Statistical Methods in Medical Research 2002; **11**: 275–292

Added value of three-way methods for the analysis of mortality trends illustrated with worldwide female cancer mortality (1968–1985)

PM Kroonenberg Leiden University, Leiden, The Netherlands, **T Murakami** University of Nagoya, Nagoya, Japan and **JWW Coebergh** Erasmus University, Rotterdam, The Netherlands

Trends in mortality rates are usually presented per tumour site or per country without an overall analysis of the complete data encompassing all three aspects (tumour sites, countries, trends). This paper presents a methodology for such an overall analysis using three-way methods applied to a data set on female mortality rates for 17 tumour sites of 43 countries for the years 1968–1985. Multivariate techniques like biplots and three-mode principal component analysis within an overall three-way analysis-of-variance framework were used. We confirmed the known patterns of comparatively high mortality for women due to cancer of the bladder, intestines, pancreas, rectum, breast, ovary, skin and leukaemia and the relatively low mortality rates for liver cancer in Western and Northern Europe, the USA, Australia and New Zealand. Also, the reverse pattern was observed for Middle and Southern Europe, Hong Kong, Singapore, and in Japan, and in some but not all Latin American countries. The relatively mortality due to cancer was high in the lungs, mouth, larynx and oesophagus in the British Isles, but was much less in other European countries. Mortality due to cancer of the thyroid, uterus, gall bladder and stomach was high in Middle European countries, as was the case in Japan, Chile and Costa Rica. Rates were low for Southern European countries, North America, Australia and New Zealand. Specific deviating patterns in the data were the more rapidly decreasing mortality rates for stomach cancer in Chile and Japan and the more rapidly increasing mortality rates for lung cancer in the USA, Scotland and Denmark. In conclusion, using three-way methods, it was feasible to analyse the cancer mortality data in their entirety. This enabled the simultaneous comparison of trends in relative mortality rates between all countries due to all tumour sites, as well as the identification of specific deviating trends for specific tumour sites in specific countries.

1 Introduction

Large-scale comparative studies of mortality from cancer have provided information on the mortality rates for several countries at several cancer sites[1–7] describing age-adjusted mortality rates as curves per cancer site for each country. Such presentations enable detailed insight into the development of the rates of individual cancers over time, and comparisons of time trends over countries. However, the comparative conclusions drawn from them are generally based on visual inspection, and comparisons between cancer sites are difficult because results are presented by cancer site, or comparisons between countries are difficult because results are presented per country. This paper shows how countries, cancer sites and trends can be analysed simultaneously within a single multivariate

Address for correspondence: PM Kroonenberg, Department of Education, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands. E-mail: kroonenb@fsw.leidenuniv.nl

 10.1191/0962280202sm287ra

framework, while paying special attention to the general relationships between countries and tumour sites. Such a general multivariate view can serve as a background for the study of patterns for individual countries, which are different from the general trends.

Exploratory three-way analysis-of-variance with one observation per cell, biplots and three-mode principal component analysis are in themselves not new, but to our knowledge they have not been used before *jointly* in the analysis of registry data.

2 Materials and methods

Database

The cancer mortality data have been collated by Kurihara *et al.*[8] from available files of the World Health Organization (WHO), except for those from the US which were derived from Vital Statistics of the United States.[8]* For the sake of conciseness and coherence we only analysed the women's data.

The dataset consists of the nine age-adjusted biennial mortality rates (i.e. deaths per 100 000) in 43 countries (of the 46 mentioned in the original publication) for 17 cancer sites over the years 1968–1985. The classification of the diseases was based on the International Classification of Diseases, ICD (1950, 1957, 1967, 1977),[9–12] and the relevant list numbers corresponding to each malignant neoplasm in the 6th to 9th Revision of the ICD, as well as the years when each Revision of the ICD was used in each country, can be found in Kurihara *et al.*,[8] Tables B and C, respectively. 'The total male and female population in 46 countries around 1950 [Dr. Segi's world population] was used as a standard population in calculating age-adjusted death rates. A . . . ten-year age-group was used [for calculating age-adjusted death rates] from 1968–69 to 1984–85' (Kurihara *et al.*,[8] p. vii). Table 2 gives a list of the cancer sites, and the countries are listed in Table 4.

Countries with unreliable or non-existing data, for example all African countries, were excluded. Independent estimates of the measurement errors across countries and sites were not available.

Missing data

Because a variety of techniques were employed, and we preferred to work with the same data throughout, it was decided to estimate the 810 missing data (12.3%) beforehand, rather than use different (model-based) procedures for each technique. If the mortality rate for a tumour site in a particular country was missing in a particular year, then the missing value was estimated from the available data. In particular, the missing values were estimated with the following analysis-of-variance model. For each cancer site $j(j = 1, . , J)$ in country $i(i = 1, . , I)$ the missing value in year $k(k = 1, . , K)$, $m_{ik}^{(j)}$, was estimated as

$$m_{ik}^{(j)} = \mu^{(j)} + \alpha_i^{(j)} + \gamma_k^{(j)} \tag{1}$$

*The data were made available to us by Prof. Hiroshi Ogawa, Aichi Mizuho University. Both the female and male data can be obtained from the first author from the website of The Three-Mode Company (www.fsw.leidenuniv.nl/ ~ kroonenb) where more easily readable colour versions of the figures can also be found. In order to include as many countries as possible, we only used the data from 1968 onwards.

where $\mu^{(j)}$ is the overall mean for site j, $\alpha_i^{(j)}$ the site main effect for country i, and $\gamma_k^{(j)}$ the site main effect of year k. If for a site j of a country i no values were present for any year k, then the site effect for that country $\alpha_i^{(j)}$ could not be evaluated and only the other two terms could be used for estimation of the missing values. The analyses to be presented are based on the completed data set including the estimates for the missing data.

Analysis

First, we considered the relative variability of the three aspects or ways (cancer site, country, year), and compared these with the variability due to the interactions between aspects. Then we examined the patterns in these interactions, especially those in the three-way interaction which contains information on which tumour sites had specific trends in specific countries. As the data are not a sample and none of the three aspects is random, hypothesis testing is not an issue. Here, the *F*-values, usually reserved for testing, serve to evaluate the relative sizes of the effects.

