BMC
Medical Research Methodology

RESEARCH ARTICLE   **Open Access**

A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures

Amalia Karahalios[1,2], Laura Baglietto[1,2], John B Carlin[2,3], Dallas R English[1,2] and Julie A Simpson[1,2]*

Abstract

Background: Retaining participants in cohort studies with multiple follow-up waves is difficult. Commonly, researchers are faced with the problem of missing data, which may introduce biased results as well as a loss of statistical power and precision. The STROBE guidelines von Elm et al. (Lancet, 370:1453-1457, 2007); Vandenbroucke et al. (PLoS Med, 4:e297, 2007) and the guidelines proposed by Sterne et al. (BMJ, 338:b2393, 2009) recommend that cohort studies report on the amount of missing data, the reasons for non-participation and non-response, and the method used to handle missing data in the analyses. We have conducted a review of publications from cohort studies in order to document the reporting of missing data for exposure measures and to describe the statistical methods used to account for the missing data.

Methods: A systematic search of English language papers published from January 2000 to December 2009 was carried out in PubMed. Prospective cohort studies with a sample size greater than 1,000 that analysed data using repeated measures of exposure were included.

Results: Among the 82 papers meeting the inclusion criteria, only 35 (43%) reported the amount of missing data according to the suggested guidelines. Sixty-eight papers (83%) described how they dealt with missing data in the analysis. Most of the papers excluded participants with missing data and performed a complete-case analysis (n = 54, 66%). Other papers used more sophisticated methods including multiple imputation (n = 5) or fully Bayesian modeling (n = 1). Methods known to produce biased results were also used, for example, Last Observation Carried Forward (n = 7), the missing indicator method (n = 1), and mean value substitution (n = 3). For the remaining 14 papers, the method used to handle missing data in the analysis was not stated.

Conclusions: This review highlights the inconsistent reporting of missing data in cohort studies and the continuing use of inappropriate methods to handle missing data in the analysis. Epidemiological journals should invoke the STROBE guidelines as a framework for authors so that the amount of missing data and how this was accounted for in the analysis is transparent in the reporting of cohort studies.

Keywords: Longitudinal cohort studies, Missing exposure data, Repeated exposure measurement, Missing data methods, Reporting

Background

A growing number of cohort studies are establishing protocols to re-contact participants at various times during follow-up. These waves of data collection provide

* Correspondence: julieas@unimelb.edu.au
[1]Cancer Epidemiology Centre, Cancer Council Victoria, Carlton, VIC, Australia
[2]Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, School of Population Health, The University of Melbourne, Parkville, VIC, Australia
Full list of author information is available at the end of the article

researchers with the opportunity to obtain information regarding changes in the participants' exposure and outcome measures. Incorporating the repeated measures of the exposure in the epidemiological analysis is especially important if the current exposure (or change in exposure) is thought to be more predictive of the outcome than the participants' baseline measurement [1] or the researcher is interested in assessing the effect of a cumulative exposure [2]. The time frames for these follow-up waves of data collection can vary from one to two years up to 20 to 30 years

or even longer post-baseline. Repeated ascertainment of exposure and outcome measures over time can lead to missing data for reasons such as participants not being traceable, too sick to participate, withdrawing from the study, refusing to respond to certain questions or death [3,4]. In this paper we focus on missing data in exposure measures that are made repeatedly in a cohort study because studies of this type (in which the outcome is often a single episode of disease or death obtained from a registry and therefore, known for all participants) are common and increasingly important in chronic disease epidemiology. Further research is needed on the consequences of and best methods for handling missing data in such study designs, but simulation and case studies have shown that missing covariate data can lead to biased results and there may be gains in precision of estimation of effects if multiple imputation is used to handle missing covariate data [5-7].

If participants with missing data and complete data differ with respect to exposure and outcome, estimates of association based on fully observed cases (known as a complete-case analysis) might be biased. Further, the estimates from these analyses will have less precision than an analysis of all participants in the absence of missing data. As well as complete-case analysis, there are other methods available for dealing with missing data in the statistical analysis [8,9]. These include *ad hoc* methods such as Last Observation Carried Forward and the missing indicator method, and more advanced approaches such as multiple imputation and likelihood-based formulations.

The STROBE guidelines for reporting of observational studies, published in 2007, state that the method for handling missing data should be addressed and furthermore, that the number of individuals used for analysis at each stage of the study should be reported accompanied by reasons for non-participation or non-response [10,11]. The guidelines published by Sterne et al. [12], an extension to the STROBE guidelines, provide general recommendations for the reporting of missing data in any study affected by missing data and specific recommendations for reporting the details of multiple imputation.

