
Theory and methods: Odd odds interactions introduced through dichotomisation of continuous outcomes

Author(s): Lutz P Breitling and Hermann Brenner

Source: *Journal of Epidemiology and Community Health* (1979-), Vol. 64, No. 4 (April 2010), pp. 300-303

Published by: BMJ

Stable URL: <https://www.jstor.org/stable/20721192>

Accessed: 01-06-2020 21:47 UTC

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/20721192?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

BMJ is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Epidemiology and Community Health* (1979-)

Odd odds interactions introduced through dichotomisation of continuous outcomes

Lutz P Breitling, Hermann Brenner

Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany

Correspondence to

Lutz P. Breitling, Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Bergheimer Str 20, D-69115 Heidelberg, Germany; l.breitling@dkfz-heidelberg.de.

Accepted 25 June 2009
Published Online First
19 August 2009

ABSTRACT

Background Dichotomisation of continuous variables before analysis has frequently been criticised but, nonetheless, remains a common approach. We were interested in the effects of dichotomisation of an outcome variable when two predictors are examined.

Methods Assuming a log-normally distributed continuous outcome, a three-level and a binary independent variable, we evaluated the results that would be obtained by logistic regression after dichotomisation. Different cut-offs, predictor effects and dispersions were examined, with a special focus on interaction terms.

Results Depending on the specific parameter combination, dichotomisation introduced sometimes substantial spurious interactions between the two predictor variables regarding their association with the outcome. These interactions could be assigned statistical significance even with modest sample sizes. Real-life data on sex \times weight as determinants of γ -glutamyltransferase provided a practical example of these issues.

Conclusions The findings presented add a new aspect to the controversy surrounding dichotomisation of continuous variables. Researchers should critically examine whether the validity of their results might be hampered by such phenomena.

INTRODUCTION

Dichotomisation of continuous variables is frequently done in many research fields, including biomedical sciences and epidemiology. This pertains to both outcome and predictor variables, and may be motivated by prevailing disease definitions (eg, hypertension¹), strongly non-linear relationships with a continuous parameter (eg, cardiovascular risk and glomerular filtration rate²) or by various conventions regarding what is seen as normal or abnormal (eg, serum γ -glutamyltransferase (gGT) concentrations³). Reporting results in terms of ORs rather than linear regression coefficients may sometimes appear advantageous, and additional arguments for dichotomisation have been discussed by others.^{4–5}

It is well known that dichotomisation (or more generally categorisation) results in a loss of information and, consequently, reduced statistical power.^{5–7} Furthermore, it has been demonstrated that dichotomisation of multiple predictor variables may lead to spurious associations and interactions regarding their associations with the dependent variable.^{4–8} However, discussion of the effects of outcome dichotomisation have largely focussed on the decreased analytical efficiency^{6,9} or the strength of associations observed, including the potential impact of changing cut-off definitions.^{6–7,10}

The aims of the present study were to examine how and to what extent dichotomisation of a continuous

outcome variable may introduce a spurious interaction between two independent variables. Given the current excitement about sex-specific genetic effects, our evaluations focussed on a scenario with a two-level stratum indicator (sex) and a three-level predictor (single nucleotide polymorphism). We demonstrate the issue with real-life example data of similar structure (sex \times normal/overweight/obese).

METHODS

Table 1 presents the nomenclature used throughout this report. We assumed the quantitative outcome variable y to be log-normally distributed with log-SD σ and log-mean μ_{ES} , where μ_{ES} depends on stratum S (considering two strata, say, male and female, this may take on the value of 0 or 1) and exposure E (considering a three categorical exposure, say, a genotype with the values GG, GT and TT, this may take on the value of 0, 1 or 2) and satisfies the simple equation $\log(\mu_{ES}) = \mu_0 + \beta_S \cdot S + \beta_E \cdot E$. Thus, the continuous data could be perfectly modelled by a semi-logarithmic linear regression model of the form $\log(y) = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + e$, and there clearly was no interaction present between S and E —that is, regardless of S , y would increase $\exp(\beta_E)$ -fold per unit increase in E .

