RESEARCH COMMUNICATION

Repeatability of Self-reported Information for Population-based Studies of Cancer

Shamshad Karatela, David M Purdie, Adèle C Green, Penelope M Webb, David C Whiteman*

Abstract

Purpose: To measure the repeatability of a cancer risk factor questionnaire in a population-based case-control study. Methods: Questionnaires were completed on two occasions by patients with cancer of the ovary (n=25) or esophagus (n=23) and by 37 controls without cancer. We assessed general cancer risk factors including height and weight (for calculating body mass index (BMI)), smoking and anti-inflammatory (NSAID) use. Risk factors specific for ovarian and esophageal cancers were also assessed. Agreement was measured by the correlation coefficient and weighted kappa statistic (kw) for continuous and categorical variables respectively. Results: We observed very high levels of agreement for BMI (kw=0.84) and smoking history, including ages at initiation and quitting (Pearson correlation = 0.87 and 0.86 respectively). There was moderate to substantial agreement for use of anti-inflammatory drugs (aspirin kw =0.52, other NSAIDS kw =0.72). Agreement for lifetime prevalence of medical conditions varied from almost perfect (e.g. history of benign breast disease (k =0.86)) to moderate (e.g. heartburn (k =0.57)). Item repeatability was not materially altered by case-control status, age or sex of respondents or interval between completions. Conclusions: Self-reported cancer risk factor information demonstrates moderate to almost perfect levels of agreement, suggesting these items are suitable for risk estimation and epidemiologic inference.

Key Words: Reproducibility of results - agreement - risk factors - case-control studies - cancer

Asian Pacific J Cancer Prev, **7**, 303-308

Introduction

Accurate measurement of exposure to putative risk factors is of fundamental importance for epidemiological studies. While case-control studies are a particularly efficient design for studying the causes of cancer and are often the only practicable method for studying rare tumors, their requirement for retrospective assessment of exposure, typically by self-report, raises concerns about the validity of measures so derived. An important property of a data collection instrument is its repeatability, that is the consistency with which respondents answer the same questions on different occasions. Estimating the level of repeatability provides an indication of the degree of random error incurred by the data collection instrument, although high levels of repeatability do not necessarily equate to high levels of validity.

Here we report the findings of a repeatability study conducted among a sample of participants in two national case-control studies of ovarian cancer and esophageal cancer. The principal aim of this analysis was to measure the level of agreement for key risk factors of interest; secondary aims were to evaluate whether case-control status, age, sex and time interval between first and second completions were associated with the repeatability of self-reported exposure to the items of interest.

Materials and Methods

Eligible participants for the repeatability study were those people who had taken part in either of two national case-control studies of cancer and who returned their primary study questionnaires between November 2002 and February 2003. Case participants in the national case-control studies were defined as patients aged between 18 and 79 years with a histologically confirmed diagnosis of either cancer of the esophagus or the ovary, newly diagnosed after July 1, 2002 in the mainland states and territories of Australia and, for ovarian cancer, Tasmania. For both case-control studies, controls were randomly sampled from the electoral roll (a compulsory adult population register) and frequency matched to the cases on age, sex and state/territory of residence

Potential participants for the repeatability study were mailed an invitation letter between March and May, 2003. If no response was received within two weeks, then follow-

*Division of Population Studies and Human Genetics, Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Queensland 4029 Australia * For Correspondence Tel: +61 7 3362 0279 Fax: +61 7 3845 3502 Email: david.whiteman@qimr.edu.au*

up telephone calls were made to elicit a response. Care was taken to explain the reason for the repeatability study, especially to reassure the participants that this was not because they had made mistakes with the original questionnaire. A second copy of the study questionnaire was mailed to those who agreed to take part in the repeatability study; completed questionnaires were returned to the study center in reply-paid envelopes.

The Survey Instrument

Participants in the repeatability study were sent the same structured, self-completed questionnaire that was completed at baseline by all those taking part in the national case-control studies. While the full questionnaire enabled comprehensive information to be collected on various established cancer risk factors including lifestyle habits, medical and occupational history and family history of cancer, the repeatability analyses described here focused on selected items considered to be either of general interest, namely body weight, smoking history, whether alcohol was consumed and use of various anti-inflammatory medications; or to be important background socioeconomic items (education, income); or considered to be of specific relevance to gynecologic or digestive cancer risk.