In this paper we: 1) give a brief review of the statistical methods that have been proposed for handling missing data and when they may be appropriate; 2) review how missing exposure data has been reported in large cohort studies with one or more waves of follow-up, where the repeated waves of exposures were incorporated in the statistical analyses; and 3) report how the same studies dealt with missing data in the statistical analyses.

Methods
Statistical methods for handling missing data
Complete-case analysis only includes in the analysis participants with complete data on all waves of data collection,

thereby potentially reducing the precision of the estimates of the exposure-outcome associations [2]. The advantage of using complete-case analysis is that it is easily implemented, with most software packages using this method as the default. The estimates of the associations of interest may be biased if the participants with missing data are not similar to those with complete data. To be valid, complete-case analyses must assume that participants with missing data can be thought of as a random sample of those that were intended to be observed (commonly referred to in the missing data nomenclature as missing completely at random (MCAR) [13]), or at least that the likelihood of exposure being missing is independent of the outcome given the exposures [5].

There are three commonly used *ad hoc* approaches for handling missing data, all of which can lead to bias [3,12,14]. The *Last Observation Carried Forward (LOCF)* method replaces the missing value in a wave of data collection with the non-missing value from the previous completed wave for the same individual. The assumption behind this approach is that the exposure status of the individual has not changed over time. The *mean value substitution* method replaces the missing value with the average value calculated over all the values available from the other waves of data collection for the same individual. Both LOCF and mean value substitution falsely increase the stated precision of the estimates by failing to account for the uncertainty due to the missing data and generally give biased results, even when the data are MCAR [7,15]. The *Missing Indicator Method* is applied to categorical exposures and includes an extra category of the exposure variable for those individuals with missing data. Indicator variables are created for the analysis, including an indicator for the missing data category [16]. This method is simple to implement, but also produces biased results in many settings, even when the data are MCAR [6,12].

Multiple Imputation (MI) begins by imputing values for the missing data multiple times by sampling from an imputation model (using either chained equations [17,18] or a multivariate normal model [19]). The imputation model should contain the variables that are to be included in the statistical model used for the epidemiological analysis, as well as auxiliary variables that may contain information about the missing data, and a "proper" imputation procedure incorporates appropriate variability in the imputed values. The imputation process creates multiple 'completed' versions of the datasets. These 'completed datasets' are analysed using the appropriate statistical model for the epidemiological analysis and the estimates obtained from each dataset are averaged to produce one overall MI estimate. The standard error for this overall MI estimate is derived using Rubin's rules, which account for variability between-and

within- the estimates obtained from the separate analyses of the 'completed datasets' [3,13]. By accounting for the variability between the completed (imputed) datasets, MI produces a valid estimate of the precision of the final MI estimate. When the imputation is performed using standard methods that are now available in many packages, with appropriate model specifications to reflect the structure of the data, the resulting MI estimate will be valid (unbiased parameter estimates with nominal confidence interval coverage) if the missing data are 'Missing At Random' (MAR) [5]. MAR describes a situation where the probability of being missing for a particular variable (e.g. waist circumference) can be explained by other observed variables in the dataset, but is (conditionally) independent of the variable itself (that is, waist circumference) [13]. On the other hand, MI may produce biased estimates if the data are 'Missing Not At Random' (MNAR), which occurs when the study participants with missing data differ from the study participants with complete data in a manner that cannot be explained by the observed data in the study [13].

MI is now implemented in many major statistical packages (including Stata [20] and SAS [21]) making it an easily accessible method. However, it can be a time-intensive process to impute multiple datasets, analyse the 'completed datasets' and combine the results; and the imputation model can be complex since it must contain the exposure and outcome variables included in the analysis model, auxiliary variables and any interactions that will be included in the final analysis model [22,23]. Sterne et al. [12] have described a number of pitfalls that can be encountered in the imputation procedure that might lead to biased results for the epidemiological analysis of interest.

Missing data can also be handled with the following more sophisticated methods: maximum likelihood-based formulations, fully Bayesian models and weighting methods. Likelihood-based methods use all of the available information (i.e. information from participants with both complete and incomplete data) to simultaneously estimate both the missing data model and the data analysis model, eliminating the need to handle the missing data directly [3,8,24,25], although in many cases the MAR assumption is also invoked to enable the missing data model to be ignored. Bayesian models also rely on a fully specified model that incorporates both the missingness process and the associations of interest [12,15,26]. Weighting methods apply weights that correspond to the inverse probability of a data observation being observed, to the observed data to account for the missing data [22,25]. These methods may improve the precision of the estimates compared with complete-case analysis. However, they are also dependent on assumptions about the missingness mechanism and in

some cases on specifying the correct missingness model. In general, these methods require tailored programming which can be time consuming and requires specialist expertise [15].