Assuming that some researchers would prefer for any of the above mentioned reasons to analyse these data after dichotomisation of the outcome, we calculated the odds of y exceeding a specified cut-off value $\exp(c)$ in each of the six stratum \times exposure categories, based on the log-normal distribution function. To reveal possible interactions between the two independent variables, we derived the ratios of the stratum-specific ORs associated with each exposure category as indicated in table 1. The resulting OR interaction terms IA_1 and IA_2 (ie, interaction terms for the stratum S with the exposure E at $E = 1$ and $E = 2$, respectively) were examined over a range of values of β_S , β_E , c and σ , with the intercept term μ_0 fixed at 0. The parameters were changed one by one, always keeping all other parameters constant at their arbitrarily chosen standard values ($\beta_S = 1$; $\beta_E = 1$; $c = 1$; $\sigma = 1$). We visualised the results by tabulation and plotting.

The most common approach to analysing the dichotomised outcome presumably would be by logistic regression. To get an impression as to the danger of assigning statistical significance to spurious interactions introduced by outcome dichotomisation, we sought the sample size for which product interaction terms between S and E would appear statistically significant ($p < 0.05$) in likelihood ratio tests with 2 degrees of freedom (E treated as a categorical predictor) or 1 degree of freedom (E treated as a continuous predictor; commonly

Table 1 Nomenclature and formulas used in this report

Stratum	Exposure		
	E = 0	E = 1	E = 2
S = 0			
log-mean	μ_0	$\mu_0 + \beta_E$	$\mu_0 + 2\beta_E$
OR	1 (ref.)	OR ₁₀	OR ₂₀
S = 1			
log-mean	$\mu_0 + \beta_S$	$\mu_0 + \beta_S + \beta_E$	$\mu_0 + \beta_S + 2\beta_E$
OR	1 (ref.)	OR ₁₁	OR ₂₁
Interaction	1 (ref.)	IA ₁ = OR ₁₁ /OR ₁₀	IA ₂ = OR ₂₁ /OR ₂₀

In each cell, the outcome is assumed to be log-normally distributed with the log-mean indicated and a constant log-SD (σ).

referred to as ‘additive genotype model’ in genetic or ‘linear trend model’ in epidemiological studies). We assumed the strata to be equal-sized (imagine a sex ratio of 1) and the exposure categories 0-1-2 to follow a distribution of $n:2n:n$ (imagine a single nucleotide polymorphism in Hardy–Weinberg equilibrium with minor allele frequency of 50%, the generally most powerful situation in genetic association studies). Simulation analyses were carried out using R 2.6.1 for Windows.¹¹

Whereas we made a somewhat arbitrary choice by conducting our evaluations using log-normal data, one would expect the principal observations to be made in the present work to be of rather general nature. An advantage of using log-normally (instead of normally) distributed data is that in this setup the linear regression of the log-transformed outcome essentially is already on a multiplicative scale, and any findings made after dichotomisation and logistic modelling consequently could not be ascribed merely to a change in scales of the regression analyses.

To demonstrate the issues discussed with a practical example, we conducted an ad hoc analysis examining whether the effect of overweight and obesity on serum gGT levels differs by sex. Data originated from the baseline examination of an epidemiological study in Germany,^{10 12} and the dichotomisation cut-off was defined either as the upper reference limit for females (38 IU/l) of the assay used or based on the literature (50 IU/l).¹³ Example data analyses were conducted using SAS V.9.2 software.¹⁴

RESULTS

The main results are summarised in figures 1 and 2. Already with the parameter combination consisting of the chosen standard values, a positive interaction for the higher exposure category was introduced by dichotomisation.

As becomes clear from figure 1, the impact of assuming different exposure or stratum effects (β_E , β_S) was qualitatively very similar: the more the effect exceeded 1, the larger became both IA₁ and IA₂. Intriguingly—and more easily seen in table 2 rather than the graphical presentation—the interaction term IA₂ (that is, the stratum \times exposure interaction at the exposure level E = 2) grew substantially faster when varying β_E than when similarly changing β_S , while the development of IA₁ (stratum \times exposure interaction at E = 1) was exactly identical for changes in these two parameters.

With either coefficient approaching 0, the interaction ORs must approach 1. For exposure effects $\beta_E < 1$, this happened somewhat more rapidly than for $\beta_S < 1$, with the difference that IA₂ remained > 1 throughout for the stratum effect approaching 0, while it dropped below 1 in the β_E evaluations (figure 1). IA₁ dropped below 1 in both cases.