Three-way ANOVA

The variability in the data set was assessed with a three-way (fixed-effects) analysis of variance with one observation per cell, i.e. the mortality rate of a country for a particular cancer site in a given year. This enabled the assessment of the relative importance of the various main and interaction effects.

Effects which required summations over sites were not interpretable because of the vast differences in mortality rates from different tumour sites. Therefore, only the site main effect, the site-by-year and the site-by-countries two-way interactions, and the three-way interaction were examined.

Two-way interactions

The site-by-year interaction represents the trends of the various cancer mortality rates across the countries, and they are presented in a straightforward graph. It provides the background for the remainder of the analyses on the relative increase and decrease of mortality rates from various cancers.

The site-by-country interaction carries information on (1) which *sites* had relatively high/low mortality rates in which countries and (2) which *countries* had relatively high mortality rates for which sites. To get an overview of the countries and sites together, we depict the countries and cancer sites in a single display (a so-called *biplot*[13–16]). The biplot provides a compact view of the relative mortality from cancers across countries and vice versa, and in particular it displays the systematic variability.

The centre of a biplot is equal to the mean mortality rate for each cancer site; a country which is located at the origin has an average mortality rate for all tumour sites. Tumour sites are represented by vectors and countries by points. Angles between the site vectors reflect their correlations, so that sites whose vectors run closely together have similar patterns in mortality across countries. Adjacent countries have similar mortality ratios across sites. The relative mortality rates of countries with respect to a site are derived from the relative sizes of their projections onto the vector of a cancer site. The product of the length of a site vector and the length of the projection of a country onto the site vector (called an *inner product*) approximates the size of the interaction between

the site and the country. A combination of a large projection with a long vector thus corresponds to a large (estimated) interaction term (cf. the Appendix[14,17]).

For the present data, the 17 sites × 43 countries two-way interaction was investigated by displaying countries and cancer sites in a biplot based on the means aggregated over years. The matrix with means was standardized per cancer site, so that the mortality rates for cancer sites were in standard deviation units, which made them independent of overall site mortality rates. To facilitate description, the cancer sites were rotated with a standard varimax procedure.

Three-way interactions

Three-way interactions in large designs generally contain a combination of sizeable systematic variance associated with a small number of degrees of freedom and residual variance associated with the remaining larger part of the degrees of freedom. The square root of the residual variance can be used as an independent estimate of the overall measurement error.[18] The structure of the systematic variance can be evaluated with three-mode principal component analysis (see Appendix[19,20]), with which specific increases in mortality rates for specific cancer sites in specific countries can be assessed. Three-mode principal component analysis is an extension of ordinary principal component analysis to three-way data. Instead of two sets of co-ordinates for rows (countries) and columns (tumour sites), there are now three sets of co-ordinates (or components), one for each way, each with its own number of components, say P for the first way (countries), Q for the second way (tumour sites) and R for the third way (years). In addition, a set of parameters, the core array, indicates the importance of a particular combination of the components of each way (cf. Appendix[20–22]).

The interpretation of a three-mode component analysis of the three-way interaction requires the joint examination (1) of a biplot (in this context generally called a *joint plot*) containing the relationships between the countries and the tumour sites and (2) the coefficients of the time components, which contain information on the changes over time in the relationships between sites and countries. Supposing that a time component show an increasing linear trend, a positive value (inner product) in the joint plot corresponds to a linear increase in the interaction over the years and a negative value a linear decrease over the years (see Appendix).

3 Results

Partitioning variability: three-way analysis of variance

Table 1 contains the partitioning of the variability in the data according to a three-way analysis-of-variance model with the three aspects as fixed factors. The degrees of freedom in the table refer to the completed data (see Appendix for some further comments on the degrees of freedom).

Main effects

Mortality differences were largest for tumour sites, followed at a great distance by country and year. Therefore, only effects which did not require summing over sites were

Table 1 Three-way fixed-effects analysis of variance of completed cancer mortality data with one observation per cell

Source	Sum of squares	Percentage of SS (total)	Degrees of freedom[b]	Mean square	F[c,d]
Non disease-site effects					
Years	75	0.04%	8	9.5	21.4
Countries	6110	3.2%	42	145.5	328.6
Years-by-countries	247	0.1%	336	0.7	1.7
Disease-site effects					
Tumour sites	129 456	68.7%	16	8091.0	18 278.4
Tumour sites-by-years	2931	1.6%	128	22.9	51.7
Tumour sites-by-countries	47 149	25.0%	672	70.2	158.5
Three-way interaction[a]					
Tumour sites-by-countries-by-years	2380	1.3%	5376	0.4	
Systematic part ($2 \times 2 \times 1$ three-mode model)	1203	0.6%	124	9.7	
Error part (residual from model)	1177	0.6%	5252	0.2	
Total	188 348	100.0%	6578	28.6	

[a]The three-way interaction is further partitioned with a three-mode principal component analysis in a systematic and an error part (for details see Appendix).
[b]No corrections were made in the degrees of freedom for the missing data. See Appendix.
[c]The mean square for the T. S. $\times$ C $\times$ Y interaction is used as an error term; using the error part of the three-way interaction would increase all F with a factor 2.
[d]No P-values are included as none of the mode can be considered stochastic.

considered for further analysis. This excluded the mortality rates per country mortality, rates per year as well as the year-by-country interaction.

Two-way interactions

In line with common knowledge, the major two-way interaction was between cancer site-by-country one (sum of squares ($SS_{s\times c}$) = 41 749; mean square ($MS_{s\times c}$) = 70). The site-by-year interaction was much smaller ($SS_{s\times y} = 2931$; $MS_{s\times y} = 23$), but still considerable.

