Epidemiologic Perspectives & Innovations

BioMed Central

Open Access

Trend tests for the evaluation of exposure-response relationships in epidemiological exposure studies

Ludwig A Hothorn*[1], Michael Vaeth[2] and Torsten Hothorn[3]

Address: [1]Institute of Biostatistics, Leibniz University Hannover, D-30419, Hannover, Germany, [2]Department of Biostatistics, University of Aarhus, Vennelyst Boulevard 6, DK-8000, Aarhus C, Denmark and [3]Institut fuer Statistik, Ludwig Maximilians Universitaet Muenchen, Ludwigstrasse 33, D-80539, Munich, Germany

Email: Ludwig A Hothorn* - hothorn@biostat.uni-hannover.de; Michael Vaeth - vaeth@biostat.au.dk;
Torsten Hothorn - Torsten.Hothorn@stat.uni-muenchen.de

* Corresponding author

Abstract

One possibility for the statistical evaluation of trends in epidemiological exposure studies is the use of a trend test for data organized in a 2 × k contingency table. Commonly, the exposure data are naturally grouped or continuous exposure data are appropriately categorized. The trend test should be sensitive to any shape of the exposure-response relationship. Commonly, a global trend test only determines whether there is a trend or not. Once a trend is seen it is important to identify the likely shape of the exposure-response relationship. This paper introduces a best contrast approach and an alternative approach based on order-restricted information criteria for the model selection of a particular exposure-response relationship. For the simple change point alternative H_1 : $\pi_1 = ... = \pi_q < \pi_{q+1} = ... = \pi_k$ an appropriate approach for the identification of a global trend as well as for the most likely shape of that exposure-response relationship is characterized by simulation and demonstrated for real data examples. Power and simultaneous confidence intervals can be estimated as well. If the conditions are fulfilled to transform the exposure-response data into a 2 × k table, a simple approach for identification of a global trend and its elementary shape is available for epidemiologists.

Introduction

Statistical trend analysis is an important component of epidemiological exposure studies. Here, "trend" simply means the demonstration of any monotone relationship between the response rate and the continuous exposure. For example, the association between all major types of childhood cancer and exposure to magnetic fields from high voltage installations was analyzed by Lausen et al. [1] using the data shown in Table 1, where the original continuous exposure data (Olsen et al., [2]) were categorized.

Although this example is seriously unbalanced, real epidemiological exposure studies with many unexposed or low-exposure cases but few high-exposure cases can be found. The appropriate evaluation of such epidemiological exposure studies is a statistical challenge. Many similar examples can be found in the literature, e.g. a case-control study for respiratory cancer possibly caused by long-term exposure to coke oven emissions [3].

In exposure studies, an unexposed group, E_1, is commonly compared with several exposure groups, $E_2,..., E_k$. The outcome of the study is the number of cases suffering from the disease being investigated, such as a specific tumor, and the number of observations without the disease (controls), i.e. the risk of disease in each category of

Table 1: Child cancer and magnetic fields

Exposure/μ Tesla	j	n_{cancer}	$n_{no\ cancer}$	n_j	p_j	RR_{j1}
0–0.05	1	1698	4759	6457	0.263	-
0.051–0.101	2	0	9	9	0	0.000
0.101-0.15	3	2	3	5	0.4	1.525
0.151-.20	4	1	3	4	0.25	0.953
0.201-0.25	5	1	3	4	0.25	0.953
0.251-0.30	6	0	4	4	0	0.000
0.301-0.35	7	0	2	2	0	0.000
0.351-0.85	8	1	0	2	0.5	1.906
0.851-1.6	9	2	0	2	1	3.812
>1.61	10	2	0	2	1	3.812

(p_j ... estimated proportion, RR_{j1} ... relative risk to unexposed)

exposure. One important objective in exposure epidemiology is causation; the demonstration of a global exposure-response relationship represents one of the causation criteria, according to Hill [4]. A global trend test leads to identification of a trend, whereas model selection allows inference of the likelihood of a particular elementary model.

The sampling strategy of epidemiological exposure studies is either a cohort study, in which a $2 \times k$ contingency table represents the data, or a case-control study, in which two multinomial distributions are compared. However, the likelihood ratio test of identical multinomials against the elementary odds ratios alternative, for a sufficient total number of observations, is equivalent to the comparison of the k independent binomial proportions against a simple ordered alternative (Agresti and Coull, [5]; Hothorn et al., [6]). Therefore, it is appropriate to evaluate both designs by means of an asymptotic trend test for a $2 \times k$ contingency table.

Numerous methods, including model-based (e.g. Royston et al., [7]) and test-based approaches (e.g. Dosemeci and Benichou, [8]), are used to analyze exposure-response relationships. A basic problem is that the shape of the exposure-response is unknown a priori and is an outcome of the study. However, the choice of model or test greatly depends on the shape of the exposure-response. Therefore, a broad class of models or tests should be used, but that, in turn, leads to a model selection dilemma. Model selection is an intricate component of statistical problems. Model selection in this case is not the objective, but is only a tool for identifying the correct trend from several possible elementary alternatives. An alternative hypothesis can be decomposed into its underlying elementary alternatives, e.g. the simple order alternative $H_1: \pi_1 \leq \pi_2 \leq \pi_3$ can be decomposed into the three elementary hypotheses H_1^1

$: \pi_1 = \pi_2 < \pi_3,\ H_1^2 : \pi_1 < \pi_2 = \pi_3,\ H_1^3 : \pi_1 < \pi_2 < \pi_3.$

The *p*-value, a commonly used outcome of a trend test, is frequently insufficient for epidemiological studies. Information concerning the shape of the exposure-response and/or a measure of the magnitude of the effect, such as relative risks or odds ratios, is desirable for a significant trend. Thus, the level of the false positive decision rate (α) should be controlled. In addition, an approach with a minimum false negative decision rate (β) (respective maximum power $\pi = 1 - \beta$) for the global test decision and a maximum correct decision rate for the selected model should be identified. The correct classification rate, the proportion of correctly identified elementary alternatives, is used as a major performance measure later on.

