one so that no/low exposure to asbestos fibres is related to no, or a low diagnosed level of asbestos while a lengthy exposure to asbestos fibres is linked to more severe asbestosis. However, this does not confirm the 20-year rule. There is also evidence of over-dispersion in Table 1 and this is visually apparent from the *dispersion plot* of Figure 2. Such plots were considered by Efron (1992) for diagnosing over-dispersion in count data and are constructed by plotting the square root of the variance of the cell counts (assuming they are Poisson random variables) against their standardised residuals. The configuration in Figure 1 shows that there is increasing variation (around zero) in the residuals as the variance increases highlighting the presence of over-dispersion in Table 1. In fact, Haberman (1973) and Agresti (2013, p. 80) show that over-dispersion exists for all two-way contingency tables indicating that, Pearson's statistic may not be the most suitable chi-squared statistic to use. The Freeman-Tukey statistic (Freeman & Tukey, 1950) is thus considered a more appropriate choice since it stabilises the variance and, for Table 1, has a value of $T^2 = 1210.04$ with p-value < 0.0001.

A visual depiction of the association can be undertaken using correspondence analysis (CA). The classical approach to CA uses Pearson's statistic as its numerical foundation (Greenacre, 1984; Beh, 2004; Beh & Lombardo, 2014, 2021) while the Hellinger Distance Decomposition (HDD) method of Cuadras & Cuadras (2006) and Cuadras, Cuadras & Greenacre (2006) uses the Freeman-Tukey statistic as it's foundation, but not by design. Beh, Lombardo & Alberti (2018) described the role of the Freeman-Tukey statistic when performing

CA while Beh and Lombardo (2022) showed the statistics link to HDD. The classical correspondence plot is presented in Figure 1a) while the plot from the HDD of Table 1 is presented in Figure 1b). Both plots clearly show that at least 93% of the association is depicted (with Pearson's statistic showing more than 99%), and that there is indeed a positive association between the two variables. While Figure 1a) shows that Grade 1 is are strongly associated with workers who were exposed to asbestos fibres for between 20 and 40 years of exposure, Figure 1b) shows that Grade 1 is more closely linked to 40+ years of exposure using the Freeman-Tukey statistic compared to what using Pearson's statistic in Figure 1a) suggests. Figure 1a) also shows that less than 10 years of exposure to asbestos fibres is strongly associated with those workers who were not diagnosed with asbestosis, Figure 1b) shows that the association isn't quite as strong. Both analyses provide visual evidence of Selikoff's "20-year rule" but the nature of this rule differs slightly depending on whether Pearson's statistic or the Freeman-Tukey statistic is used to quantify the association.



(a) (b)

Figure 1. Plots from the CA of Table 1 where (a) Pearson's statistic and (b) the Freeman-Tukey statistic are used to assess the association structure

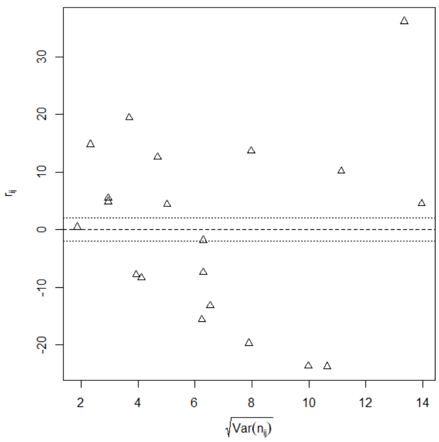


Figure 2. Dispersion plot of Efron (1992) showing over-dispersion being present in Table 1