Criteria for considering studies in this review

For this review we selected prospective cohort studies that analysed exposure data collected after initial recruitment during the follow-up period (i.e. studies looking at a change in exposure or at a time varying covariate). We restricted our review to cohort studies with more than 1,000 participants, as we thought it was more likely for there to be more missing data in follow-up measurements of exposures in large cohort studies (typically population based studies) compared to small cohorts (often based on a specific clinical population). For cohort studies reported in multiple papers, we included only the most recent original research article. Studies that only used data collected at baseline or at one of the follow-up waves in the analysis, and studies that newly recruited participants at one of the waves after baseline were excluded. We did not place any restrictions on the types of exposures or outcomes studied or the type of statistical analysis performed.

Search strategy

PubMed was searched for English language papers published between January 2000 and December 2009. We chose January 2000 as a starting date because the first widely available statistical software package for implementing MI, the NORM package [27], was developed in 1997 and updated in 1999. Search terms included: "Cohort Studies"[MeSh] AND ("attrition" OR "drop out" OR "longitudinal change" OR "missing data" OR "multiple exposure" OR "multiple follow-up" OR "multiple waves" OR "repeated exposure" OR "repeated follow-up" OR "repeated waves" OR "repeated measures" OR "time dependent covariates" OR "time dependent" OR "time varying covariate" OR "cumulative average").

We carried out a further search of cohort studies listed in the web appendix of the paper by Lewington et al. [28], to ensure that any known large cohort studies were not missed in the original PubMed search. These cohort studies were established in the 1970s and 1980s, allowing them time to measure repeated waves of exposure on their participants and to publish these results during our study period (i.e. between 2000 and 2009).

Methods of the review

AK reviewed all articles; any uncertainties regarding the statistical method used to handle the missing data were resolved by discussion with JAS, and AK extracted the data. Additional tables and methods sections from journal websites were checked if referred to in the article.

Table 1 Summary of cohort studies included in the review

Characteristics	Number of studies (N = 82)	(%)
Publication year		
2000	2	(2)
2001	4	(5)
2002	4	(5)
2003	5	(6)
2004	6	(7)
2005	8	(10)
2006	11	(13)
2007	13	(16)
2008	13	(16)
2009	16	(20)
Number of participants recruited at baseline		
Not stated in paper	4	(5)
1,000–2,000	15	(18)
2,000–3,000	13	(16)
3,000–5,000	14	(17)
5,000–10,000	13	(16)
10,000–20,000	14	(17)
20,000 +	9	(11)
Date of baseline recruitment (decade recruitment started)		
Not stated in paper	5	(6)
Before 1970	9	(11)
1970–1979	9	(11)
1980–1989	25	(30)
1990–1999	30	(37)
2000–2009	4	(5)
Number of follow-up waves used in the analysis		
Number not stated	12	(15)
Mean or range given	7	(9)
1	10	(12)
2	9	(11)
3	18	(22)
4	12	(15)
≥5	14	(17)
Statistical methods for epidemiological analysis†		
Cox proportional hazards regression^	37‡	
Time-varying covariates	35	
Time-invariant covariates	3	
Generalised Estimating Equations	12*	
Linear regression	3	
Logistic regression	10	
Generalised linear mixed-effects modelling	16	
Linear regression	13	

Table 1 Summary of cohort studies included in the review (Continued)

Logistic regression	3	
Standard linear regression	3	
Standard logistic regression	9	
Other methods	6	

† The total number of papers is 83 because one paper used two analysis models: a Cox proportional hazards model for a time to event outcome and a linear mixed effects model for a numerical outcome.
^ Note one paper used a parametric survival model.
‡ One paper incorporated their repeated measures of exposures using both a time varying covariate and time-invariant covariates.
*One paper had a numerical and binary outcome.

Our aim was to assess the reporting of missing data and the methods used to handle the missing data according to the recommendations given by the STROBE guidelines [10,11] and Sterne et al. [12]. The information extracted is summarised in Tables 1 and 2 and Additional file 1: Table S1.

Results
Study selection
We identified 4,277 articles via the keyword search. A total of 3,684 articles were excluded based on their title and abstract, leaving 543 articles for further evaluation. Of these, 471 articles were excluded and 72 articles were found to be appropriate for the review. A further ten studies were identified from the reference list of Lewington et al. [28] (Figure 1), giving 82 studies included in this review. The reasons for excluding studies are outlined in Figure 1, the most common reasons were sample size of less than 1,000 participants (54%), study design was not a prospective cohort (19%), and did not report original research findings (13%).