The exposure in case-control studies is frequently measured on a continuous scale. Categorization at pre-selected cut-off points of a small number of ordered categories is common; for example, four categories of trihalomethane exposure (Jones et al., [9]), or three categories of lifetime dose of hair dye (Benavente et al., [10]). Inappropriately chosen cut-off points dramatically reduce the power of the trend test (Greenland, [11]). Some exposures are naturally grouped, for example 2–3 cups of coffee per day, by the impreciseness of the definitions, such as "cup" and "coffee" (Ascherio et al., [12]). An example of ordinal definition of the exposure is given in a case-control study of Norwegian nickel refinery workers (Grimsrud et al., [13]). The exposure-related associations between smoking-adjusted lung cancer rates and cumulative exposure to different forms of nickel used the categories "low," "medium," and "high."

The best approach, in terms of both power and interpretation, occurs when a single cut-off point exists and is known a priori, resulting in a two-sample test "above" vs. "below" the cut-off point. This is because an odds ratio and its one-sided confidence interval can be estimated. The trend test approach discussed here is designed for naturally grouped exposure with a single change point. For continuous exposure models a continuous covariate can be used. However, the choice of an appropriate model – such as linear, logistic, or other – remains open and model selection influences the inference.

In this paper, a trend test for the comparison of k ordered binomial proportions using a change point alternative is presented. Either a single change point is directly of interest or the change point alternative is pivotal, i.e. many other elementary monotone alternatives can be generated from it. The concept of multiple contrasts is used because of the simplicity and the availability of the distribution under the alternative. After a significant trend test, information is provided that determines which contrast was the "best," and therefore, which exposure-response shape describes the data most accurately. Alternatively, an information criterion-based approach for the likelihood ratio

test under monotone order-restriction according to Anraku [14] is examined.

Therefore, the primary objective of this paper is not just describing the exposure-response relationship but also identifying the most likely elementary exposure-response model with a control of the false model classification rate.

Analysis
Global tests on exposure-response relationships
The number of diseased and healthy persons for each exposure group, Ej, are organized in the following $2 \times k$ table, where Index 1 denotes the group without exposure.

The estimator for the proportions per exposure group is $p_j = n_{j1}/n_{j.}$ $j = 1,..., k$, the total is $p = n_{.1}/n_{..}$, and the expected values for the proportions are denoted as π_j. The hypotheses system for a monotone order is:

H_0: $\pi_1 = \pi_2 = ... = \pi_k$ against

H_1: $\pi_1 \le \pi_2 \le ... \le \pi_k$ with at least one strict inequality.

For simplicity, assume increasing effects with increased exposure; analogously, a directional decision for a decrease is possible.

There are an extensive number of publications concerning order-restricted tests, including the analysis of $2 \times k$ contingency tables (e.g. Agresti and Coull, [5]; Leuraud and Benichou, [15]). However, no uniformly powerful trend test exists for all possible alternative shapes. The possible shapes can be seen as different equality-inequality patterns of H_1. This can be seen for an extreme convex shape $\{0, 0, 0, \pi\}$. Clearly, the "Helmert's contrast" is most powerful because of the optimal pooling of all the lower exposures and the comparison with the high exposure: $p_4 - (p_1 + p_2 + p_3)/3$. However, power for Helmert's contrast is greatly reduced for the extreme concave shape $\{0, \pi, \pi, \pi\}$. The shape of the exposure-response relationship is unknown a priori. Irrespective of numerous recent alternative proposals, the likelihood ratio test represents an appropriate solution for this situation. This test is numerically complicated, particularly concerning its distribution

Table 2: Principle of 2 by *k* tables for epidemiological exposure studies

	E_1	....	E_k	Total
Disease	n_{11}	...	n_{k1}	$n_{.1}$
No disease	n_{10}	...	n_{k0}	$n_{.0}$
Sample size	$n_{1.}$	...	$n_{k.}$	$n_{..}$

under the alternative, which is needed for power/sample size calculations (Robertson et al., [16]). The multiple contrast test according to Bretz and Hothorn [17] approximates its power and is simpler. There are 2^{k-1} different shapes for k exposure groups, and for each shape a contrast with a minimum false negative rate (β) can be defined. The idea is to select the best contrast, which is sensitive for a certain shape. The best contrast is simply tested by a maximum test. Because the proportions p_j are asymptotically normally distributed, their linear combination (denoted as contrast) $\sum_{j=1}^{k} c_j p_j$ is also normally distributed, and therefore, the single contrast test statistic

$$t_{SingleC} = \frac{\sum_{j=1}^{k} c_j p_j}{\sqrt{p(1-p) \sum_{j=1}^{k} c_j^2 / n_{j.}}}$$ is asymptotically normally

distributed, where $\sum_j c_j = 0$ guarantees a level α test under the null hypothesis. Different variance estimators can be used, but to keep the problem simple, the commonly used pooled estimator p is used here. Asymptotic test versions are used throughout. The contrast coefficients, c_j, are specific for each contrast test; for example the Helmert's contrast [$c_i = -1$; $j = 1,..., k - 1$ and $c_k = k$]. A multiple contrast test is the maximum of s pre-defined single contrast tests $t_{MultipleC} = \max_{i \in (1,...,s)} (t_{SingleC}(\mathbf{c}_i))$, $i = 1,..., s$ where $\mathbf{c}_i = (c_{i1},..., c_{ik})$ is a k vector of contrasts. Under the null hypotheses, the joint distribution of the linear contrast tests $t_{SingleC}(\mathbf{c}_i)$ $i = 1,..., s$ is an s-variate normal distribution with a zero vector of means and a non-product-moment correlation matrix. The correlation between two arbitrary contrasts, $\mathbf{a} = (a_1,..., a_k)$ and $\mathbf{b} = (b_1,..., b_k)$, is

$$\rho_{\mathbf{a},\mathbf{b}} = \frac{\sum_{j=1}^{k} a_j b_j \frac{p_j(1-p_j)}{n_{j.}}}{\sqrt{\left(\sum_{j=1}^{k} a_j^2 \frac{p_j(1-p_j)}{n_{j.}} \right) \left(\sum_{j=1}^{k} b_j^2 \frac{p_j(1-p_j)}{n_{j.}} \right)}} .$$

This so-called isotonic contrast approach, based on s = 7 contrasts, for the balanced design with four exposure groups is demonstrated in Table 3.