To explore this association further, this paper is divided into three further sections. Section's 2.1 and 2.2 briefly show how the Cressie-Read family of divergence statistics can be used to a perform CA. Such an analysis is based on the work of Beh and Lombardo (2022) where Figure's 1a) and 1b) are graphical features that can be gained from two special cases of this analysis. Section 2.3 provides some insight into whether one or more of the categories of Table 1 contribute to the association structure that exists between the variables. This is done by showing how elliptically shaped $100(1-\alpha)\%$ confidence regions can be constructed for each category of a two-way contingency table. Such a method is based on the confidence regions of Beh (2010) with Pearson's statistic at its heart while Alzahrani, Beh and Stojanovski (2023) show that these regions can be extended to incorporate a CA using the Cressie-Read family of divergence statistics. Alzahrani, Beh and Stojanovski (2023) also examined a range of features of the regions derived from the family of divergence statistics (including the eccentricity and area of each region) which will not be discussed further in this paper. The authors also did not examine the association structure of Table 1 and so that shall be the focus of Section 3. Some final remarks on this analysis will be provided in Section 4.

2. THE METHOD

2.1. The Cressie-Read family of divergence statistics

Consider an $r \times c$ contingency table, $\mathbf{N}$, where the $(i,\ j)$'th cell entry is given by n_{ij} for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$. Denote the grand total of $\mathbf{N}$ by n and the $(i,\ j)$'th relative frequency by $p_{ij} = n_{ij}/n$. Define the i'th row relative marginal frequency by $p_{i\bullet} = \sum_{j=1}^{c} p_{ij}$ and the j'th column relative marginal frequency by $p_{\bullet j} = \sum_{i=1}^{r} p_{ij}$. To quantify whether there is a statistically significant association between the row and column variables, there are a range of statistics that can be calculated. The five common choices are Pearson's chi-squared statistic, X^2 and the Freeman-Tukey statistic, T^2 (Freeman and Tukey, 1950),

$$X^2 = n\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(p_{ij} - p_{i\bullet}p_{\bullet j}\right)^2}{p_{i\bullet}p_{\bullet j}}, \qquad T^2 = 4n\sum_{i=1}^{r}\sum_{j=1}^{c}\left(\sqrt{p_{ij}} - \sqrt{p_{i\bullet}p_{\bullet j}}\right)^2$$

respectively, as well as the log-likelihood ratio statistic, G^2, the modified chi-squared statistic, N^2, and the modified log-likelihood ratio statistic, M^2. Each of these measures is a chi-squared random variable with $(r-1)(c-1)$ degrees of freedom. These measures are also special cases of the Cressie-Read family of divergence statistics (Cressie and Read, 1984) which is defined as

$$CR^*(\delta) = \frac{2n}{\delta(\delta+1)}\sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij}\left\{\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}}\right)^{\delta} - 1\right\} \tag{1}$$

for $\delta \in (-\infty, \infty)$. A second order Taylor series approximation of (1) around $\left(p_{ij}/\left(p_{i\bullet}p_{\bullet j}\right)\right)^{\delta} = 1$ is

$$CR^*(\delta) \approx CR(\delta) = n\sum_{i=1}^{r}\sum_{j=1}^{c} p_{i\bullet}p_{\bullet j}\left[\frac{1}{\delta}\left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}}\right)^{\delta} - 1\right)\right]^2 \tag{2}$$

where exact measures of some of the above chi-squared statistics can be obtained; for example, $X^2 = CR^*(1) = CR(1)$, $T^2 = CR^*(-1/2) = CR(1/2)$ and $M^2 = CR^*(-1) = CR(0)$. This approximation is ideal for performing a CA on a contingency table and we briefly outline how this analysis can be performed in Section 2.2. We invite the interested author to peruse the pages of Beh and Lombardo (2022) for further information on this CA method.