Characteristics of included studies
The characteristics of the 82 studies included are summarised in Table 1 and further details can be found in the additional table (see Additional file 1: Table S1). The studies included ranged from smaller studies that recruited 1,000 to 2,000 participants at baseline to larger studies with more than 20,000 participants, and the number published annually increased steadily from two papers in 2000 to 16 papers in 2009. The majority of studies recruited their participants in the decades 1980 to 1989 (n = 25), and 1990 to 1999 (n = 30). Cox proportional hazards regression was the most common statistical method used for the epidemiological analysis (n = 37) to analyse the repeated measures of exposure, with 35 of these papers incorporating the repeated exposure(s) as a time varying covariate and the remaining two papers including a single measure of the covariate derived from repeated assessments. Generalised Estimating Equations with a logistic (n = 10) or linear regression (n = 3) and generalised linear

Table 2 Missing data features reported by the studies

Features reported	Number of papers (N = 82) (%)
Was the amount of missing data reported?	
Yes	66 (80%)
Missing data reported for each follow-up wave used in the analysis	35
A general statement was made about the amount of missing data or the amount of completed follow-up (how many participants attended at least one wave or only the final follow-up wave)	22
Indicated number that completed all waves of follow-up (i.e. number included in final sample)	6
Indicated amount missing for certain (key) variables	3
No	16 (20%)
Assessed differences between individuals with complete data and those with incomplete data?	
Yes	26 (32%)
Provided a table comparing distributions of key exposures and outcome variables for those with missing and non-missing information	6
Table not provided but some summary statistics included in text	4
General comment provided (did not include a table or summary statistics or included p-values only)	16
No	56 (69%)
Reasons were given for the missing data	13 (16%)
Statistical method for handling missing data†	
Method not stated	14 (16%)
Complete-case analysis assumed	9 (11%)
Complete-case analysis	54 (66%)
Weighted	1
Unweighted	53
Exclude participants with missing data at any repeated waves of exposure	38
Exclude participant data record for waves of data collection with missing exposure data††	15
Missing Indicator Method	1 (1%)
Mean value substitution	3 (4%)
Last Observation Carried Forward	7 (9%)
Multiple Imputation	5 (6%)
Details provided for the multiple imputation:	
Indicated how many imputations were performed	4
Indicated which variables were included in the imputation model	2
Compared results from multiple imputation with complete case analysis	3
Performed a sensitivity analysis under different assumptions for missing data	4

Table 2 Missing data features reported by the studies
(Continued)

Fully Bayesian Model	1 (1%)

† Three papers used more than one method to handle their missing exposure data.
†† These studies assessed both exposure and outcome measures repeatedly over the waves of data collection. The data were analysed using either Generalised Estimating Equations (GEEs) or mixed-effects models to account for the correlated outcome data within an individual and excluded participant data records for waves where the exposure data were missing.

mixed-effects models (logistic regression (n = 3) and linear regression (n = 13)) were the next most common epidemiological analyses used.

### Missing covariate data at follow-up

The methods used by the selected papers for handling missing data are summarised in Table 2. Sixty-six papers (80%) commented on the amount of missing data at follow-up. Of these, only 35 papers provided information about the proportion of participants lost to follow-up at each wave. The remaining 31 papers provided incomplete details about the amount of missing data at each wave: 22 papers made a general comment about the amount of missing data; six papers reported the amount of missing data for the final wave but gave no detail regarding the number of participants available at previous waves of data collection (including baseline); and three papers only reported the amount of missing data for a few of the variables.

Of the 29 papers published after 2007, nine papers did not state the proportion of missing data at each follow up wave, three papers provided a comment as to why the data were missing and eight papers compared the baseline covariates for those with and without missing covariate data at the repeated waves of follow up.

Among those papers that provided information on missing data, the proportion of covariate data missing at any follow-up wave ranged from 2% to 65%. Twenty-six papers (32%) compared the key variables of interest for those who did and did not have data from post-baseline waves, but only six of these presented the results in detail while the rest commented briefly in the text on whether or not there was a difference.