However, the correct classification rates for the most likely elementary alternative (shape of the exposure-response) were found to be unsatisfactory for isotonic contrasts (Hothorn et al., [6]). Therefore, a special case of order-

restricted inference is considered for step shapes only and denoted as a change point alternative (Hirotsu and Marumo, [18]). Two situations should be considered: i) threshold level studies assuming that an exposure-response reveals a single change point, which can be characterized by a lower part, an upper part, and an abrupt change between both; and ii) exposure-response studies with continuous exposure data where the change point alternative is a special and substantial component of the all-pattern alternative, which can simplify the evaluation. In some epidemiological problems this question arises. An example of a threshold level study is a diabetes study (Pastor-Barriuso et al., [19]) with the relationship between 2-hour plasma glucose and mortality, where the following questions were formulated: i) Does a certain glucose level exists that markedly increases the mortality risk? ii) Can this change point be estimated? Proposals in the literature are directed only at proof of the existence of such a change point. However, epidemiologists not only want to know that such a change exists, but also where this change is located. Here it is demonstrated that the estimation of the change point q is characterized by its correct classification rate by means of multiple contrast tests, that is, in a testing framework. The hypotheses system for a change from q to $q+1$ is:

$$H_0: \pi_1 = \pi_2 = ... = \pi_k$$

$$H_1: \pi_1 = ... = \pi_q < \pi_{q+1} = ... \pi_k \quad q \in (1,..., k-1)$$

The above hypotheses system can be tested by multiple step contrasts. Exactly (k-1) step contrasts are appropriate for testing the above hypothesis:

$$
\begin{pmatrix}
-k, & 1, & 1, & ... & 1) \\
-(k-1), & -(k-1), & 2, & .... & 2) \\
... & ... & ... & ... & ... \\
-1, & -1, & -1, & ... & k)
\end{pmatrix}
$$

Exactly three possible change points, q, exist for the simple design with one unexposed and three exposure groups. Exactly one contrast is power-optimal for the balanced design of each change point:

q	c_1	c_2	c_3	c_4
1	(−3	1	1	1)
2	(−2	−2	2	2)
3	(−1	−1	−1	3)

"Power-optimal" simply means the maximum test statistics because the $t^i_{SingleC}$ is normally distributed, and therefore, standardized. The $t_{MultipleC}$ is q-variate normally distributed. The contrast coefficients, **c**, for q contrasts are defined for the general unbalanced design (Hirotsu et al., [20]):

$$
c_{qj} = \begin{cases}
-n_{j.} / \sum_{l=1}^{j} n_{l.} & if \quad j = 1,...,q \\
n_{j.} / \sum_{l=j}^{k} n_{l.} & if \quad j = q+1,...,k
\end{cases}
$$

These step contrasts reveal a nice ability to transform the k-sample problem into an unbalanced two-sample problem, which can be used later for estimation of the unadjusted relative risk (or odds ratio) "above/below" the change point. Moreover, the step contrasts belong to a broader class of multiple contrasts. Isotonic contrasts approximate the power of the likelihood ratio test for the monotone ordered hypothesis. The bivariate up/down proposals (Neuhaeuser and Hothorn, [21]; Stewart and Ruberg, [22]) only use the two extreme contrasts (Table 3). Therefore, the change point alternative represents a compromise for testing trends. It is much less dependent on the power of the shape compared with the frequently used single linear contrast test, although only k instead of 2^k - 1 isotonic contrasts were used. The multiple contrast test (above) is defined for differences of proportions, but can be re-formulated for the relative risk, commonly used in epidemiology (see Appendix A).

It seems that a multiple contrast test may be a different approach to the commonly used logistic model. However, a strong relationship between the multiple contrast test and the score test in a logistic model exists, which allows the correction for additional confounders (Hothorn et al., [6]).

Table 3: Contrast coefficients for the balanced design with four exposures groups

Type of contrasts	No. of contrasts	Alternative	Contrast c_j
Isotonic	2^k-1	$\pi_1 < \pi_2 = \pi_3 = \pi_4$	{-3 1 1 1}
		$\pi_1 = \pi_2 < \pi_3 = \pi_4$	{-1 -1 1 1}
		$\pi_1 = \pi_2 = \pi_3 < \pi_4$	{-1 -1 -1 3}
		$\pi_1 < \pi_2 < \pi_3 < \pi_4$	{-3 -1 1 3}
		$\pi_1 = \pi_2 < \pi_3 < \pi_4$	{-1 -1 0 2}
		$\pi_1 < \pi_2 = \pi_3 < \pi_4$	{-1 0 0 1}
		$\pi_1 < \pi_2 < \pi_3 = \pi_4$	{-2 0 1 1}
Change point	k-1	$\pi_1 < \pi_2 = \pi_3 = \pi_4$	{-3 1 1 1}
		$\pi_1 = \pi_2 < \pi_3 = \pi_4$	{-1 -1 1 1}
		$\pi_1 = \pi_2 = \pi_3 < \pi_4$	{-1 -1 -1 3}
Up/down	2	$\pi_1 < \pi_2 = \pi_3 = \pi_4$	{-3 1 1 1}
		$\pi_1 = \pi_2 = \pi_3 < \pi_4$	{-1 -1 -1 3}
Single (linear)	1	$\pi_1 < \pi_2 < \pi_3 < \pi_4$	{-3 -1 1 3}

Identification of the exposure-response shape
The trend tests distinguish only globally between the null hypothesis and alternative hypotheses, based on the

asymptotic distribution of the test statistics under the null hypothesis. That is, either a trend exists or it does not. However, the alternative hypothesis is not unique. For example, the following three hypotheses are possible for the change point alternative for a design with one unexposed and three exposure groups:

$$H_1^1 : \pi_1 < \pi_2 = \pi_3 = \pi_4, \ H_1^2 : \pi_1 = \pi_2 < \pi_3 = \pi_4, \ H_1^3 : \pi_1 = \pi_2 = \pi_3 < \pi_4$$

However, the global trend tests provide no answer as to which particular alternative exists. Two different approaches can be used to answer this question: i) the best contrast approach; and ii) a model selection approach based on the information criterion for order restriction. This paper explores the identification of one of the possible k - 1 elementary alternatives; that is, a classification into $H_1^1, ..., H_1^{k-1}$. Consequently, the correct classification rate, or the proportion of correctly identified elementary alternatives, is used as a performance measure later on.