2.2. The divergence residual

Define the $(i,\ j)$'th *divergence residual* of the contingency table by

$$r_{ij}(\delta) = \frac{1}{\delta}\left(\left(\frac{p_{ij}}{p_{i\bullet}p_{\bullet j}}\right)^{\delta} - 1\right) \tag{3}$$

so that (2) can be expressed in terms of (3) as $CR(\delta) = n\sum_{i,j} p_{i\bullet}p_{\bullet j}r_{ij}^2(\delta)$. Beh and Lombardo (2022) showed that a CA of a $r \times c$ contingency table can be performed by applying a generalised singular value decomposition (GSVD) to the matrix of divergence residuals, (3), so that, for the $(i,\ j)$'th,

$$r_{ij}(\delta) = \sum_{m=1}^{M^*} a_{im}(\delta)\,\lambda_m(\delta)\,b_{jm}(\delta)$$

where $\lambda_m(\delta)$ is the m'th singular value of the matrix of $r_{ij}(\delta)$ values such that $1 > \lambda_1(\delta) > \lambda_2(\delta) > \cdots > \lambda_{M^*}(\delta) > 0$. The square of these singular values gives the *principal inertia* associated with the m'th dimension of the correspondence plot. The optimal correspondence plot consists of M^* dimensions; here $M^* = \min(r,\ c) - 1$ for all δ except $\delta = 0,\ 1/2$ where $M^* = \min(r,\ c)$. From the GSVD of (3), the components $a_{im}(\delta)$ and $b_{jm}(\delta)$ have the property

$$\sum_{i=1}^{r} p_{i\bullet}\, a_{im}(\delta)a_{im'}(\delta) = \begin{cases}1, & m = m'\\ 0, & m \neq m'\end{cases} \qquad \sum_{j=1}^{c} p_{\bullet j}\, b_{jm}(\delta)b_{jm'}(\delta) = \begin{cases}1, & m = m'\\ 0, & m \neq m'\end{cases}.$$

A visual summary of the association between the row and column variables of **N** can then be obtained by jointly depicting along the m'th dimension of a correspondence plot

$$f_{im}(\delta) = a_{im}(\delta)\lambda_m(\delta), \qquad g_{jm}(\delta) = b_{jm}(\delta)\lambda_m(\delta)$$

which is the principal coordinate for the i'th row and j'th column, respectively. For example, Figure 1a) is produced by simultaneously plotting $\left(f_{i1}(\delta), f_{i2}(1)\right)$ and $\left(g_{j2}(1), g_{j2}(1)\right)$ while Figure 1b) is obtained by plotting $\left(f_{i1}(0.5), f_{i2}(0.5)\right)$ and $\left(g_{j2}(0.5), g_{j2}(0.5)\right)$, for $i = 1, 2, \cdots, I$ and $j = 1, 2, \cdots, J$.

2.3. $100(1-\alpha)\%$ confidence regions

Alzahrani, Beh and Stojanovski (2023) showed using (2) how each category of a two-way contingency table can be identified as contributing (or not) to the statistically significant association between its variables using elliptically shaped $100(1-\alpha)\%$ confidence regions that are akin to those derived by Beh (2010). Such regions are interpreted as follows: if any part of a region overlaps the origin, then there is evidence that its category is not a statistically significant contributor to the association and if a region doesn't overlap the origin then its category is such a contributor. While Beh (2010) was concerned only with Pearson's statistic, Alzahrani, Beh and Stojanovski (2023) generalised the regions for (2). They showed that the semi-minor and semi-major axis lengths of the $100(1-\alpha)\%$ confidence region for the i'th row category are defined as

$$x_{i(\alpha)}(\delta) = \lambda_1(\delta)\sqrt{\frac{\chi^2_\alpha}{CR(\delta)}\left(\frac{1}{p_{i\bullet}} - \sum_{m=3}^{M^*} a^2_{im}(\delta)\right)}, \quad y_{i(\alpha)}(\delta) = \lambda_2(\delta)\sqrt{\frac{\chi^2_\alpha}{CR(\delta)}\left(\frac{1}{p_{i\bullet}} - \sum_{m=3}^{M^*} a^2_{im}(\delta)\right)}$$

respectively. Similar lengths can also be defined for the j'th column category; note that these lengths produce an elliptically shaped region that is dominated along the first principal axis since $\lambda_1(\delta) > \lambda_2(\delta), \forall\delta$ and can be generated for all principal axes of the optimal (M^*) correspondence plot. Therefore, when a CA is performed using Pearson's chi-squared statistic, so that $\delta = 1$, these lengths are