Three-way interaction

The absolute size of the three-way interaction sum of squares was similar in size to that of the site-by-year interaction: $SS_{s\times c\times y} = 2380$ versus $SS_{s\times y} = 2931$, but less so when the mean squares are compared: $MS_{s\times c\times y} = 0.4$ versus $MS_{s\times y} = 22.9$. Table 1 shows that the sum of squares of the three-way interaction could be equally split between a systematic and a residual part ($SS_{sys} = 1203$, and $SS_{res} = 1177$, respectively), but that the number of degrees of freedom associated with systematic part was very small ($df_{sys} = 124$) compared to that of the residual part ($df_{res} = 5252$). The ensuing systematic mean square ($MS_{\text{sys}} = 9.7$) was sufficiently large to warrant examination of the systematic part of the three-way interaction for specific patterns.

Mortality rates per site

Table 2 contains the overall mortality rates per site. Breast cancer was the most frequently occurring cause of death due to cancer in the 43 countries together, followed

Table 2 Tumour site-specific female mortality rates (averaged over countries and years, based on valid data, ordered according to their mortality rate)

Tumour site	Mortality rate (deaths per 100 000)
Breast	17.25
Stomach	10.99
Uterus (all parts)	10.13
Lung, trachea and bronchus	8.00
Intestines, except rectum	8.58
Ovary, fallopian tube and broad ligament	6.19
Pancreas	4.43
Leukaemia	3.82
Rectum and rectosigmoid junction	3.70
Liver	3.09
Gall bladder and bile ducts	2.56
Oesophagus	1.66
Bladder	1.43
Skin	1.32
Mouth and pharynx	1.30
Thyroid gland	0.70
Larynx	0.35

at a distance by stomach and uterus and lung cancer, while death due to cancer of the thyroid gland and larynx was rare.

Mortality rates of cancer sites over years

Figure 1 shows, for all countries combined, the cancer site-by-year interaction, i.e. the trends in site mortality rates with the mean site mortality rate added in. The curves

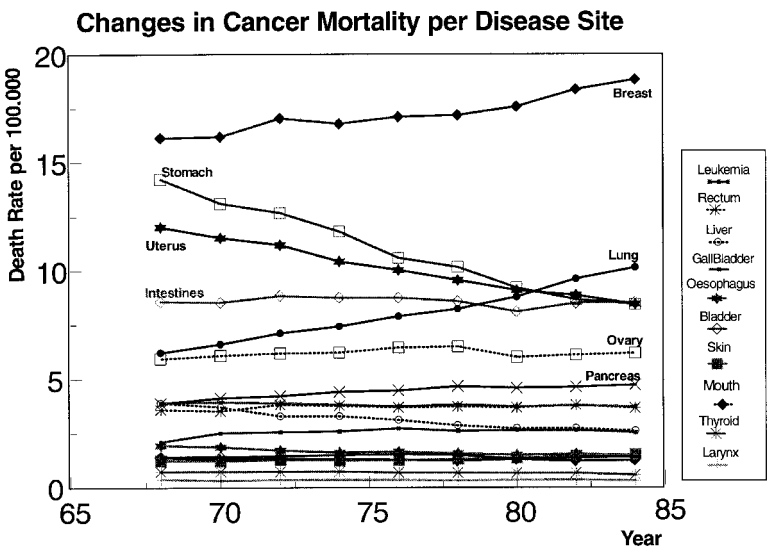

Figure 1 Trends in mortality rates for cancer sites between 1968 and 1985 based on biannual rates.

representing these trends, however, disguised sizeable differences between countries. The mostly increasing (already high) mortality rate was that of lung cancer, with breast cancer not far behind. Mortality from stomach cancer, uterus cancer, and – to a certain extent liver cancer – declined considerably. For most of the other cancer sites, the mortality rates were fairly stable over the period of investigation.

Mortality rates of cancer sites for countries

The three-dimensional solution accounted for 64% of the variability in the standardized means with 35, 15 and 14% for the first three axes of the biplot, respectively. In the biplot (Figures 2 and 3) three reasonably tight bundles of vectors, representing three groups of cancer sites could be distinguished (see also Table 3). To assist the interpretation of the three-dimensional figure Table 4 was constructed from Figures 2 and 3.

Cancer sites (Table 3; Figures 2 and 3)

Bladder, pancreas, intestines and breast cancer sites (group A1) formed a close bundle of vectors, as did ovary, rectum, skin sites and leukaemia, (group A2) while liver was