The global test decision for the multiple contrast approach is based on the maximum of all included single contrasts $t_{MultipleC} = \max_{i \in (1,...,s)} (t_{SingleC}(\mathbf{c}_i))$, $i = 1,..., s$, where each single contrast is power optimal for a particular type of alternative (Table 3). Therefore, this maximum contrast approach can be used as an estimator for the exposure-response shape, where the classification is performed after a significant trend test for control α. For example, two alternatives are possible for a design with three exposure groups: $H_1^1 \ \pi_1 = \pi_2 < \pi_3$ or $H_1^2 : \pi_1 < \pi_2 = \pi_3$. Assume that the number of diseased cases, $n_{11}, ..., n_{k1}$, is drawn from k binomial random variables with parameters π_j and $n_{j.}$. A possible exposure-response is described by a contrast vector, $\mathbf{c} = (c_1, ..., c_k)$. The problem is to estimate the underlying exposure-response relationship when s contrast vectors are given. A simple estimator is the function Ψ : $(n_{11}, ..., n_{k1}) \rightarrow \{1, ..., s\}$ which can be derived from the associated contrast test, i.e. $\psi_1 = \psi(n_{11}, ..., n_{k1}) = \arg\max_{i \in (1,...,s)} |t_{SingleC}(\mathbf{c}_i)|$. Then explore variability of the simple estimator, Ψ_1. How likely is each of the s possible values under the observed data? This question can be addressed via the parametric bootstrap. Repeated realizations from k binomial distributions with sample sizes $n_{j.}$ and the estimated success parameter $p_j = n_{j1}/n_{j.}$ for j = 1,..., k are drawn.

- Draw B bootstrap samples $n_{11}^{*b}, ..., n_{k1}^{*b}, \ n_{j1}^{*b} \sim B(n_{j.}, p_j), \ b = 1, ..., B$

- Compute $\psi_1^b = \psi_1(n_{11}^{*b}, ..., n_{k1}^{*b})$

- Compute the relative frequency of each possible value from 1,..., s

This is a measure for the variance of the estimator. Under special circumstances, an improved estimator can be computed by a majority voting according to Breiman [23] over $\psi_1^b : \psi_2 = \arg\max_{i \in \{1,...,s\}} \sum_b I(\psi_1^b = i)$, where I denotes the indicator function. This approach is designated the "parametric bootstrap best contrast" approach.

The model selection approach, based on the information criterion for order-restriction of normally distributed variables according to Anraku [14], can be modified for proportions and the change point alternative. The AIC criterion for the unrestricted maximum likelihood estimator $\hat{\theta} : AIC(\hat{\theta}) = l(\hat{\theta}) - p$: (with $l(\hat{\theta})$ = log-likelihood, p = dimension of θ) was modified for order-restricted maximum likelihood estimators: $ORIC(\tilde{\theta}) = l(\tilde{\theta}) - penalty(k, n_i)$. The penalty term is calculated for each model using the level of probabilities under an order-restriction. The explicit formulas for a design with three exposure groups, such as the null-model M_0 and the two change point models M_1 and M_2, are given in Appendix B. The ORIC-approach represents a model estimation approach, where model $M_0\{H_0: \pi_1 = \pi_2 = \pi_3\}$, model $M_1 \{ H_{M_1} : \pi_1 = \pi_2 < \pi_3\}$, or model $M_2 \{ H_{M_2} : \pi_1 = \pi_2 < \pi_3\}$ will be estimated as a "best fitted" model.

Simulation study
The simulation study is structured in two parts: i) empirical comparison between the best-contrast approach and the ORIC approach for a design with three groups; and ii) investigation of the best contrast approach for more general designs. Fifty thousand pseudo-random 2 × k tables (k ranging from 3 to 7) were generated and 10,000 bootstrap samples were drawn. Two criteria are used, the correct classification rate – the empirical decision rate for the correct model – and the power.

Part I
The correct classification rates for the ORIC approach, ORIC (M_0, M_1, M_2), and the parametric bootstrap best

contrast approach, Max(H^1, H^2), were compared for a design with three exposure groups (in Table 4) for the change point alternatives with different unexposed rates, π_1. From the first row in Table 4, where no differences between the proportions were investigated, the main difference between both approaches becomes clear. The ORIC approach, as an estimation approach, did not control for α. Only in 76% of the cases, not 95%, was M_0 selected under the null hypothesis. On the other hand, the best contrast test approach does control for α. Both approaches reveal high correct classification rates, greater than 90%, as long as the power is sufficient: either small unexposed rates, π_1, or large non-centrality parameters Δ (Table I in Appendix C (available as additional file 1) and larger sample sizes in Table II in Appendix C). This behavior is similar to the power of trend tests of proportions (Bretz and Hothorn, [17]). Due to the fact that the correct classification rates of the best contrast approach are similar or superior to those of the ORIC approach with decreasing π_1, increasing Δ, and n_j, the best contrast approach is recommended because of its simplicity and generalizability for use within the generalized linear model.

Part II
For one selected change point alternative { π_1, π_1, π_1, π_1, π_1 + Δ } the best contrast approach was investigated for the different dimensions k, different unexposed rates π_1, and several non-centrality parameters Δ, shown in Table 5. With an increasing number of exposure groups, a slight decrease of the correct classification rate occurs where the power is slightly increasing. With a decreasing sample size, a slight decrease of the correct classification rate occurs where the power is substantially decreasing. The well-known decrease of sensitivity with an increasing unexposed rate from 2 × 2 table analysis holds true for power and, less markedly, for the correct classification

rate. The effect size (non-centrality Δ) has much less impact on the correct classification rate compared with its well-known impact on power.

Table 6 demonstrates the decreasing correct classification rate for change points $q \ll k$. More important, from an epidemiological point of view, are the asymmetrical cumulative false classification rates. False classification is primarily from an overestimation instead of an underestimation of the true change point, that is, it is very unlikely to mistake a lower change point for the true one.