The picture was rather different for varying values of the dichotomisation cut-off parameter c (figure 2A). Here, positive

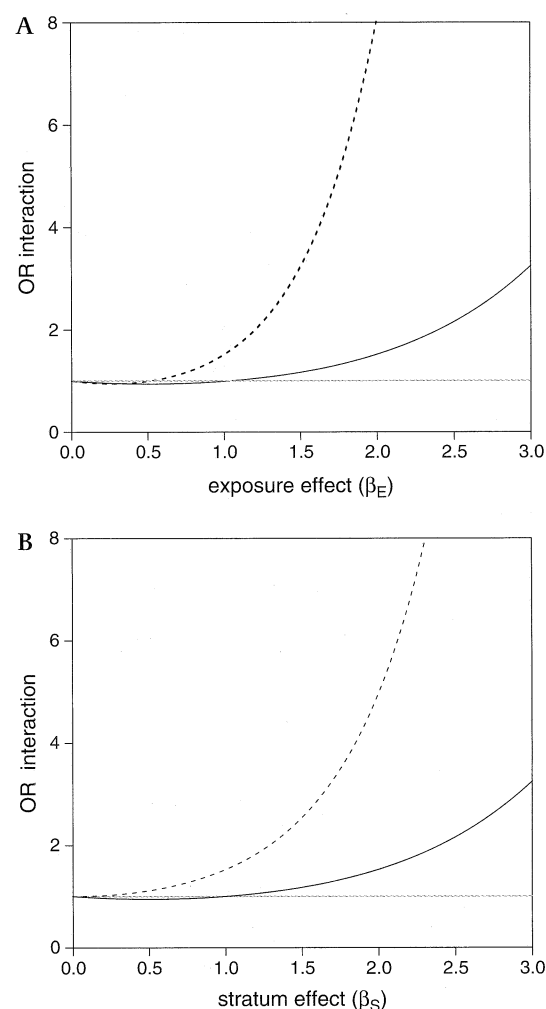


Figure 1 Interaction ORs IA₁ (OR associated with exposure level E = 1 in stratum S = 1 divided by the corresponding OR in stratum S = 0; solid black line) and IA₂ (equivalent to IA₂, but at exposure level E = 2; dashed line) observed upon dichotomising a continuous outcome as in the scenario laid out in the text, across a range of values of the assumed exposure effect (β_E ; panel A; $\beta_S = 1$; $c = 1$; $\sigma = 1$) or stratum effect (β_S ; panel B; $\beta_E = 1$; $c = 1$; $\sigma = 1$). Grey line: reference null value.

interactions were introduced when the cut-off was shifted to the left—that is, c approached 0. However, when the cut-off increased beyond 1, both IA₁ and IA₂ pronouncedly dropped below 1. Whereas IA₂ had generally exceeded IA₁ in the prior evaluations, this relationship was clearly reversed for larger values of c .

Figure 2B finally presents the effect of varying the dispersion of the continuous outcome dichotomised. Given that IA₁ would be constantly 1 if using only default values for the distribution shifts (table 2), a stratum effect of $\beta_S = 1.5$ was used for this evaluation. With increased dispersion diluting associations and rendering the y -distribution more and more uniform, the interaction terms unsurprisingly approached 1 for larger values of σ . In contrast, lowering σ lead to an extremely rapid growth of both IA₁ and IA₂.

Note that all of the results above are identical to those that would be obtained for a normally distributed outcome with mean $\mu_{ES} = \mu_0 + \beta_S \cdot S + \beta_E \cdot E$, SD σ and cut-off c . Similarly, this would extend to a variable whose logit is normally distributed with means determined as above, if the cut-offs were applied on the logit scale.

When we evaluated the number of subjects required to obtain a statistically significant interaction finding in logistic regression