Information about height (cm) and weight (kg) was obtained from which we calculated body mass index (BMI) at current age and at the time of greatest weight by dividing weight in kilograms by the square of height in meters. We used standard BMI categories for analysis (<18.5 kg/m2 "underweight", 18.5 to 24.9 kg/ m2 "normal", 25 to 29.9 kg/ m2 "overweight", >30 kg/ m2 "obese").

Regarding smoking, participants were asked whether, over their whole life, they had ever smoked more than 100 cigarettes, cigars or pipes; positive responses elicited further questions about how much they usually smoked on a typical day, how many years they had smoked and the ages they first started to smoke and stopped smoking. Among smokers, we derived the number of pack-years of tobacco exposure by dividing the number of cigarettes smoked on a typical day by 20, and multiplying by the total number of years smoked. For analysis (restricted to smokers), pack-years were categorized as up to 9 pack-years, 10 to 29 pack-years, 30 to 49 pack-years and 50 or more pack-years. As well we studied information about the frequency of use of each of the following classes of medications during the past fiveyears: aspirin, aspirin and codeine, acetaminophen and codeine, non-steroidal anti-inflammatory drugs [NSAIDs] and antihistamines. Each medication item was followed by a list of at least four trade names of formulations of these medications that were commonly available in Australia. Participants were asked to select from a semi-quantitative scale of eight usage frequencies "never", "occasionally", "less than once a month", "2-3 times a month", "once a week", "2-3 times a week", "4-7 times a week" or "twice or more per day".

Regarding socioeconomic status, participants were asked "how old were you when you left school", "have you completed any further study since leaving school", and "what was your household income (including pensions and allowances) before tax".

Digestive health was assessed by asking participants if they had experienced acid reflux, defined as "a sour taste from acid or bile rising up into the mouth or throat" and heartburn ("a burning pain behind the breastbone after eating"). Participants were also asked whether they had ever had medical procedures such as endoscopy or surgery to the esophagus, stomach or gall bladder. Gynecologic information asked of women included whether they had ever had a hysterectomy, or whether "their tubes were tied or clipped for contraception". We also asked women about their use of oral contraceptive preparations and hormone replacement therapy.

Statistical Analysis

We measured the repeatability of the data collection instrument by comparing each participant's responses to the same questionnaire items at two time points. For categorical and ordinal variables, we used the kappa (k) and weighted kappa (kw) statistics, respectively, to quantify the degree of agreement between responses (Landis and Koch 1977) . Kappa values range from -1.0 to 1.0 with values of 0.8 to 1.0 indicating almost perfect agreement; 0.6 to 0.8 substantial agreement; 0.4 to 0.6 moderate agreement; 0.2 to 0.4 fair agreement; 0 to 0.2 slight agreement; and less than 0 poor agreement (Landis and Koch 1977) . For continuous variables, Pearson's correlation coefficient was used as a measure of repeatability. We also calculated the concordance for responses to each question, which was the percentage of responses in exact agreement on the first and second occasions.

For items with concordance less than 90%, we estimated the effect of time interval between completion of interviews, age, gender and cases-control status as potential predictors of concordance. For categorical variables, this was done by fitting logistic regression models in which the outcome was whether or not the respondent had answered the same on both occasions. For continuous variables, linear regression was used with the outcome being the difference between the two responses.

Ethics

Ethical approval was obtained from the Queensland Institute of Medical Research Humans Research Ethics Committee and from all other collaborating institutions and hospitals from where cancer patients were recruited. Permission was given by the Australian Electoral Commission to obtain a random sample of potential controls from the Commonwealth Electoral Roll. Ethical approval for the repeatability study was obtained from the University of Queensland. All participants provided written consent before they were enrolled within each study. Ethical standards were in accordance with the Australian national statement on the ethical conduct of research involving humans and the Helsinki Declaration.

Table 1. Characteristics of Respondents in the Repeatability Study

	Ovarian cancer cases	Esophageal cancer cases	Controls	Total
Number of respondents	25	23	37	85
Mean age at 1st interview (yrs, range)	54.8 (23 – 75)	63.7 (51 – 75)	62.2 (38 - 77)	60.4 (23 – 77)
% male	0%	74%	57%	45%
Mean interval between 1st and 2nd interviews (months, range)	4.9 (2.4 – 8.9)	4.0 (1.3 – 6.6)	4.3 (1.8 – 7.1)	4.4 (1.3 – 8.9)

Results

A total of 128 case participants were selected for the repeatability study (63 women with ovarian cancer and 65 patients with esophageal cancer), of whom 4 declined the invitation to take part, 49 were too ill due to deterioration of their condition, 6 had died and 7 could not be contacted. Thus 62 cases were sent repeat questionnaires (34 patients with ovarian cancer, 28 patients with esophageal cancer). Of these, 48 (77%) were returned (25 patients with ovarian cancer, 23 patients with esophageal cancer).