Extreme unbalanced exposure data
Particularly for environmental studies, much of the data is for unexposed and low-to-medium exposures; only rarely does data for high exposure exist. This is quite fortunate from an ethical point of view. However, this results in extremely unbalanced 2 × k tables and the statistical outcome depends on the rare, high-level exposure data. In a case-control study for respiratory cancer possibly caused by long-term exposure to coke oven emissions, the sample size was 10,198 in the unexposed group, but only 487 were in the highest exposure group (Costantino et al., [3]). A more extreme example was the study evaluating the connection between childhood cancer and magnetic fields from high voltage installations. The sample size was 2 in the highest exposure group, but 6,457 in the unexposed group (Table 1). The power decreases greatly for extremely unbalanced designs and accordingly the correct classification rate also decreases. If the total sample size is increased to achieve the same power, then the correct classification would be of the same magnitude as the balanced case, see Table 7. The identification of a trend in such a highly unbalanced design is complicated. A significant trend may depend on only these few cases, and the size and power of unbalanced designs differ greatly from those in balanced designs. In unbalanced designs with smaller change points, the correct classification rate increases if the resulting two-sample test is less unbalanced (as a result of the related step contrast). A change point at a high exposure that is based on rare data is very vague, however it becomes more stable when medium-to-high exposure from additional data are obtained.

Unbalanced designs, where the smallest sample size occurs in the informative groups (large change point s), reveal a clearly reduced classification rate. However, that decrease, compared with the balanced design, is much weaker than the related power loss. A further reduction occurs for the "in-between" change points as long as the sample size of the pooled informative groups is still smaller than the lower exposure groups. A further substantial increase of the sample size for the unexposed group had almost no influence on the classification rate.

Table 4: Correct classification rates for several spontaneous rates π_0

π_j	True Change q	ORIC(M_0, M_1, M_2)			Max(H^1, H^2)	
		M_0	M_1	M_2	H^1	H^2
0.3/0.3/0.3	*0*	*.758*	*.112*	*.129*	*.514*	*.486*
0.1/0.1/0.3	2	.001	**.979**	.021	**.987**	.004
0.1/0.3/0.3	1	.001	.020	**.980**	.030	**.961**
0.2/0.2/0.4	2	.002	**.958**	.041	**.936**	.023
0.2/0.4/0.4	1	.005	.029	**.967**	.040	**.926**
0.3/0.3/0.5	2	.006	**.940**	.054	**.906**	.034
0.3/0.5/0.5	1	.004	.053	**.943**	.044	**.882**
0.4/0.4/0.6	2	.009	**.940**	.052	**.887**	.036
0.4/0.6/0.6	1	.009	.053	**.940**	.039	**.885**

(n_j = 100, $H_1^1 : \pi_0 = \pi_1 < \pi_2$, $H_1^2 : \pi_0 < \pi_1 = \pi_2$) *(bold indicate correct classification)*

Table 5: Correct classification rates and power for several dimensions, sample sizes, unexposed rates, and non-centralities

Dimension	k			3	4	5	6	7
	Correct classif. rate			.992	.987	.977	.971	.971
	Power			*.828*	*.845*	*.861*	*.899*	*.889*
Sample size	n_j			25	50	75	100	125
	Correct classif. rate			.809	.973	.978	.987	.989
	Power			*.393*	*.618*	*.742*	*.845*	*.903*
Unexpos. rate	Π_1			.01	.06	.11	.16	.20
	Correct classif. rate			.987	.903	.817	.767	.766
	Power			*.845*	*.488*	*.373*	*.312*	*.266*
Non-centrality	Δ			0.03	0.05	0.07	0.09	0.11
	Correct classif. rate			*.953*	*.973*	*.985*	*.994*	*.998*
	Power			*.479*	*.773*	*.904*	*.972*	*.991*

Since a sample size of $n_j = 1$ is possible, in principle, for this approach, the impact of the continuous exposure categorization can be demonstrated quantitatively with respect to power and classification rate. When a single change point exists, the best approach is the categorization below or above this change point. The true alternative is never known a priori when dealing with real data. Therefore, appropriate categorization may be helpful and inappropriate categorization can greatly reduce the sensitivity.

The asymptotic power for the change point alternative is available (Bretz and Hothorn, [17]). Based on an R-code, the power can be calculated for an arbitrary sample size pattern, which shapes the exposure response and dimensions k. Power estimation for unbalanced designs can be found in [6] whereas a serious power loss can be observed when the sample size in the informative high exposure groups is very small compared with the sample size in the unexposed or low exposure groups.

Evaluation of the example
The *p*-value for the global trend test (change point alternative) and the classification rate of the best contrast approach is determined using an implementation of the proposed procedures in R (R Development Core Team, [24]). The most likely change point, q, and simultaneous confidence intervals for the related change point contrasts

can be calculated for the $2 \times k$ contingency table data. A marginal confidence interval can be estimated for each elementary contrast because it represents a linear combination of the proportions p_j. Simultaneous confidence intervals for the maximum of several contrasts can be estimated using a multivariate normal distribution. A detailed description for the estimation of simultaneous confidence intervals for several multiple contrast tests can be found in [25] where the particular problems for binomial data were described recently [26]. The software is available as the R library *bindosres* as additional file 2. This file can be installed in the private R program via "Install packages from local zip files",

The magnet field cancer data in Table 8 revealed a change point $q = 8$ with a classification rate of 0.74 (*p*-value for a global trend = 0.002). The cumulative false classification of 0.26 is nearly concentrated on $q = 7$. The maximum simultaneous lower confidence limit is for the sub-set [10 vs. {1, 2, 3, 4, 5, 6, 7, 8, 9}] and seems to be medically relevant with 0.563, but differs only a little from that of subset {10, 9} vs. {1, 2, 3, 4, 5, 6, 7, 8} that is related to the change point. The analysis of the continuous data using maximally selected rank statistics gave a cut-point of 0.45 μTesla[1]. However, above this cut-point only six cancer cases with an exposure of 0.51, 0.73, 1.0, 1.59, 1.66, and 1.72, and two cases without cancer with exposures 0.73 and 0.83 μ Tesla were available. A careful interpretation is