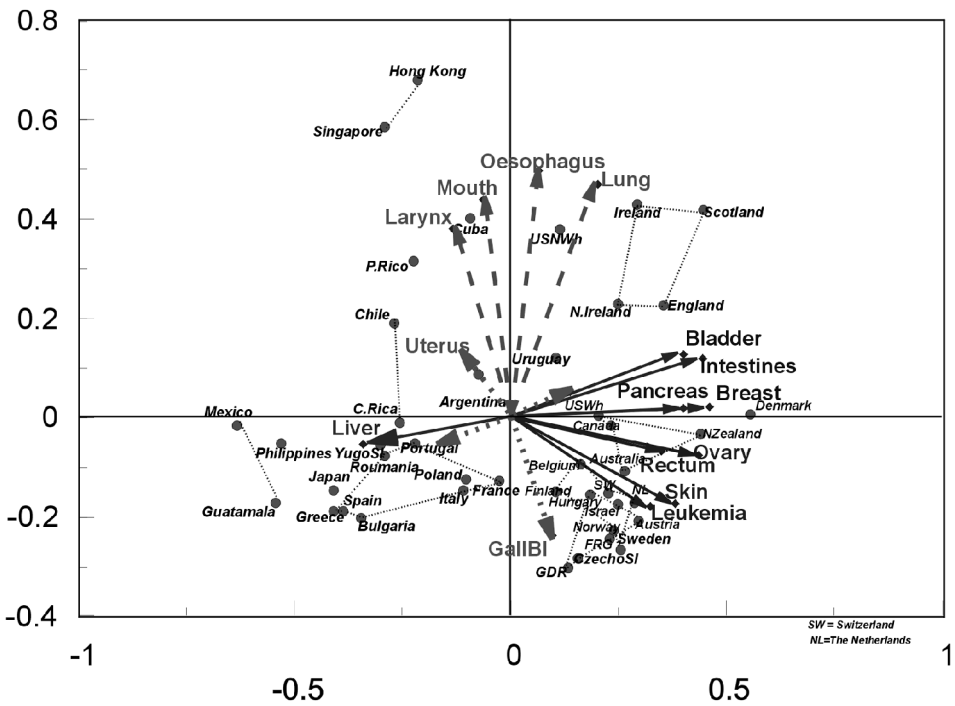

Figure 2 Two-way interaction between countries and cancer sites. Horizontal axis is *first* component; vertical is *second* component. Circles are countries; vectors towards diamonds are cancer sites. The sites fall into three major groups (see also Tables 3 and 4). Sites belonging to the same group have vectors pointing in the same direction in three-dimensional graph. Group A sites lie along the first axis, Group B sites along the second axis and Group C sites point downward on the third axis (see Figure 3). Vectors of sites that have small angles with each other are highly positively correlated. Countries which have similar projections on a site vector have a similar relative mortality rate for that site.

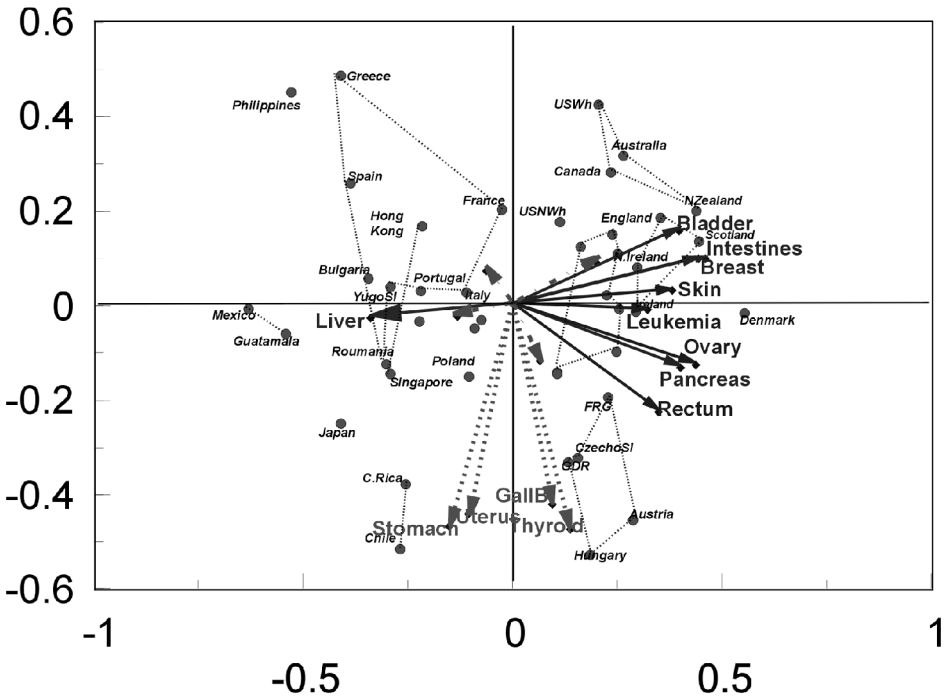

Figure 3 Two-way interaction between countries and cancer sites. Horizontal axis is *first* component; vertical is *third* component. Circles are countries; vectors towards diamonds are cancer sites. The sites fall into three major groups (see also Tables 3 and 4). Sites belonging to the same group have vectors pointing in the same direction in three-dimensional graph. Group A sites lie along the first axis, Group B sites along the second axis (see Figure 2) and Group C sites point downward on the third axis. Vectors of sites that have small angles with each other are highly positively correlated. Countries which have similar projections on a site vector have a similar relative mortality rate for that site.

situated on the other end of the same axis. The smoking-related cancer sites, such as lung, mouth, larynx and oesophagus (group B) all pointed in the same direction along the second axis, and finally thyroid, stomach, uterus and gall bladder (group C) were located along the third axis.

Countries

Countries were roughly grouped into five clusters: (I) British Isles (England and Wales, Ireland, Northern Ireland and Scotland); (II) predominantly 'Anglo-Saxon' countries outside Europe (Australia, Canada, New Zealand and US white population); (III) 'Northern' Europe (Belgium, Finland, Norway, The Netherlands, Sweden) and Israel; (IV) Central Europe (Austria, Czechoslovakia, Federal Republic of Germany, German Democratic Republic, Hungary, Switzerland); and (V) Southern Europe (Bulgaria, France, Greece, Italy, Portugal, Romania, Spain, Yugoslavia). Several other countries also exhibited especially large interactions, such as Mexico + Guatemala, Costa Rica + Chile, Japan, Hong Kong + Singapore, and Denmark.

Table 3 Varimax rotated co-ordinates for tumour sites in country-by-tumour site biplot (Figure 2). The grouping is an empirical one and is based on the clustering of the tumour sites in Figure 2. Each (sub)group corresponds with a relatively tight bundle of vectors in that figure

Tumour site	Component		
	1	2	3
Group A			
Bladder	**0.400**	0.128	0.157
Intestines (except rectum)	**0.445**	0.118	0.098
Breast	**0.462**	0.021	0.098
Pancreas	**0.402**	0.018	−0.131
Rectum and rectosigmoid junction	**0.349**	−0.066	−0.224
Ovary	**0.437**	−0.077	−0.126
Skin	**0.383**	−0.173	0.031
Leukaemia	**0.324**	−0.180	0.009
Liver	−**0.339**	−0.054	0.025
Group B			
Oesophagus	0.067	**0.499**	−0.118
Lung, trachea, bronchus	0.202	**0.469**	0.091
Mouth, pharynx	−0.062	**0.440**	0.074
Larynx	−0.132	**0.381**	−0.022
Group C			
Thyroid	0.138	0.052	−**0.472**
Stomach	−0.153	−0.052	−**0.466**
Uterus (all parts)	−0.105	0.129	−**0.441**
Gall bladder, bile ducts	0.097	−0.236	−**0.419**

Countries and cancer sites

The mortality rate for liver cancer was relatively highest for Mexico, Guatemala and the Philippines, while it was relatively lowest for the British Isles (cluster I) and other 'Anglo-Saxon' countries (cluster II), and especially Denmark. Mortality due to lung cancer (and other group B sites) was relatively prevalent in the British Isles (cluster I), and less so for Hong Kong and Singapore. On the other hand, middle European countries (cluster IV) had relatively low rates compared to other countries. Stomach and uterus (and other group C sites) had relatively high mortality rates for middle European countries (cluster IV) while they had low rates for non-European Anglo-Saxon countries (cluster II) and Southern Europe (cluster V).