Methods used to deal with missing data at follow-up

The most common methods used to deal with missing data were complete-case analysis (n = 54), LOCF (n = 7) and MI (n = 5). Of the 54 papers that used complete-case analysis: 38 excluded participants who were missing exposure data at any of the waves of data collection from the analysis; one paper also excluded participants with any missing exposure data but used a weighted analysis to deal with the missing data; and the remaining 15 papers, where both the exposure and outcome measures were assessed repeatedly at each wave of data collection,

Figure 1 Search results.
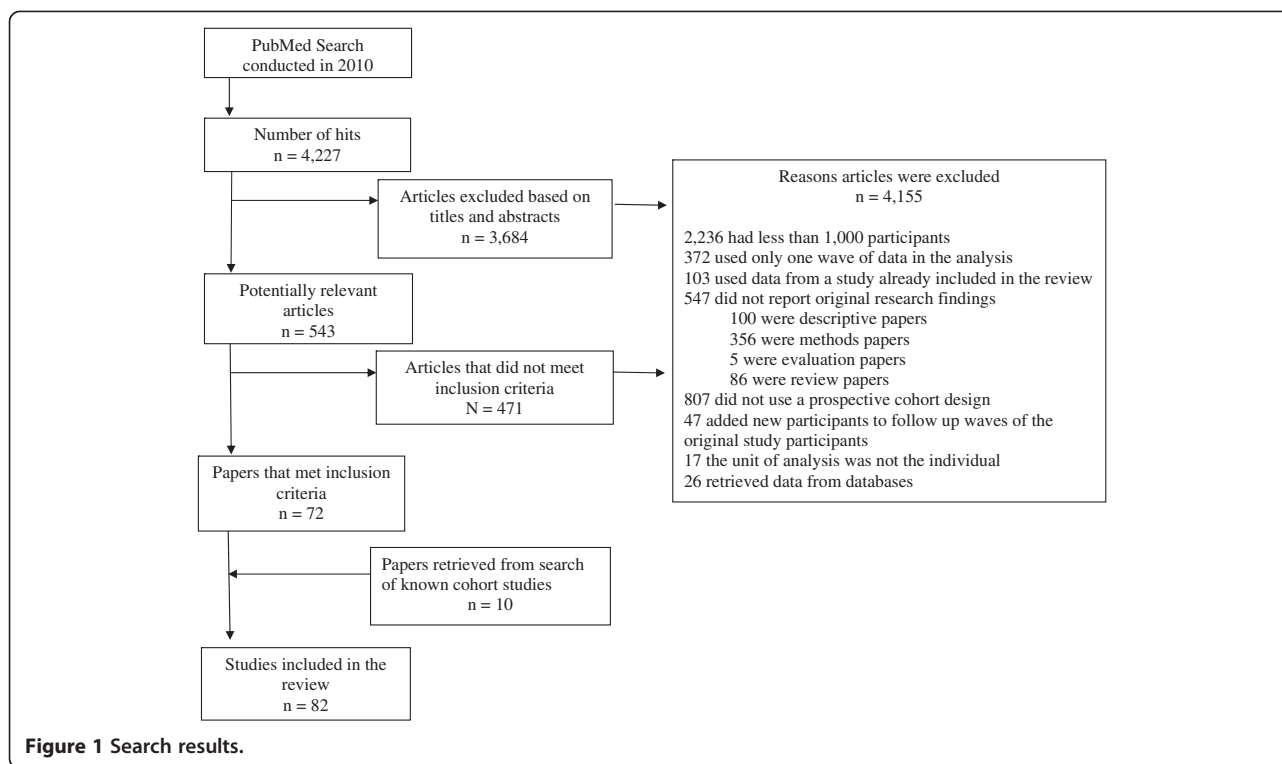
excluded participant data records for waves where the exposure data were missing. Fourteen papers did not state the method used to deal with the missing data, although nine of these papers performed a Cox regression model using SAS [21] or Stata [20] and we therefore assumed that they used a complete-case analysis (Table 2). Both papers published in 2000 used complete-case analysis. From 2001 to 2009, the proportion of papers using complete-case analysis ranged from 25% to 65%. Methods known to produce biased results (i.e. LOCF, the missing indicator method and mean value substitution) continue to be used, with four papers using these methods in 2009.

Of the five papers that used MI [29-33], two papers [29,30] compared the characteristics of the participants with and without missing data. For the MI, three of the five papers [30,31,33] provided details of the imputation process including the number of imputations performed and the variables included in the imputation model, and compared the results from the MI analysis to results from complete-case analysis. The other two papers [29,32] provided details about the number of imputations performed but did not describe the variables included in their imputation model and did not compare the MI results to the complete-case analysis.

Discussion

We identified 82 cohort studies of 1,000 or more participants that were published from 2000 to 2009 and which

analysed exposure data collected from repeated follow-up waves. The reporting of missing data in these studies was found to be inconsistent and generally did not follow the recommendations set out by the STROBE guidelines [10,11] or the guidelines set out by Sterne et al. [12]. The STROBE guidelines recommend that authors report the number of participants who take part in each wave of the study and give reasons why participants did not attend a wave. Only three papers [30,34,35] followed the STROBE guidelines fully. The majority of papers did not provide a reason or comment for why study participants did not attend each wave of follow-up. Sterne et al. [12] recommend that the reasons for missing data be described with respect to other variables and that authors investigate potentially important differences between participants and non-participants.