Table 6: Asymmetrical cumulative false classification rates

Alternative	True Change	H[1]	H[2]	H[3]	H[4]	H[5]	Cum. over.	Cum. under.
.01/.01/.01/.01/.01/.07	5	.000	.000	.001	.027	**.972**	-	0.028
.01/.01/.01/.01/.07/.07	4	.000	.002	.012	**.847**	.139	0.139	0.014
.01/.01/.01/.07/.07/.07	3	.000	.011	**.819**	.119	.051	0.17	0.011
.01/.01/.07/.07/.07/.07	2	.004	**.809**	.117	.038	.032	0.187	0.004
.01/.07/.07/.07/.07/.07	1	**.711**	.135	.052	.050	.053	0.29	-

($n_j = 100$; *bold indicate correct classification*)

Table 7: Correct classification rates for extreme unbalanced designs

Sample sizes	N	Alternative	Power	Correct classif. rate
200/200/200/200	800	.05/.05/.05/.10	.682	**.935**
540/200/40/20	800	.05/.05/.05/.10	*.251*	**.758**
200/200/200/200	800	.05/.05/.10/.10	.792	**.831**
540/200/40/20	800	.05/.05/.10/.10	.425	**.687**
200/200/200/200	800	.05/.10/.10/.10	.603	**.783**
540/200/40/20	800	.05/.10/.10/.10	.755	**.854**
400/400/400/400	1600	.05/.05/.05/.10	.915	**.971**
1340/200/40/20	1600	.05/.05/.05/.10	.266	**.749**
400/400/400/400	1600	.05/.05/.10/.10	.968	**.916**
1340/200/40/20	1600	.05/.05/.10/.10	*.438*	**.667**
400/400/400/400	1600	.05/.10/.10/.10	.903	**.904**
1340/200/40/20	1600	.05/.10/.10/.10	.832	**.883**
9740/200/40/20	10000	.05/.05/.05/.10	*.252*	**.702**

recommended: i) the correct classification rate is not high, ii) a high change point was identified, iii) above the change point are only 4 of 6,491 cases, and iv) the spontaneous rate of 0.263 is rather high. More examples and their interpretation can be found in Hothorn et al., [6].

Conclusion

Trend tests for the analysis of $2 \times k$ tables using epidemiological exposure data are described to identify the change point alternatives. Not only is the identification of a trend of interest important, but also the information regarding the particular types of alternatives. The best contrast approach for the multiple contrast test is useful for identifying the type of alternative or the change point, whereas a parametric bootstrap is suitable for an assessment of the variability. Both the bootstrapped best contrast and the ORIC approach are appropriate for different dimensions, non-centralities, sample sizes, and the unexposed group rates (due to the asymmetry in binomial testing). The consequences of unbalanced designs – of a large number in the unexposed or low exposure groups and a small number in the high exposure groups – can be calculated

depending on the expected shape. Simultaneous confidence intervals for the change point alternative are also available.

Approaches that test a global trend in epidemiological exposure data and also provide information on the pattern of the exposure-response relationship are rare. The most competitive approach is the fractional polynomials model [7], which is a specific multivariable regression approach.

Most epidemiological studies are characterized not only by the primary exposure factor but also by several covariates, such as gender, age, occupational status, and competing risk characteristics. Therefore, the best contrast approach within the framework of the generalized linear model is recently available [27]. Using the related R library (multcomp), real data can be evaluated using the contrast option "Changepoint" [28].

The suitability of such a simple change point alternative in epidemiological exposure studies should be critically dis-

Table 8: Child cancer and magnetic fields

Exposure/μ Tesla	j	p_j	Pattern	Lower confidence limit
0–0.05	1	0.263	{10,9,8,7,6,5,4,3,2} vs.1	-.716
0.051–0.101	2	0	{10,9,8,7,6,5,4,3} vs.{1,2}	-.410
0.101-0.15	3	0.4	{10,9,8,7,6,5,4} vs.{1,2,3}	-.327
0.151-.20	4	0.25	{10,9,8,7,6,5} vs.{1,2,3,4}	-.246
0.201-0.25	5	0.25	{10,9,8,7,6} vs.{1,2,3,4,5}	-.139
0.251-0.30	6	0	{10,9,8,7} vs.{1,2,3,4,5,6}	.108
0.301-0.35	7	0	{10,9,8} vs.{1,2,3,4,5,6,7}	.343
0.351-0.85	8	0.5	{10,9} vs.{1,2,3,4,5,6,7,8}	.534
0.851-1.6	9	1	10 vs.{1,2,3,4,5,6,7,8,9}	.563
>1.61	10	1		

cussed and some real data examples tested. Clearly, such a change point test describes the exposure-response of the population only. Further investigations are required to demonstrate that this simple approach can be utilized to estimate the center of the individual-level change point distribution. Moreover, the above approach is not limited to change point alternatives: other trend alternatives, such as Williams-type trends [29], can be assumed as well.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
LAH adapted the multiple contrast tests on epidemiological case-control studies and performed part of the simulation study. MV selected, analyzed, and interpreted the epidemiological examples and designed the simulation study. TH developed the majority voting algorithm and wrote the R program.