Overall 52 control participants were eligible to take part, of whom 6 declined the invitation, 2 were too ill and 7 could not be contacted, leaving 37 (71%) control participants all of whom returned repeat questionnaires.

Respondents had a mean age at the time of the first interview of 60 years (range 23 – 77 years), although women with ovarian cancer were 8 years younger on average than esophageal cancer patients and controls (Table 1). 74% of esophageal cancer patients were male, as were 57% of controls. The mean interval between completion of the first and second questionnaires was 4.4 months (range 1.3 – 8.9 months) with little difference in the interval across the two case groups or controls.

Social Factors

Questions about level of education, household income and employment had high levels of agreement with kappa values ranging between 0.70 and 0.84 (Table 2). While exact agreement for weight (within 1 kg) and height (within 1 cm) was low (concordances of 27% and 53% respectively within these narrow restrictions), correlations between the values reported at each time point were high (0.96 and 0.70 respectively). Body mass index was calculated for each participant at the two time points, with an overall correlation of 0.95 (95% CI 0.92-0.97). When categorized according to internationally accepted cut points for BMI, we observed a concordance of 0.83 and a weighted kappa of 0.84 (95% CI 0.76-0.93).

Among the range of questions about smoking history, we found almost perfect agreement for current smoking status (k= 0.97; 95% CI: 0.91- 1.00), and items asking about age started smoking, smoking status one year ago, and age stopped smoking also demonstrated high levels of agreement (Table 3). Other measures of smoking behavior such as total duration of smoking and number of days per week that a person smoked were also reported reasonably consistently. Cumulative smoking history (measured in pack-years) demonstrated high levels of correlation for the continuous

Table 2. Repeatability of Social and Anthropometric Characteristics

Variable	Total number	Concordance (%)	Measure	Value	95% CI[1]
Age left school	85	89	Pearson correlation	0.97	0.95-0.98
Any further study after school[2]	85	78	Kappa	0.70	0.58-0.82
Annual household income[3]	69	77	Weighted kappa	0.81	0.70-0.91
Employment status[4]	85	88	Kappa	0.84	0.75-0.93
Weight (in kg)	81	27	Pearson correlation	0.96	0.94-0.97
Height (in cm)	74	53	Pearson correlation	0.70	0.56-0.80

[1]Confidence interval [2]5 categories: No, Technical/College diploma, Trade certificate/apprenticeship, University degree, Other [3]7 categories: less than $15,000 p.a., 15,000-$29,999 p.a, $30,000-$44,999 p.a, $45,000-$59,999 p.a, $60,000-$79,999 p.a., $80,000-$99,999 p.a., $100,000 or more p.a [4]7 categories: Full-time worker, Part-time worker, Home duties, Student, Unemployed, Retired, Other (eg sickness benefit)

Table 3. Repeatability of Drinking Status and Smoking Attributes

Variable	N	Concordance (%)	Measure	Value	95% CI
Drinking status[1]	78	92	Kappa	0.82	0.66-0.92
Smoking status[2]	82	99	Kappa	0.97	0.91-1.03
Age first started to smoke	58	36	Pearson correlation	0.87	0.79-0.92
Smoking days per week[3]	57	86	Weighted kappa	0.59	0.32-0.86
Number of cigarettes smoked/day	55	49	Pearson correlation	0.92	0.86-0.95
Smoking this time last year	58	91	Kappa	0.79	0.63-0.96
Age stoped smoking	37	43	Pearson correlation	0.86	0.76-0.92
Total years smoked	56	21	Pearson correlation	0.83	0.73-0.90
Pack-years of smoking	53	66	Weighted kappa	0.68	0.54-0.81

[1]Categorized as lifelong non-drinker, current drinker, ex-drinker [2]Categorized as never smoker, ever smoker [3]Categorized as 0-7 days

Table 4. Repeatability of Contraceptive Pill Use, Hormone Replacement Therapy and Medication Use