The STROBE guidelines were published in 2007. Of the nine papers published after 2007, only one followed the STROBE guidelines fully. This suggests that either journal editors are not using these guidelines or authors are not considering the impact of missing covariate data in their research.

A review of missing data in cancer prognostic studies published in 2004 by Burton et al. [36] and a review of developmental psychology studies published in 2009 by Jelicic et al. [3] reported similar findings to ours. Burton et al. [36] found a deficiency in the reporting of missing covariate data in cancer prognostic studies. After reviewing 100 articles, they found that only 40% of articles

provided information about the method used to handle missing covariate data and only 12 articles would have satisfied their proposed guidelines for the reporting of missing data. We observed in our review, of articles published from 2000 to 2009, that a larger proportion of articles reported the method used to handle the missing data in the analysis but that many articles were still not reporting the amount of missing data and the reasons for missingness.

The cohort studies we identified used numerous methods to handle missing data in the exposure-outcome analyses. Although some studies used advanced statistical modelling procedures (e.g. MI and Bayesian), the majority removed individuals with missing data and performed a complete-case analysis; a method that may produce biased results if the missing data are not MCAR. Jelicic et al. also found in their review that a large proportion of studies used complete-case analysis to handle their missing data [3]. For studies with a large proportion of missing data, excluding participants with missing data may also reduce the precision of the analysis substantially. *Ad hoc* methods (e.g. LOCF, the missing indicator method and mean value substitution), which are generally not recommended [16,25] because they fail to account for the uncertainty in the data and may produce biased estimates [12], continue to be used. Although MI is becoming more accessible, only five studies used this method. The reporting of the imputation procedure was inconsistent and often incomplete. This was also observed by two independent reviews of the reporting of MI in the medical journals: *BMJ, JAMA, Lancet* and the *New England Journal of Medicine* [12,37]. Future studies should follow the recommendations outlined by Sterne et al. [12] to ensure that enough details are provided about the MI procedure, especially the implementation and details of the imputation modelling process.

Strengths and limitations of the literature review

We aimed to complete a comprehensive review of all papers published that analysed exposure variables measured at multiple follow-up waves. Several keywords were used in order to obtain as many articles as possible. The keyword search was then supplemented with cohort studies identified from a pooled analysis of 61 cohort studies. Although a large number of abstracts and studies were identified, some cohort studies might have been missed. If multiple papers were identified from one study, the most recent article was included in the review, which might have led us to omit papers from the same study that used a more appropriate missing data method. Our search criteria only included papers written in English and only PubMed was searched. Our search strategy was limited to articles published between 2000

and 2009. On average three papers of the type we focussed on were published each year from 2000 to 2002 and the number has increased since then, so it seems unlikely that many papers were published before this time. Also, MI was not as accessible prior to 1997, so papers published before 2000 were more likely to have used complete case analysis or other *ad hoc* methods.

Conclusions

With the increase in the number of cohort studies analysing data with multiple follow-up waves it is essential that authors follow the STROBE guidelines [10,11] in conjunction with the guidelines proposed by Sterne et al. [12] to report on the amount of missing data in the study and the methods used to handle the missing data in the analyses. This will ensure that missing data are reported with enough detail to allow readers to assess the validity of the results. Incomplete data and the statistical methods used to deal with the missing data can lead to bias, or be inefficient, and so authors should be encouraged to use online supplements (if necessary) as a way of publishing both the details of the missing data in their study and the details of the methods used to deal with the missing data.

Additional file

Additional file 1: Table S1. Detailed characteristics of the studies included in the systematic review. Details of studies included in the systematic review and the corresponding reference list [29-35, 38-112].

Author details
[1]Cancer Epidemiology Centre, Cancer Council Victoria, Carlton, VIC, Australia. [2]Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, School of Population Health, The University of Melbourne, Parkville, VIC, Australia. [3]Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, VIC, Australia.

Authors' contributions
AK drafted the protocol for the review, reviewed the articles and drafted the manuscript. JAS conceived of the review, resolved any discrepancies encountered by AK when reviewing the articles and helped with drafting the manuscript. LB, JBC and DRE provided feedback on the design of the protocol and drafts of the manuscript. All authors read and approved the final manuscript.