$$x_{i(\alpha)}(1) = \lambda_1(1)\sqrt{\frac{\chi^2_\alpha}{X^2}\left(\frac{1}{p_{i\bullet}} - \sum_{m=3}^{M^*} a^2_{im}(1)\right)} \quad \text{and} \quad y_{i(\alpha)}(1) = \lambda_2(1)\sqrt{\frac{\chi^2_\alpha}{X^2}\left(\frac{1}{p_{i\bullet}} - \sum_{m=3}^{M^*} a^2_{im}(1)\right)} \tag{4}$$

and are equivalent to the semi-axis lengths of Beh (2010, eq. (5)). Similarly, when $\delta = 0.5$, this produces the semi-axis lengths when the Freeman-Tukey statistic is used as the measure of association between the variables. In this case, these lengths are

$$x_{i(\alpha)}\left(\frac{1}{2}\right) = \lambda_1\left(\frac{1}{2}\right)\sqrt{\frac{\chi^2_\alpha}{T^2}\left(\frac{1}{p_{i\bullet}} - \sum_{m=3}^{M^*} a^2_{im}\left(\frac{1}{2}\right)\right)} \quad \text{and} \quad y_{i(\alpha)}\left(\frac{1}{2}\right) = \lambda_2\left(\frac{1}{2}\right)\sqrt{\frac{\chi^2_\alpha}{T^2}\left(\frac{1}{p_{i\bullet}} - \sum_{m=3}^{M^*} a^2_{im}\left(\frac{1}{2}\right)\right)} \tag{5}$$

and so are appropriate to use with the HDD variant of CA described by Cuadras & Cuadras (2006) and Cuadras, Cuadras & Greenacre (2006) and the approach of Beh, Lombardo & Alberti (2018).

The semi-axis lengths defined by (4) may also be expressed in terms of the principal coordinates such that

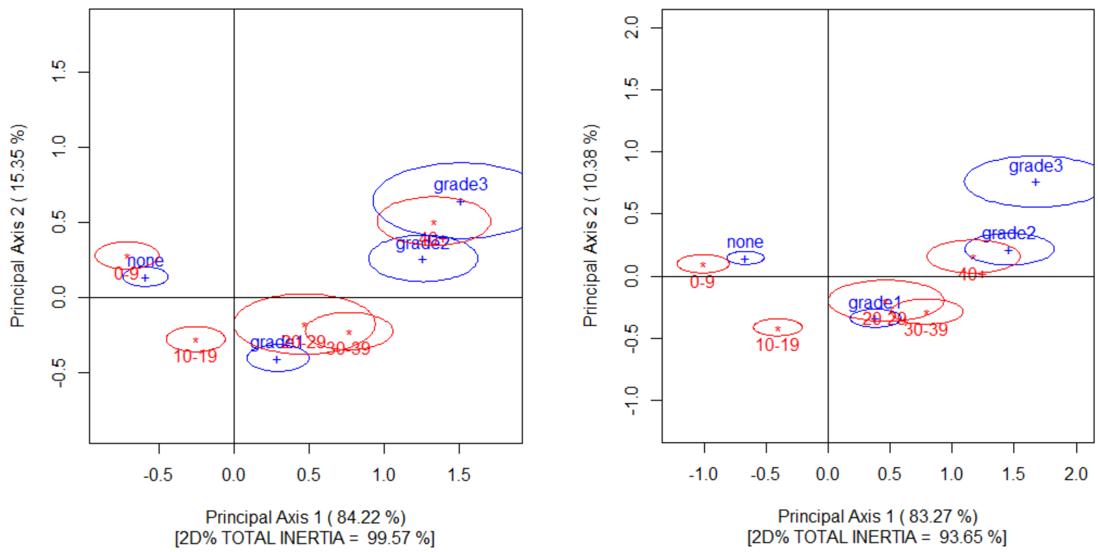
$$x_{i(\alpha)}(\delta) = \lambda_1(\delta)\sqrt{\frac{\chi^2_\alpha}{CR(\delta)}\left(\frac{1}{p_{i\bullet}} - \sum_{m=3}^{M^*} \frac{f^2_{im}(\delta)}{\lambda^2_m(\delta)}\right)} \quad \text{and} \quad y_{i(\alpha)}(\delta) = \lambda_2(\delta)\sqrt{\frac{\chi^2_\alpha}{CR(\delta)}\left(\frac{1}{p_{i\bullet}} - \sum_{m=3}^{M^*} \frac{f^2_{im}(\delta)}{\lambda^2_m(\delta)}\right)}.$$