US samples

Figures 2 and 3 showed that the smoking-related group B cancer sites have much higher mortality rates for non-white US females, while on group A cancer sites there was not much difference. White US females exhibited much lower mortality for group C sites, such as thyroid and stomach cancers, than non-white US females who had average mortality rates for these sites.

Specific changes over time in the country–cancer site combinations

The three-mode principal component model with which the systematic part of the three-way interaction was analysed, required a solution with two components for the

Table 4 Relative mortality rates per cluster of countries and per group of tumour sites summed over years (based on country-by-tumour site biplot; Figure 2)

Cluster	Countries	Mortality rates for tumour sites				
		Group A			Group B	Group C
		A1. Bladder, breast, intestines, pancreas	A2. Ovary, skin, rectum, leukaemia	Liver	Lung, mouth, larynx, oesophagus	Thyroid, uterus, gall bladder, stomach
	Denmark	Very high		Very low	Average	Average
I	England and Wales, Ireland, Northern Ireland, Scotland (*British Isles*)	High		Low	High	Average
II	Australia, Canada, New Zealand, US white population	High		Low	Average	Low
III	Belgium, Finland, The Netherlands, Norway, Sweden, Switzerland (*'Northern' Europe*) + Israel	Rather high	High	Rather low	Low	Average
IV	Austria, Czechoslovakia, Federal Republic of Germany, German Democratic Republic, Hungary (*Middle Europe*)	Rather high	High	Rather low	Low	High
V	Bulgaria, France, Greece, Italy, Portugal, Romania, Spain, Yugoslavia (*Southern Europe*)	Low		High	Rather low	Rather low (Greece very low)
VI	Hong Kong, Singapore	Average	Very low	Rather high	Very high	Average
	Japan	Low		High	Rather low	High
	Chile, Costa Rica	Low		High	Rather high	High
	Mexico, Guatemala	Very low		Very high	Average	Average
	Argentina	Average		Average	Average	Average

Countries not mentioned in table: Cuba, Philippines, Poland, Puerto Rico, Uruguay, US non-white population. The qualifications 'very high', 'high', 'low' etc. are based on the average inner products between sites and countries; these products represent the approximation of the average site-by-country interaction in a group of countries on the cluster of tumour sites. They provide a qualitative indication of the relative differences between countries.

countries, two components for the cancer sites and one for the years (a $2 \times 2 \times 1$ solution). Due to the single component for the years, the three-way core array simplified to a 2×2 diagonal matrix. This three-mode component model described 50.6% of the sum of squares for the three-way interaction with only 2.3% (or 124) of the degrees of freedom.

Components of the three ways

Time

The coefficients of the single time component (Table 5) show a virtually linearly increasing trend over the years, indicating that the specific relationships between the sites and countries were linearly increasing or decreasing over time.

Countries and tumour sites

The joint plot of Figure 4 displays the linearly increasing and decreasing relationships between the countries and sites in the three-way interaction (see also Table 6). Of the cancer sites only lung cancer and stomach cancer gave rise to country-specific trends as can be seen from their long vectors in the plot. In particular, Chile and Japan had a specific decreasing trend in mortality due to stomach cancer. Therefore, in addition to the general decreasing mortality trend for stomach cancer (see Figure 1) and the stable cancer mortality rate in Japan, this country showed an additional decrease in mortality due to stomach cancer. In other words, the mortality due to stomach cancer in Japan was declining faster than in other countries.* A similar but weaker trend occurred in Austria, Hungary and Costa Rica. Figure 4 also shows that comparatively Guatemala's mortality rate appeared to be increasing. However, in combination with the general decreasing trend for stomach cancer this means that its mortality rate was not declining as fast as in other countries.

Table 5 Analysis of three-way interaction: time component of the three-way analysis

Years	Component 1
1968/69	−0.52
1970/71	−0.42
1972/73	−0.27
1974/75	−0.11
1976/77	0.02
1978/79	0.17
1980/81	0.26
1982/83	0.37
1984/85	0.49

The increasing time trend in this table is one part of the information necessary for the interpretation of the three-way interaction and can only be evaluated together with the information contained in Figure 4 (see also explanation of Figure 4).

*Our results cannot visually be compared with the graphs presented in Kurihara *et al.*[8] as their graphs have a logarithmic vertical scale.