References
1. Cupples LA, D'Agostino RB, Anderson K, Kannel WB: **Comparison of baseline and repeated measure covariate techniques in the Framingham Heart Study.** *Stat Med* 1988, **7**:205–222.
2. Shortreed SM, Forbes AB: **Missing data in the exposure of interest and marginal structural models: a simulation study based on the Framingham Heart Study.** *Stat Med* 2010, **29**:431–443.

3. Jelicic H, Phelps E, Lerner RM: **Use of missing data methods in longitudinal studies: the persistence of bad practices in developmental psychology.** *Dev Psychol* 2009, **45**:1195–1199.
4. Kurland BF, Johnson LL, Egleston BL, Diehr PH: **Longitudinal data with follow-up truncated by death: match the analysis method to research aims.** *Stat Sci* 2009, **24**:211.
5. White IR, Carlin JB: **Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values.** *Stat Med* 2010.
6. Knol MJ, Janssen KJ, Donders AR, Egberts AC, Heerdink ER, Grobbee DE, Moons KG, Geerlings MI: **Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example.** *J Clin Epidemiol* 2010, **63**:728–736.
7. Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, Carroll RJ: **Analyzing incomplete longitudinal clinical trial data.** *Biostatistics* 2004, **5**:445–464.
8. Schafer JL, Graham JW: **Missing data: our view of the state of the art.** *Psychol Methods* 2002, **7**:147–177.
9. Carpenter J, Kenward MG: *A critique of common approaches to missing data, Missing data in randomised controlled trials — a practical guide.* Birmingham, AL: National Institute for Health Research; 2007.
10. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP: **The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies.** *Lancet* 2007, **370**:1453–1457.
11. Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M: **Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration.** *PLoS Med* 2007, **4**:e297.
12. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR: **Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls.** *BMJ* 2009, **338**:b2393.
13. Little RJA, Rubin DB: *Statistical analysis with missing data.* Hoboken, New Jersey: John Wiley & Sons, Inc; 2002.
14. Rubin DB: *Multiple imputation for nonresponse in surveys.* New York: John Wiley & Sons; 1987.
15. Buhi ER, Goodson P, Neilands TB: **Out of sight, not out of mind: strategies for handling missing data.** *Am J Heal Behav* 2008, **32**:83–92.
16. Greenland S, Finkle WD: **A critical look at methods for handling missing covariates in epidemiologic regression analyses.** *Am J Epidemiol* 1995, **142**:1255–1264.
17. Azur MJ, Stuart EA, Frangakis C, Leaf PJ: **Multiple imputation by chained equations: what is it and how does it work?** *Int J Methods Psychiatr Res* 2011, **20**:40–49.
18. White IR, Royston P, Wood AM: **Multiple imputation using chained equations: issues and guidance for practice.** *Stat Med* 2011, **30**:377–399.
19. Lee KJ, Carlin JB: **Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation.** *Am J Epidemiol* 2010, **171**:624–632.
20. StataCorp: Stata Statistical Software: *Release 11.* College Station, TX: StataCorp LP; 2009.
21. SAS Insitute Inc: *SAS OnlineDoc, Version 8.* Cary, NC: SAS Institute, Inc.; 2000.
22. Carpenter JR, Kenward MG, Vansteelandt S: **A comparison of multiple imputation and doubly robust estimation for analyses with missing data.** *J R Stat Soc: Series A (Stat Soc)* 2006, **169**:571–584.
23. Graham JW: **Missing data analysis: making it work in the real world.** *Annu Rev Psychol* 2009, **60**:549–576.
24. Enders CK: **A primer on maximum likelihood algorithms available for use with missing data.** *Struct Equ Model* 2001, **8**:128–141.
25. Horton NJ, Kleinman KP: **Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models.** *Am Stat* 2007, **61**:79–90.
26. Baraldi AN, Enders CK: **An introduction to modern missing data analyses.** *J Sch Psychol* 2010, **48**:5–37.
27. Schafer J, Yucel R: *PAN: Multiple imputation for multivariate panel data (Software)* 1999, **86**:949–955.
28. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R: **Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies.** *Lancet* 2002, **360**:1903–1913.
29. Bond GE, Burr RL, McCurry SM, Rice MM, Borenstein AR, Larson EB: **Alcohol and cognitive performance: a longitudinal study of older Japanese Americans. The Kame Project.** *Int Psychogeriatr* 2005, **17**:653–668.