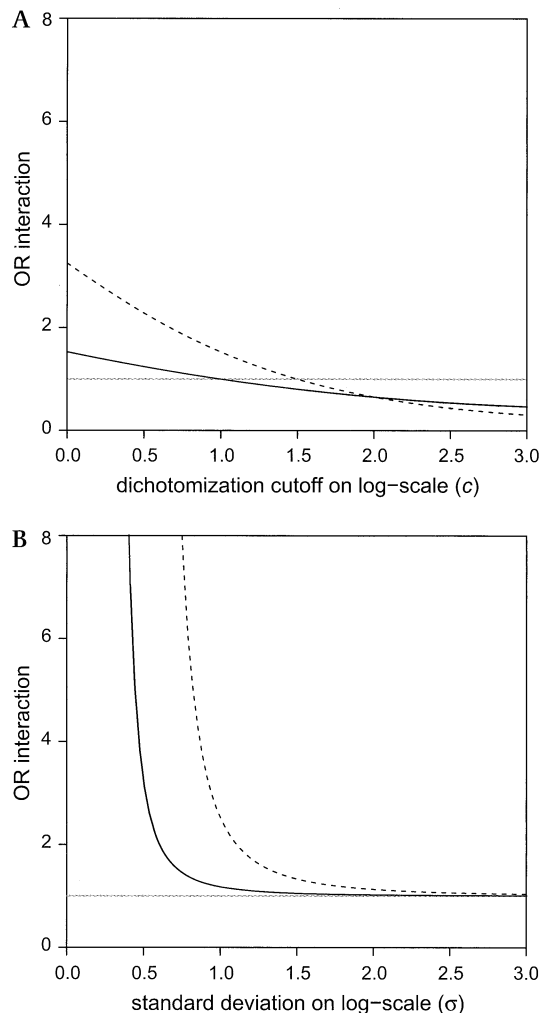


Figure 2 Interaction ORs IA_1 (solid black line) and IA_2 (dashed line) observed upon dichotomising a continuous outcome as in the scenario laid out in the text, across a range of values of the dichotomisation parameter used (c ; panel A; $\beta_S = 1$; $\beta_E = 1$; $\sigma = 1$) or dispersion of the underlying distribution (σ ; panel B; $\beta_S = 1.5$; $\beta_E = 1$; $c = 1$). Grey line: reference null value.

Table 2 Introduction of interactions by dichotomising the outcome

Interaction*	Value of the parameter indicated							
	0.25	0.50	0.75	1.00	1.25	1.50	2.00	3.00
Exposure effect (β_E)								
IA_1	0.96	0.95	0.96	1.00	1.07	1.18	1.53	3.25
IA_2	0.95	1.00	1.18	1.53	2.16	3.25	8.05	54.8
Stratum effect (β_S)								
IA_1	0.96	0.95	0.96	1.00	1.07	1.18	1.53	3.25
IA_2	1.03	1.11	1.27	1.53	1.93	2.54	4.96	26.1
Dichotomisation cut-off on log-scale (c)								
IA_1	1.38	1.24	1.11	1.00	0.90	0.81	0.65	0.47
IA_2	2.74	2.28	1.87	1.53	1.24	1.00	0.65	0.31
SD on log-scale (σ)								
IA_1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
IA_2	1.5×10^6	17.1	2.62	1.53	1.25	1.14	1.06	1.02

Unless stated otherwise, the cut-off parameter c , as well as σ , β_E and β_S , all equal 1.

*Interaction terms are labelled as indicated in table 1—that is, IA_1 equals the OR associated with exposure level $E = 1$ in stratum $S = 1$ divided by the corresponding OR in stratum $S = 0$, whereas IA_2 equals the equivalent ratio at exposure level $E = 2$.

models, the lowest number of subjects per homozygotes cell for the parameter combinations shown in table 2 occurred at $\beta_S = 2$ ($LRT_{1df} = 312$; $LRT_{2df} = 665$). Assuming this combination of values and a stratum and exposure distribution as detailed in the methods section, a 1 degree of freedom likelihood ratio test for a stratum \times exposure trend interaction term could assign significance to the spurious interaction with a total study size as small as $2 \times 312 + 2 \times 624 + 2 \times 312 = 2496$ participants.

Results of our real-life example on gGT, sex and body mass index are presented in table 3. In the population studied, gGT was—as generally the case for this serum marker—highly skewed with median (interquartile range) of 37.3 (26.8–57.1) IU/l in males and 24.8 (18.5–36.5) IU/l in females, and natural log-transformation generated reasonable normality (not shown). Modelling the transformed gGT by linear regression, the effect estimates for overweight and obesity as determinants of gGT appeared to differ somewhat between males and females, but the statistical interaction test did not reach significance at $\alpha = 0.05$. However, the interaction p value was already smaller for the cut-off of 38 IU/l and ultimately dropped below the significance threshold for the cut-off of 50 IU/l. Thus, in contrast to the analysis of the continuous data, modelling the arbitrarily dichotomised outcome here would lead to a finding of statistically significant effect heterogeneity, in particular an interaction of weight and sex with respect to their role as risk factors for elevated serum gGT, depending on the choice of cut-off.

DISCUSSION

The elaborations presented above demonstrate that the dichotomisation of a continuous outcome variable can introduce spurious interactions between two predictor variables. Such interaction was caused in essentially all parameter scenarios examined, and in some cases could be assigned statistical significance assuming sample sizes often analysed in modern epidemiological studies.