Therefore, the confidence regions get smaller if the information contained in the third and higher dimensions is incorporated into the analysis. Small elliptical regions will also arise if the association depicted by the m'th axis (for m > 3) of the correspondence plot is large.

3. SELIKOFF'S DATA REVISITED

Attention is now given to the construction of the 95% confidence regions for each of the categories of Table 1 that are depicted in Figure 1. The regions for the row categories of Figure 3a) are produced using (4) (where $\delta = 1$ so that Pearson's statistic is used to quantify the association) while those for Figure 3b) are calculated using (5) (where $\delta = 0.5$ so that the Freeman-Tukey statistic is used to quantify the association). Since none of the confidence regions in the two figures overlap the origin of their respective correspondence plots, for both analyses, all categories of Table 1 are statistically significant contributors to the association structure that is present between the variables. This is highly informative since it shows that all of the New York workers who were diagnosed (or not) with asbestosis are deemed important for understanding the link between asbestos exposure and diagnosed grade of asbestosis. It also shows that all years of exposure are also important to the analysis.

The following remarks can also be made. Since there is some overlap between the ellipses for the "20 – 29" and "30 – 39" years of exposure for $\delta = 1$ and $\delta = 0.5$, this provides some indication that there is very little difference in the diagnosed grade of asbestosis for those who were exposed to the fibres for these two periods, suggesting that the two categories can be combined to form a single category for any subsequent analysis performed. We must, however, not be tempted to interpret directly whether the confidence region of a category overlaps the position of the principal coordinate of another category.



(a) (b)

Figure 3. Plots from the CA of Table 1 with 95% confidence regions superimposed where (a) Pearson's statistic and (b) the Freeman-Tukey statistic are used to assess the association structure

4. DISCUSSION

Selikoff's study of the side effects of exposure to asbestos fibres has had a profound impact around Australia and around the world and one may now view his "20-year rule" to be not as dire as we now understand. For example, Figures 1 and 2 show that asbestosis becomes diagnosable after approximately 10 years of exposure to asbestos fibres. In addition to damage that asbestos fibres to do the lungs, the Asbestos Diseases Society of Australia Inc. also describes exposure to the fibres as a cause of mesothelioma, laryngeal cancer, ovarian cancer, testes cancer, and pleural disorders. Beh and Lombardo (2014, Section 1.4.1) provide an outline of several major international events involving asbestos that have negatively impacted upon communities around the world. The Asbestos Safety and Eradication Agency, under the preview of the Australian Government, notes that the production and importation of products made with asbestos has been banned in 61 countries. The first of these was Iceland in 1983 with some members of the European community imposing bans the same year while Australian bans went into effect on December 31, 2003. While the international community have acted to minimise the exposure and production of asbestos, new dangers from breathing small, dangerous fibrous materials have appeared in the media. Recent attention has been given to the dangers of silica dust. The

UK's Health and Safety Executive website notes that "silica is the biggest risk to construction workers after asbestos". The dangers have also received considerable attention in Australia recently. For example, the Australian Workers' Union reported in July 2022 that "about 600,000 Australian workers are currently exposed to silica dust" and "we will see a tsunami of silicosis in the coming years and decades if swift preventative, regularly and compensatory measures are not quickly adopted by governments around the nation to protect workers from exposure to silica dust". Therefore, visualisations of the kind presented in Figures 1 and 2 help the analyst to gain a quick, meaningful and informative understanding of the associations that exist in their data. However, further work can be undertaken to expand its utility. For example, the ordinal structure of the two variables of Table 1 have not been incorporated. Doing so for CA using (2) will allow for more information of the data structure to be incorporated into the analysis.