	Total number	Concordance (%)	Weighted Kappa	95% CI
Oral Contraceptive pill1	42	83	0.67	0.44-0.89
Hormone replacement therapy[1]	47	83	0.66	0.45-0.87
Aspirin2	79	52	0.52	0.37-0.68
Aspirin and Codeine[2]	69	71	0.59	0.38-0.79
Acetaminophen (Paracetamol)[2]	78	59	0.58	0.43-0.73
Acetaminophen and Codeine[2]	70	71	0.65	0.51-0.80
Anti-inflammatory drugs[2]	75	76	0.72	0.57-0.86
Antihistamines[2]	70	69	0.42	0.21-0.63

[1]Unweighted kappa statistic as exposure was recorded as "ever taken (yes vs no)" [2]Frequency of use in the last 5 years, categorized as "never", "occasionally", "less than once a month", "2-3 times a month", "once a week", "2-3 times a week", "4-7 times a week", "twice or more per day"

Table 5. Repeatability of Medical History and Symptoms

	N	Prevalence at Time 1	Concordance (%)	Kappa	95% CI
Benign breast disease	46	11.0%	96	0.86	0.67-1.00
Endometriosis	46	5.9%	98	0.88	0.64-1.00
Ovarian cyst	47	22.0%	91	0.82	0.65-0.99
Barrett's esophagus	74	4.7%	97	0.65	0.19-1.00
Oesophagitis	74	14.0%	91	0.54	0.24-0.84
Gastritis	75	8.2%	89	0.44	0.12-0.76
Peptic ulcer	73	8.2%	94	0.92	0.76-1.00
Hiatus hernia	77	17.0%	99	0.79	0.61-0.97
Heart burn	82	60.0%	79	0.57	0.39-0.75
Acid reflux	82	49.0%	74	0.49	0.30-0.68
Food regurgitation	81	19.0%	83	0.50	0.27-0.73
Helicobacter pylori infection	85	9.0%	95	0.72	0.46-0.98

measure (Pearson's correlation 0.87, 95% CI 0.79-0.93) and substantial levels of agreement for the categorical measure (kw= 0.68, 95% CI 0.54–0.81).

We asked participants to report their frequency of use during the past 5 years of a range of commonly used medications using an 8-point scale (Table 4). Aspirin, the medication most commonly reported in our sample, had modest levels of agreement between the two time points (kw= 0.52). We observed higher levels of agreement for all other common medications except anti-histamines (kw= 0.42). Highest agreement was observed for non-steroidal anti-inflammatory drugs (kw= 0.72), and this was similar for cases and controls (data not shown). Responses to questions about level of education, household income and employment had high levels of agreement with kappa values ranging between 0.70 and 0.84 (Table 2).

Digestive and Gynecologic Risk Factors

Repeatability of medical history and symptoms was generally high (Table 5) with concordance values ranging from 74% for history of acid reflux to 99% for hiatus hernia. Corresponding kappa values were somewhat lower,

Table 6. Odds ratio (and 95% confidence interval) for Agreement of Key Variables in a Repeatability Study[1]

	Case vs control	Age group 60+ vs <60	Sex (M vs F)	Time between interview completions (per month)
Any further study after school	2.33 (0.75-7.23)	0.55 (0.15-2.01)	1.63 (0.48-5.59)	0.68 (0.46-0.98)
Annual household income	1.14 (0.32-4.15)	6.74 (1.65-27.6)	0.90 (0.22-3.63)	1.49 (0.96-2.33)
Smoking days per week	0.94 (0.17-5.15)	0.95 (0.13-7.23)	0.28 (0.04-2.07)	1.41 (0.71-2.80)
Heart burn	0.70 (0.22-2.21)	1.01 (0.31-3.32)	1.95 (0.55-6.94)	0.97 (0.68-1.38)
Acid reflux	0.45 (0.15-1.38)	1.41 (0.47-4.25)	1.10 (0.35-3.40)	1.05 (0.75-1.47)
Food regurgitation	0.25 (0.06-1.07)	0.91 (0.23-3.54)	0.36 (0.10-1.36)	0.87 (0.59-1.29)
Oral contraceptive pill	1.54 (0.26-9.27)	2.07 (0.34-12.5)	-	0.96 (0.58-1.57)
Hormone replacement therapy	0.75 (0.15-3.78)	2.42 (0.41-14.3)	-	1.15 (0.74-1.78)
Aspirin	0.66 (0.25-1.72)	1.14 (0.41-3.21)	0.49 (0.18-1.35)	0.96 (0.71-1.31)
Aspirin and codeine	0.63 (0.19-2.06)	2.12 (0.61-7.37)	1.04 (0.28-3.84)	1.14 (0.79-1.65)
Acetaminophen (Paracetamol)	2.39 (0.83-6.86)	2.64 (0.87-7.99)	0.73 (0.24-2.17)	0.65 (0.46-0.92)
Acetaminophen and codeine	1.26 (0.42-3.78)	1.02 (0.31-3.39)	1.26 (0.37-4.29)	0.87 (0.61-1.23)
Anti-inflammatory drugs	1.14 (0.37-3.45)	0.99 (0.31-3.2)	1.59 (0.48-5.31)	0.96 (0.67-1.37)
Antihistamines	0.98 (0.33-2.93)	1.57 (0.52-4.75)	1.55 (0.48-5.00)	1.05 (0.75-1.49)