Appendix A
Formulation contrast tests for the relative risk
The estimators for the relative risk (RR) of each exposure group versus unexposed ($j = 1$) are:
$RR_{j1} = \dfrac{n_{j1}/n_{j.}}{n_{11}/n_{1.}}$ $j = 2, ..., k$ The single contrast tests can be formulated for relative risks, for example for the reverse

Helmert's contrast: $t_{revHelmert} = \dfrac{-kp_1 + \sum\limits_{j=2}^{k} p_j}{\sqrt{p(1-p)(\sum\limits_{j=2}^{k} 1/n_{j.} + k^2/n_1)}}$

$$t_{revHelmert}^{RR} = \frac{-k\frac{n_{11}}{n_{1.}} + \sum\limits_{j=2}^{k}\frac{n_{j1}}{n_{j.}}}{\sqrt{p(1-p)(\sum\limits_{j=2}^{k} 1/n_{j.} + k^2/n_1)}} = \frac{-k + \sum\limits_{j=2}^{k}\frac{n_{j1}}{n_{j.}}\frac{n_{1.}}{n_{11}}}{\frac{n_{1.}}{n_{11}}\sqrt{p(1-p)(\sum\limits_{j=2}^{k} 1/n_{j.} + k^2/n_1)}} = \frac{-k + \sum\limits_{j=2}^{k} RR_{j1}}{\frac{n_{1.}}{n_{11}}\sqrt{p(1-p)(\sum\limits_{j=2}^{k} 1/n_j + k^2/n_1)}}$$

For general contrasts hold true

$$t_{SingleContrast}^{RR} = \frac{c_1 + \sum\limits_{j=2}^{k} c_j RR_{j1}}{\frac{n_{1.}}{n_{11}}\sqrt{p(1-p)(\sum\limits_{j=1}^{k} c_j^2/n_{j.})}}$$

Appendix B
The ORIC approach for three binomials and the change point alternative. The three models are:

$M_0\{H_0: \pi_1 = \pi_2 = \pi_3\}$, $M_1\{H_{M_1}: \pi_1 = \pi_2 < \pi_3\}$, $M_2\{H_{M_2}: \pi_1 < \pi_2 = \pi_3\}$. The likelihood is

$$L(\pi) = \frac{n_{1.}!}{n_{11}!(n_{1.}-n_{11})!}\pi_1^{n_{11}}(1-\pi_1)^{n_{1.}-n_{11}}\frac{n_{2.}!}{n_{21}!(n_{2.}-n_{21})!}\pi_2^{n_{21}}(1-\pi_2)^{n_{2.}-n_{21}}\frac{n_{3.}!}{n_{31}!(n_{3.}-n_{31})!}\pi_3^{n_{31}}(1-\pi_3)^{n_{3.}-n_{31}}$$

. With the expected values π_j and their crude estimators:
$p_1 = {n_{11}}/{n_{1.}}$, $p_2 = {n_{21}}/{n_{2.}}$, $p_3 = {n_{31}}/{n_{3.}}$

The $\tilde{p}_j$ are the maximum likelihood estimates under the

simple order restriction: $\tilde{p}_j = \min\limits_{t:t\geq j} \max\limits_{s:s\leq i} \dfrac{\sum\limits_{j=s}^{t} w_j p_j}{\sum\limits_{s=j}^{t} w_j}$. The like-

lihood for the null-model M_0 is:

$$L(\tilde{p}_{H_0}) = \prod\limits_{j=1}^{3} \frac{n_{j.}!}{n_{j1}!(n_{j.}-n_{j1})!}\tilde{p}_{H_0}^{n_{11}+n_{21}+n_{31}}(1-\tilde{p}_{H_0})^{n_{1.}+n_{2.}+n_{3.}-(n_{11}+n_{21}+n_{31})}$$

where $\tilde{p}_{H_0} = \dfrac{n_{11}+n_{21}+n_{31}}{n_{1.}+n_{2.}+n_{3.}} = \dfrac{w_1 p_1 + w_2 p_2 + w_3 p_3}{w_1+w_2+w_3}$ provided $w_j = n_j$

The likelihood for the model M_1 is:

$$L(\tilde{p}_{M_1}) = \prod\limits_{j=1}^{3} \frac{n_{j.}!}{n_{j1}!(n_{j.}-n_{j1})!}\tilde{p}_{(12)}^{n_{11}+n_{21}}(1-\tilde{p}_{(12)})^{n_{1.}+n_{2.}-(n_{11}+n_{21})}\tilde{p}_3^{n_{31}}(1-\tilde{p}_3)^{n_{3.}-n_{31}}$$

where $\tilde{p}_{(12)} = \dfrac{n_{11}+n_{21}}{n_{1.}+n_{2.}}$, for

$p_{(12)} < p_3 \Rightarrow \tilde{p}_{12} = p_{(12)}, \tilde{p}_3 = p_3$

and

$\tilde{p}_{(12)} \geq p_3 \Rightarrow \tilde{p}_{12} = \tilde{p}_3 = \dfrac{w_{(12)}p_{(12)} + w_3 p_3}{w_{(12)}+w_3} = \dfrac{n_{11}+n_{21}+n_{31}}{n_{1.}+n_{2.}+n_{3.}}$

The likelihood for the model M_2 is:

$$L(\tilde{p}_{M_2}) = \prod\limits_{j=1}^{3} \frac{n_{j.}!}{n_{j1}!(n_{j.}-n_{j1})!}\tilde{p}_{(23)}^{n_{21}+n_{31}}(1-\tilde{p}_{(23)})^{n_{2.}+n_{3.}-(n_{21}+n_{31})}\tilde{p}_1^{n_{11}}(1-\tilde{p}_1)^{n_{1.}-n_{11}}$$

where $\tilde{p}_{(23)} = \dfrac{n_{21}+n_{31}}{n_{2.}+n_{3.}}$ for

$p_1 < \tilde{p}_{(23)} \Rightarrow \tilde{p}_1 = p_1, \tilde{p}_{(23)} = \tilde{p}_{(23)}$

and

$p_1 \geq \tilde{p}_{(23)} \Rightarrow \tilde{p}_1 = \tilde{p}_{(23)} = \dfrac{w_1 p_1 + w_{(23)}\tilde{p}_{(23)}}{w_1+w_{(23)}} = \dfrac{n_{11}+n_{21}+n_{31}}{n_{1.}+n_{2.}+n_{3.}}$

The model-specific ORIC are: $ORIC(M_r) = \log L(\tilde{\pi}_M) - penalty(M_r)$.