REFERENCES

Agresti, A. 2013. Categorical Data Analysis, 3rd ed, New York.

Alzahrani, A., Beh, E.J., Stojanovski, E. 2023. Confidence regions for simple correspondence analysis using the Cressie-Read family of divergence statistics, in review.

Asbestos Diseases Society of Australia Inc. https://asbestosdiseases.org.au/ (last accessed 1/3/2023).

Asbestos Safety and Eradication Agency, under the preview of the Australian Government. https://www.asbestossafety.gov.au/ (last accessed 1/3/2023)

Asbestos Workers' Recovery Act, 1995, Hearing before the Subcommittee on Health of the Committee on Finance, United States Senate. Ninety-ninth Congress, First Session on S.1265 (September 9, 1985), https://www.finance.senate.gov/imo/media/doc/HRG99-457.pdf (last accessed 1/3/2023)

Australian Workers' Union, 2022, Curtin University study backs AWU's alarm at extent of workplace silicosis health risk. https://www.awu.net.au/national/news/2022/07/17515/curtin-university-study-backs-awus-alarm-at-extent-of-workplace-silicosis-health-risk/ (last accessed 1/3/2023)

Beh, E.J. 2004. Simple correspondence analysis: a bibliographic review. International Statistical Review 72(2), 257–284.

Beh, E.J. 2010. Elliptical confidence regions for simple correspondence analysis. Journal of Statistical Planning and Inference 140, 2582–2588.

Beh, E.J., Lombardo, R. 2014. Correspondence Analysis: Theory, Methods and New Strategies, Wiley.

Beh, E.J., Lombardo, R. 2021. An Introduction to Correspondence Analysis, Wiley.

Beh, E.J., Lombardo, R. 2022. Correspondence analysis and the Cressie-Read family of divergence statistics. Working Paper Series 06-22, National Institute for Applied Statistics Research Australia (NIASRA), University of Wollongong, Australia.

Beh, E.J., Lombardo, R., Alberti, G. 2018. Correspondence analysis and the Freeman–Tukey statistic: A study of archaeological data. Computational Statistics & Data Analysis, 128, 73–86.

Cressie, N.A.C., Read, T.R.C. 1984. Multinomial goodness-of-fit tests. Journal of the Royal Statistical Society, Series B 46(3), 440–464.

Cuadras, C.M., Cuadras, D. 2006. A parametric approach to correspondence analysis. Linear Algebra and Its Applications 417, 64–74.

Cuadras, C.M., Cuadras, D., Greenacre, M.J. 2006. A comparison of different methods for representing categorical data. Communications in Statistics – Simulation and Computation 35(2), 447–459.

Efron, B. 1992. Poisson overdispersion estimates based on the method of asymmetric maximum likelihood. Journal of the American Statistical Association 87, 98–107.

Freeman, M.F., Tukey, J.W. 1950. Transformations related to the angular and square root. The Annals of Mathematical Statistics 21(4), 607–611.

Greenacre, M.J. 1984. Theory and Applications of Correspondence Analysis. Academic Press.

Haberman, S.J. 1973. The analysis of residuals in cross-classified tables. Biometrics 29(1), 205–220.

Hooper, J. 1990, The Asbestos Mess, The New York Times, November 25, 1990. https://www.nytimes.com/1990/11/25/magazine/the-asbestos-mess.html (last accessed 1/3/2023)

Health and Safety Executive. Cancer and construction: Silica. https://www.hse.gov.uk/construction/healthrisks/cancer-and-construction/silica-dust.htm (last accessed 1/3/ 2023).

Selikoff, I.J. 1981. Household risks with inorganic fibers. Bulletin of the New York Academy of Medicine 57, 947–961.