[1]Logistic regression models with terms for age, sex and case-control status and time between interview completions

nevertheless agreement was rated between moderate to almost perfect for all of these items. Female participants were asked whether they had ever taken the oral contraceptive pill or hormone replacement therapy. Both items were reported consistently (concordance 83% for both) and the kappa statistics indicated substantial levels of agreement (0.67 and 0.66 respectively).

Determinants of Item Repeatability

We assessed whether the interval between responses or the respondents' age, sex or case-control status influenced item repeatability (Table 6). For most items surveyed, concordance rates were similar for cases and controls. Cases were somewhat more likely than controls to have concordant reports of further study after school and use of acetaminophen (OR > 1.0), and less likely than controls to have concordant reports of regurgitation and acid reflux (OR < 1.0). Analyses comparing item repeatability among participants stratified by age found that those older then 60 years were significantly more likely to have concordant reports of household income than younger participants (OR>6). Although not statistically significant, we also found that those older than 60 were about twice as likely to consistently report their use of the oral contraceptive pill, hormone replacement therapy and acetaminophen. We found no evidence that concordance for any items was influenced by sex or time between questionnaire completions.

Assessment of Random Error

Assuming similar levels of random misclassification for cases and controls as suggested by our data, we estimated the likely effects of such error on the odds ratio. For a true odds ratio of 2.0 and an exposure prevalence of 25%, then 10% random misclassification of a dichotomous exposure (90% concordance of measures) would reduce the observed odds ratio to 1.7, and 20% random misclassification (80% concordance of measures) would reduce the observed odds ratio to 1.5.

Discussion

Epidemiologic inference depends upon the accurate measurement of risks of an outcome according to the level of exposure to putative causal factors. Repeatability, that is the demonstration of high levels of agreement between measures taken independently at two separate time points, is a necessary (although not sufficient) attribute for a valid measurement of exposure.

We examined the repeatability of a self-completed questionnaire instrument for measuring exposure to some health and lifestyle factors of general interest as well as factors relevant to risk of ovarian and esophageal cancer. We found empirical evidence of high levels of repeatability for the majority of questionnaire items investigated, including measures of socio-economic status, body mass, consumption of tobacco and alcohol, medical history and use of medications. These findings suggest that a comprehensive risk factor questionnaire such as the data collection instrument described here is unlikely to introduce large amounts of random error to the measurement of key exposures for these cancers, although even these relatively modest amounts of error will result in discernible bias in estimates of the odds ratio. Importantly, we found no evidence that repeatability for key exposure measures was significantly altered by case-control status or age.

Demographic characteristics such as level of education, employment status and annual household income showed excellent agreement in this setting, as has been reported in other epidemiologic studies (Langer et al. 2003; Rohan et al. 1988) . While self-reports of height and weight have been observed to have modest levels of agreement in previous studies, we found acceptable levels of repeatability in this series of participants. Height and weight measures are typically recorded for the purpose of calculating the body mass index as a measure of adiposity, a phenotypic attribute of increasing interest in epidemiological studies of cancer (Adami and Trichopoulos 2003; Calle and Kaaks 2004; Calle et al. 2003). After calculating BMI and categorizing the participants according to international criteria, we found very high levels of agreement for this measure.

Earlier studies have found that smoking history tends to be reliably reported by most participants (Kenkel et al., 2003). Our finding of very high levels of agreement for ages at initiation and quitting, number of days per week the participant smoked and numbers of cigarettes smoked per day therefore accords with findings from case-control studies conducted in a variety of settings (Donato et al. 1998; Gartner et al. 2005). As reported by these other studies, we found no evidence of differences for smoking-related variables between cases and controls.