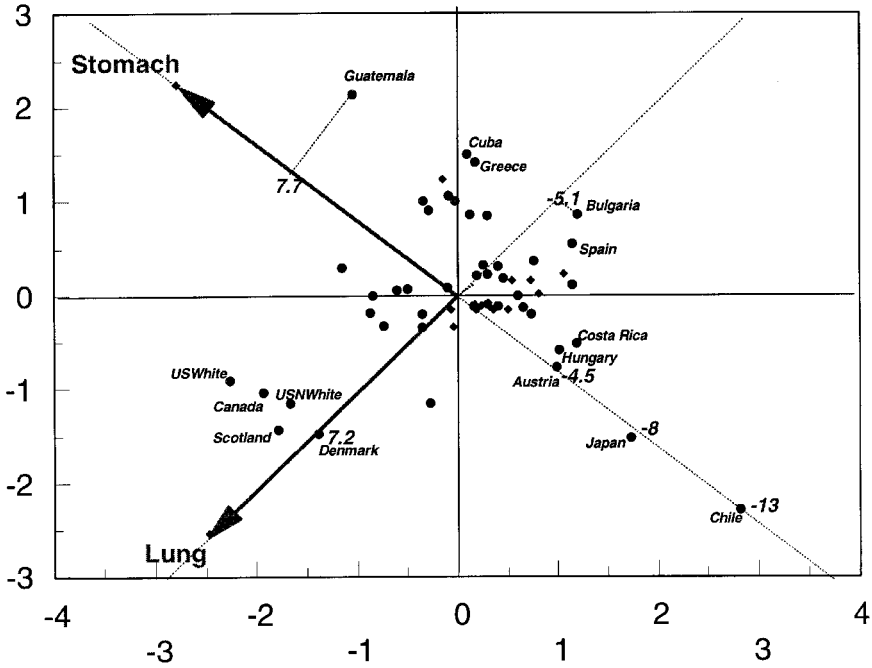

Figure 4 Three-way interaction: biplot of cancer site and country components. Circles indicate countries; diamonds indicate cancer sites. Only countries and cancer sites which have sizeable interaction terms with each other are marked. Numbers on the (biplot) axes represent approximated interaction terms, e.g. the value of the approximation for the interaction term between Japan and stomach cancer is −8.

The three-way interaction also showed an increasing trend for mortality due to lung cancer in Scotland, USA, Canada and Denmark. Thus the mortality rates for women in these countries were increasing at an even faster rate than shown in Figure 1; on the other hand those of Spain, Cuba, Bulgaria and Greece are increasing somewhat more slowly than shown in Figure 1.

Summary of substantive results

First, the relative mortality rates due to various cancers averaged over countries and the changes in mortality over time showed, in particular, that mortality due to breast cancer (increasing) was the highest followed by stomach cancer (decreasing), uterus (decreasing), intestines (stable) and lung (strongly increasing).

Secondly, in Western and Northern Europe, the USA, Australia and New Zealand, women have comparatively high mortality due to cancer of the bladder, intestines, pancreas, rectum, breast, ovary, skin and leukaemia, and relatively low mortality rates for liver cancer. The reverse pattern is true in Middle and Southern Europe, Hong Kong, Singapore, Japan, and in some but not all Latin American countries. The relative mortality due to cancer of the lungs, mouth, larynx and oesophagus was high in the British Isles, but was much lower in other European countries. Mortality due to cancer of the thyroid, uterus, gall bladder and stomach was comparatively high in Middle

Table 6 Analysis of three-way interactions: size of interactions between stomach cancer and lung cancer with country (see Figure 4 for a pictorial representation)

Country	Stomach	Lung
Guatemala	**7.7**	−2.8
USA white population	**4.3**	**7.9**
Canada	3.1	**7.4**
USA non-white population	2.1	**7.0**
Scotland	1.8	**8.0**
Denmark	0.5	**7.2**
Greece	2.7	**−4.0**
Cuba	3.1	**−4.0**
Spain	−1.9	**−4.2**
Bulgaria	−1.4	**−5.1**
Hungary	**−4.1**	−1.0
Costa Rica	**−4.4**	−1.6
Austria	**−4.5**	−0.5
Japan	**−8.2**	−0.4
Chile	**−13.0**	−1.2

Only countries and sites with at least one inner product larger than 4.0 are included in the table. The values presented in this table represent the product of the length of the projection of a country on the vector of a tumour site multiplied by the length of the tumour site vector. This product indicates whether a country has a particularly large interaction with a site over and above the pattern shown in Figures 2 and 3. When multiplied with the time co-ordinates of Table 5, one gets an approximation of the three-way interaction. Large positive values in this table indicating an increasing size of the (approximate) three-way interaction (Guatemala for stomach cancer), and a large negative value indicates a decreasing size of the three-way interaction (Chile and Japan for stomach cancer).

European countries, as was the case in Japan, Chile and Costa Rica. Rates were relatively low for Southern European countries, North America, Australia and New Zealand.

Thirdly, specific deviating patterns from the general mortality trends were found for specific sites and specific countries. In particular, the mortality rates for stomach cancer in Chile and Japan (and to a lesser degree also in Austria, Hungary and Costa Rica) were more rapidly decreasing than in other countries. On the other hand, the general decrease was tempered in Guatemala. Moreover, mortality due to lung cancer in females was more rapidly increasing in the USA, Scotland and Denmark, in contrast with Cuba, Greece, Bulgaria and Spain, where the rates for female lung cancer were not increasing as much as the average trend.

4 Discussion and conclusions

In this paper we have presented an overall analysis of the changes in female cancer mortality rates in several countries. The results refer to four facets: (1) the relative sizes

of the variability due to the three aspects of the data (cancer sites, countries and time) and the variability due to their interactions; (2) the general trends in mortality rates for different sites; (3) the relative sizes of the mortality in different countries for the cancer sites; and (4) the deviating trends for particular sites in particular countries. Given the extensive knowledge about cancer mortality, none of these findings by themselves were particularly new, but in the past they were derived in isolation and could not be assessed with respect to each other. Furthermore, due to the high level of condensation of a large amount of information, our overviews and graphs provide a useful addendum to the information already presented in volumes such as *Cancer Incidence in Five Continents*, produced every five years by the International Agency for Research on Cancer (IARC).[23–26]

Methodological conclusions

The procedures outlined provide numerical information and maps which can be used as a backdrop for country-by-country analyses such as in the IARC's cancer incidence reports. These maps may serve a similar purpose as the map John Snow made to search for the source of cholera in London in 1854.[27] That map was used to put the epidemiological data in perspective, but was not an explanation in itself. The patterns found through analyses, such as those presented in this paper, can be related to further characteristics of the countries to explain why they take the form they do. The advantage of the present approach is that it can be shown that certain patterns are indeed very specific, such as the rapid decline of mortality due to stomach cancer in Chile and Japan. Without the integral approach, it is known that such patterns exist but it is impossible to tell how specific they are and whether they are more specific than other patterns.