Clinical dichotomisation thresholds, rightly, are widely applied and often of clear advantage for classifying, diagnosing and appropriately treating patients—for example, for obesity (body mass index above a certain cut-off) or hypertension (blood pressure above a certain cut-off)—but for analyses trying to understand the determinants of such characteristics, arbitrary dichotomisation may be not only unfounded but even hazardous. While multiple reasons have previously been brought forward against the widespread habit of dichotomising continuous variables prior to analysis,^{5–7} our findings add an additional facet to the list. The impact of only varying the dichotomisation cut-off on effect estimates and statistical power has been evaluated previously for situations with a single independent variable.⁶ Realising that our stratified scenario can be seen as a combination of two such bivariate situations, in which the single dichotomisation cut-off appears shifted in relation to the underlying distribution in one of the strata, our intriguing findings probably could have been anticipated. Somewhat surprisingly, however, the implications of the cut-point dependence for assessment of interactions have received little if any attention to date.

The phenomenon described would be of little relevance, if realistic studies never reached the sample size required to detect these interactions with any precision—for example because extreme cut-offs would make the dichotomised outcome too rare. It was not the objective of this study to evaluate all possible study designs and our simulations evaluated only the situation in which the genotypes (our three-level predictors) are in perfect Hardy–Weinberg equilibrium, the variant genotype is as frequent as possible (minor allele frequency 50%) and strata are

Table 3 Example analysis of body mass index (BMI) as a sex-specific determinant of γ -glutamyltransferase (gGT) in an epidemiological study*

Exposure	Linear regression of log-transformed gGT β (95% CI)	Logistic regression of elevated gGT	
		>38 IU/L OR (95% CI)	>50 IU/L OR (95% CI)
BMI in females			
≤25 kg/m ² (normal)	0 ref	1 ref	1 ref
>25–30 kg/m ² (overweight)	0.17 (0.13 to 0.21)	1.81 (1.51 to 2.17)	1.73 (1.39 to 2.16)
>30 kg/m ² (obese)	0.32 (0.28 to 0.37)	2.90 (2.40 to 3.50)	2.68 (2.13 to 3.37)
BMI in males			
≤25 kg/m ² (normal)	0 ref	1 ref	1 ref
>25–30 kg/m ² (overweight)	0.11 (0.06 to 0.16)	1.41 (1.20 to 1.66)	1.24 (1.03 to 1.48)
>30 kg/m ² (obese)	0.24 (0.19 to 0.30)	2.18 (1.81 to 2.62)	1.84 (1.50 to 2.24)
Interaction terms†			
for >25–30 kg/m ²	−0.06 (−0.12 to 0.01)	0.78 (0.61 to 0.99)	0.71 (0.54 to 0.95)
for >30 kg/m ²	−0.08 (−0.15 to −0.01)	0.75 (0.58 to 0.98)	0.69 (0.51 to 0.93)
Likelihood ratio test for interaction	<i>P</i> _{2df} = 0.071	<i>P</i> _{2df} = 0.068	<i>P</i> _{2df} = 0.031

*n=8614 participants of the ESTHER cohort study conducted in the German state of Saarland; models adjusted for intensity of regular alcohol consumption.

†For logistic models, interaction ORs (equivalent to IA_1 and IA_2 in the simulation studies) are given, whereas untransformed β -coefficients of the interaction terms are given for log(gGT) models.

of equal size. Nonetheless, the minimum relevant sample size found was well below the number of subjects that nowadays are routinely included especially in genetic epidemiology—a field heavily relying on the concept of statistical significance and p values. In this field, large sample sizes are typically needed to detect the mostly very small main effects of specific polymorphisms or haplotypes. Our findings, paired with the fact that sex heterogeneities of the effects of genetic predictors in particular of quantitative traits¹⁵ are currently very much en vogue, urge caution when interpreting pertinent results.

Our real-life example presented a situation in which informed—yet, nonetheless arbitrary—usage of cut-offs could result in different conclusions regarding the statistical significance of interactions between two predictors. As real life goes, the data chosen potentially suggested some effect heterogeneity also in the continuous model, even though these analyses conducted purely for instructional purposes clearly would require more extensive considerations and more detailed modelling before drawing conclusions beyond potential pitfalls of dichotomisation.