We found some variability in the repeatability of self-reports of medical history and symptoms. Some conditions of low to moderate prevalence, including benign peptic ulcers, breast disease, endometriosis and ovarian cysts, were very reliably reported by study participants, similar to levels of agreement observed for similar items in surveys of US women (Langer et al., 2003). In contrast, more prevalent conditions such as heartburn and acid reflux had lower levels of agreement in our sample than those reported in US populations (Locke et al., 1994), although were similar to those observed in other settings in Australia (Talley et al., 1995). An important difference between our study and the US study was the considerably shorter time interval between survey completions for that study (mean 4 weeks) compared with this study (mean 4.4 months), during which time symptoms may have altered.

A potential source of error was the less than ideal participation rate in this repeatability study (48% for cases and 71% for controls). This would lead to an over-estimation of levels of agreement for our survey instrument if participants in the repeatability study were more competent at filling out the questionnaire than those who were selected but did not take part. However, the most common reasons for non-participation in this repeatability survey were ill

health or death, an unavoidable consequence of re-surveying people suffering from cancers associated with high levels of morbidity and mortality.

Although it is possible that participants may have attempted to remember their responses to the original questionnaires, we consider this an unlikely explanation for the high levels of agreement that we observed. Firstly, the interval between completing the two questionnaires was 4.4 months on average. Secondly, the size and scale of the questionnaire (22 pages for women, 18 pages for men) argues against recall of previous responses to particular items. We therefore consider the responses to the questionnaire at the two time points to be independent. However, given the long interval between for some participants, we cannot exclude the possibility that some discordant responses may actually reflect real changes in exposure.

In summary, we have demonstrated substantial levels of agreement for questionnaire items measuring body weight, consumption of tobacco and alcohol and reasonable agreement for use of various anti-inflammatory medications and common digestive and gynecologic conditions. The impact of modest random error on risk estimates is discernible but not large. We conclude that these items are acceptable for the purposes of risk estimation and epidemiological inference.

Acknowledgements

This research was supported by grants from the United States Department of Defense (OC000109), the Queensland Cancer Fund and the National Health and Medical Research Council of Australia (199600). Its contents are solely the responsibility of the authors. Penelope Webb and David Whiteman are Senior Research Fellows of the Queensland Cancer Fund and National Health and Medical Research Council of Australia respectively. Women with ovarian cancer were recruited through the Australian Ovarian Cancer Study. We thank the participants for their time and efforts in completing the study questionnaires on two occasions and the attending clinicians who assisted this research.

References

Adami HO, Trichopoulos D (2003). Obesity and mortality from cancer. *N Engl J Med*, **348**, 1623-4.

Calle EE, Kaaks R (2004). Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat Rev Cancer*, **4**, 579-91.

Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ (2003). Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med*, **348**, 1625-38.

Donato F, Boffetta P, Fazioli R, et al (1998). Reliability of data on smoking habit and coffee drinking collected by personal interview in a hospital-based case-control study. *Eur J Epidemiol*, **14**, 259-67.

Gartner CE, Battistutta D, Dunne MP, et al (2005). Test-retest repeatability of self-reported environmental exposures in Parkinson's disease cases and healthy controls. *Parkinsonism Relat Disord*, **11**, 287-95.

Kenkel D, Lillard DR, Mathios A (2003). Smoke or fog? The usefulness of retrospectively reported information about smoking. *Addiction*, **98**, 1307-13.

Landis JR, Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics*, **34**, 159-74.

Langer RD, White E, Lewis CE, et al (2003). The Women's Health Initiative Observational Study: baseline characteristics of participants and reliability of baseline measures. *Ann Epidemiol*, **13**, S107-21.

Locke GR, Talley NJ, Weaver AL, Zinsmeister AR (1994). A new questionnaire for gastroesophageal reflux disease. *Mayo Clin Proc*, **69**, 539-47.

Rohan TE, Record SJ, Cook MG (1988). Repeatability of interview-derived socio-demographic and medical information. *J Clin Epidemiol*, **41**, 763-70.

Talley NJ, Boyce PM, Owen BK, et al (1995). Initial validation of a bowel symptom questionnaire and measurement of chronic gastrointestinal symptoms in Australians. *Aust N Z J Med*, **25**, 302-8.