Overall analyses allow for rational decisions about the clustering countries with similar patterns and about grouping cancer sites whose patterns over countries are comparable. For instance, Franceschi *et al.*[2] presented an analysis of cancer mortality rates of European countries dividing them into three clusters without explicitly stating the basis for the subdivision into EEC countries, non-EEC Western countries and Eastern European countries. Using analyses such as ours, their clustering of countries could have been made on empirical grounds.

To show that large data sets from registries can be handled as a whole, we have preferred to include as large a number of countries as possible, at the cost of introducing some inaccuracies in the results and their interpretation due to imperfect data. Overall analyses are especially vulnerable to missing data. In the present data set we had to estimate some 12% of them. We expect that this did not distort the results too much, but higher percentages of missing data will inevitably make the results more unreliable.

In summary, by analysing the data within the single framework of three-way analysis of variance, a better insight is acquired in differences between changes over time in mortality rates for different cancer sites in different countries. In particular the joint representation of the tumour sites and countries in a biplot and the detailed analysis of the three-way interaction form the basis for this claim. Together they provide a better overall view of the worldwide dimension of the epidemiological development of various types of cancers and of specific patterns in this development.

Acknowledgements

The research of the first author was financially supported by a grant from the Nissan Fellowship Programme of the Netherlands Organization of Scientific Research (NWO), a grant awarded to both first authors by the Japanese Society for the Promotion of Science, and the Graduate School of Education and Human Development, Nagoya University, Nagoya, Japan.

References

1 Davis DL, Hoel D, Fox J, Lopez AD. International trends in cancer mortality in France, West Germany, Italy, Japan, England and Wales, and the United States. *Annals of the New York Academy of Science* 1990; **609**: 5–48.

2 Franceschi S, Levi F, La Vecchia C, Lucchini F, Negri E. Comparison of cancer mortality trends in major European areas, 1960–1989. *European Journal of Cancer Prevention* 1994; **3**: 145–206.

3 Ghadirian P, Thouez JP, Simard A. La géographie du cancer de l'oesophage. *Social Science and Medicine* 1988; **27**: 971–85.

4 Groves FD, Zavala PD, Correa P. Variation in cancer mortality: factor and cluster analysis. *International Journal of Epidemiology* 1987; **16**: 501–508.

5 Negri E, La Vecchia C, Franceschi S, Levi F. Patterns of mortality from major cancers in Europe. *Cancer Epidemiology Biomarkers and Prevention* 1994; **3**: 531–36.

6 Parkin DM, Läärä E, Muir CS. Estimates of the world-wide frequency of sixteen major cancers in 1980. *International Journal of Cancer* 1988; **41**: 184–97.

7 Coleman MP, Estève J, Damiecki P, Arslan A, Renard H. *Trends in Cancer Incidence and Mortality*. IARC Scientific Publication no. 121. Lyon: International Agency for Research on Cancer, 1993.

8 Kurihara M, Aoki K, Hisimachi S (eds). *Cancer Mortality Statistics in the World 1950–1985*. Nagoya: University of Nagoya Press, 1989.

9 World Health Organization. *International Classification of Disease: 6th Revision*. Geneva: WHO, 1950.

10 World Health Organization. *International Classification of Disease: 7th Revision*. Geneva: WHO, 1957.

11 World Health Organization. *International Classification of Disease: 8th Revision*. Geneva: WHO, 1967.

12 World Health Organization. *International Classification of Disease: 9th Revision*. Geneva: WHO, 1977.

13 Gabriel KG. The biplot graphical display of matrices with applications to principal component analysis. *Biometrika* 1971; **58**: 453–67.

14 Gabriel KG, Odoroff CL. Biplots in medical research. *Statistics in Medicine* 1990; **9**: 469–85.

15 Gower JC, Hand DJ. *Biplots*. London: Chapman and Hall, 1996.

16 Osmond C. Biplot models applied to cancer mortality rates. *Applied Statistics* 1985; **34**: 63–70.

17 Kroonenberg PM. *Introduction to Biplots for $G \times E$ Tables*. Research Report no. 51. Brisbane: Centre of Statistics, University of Queensland, Australia, 1995 (also available as downloadable document via: www.fsw.leidenuniv.nl/~kroonenb).

18 Van Eeuwijk FA, Kroonenberg PM. Multiplicative decompositions of interactions in three-way ANOVA, with applications to plant breeding. *Biometrics* 1998; **54**: 1315–33.

19 Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966; **31**: 279–311.

20 Kroonenberg PM. *Three-mode principal analysis: theory and applications*. Leiden: DSWO Press, 1983.

21 Kiers HAL. Hierarchical relations among three-way methods. *Psychometrika* 1991; **56**: 449–70.

22 Kroonenberg PM. The TUCKALS line. A suite of programs for three-way analysis. *Computational and Statistical Data Analysis* 1994; **18**: 73–96.

23 Waterhouse JAH, Muir CS, Shanmugaratnam K, Powell J. *Cancer incidence in five continents*, Vol. IV. IARC Scientific Publication no. 42. Lyon: International Agency for Research on Cancer, 1982.

24 Muir CS, Waterhouse JAH, Mack T, Powell J. *Cancer incidence in five continents*. Vol. V. IARC Scientific Publication no. 88. Lyon: International Agency for Research on Cancer, 1987.
25 Parkin DM, Muir CS, Whelan SL, Gao YT, Ferlay J, Powell J. *Cancer incidence in five continents*. Vol. VI. IARC Scientific Publication no. 120. Lyon: International Agency for Research on Cancer, 1992.
26 Parkin DM, Whelan SL, Ferlay J, Raymond L, Young J. *Cancer incidence in five continents*. Vol. VII. IARC Scientific Publication no. 143. Lyon: International Agency for Research on Cancer, 1997.
27 Tufte ER. *Visual explanations: images and quantities, evidence and narrative*. Cheshire, CT: Graphics Press, 1997.
28 Ten Berge JMF, De Leeuw J, Kroonenberg PM. Some additional results on principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* 1987; **52**: 183–91.
29 Heiser WJ, Kroonenberg PM. *Triadic algorithms for the PARAFAC model with constraints*. Internal report. Leiden: Department of Psychometrics and Research Methods, Leiden University, 1997.
30 Bro R. *Multi-way analysis in the food industry*. Unpublished doctoral thesis, Amsterdam: University of Amsterdam, 1998 (available as downloadable document from: www.models.kvl.dk/users/rasmus/thesis/thesis.html).