Where the penalty terms are $penalty(M_r) = \sum iP\{i, j, w(M_r)\}$

With

$$w(M_0) = n_{1.} + n_{2.} + n_{3.}$$

$$w(M_1) = n_{1.} + n_{2.}, n_{3.}$$

$$w(M_2) = n_{1.}, n_{2} + n_{3.}$$

Because $\quad P\{1, 1, \quad w(M_0)\} \quad = \quad 1 \quad ORIC(M_0) \quad =$ $L(\, P\{1, 1, w(M_0)\} = 1 \ ORIC(M_0) = L(\tilde{p}_{H_0}) - 1 \,)$ - 1

Because
$P\{1, 2, w(M_1)\} = \frac{1}{2} \quad P\{2, 2, w(M_1)\} = \frac{1}{2} \quad ORIC(M_1) = L(\tilde{p}_{M_1}) - \frac{3}{2}$

Because
$P\{1, 2, w(M_2)\} = \frac{1}{2} \quad P\{2, 2, w(M_2)\} = \frac{1}{2} \quad ORIC(M_2) = L(\tilde{p}_{M_2}) - \frac{3}{2}$

Additional material

Additional file 1
Additional simulation results. Contains three tables with simulated correct classification results
Click here for file
[http://www.biomedcentral.com/content/supplementary/1742-5573-6-1-S1.pdf]

Additional file 2
R package bindosres. The R package bindosres which can be installed in R via "install packages from local zip files"
Click here for file
[http://www.biomedcentral.com/content/supplementary/1742-5573-6-1-S2.gz]

Acknowledgements

References

1. Lausen B, Lerche R, Schumacher M: **Maximally selected rank statistics for dose-response problems.** *Biometrical Journal* 2002, **44:**131-147.
2. Olsen JH, Nielsen A, Schulgen G: **Residence near high voltage facilities and risk of cancer in children.** *British Medical Journal* 1993, **307:**891-895.
3. Costantino JP, Redmond CK, Bearden A: **Occupationally related cancer risk among coke oven workers: 30 years of follow-up.** *Journal of Occupational and Environmental Medicine* 1995, **37:**597-604.
4. Hill AB: **Principles of medical statistics.** 9th edition. Oxford University Press, New York; 1971:309-323.
5. Agresti A, Coull BA: **The analysis of contingency tables under inequality constraints.** *Journal of Statistical Planning and Inference* 2002, **107:**45-73.
6. Hothorn LA, Vaeth M, Hothorn T: **Trend tests for the evaluation of dose-response relationships in epidemiological exposure studies.** *Research Report 2003–3* 2003 [http://www.biostat.uni-hannover.de/report/ReportAarhus2003.pdf]. Department of Biostatistics, University of Aarhus
7. Royston P, Ambler G, Sauerbrei W: **The use of fractional polynomials to model continuous risk variables in epidemiology.** *International Journal of Epidemiology* 1999, **28:**964-974.
8. Dosemeci M, Benichou J: **An alternative test for trend in exposure-response analysis.** *J Expo Anal Environ Epidemiol* 1998, **8(1):**9-15.
9. Jones AQ, Dewey CE, Dore K, Majowicz SE, McEwen SA, Waltner-Toews D: **Exposure assessment in investigations of waterborne illness: a quantitative estimate of measurement error.** *Epidemiologic Perspectives & Innovations* 2006, **3:**6.
10. Benavente Y, Garcia N, Domingo-Domenech E: **Regular use of hair dyes and risk of lymphoma in Spain.** *International Journal of Epidemiology* 2005, **34:**1118-1122.
11. Greenland S: **Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis.** *Epidemiology* 1995, **6:**450-454.
12. Ascherio A, Weisskopf MG, O'Reilly EJ: **Coffee consumption, gender, and Parkinson's disease mortality in the Cancer Prevention Study II cohort: The modifying effects of estrogen.** *American Journal of Epidemiology* 2004, **160:**977-84.
13. Grimsrud TK, Berge SR, Haldorsen T, Andersen A: **Exposure to different forms of nickel and risk of lung cancer.** *American Journal of Epidemiology* 2002, **156:**1123-1132.
14. Anraku K: **An information criterion for parameters under a simple order restriction.** *Biometrika* 1999, **86:**141-152.
15. Leuraud K, Benichou J: **Tests for monotonic trend from case-control data: Cochran-armitage-mantel trend test, isotonic regression and single and multiple contrast tests.** *Biometrical Journal* 2004, **46:**731-749.
16. Robertson T, Wright FT, Dykstra RL: **Order restricted statistical inference.** New York. Wiley; 1988.
17. Bretz F, Hothorn LA: **Detecting dose-response using contrasts: asymptotic power and sample size determination for binomial data.** *Statistics in Medicine* 2002, **21:**3325-3335.
18. Hirotsu C, Marumo K: **Changepoint analysis as a method for isotonic inference.** *Scandinavian Journal of Statistics* 2002, **29:**125-138.
19. Pastor-Barriuso R, Guallar E, Coresh J: **Transition models for change-point estimation in logistic regression.** *Statistics in Medicine* 2003, **22:**1141-1162.
20. Hirotsu C, Kuriki S, Hayter AJ: **Multiple comparison procedures based on the maximal component of the cumulative chi-square statistic.** *Biometrika* 1992, **79:**381-392.
21. Neuhaeuser M, Hothorn LA: **Trend tests for dichotomous endpoints with application in carcinogenicity studies.** *Drug Information Journal* 1997, **30:**463-469.
22. Stewart WH, Ruberg SJ: **Detecting dose response with contrasts.** *Statistics in Medicine* 2000, **19:**913-921.
23. Breiman L: **Bagging Predictors.** *Machine Learning* 1996, **24:**123-140.
24. R Development Core Team. R: *A language and environment for statistical computing* 2005 [http://www.R-project.org]. R Foundation for Statistical Computing, Vienna, Austria
25. Hothorn LA: **Multiple comparisons and multiple contrasts in randomized dose-response trials – confidence interval orient approaches.** *J Biopharm Stat* 2006, **16:**7111-731.
26. Schaarschmidt F, Sill M, Hothorn LA: **Approximate simultaneous confidence intervals for multiple contrasts of binomial proportions.** *Biometrical Journal* 2008, **50:**782-792.
27. Hothorn T, Bretz F, Westfall P: **Simultaneous inference in general parametric models.** *Biometrical Journal* 2008, **50:**346-363.
28. Bretz F, Hothorn T, Westfall P: **On multiple comparisons in R.** *R News* 2002, **3:**14-17.
29. Williams DA: **Comparison of several dose levels with a zero dose control.** *Biometrics* 1972, **28:**519-531.