We only examined a log-normally distributed outcome and focussed on varying effect sizes and cut-off definitions without changing the general shape of the underlying distribution. For the sake of feasibility, we did not attempt to cover all possible parameter combinations or alternative distributions. One, consequently, cannot extend the results to the entirety of continuous outcomes of interest in the health sciences or other fields. These may sometimes not even lend themselves to continuous modelling; for example, because no statistical distribution fitting the outcome well can be identified and consequently no appropriate regression model can be specified. Our findings can only stipulate any researcher to critically question his/her results if obtained after outcome dichotomisation, investigate whether phenomena such as demonstrated here might have endangered

the validity of the results given the particular real data at hand and determine if the pros of dichotomisation actually outweigh the cons in the individual analysis. If the latter is the case, stratum-specific cut-offs at equivalent positions relative to the stratum-specific distributions could prevent the exposure effects from spuriously varying between the strata, if the shape of the underlying distributions otherwise was equivalent in the strata (excluding the shift accounted for by the stratum-specific cutoff). Alternatively, one should consider the possibility of presenting appropriate analyses of the continuous outcome alongside the artificially binary analyses in order to reveal discrepancies possibly due to spurious interactions and probably worth discussion. It should also be noted that efficient methods have been developed to model the risk of exceeding a dichotomisation cut-off without discarding the information inherent to the underlying continuous data,^{9 16} though their performance regarding the scenario described here presumably needs further investigation.

Others have mentioned (or sung, for that matter) that, ‘Breaking Up is Hard to Do.’¹⁷ As shown in this report, making sense of relationships found after breaking up might be even harder and interactions observed might well be spurious oddities. This does not preclude consciously splitting up, but it contributes a new aspect to the controversy surrounding this common and sometimes unavoidable practice.

Acknowledgements This work was supported by a grant within the German Research Foundation (DFG) priority programme SPP1226 ‘Nicotine’ (Br1704/11-1). The real-life example data were collected as part of the ESTHER cohort baseline examination, which was funded by the Baden Württemberg Ministry of Research, Science and Arts. Over the years, many individuals have contributed to ESTHER and their efforts are gratefully acknowledged.

Funding Deutsche Forschungsgemeinschaft, Germany.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- Messerli FH, Williams B, Ritz E. Essential hypertension. *Lancet* 2007;**370**:591–603.
- Foster MC, Hwang SJ, Larson MG, et al. Cross-classification of microalbuminuria and reduced glomerular filtration rate: associations between cardiovascular disease risk factors and clinical outcomes. *Arch Intern Med* 2007;**167**:1386–92.
- Yamada Y, Noborisaka Y, Suzuki H, et al. Alcohol consumption, serum gamma-glutamyltransferase levels, and coronary risk factors in a middle-aged occupational population. *J Occup Health* 2003;**45**:293–9.
- MacCallum RC, Zhang S, Preacher KJ, et al. On the practice of dichotomisation of quantitative variables. *Psychol Methods* 2002;**7**:19–40.
- Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006;**25**:127–41.
- Ragland DR. Dichotomizing continuous outcome variables: dependence of the magnitude of association and statistical power on the cutpoint. *Epidemiology* 1992;**3**:434–40.
- Streiner DL. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Can J Psychiatry* 2002;**47**:262–6.
- Maxwell SE, Delaney HD. Bivariate median splits and spurious statistical significance. *Psychol Bull* 1993;**113**:181–90.
- Suissa S, Blais L. Binary regression with continuous outcomes. *Stat Med* 1995;**14**:247–55.
- Brenner H, Rothenbacher D, Weck MN. Epidemiologic findings on serologically defined chronic atrophic gastritis strongly depend on the choice of the cutoff-value. *Int J Cancer* 2007;**121**:2782–6.
- R Foundation of Statistical Computing. R v2.6.1. Vienna, 2007.
- Breitling LP, Raum E, Muller H, et al. Synergism between smoking and alcohol consumption with respect to serum gamma-glutamyltransferase. *Hepatology* 2009;**49**:802–8.
- Poikolainen K, Vartiainen E. Determinants of gamma-glutamyltransferase: positive interaction with alcohol and body mass index, negative association with coffee. *Am J Epidemiol* 1997;**146**:1019–24.
- SAS Institute. Statistical analysis software, release 9.2. Cary (NC): SAS Institute, 2008.
- Weiss LA, Pan L, Abney M, et al. The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet* 2006;**38**:218–22.
- Suissa S. Binary methods for continuous outcomes: a parametric alternative. *J Clin Epidemiol* 1991;**44**:241–8.