Appendix

Technical notes

Data representations

The cancer mortality data can be conceived in two ways. One is as a three-factorial design with site (j), year (k) and country (i) as factors with mortality rate as dependent variable. This implies that the mortality rates can be modelled as

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} \tag{A1}$$

The three-way interaction term has to serve as error term as there is only one observation per cell. By modelling the three-way interaction as 'three-way model + error', a smaller error term can be obtained. This is worked out in more detail by Van Eeuwijk and Kroonenberg.[18]

The second way is to conceive the data as a multivariate two-factorial design with mortality rates per site as dependent variables and country and year as factors. The model is then

$$x^j_{ik} = \mu^j + \alpha^j_i + \gamma^j_k + \alpha\gamma^j_{ik} \tag{A2}$$

This formulation implies that the mortality rates across sites are not really comparable and should be treated as separate variables.

Modelling two-way interactions

The second approach was taken when looking at the site by country interaction. Adding over time (k) gave

$$\bar{x}^j_i = \mu^j + \alpha^j_i \tag{A3}$$

To eliminate the effect of the different variability of the mortality rates per site, (A3) was divided by sj, so that the biplot in Figures 2 and 3 represent standard scores for each site. The standardized deviations from the site means were decomposed with the singular value decomposition of which we have used an approximation with the first three components of the decomposition

$$\tilde{x}_{ij} = \frac{\bar{x}_i^j}{S_j} = \sum_{p=1}^{3} g_{pp} a_{ip} b_{jp} + e_{ij} \tag{A4}$$

where $\mathbf{a}_p$ are the co-ordinates of the countries and $g_{pp}\mathbf{b}_p$ those of the sites and e_{ij} the errors of approximation. In this way, the two-way means are split into a systematic part and an error part. It can be shown that the choice of standardization causes the angles of the vectors from the origin to the site points to be the correlations between sites in the full space. The inner product between a site vector and a country vector approximates the value of $\tilde{x}_{ij}$. The quality of the approximation depends on the overall degree of approximation of (A4) by the three components and the fit of a site and country combination. The second representation was chosen to show the relative mortality rates across all sites independent of the overall level of a site and the variability within a site.

Modelling the three-way interaction

To investigate the three-way interaction, the first representation was chosen, in other words the raw three-way interaction $\alpha\beta\gamma_{ijk}$ was modelled because here the interest centred around the remainder of the variability after all other effects had been taken care of. In investigating the three-way interaction, we are looking for outlying observations or patterns, and only those which are sizeable are of interest. The three-way interaction was modelled with Tucker's[19] three-mode component model, which is a generalization of the singular value decomposition used in (A4) to analyse the two-way interaction. Because there is an additional way, (A4) is expanded with a term for the third way, and it turns out that the three-way singular values have three indices (p, q and r) rather than one index (p).

$$\alpha\beta\gamma_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk} \tag{A5}$$

with a_{ip}, b_{jq}, c_{kr} the co-ordinates of country i, site j and year k on their pth, qth and rth component, respectively, and g_{pqr} the elements of the core array and the e_{ijk} are the errors of approximation. Again the interaction is split into a systematic part, modelled with the three-mode principal component model, and an error part. In the present case, $P = 2$, $Q = 2$, and $R = 1$, so that (A5) reduces to

$$\alpha\beta\gamma_{ijk} \approx c_k \sum_{p=1}^{2}\sum_{q=1}^{2} a_{ip} b_{jq} g_{pq} = c_k \sum_{p=1}^{2} a_{ip} b_{jp} g_{pp} = c_k d_{ij} \tag{A6}$$

where the first equality was proven by Ten Berge *et al.*[28] Equation (A6) shows that, in our special case, Figure 4 is an ordinary biplot (see (A4)). The c_k are the weights of the years with which the inner product d_{ij} of site j and country i has to be multiplied to approximate the three-way interaction terms $\alpha\beta\gamma_{ijk}$. As these c_k are increasing with increasing years (see Table 5), they cause the country–site inner products to increase over time if they are positive (e.g. lung cancer and Denmark) and to decrease over time if they are negative (e.g. Japan and stomach cancer), indicating an increase and decrease of the mortality rate, respectively.

Degrees of freedom

In Table 1 we have presented the three-way analysis of variance of the completed data and have ignored the fact that 810 data point were estimated via the model of (1). Detailed inspection of the mean structure showed that of the 731 disease by country marginal means that 41 were not available and that, of the 387 year-by-country means, nine means were not available. This means that the proper degrees of freedom for the disease-by-country interaction are not 672, but $672 - 41 = 631$, the degrees of freedom for the year-by-country interaction $336 - 9 = 327$ and the three-way interaction is $5376 - 50 = 5326$. For simplicity we have not included these values in Table 1.

Residuals

In this analysis there are two types of residuals to be considered: those arising from fitting the country-by-tumour site interaction (see (A4)) and those arising from fitting the three-way interaction (see (A5)). Ideally, one should have a comprehensive loss function for the entire problem, but techniques for this are still under development.[29,30] The residuals for the two-way interaction had a skew of 0.52 (standard error = 0.09), and the distribution of residuals is somewhat leptokurtic (kurtosis = 1.61 with standard error = 0.81). However, the distribution and the plot of the standardized residuals versus the predicted values did not suggest that logarithmic transformations of the mortality rates were necessary. The residuals from the three-way interaction after fitting the three-mode component model had a lower skewness of 0.16 (standard error = 0.030), but the distribution was extremely leptokurtic with a kurtosis of 10.58 (standard error = 0.60). Of the 6579 residuals, there are 129 standardized residuals greater in absolute value than 3 of which six have a standardized residual ≤ -5.0 and $12 \geq 5.0$, but the vast majority of the standardized residuals are located around 0.0. The missing data substitition is partly to blame for this.