30. Kivimaki M, Lawlor DA, Singh-Manoux A, Batty GD, Ferrie JE, Shipley MJ, Nabi H, Sabia S, Marmot MG, Jokela M: **Common mental disorder and obesity: insight from four repeat measures over 19 years: prospective Whitehall II cohort study.** *BMJ* 2009, **339**:b3765.
31. McCormack VA, Dos Santos Silva I, De Stavola BL, Perry N, Vinnicombe S, Swerdlow AJ, Hardy R, Kuh D: **Life-course body size and perimenopausal mammographic parenchymal patterns in the MRC 1946 British birth cohort.** *Br J Cancer* 2003, **89**:852–859.
32. Sugihara Y, Sugisawa H, Shibata H, Harada K: **Productive roles, gender, and depressive symptoms: evidence from a national longitudinal study of late-middle-aged Japanese.** *J Gerontol B Psychol Sci Soc Sci* 2008, **63**:P227–P234.
33. Wiles NJ, Haase AM, Gallacher J, Lawlor DA, Lewis G: **Physical activity and common mental disorder: results from the Caerphilly study.** *Am J Epidemiol* 2007, **165**:946–954.
34. Fuhrer R, Dufouil C, Dartigues JF: **Exploring sex differences in the relationship between depressive symptoms and dementia incidence: prospective results from the PAQUID Study.** *J Am Geriatr Soc* 2003, **51**:1055–1063.
35. Sogaard AJ, Meyer HE, Tonstad S, Haheim LL, Holme I: **Weight cycling and risk of forearm fractures: a 28-year follow-up of men in the Oslo Study.** *Am J Epidemiol* 2008, **167**:1005–1013.
36. Burton A, Altman DG: **Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines.** *Br J Cancer* 2004, **91**:4–8.
37. Mackinnon A: **The use and reporting of multiple imputation in medical research - a review.** *J Intern Med* 2010, **268**:586–593.
38. Agrawal A, Grant JD, Waldron M, Duncan AE, Scherrer JF, Lynskey MT, Madden PA, Bucholz KK, Heath AC: **Risk for initiation of substance use as a function of age of onset of cigarette, alcohol and cannabis use: findings in a Midwestern female twin cohort.** *Prev Med* 2006, **43**:125–128.
39. Anstey KJ, Hofer SM, Luszcz MA: **Cross-sectional and longitudinal patterns of dedifferentiation in late-life cognitive and sensory function: the effects of age, ability, attrition, and occasion of measurement.** *J Exp Psychol Gen* 2003, **132**:470–487.
40. Arifeen S, Black RE, Antelman G, Baqui A, Caulfield L, Becker S: **Exclusive breastfeeding reduces acute respiratory infection and diarrhea deaths among infants in Dhaka slums.** *Pediatrics* 2001, **108**:E67.
41. Bada HS, Das A, Bauer CR, Shankaran S, Lester B, LaGasse L, Hammond J, Wright LL, Higgins R: **Impact of prenatal cocaine exposure on child behavior problems through school age.** *Pediatrics* 2007, **119**:e348–359.
42. Beesdo K, Bittner A, Pine DS, Stein MB, Hofler M, Lieb R, Wittchen HU: **Incidence of social anxiety disorder and the consistent risk for secondary depression in the first three decades of life.** *Arch Gen Psychiatry* 2007, **64**:903–912.
43. Berecki-Gisolf J, Begum N, Dobson AJ: **Symptoms reported by women in midlife: menopausal transition or aging?** *Menopause* 2009, **16**:1021–1029.
44. Blazer DG, Sachs-Ericsson N, Hybels CF: **Perception of unmet basic needs as a predictor of depressive symptoms among community-dwelling older adults.** *J Gerontol A Biol Sci Med Sci* 2007, **62**:191–195.
45. Bray JW, Zarkin GA, Ringwalt C, Qi J: **The relationship between marijuana initiation and dropping out of high school.** *Health Econ* 2000, **9**:9–18.
46. Breslau N, Schultz LR, Johnson EO, Peterson EL, Davis GC: **Smoking and the risk of suicidal behavior: a prospective study of a community sample.** *Arch Gen Psychiatry* 2005, **62**:328–334.
47. Brown JW, Liang J, Krause N, Akiyama H, Sugisawa H, Fukaya T: **Transitions in living arrangements among elders in Japan: does health make a difference?** *J Gerontol B Psychol Sci Soc Sci* 2002, **57**:S209–220.
48. Bruckl TM, Wittchen HU, Hofler M, Pfister H, Schneider S, Lieb R: **Childhood separation anxiety and the risk of subsequent psychopathology: Results from a community study.** *Psychother Psychosom* 2007, **76**:47–56.
49. Cauley JA, Lui LY, Barnes D, Ensrud KE, Zmuda JM, Hillier TA, Hochberg MC, Schwartz AV, Yaffe K, Cummings SR, Newman AB: **Successful skeletal aging: a marker of low fracture risk and longevity. The Study of Osteoporotic Fractures (SOF).** *J Bone Miner Res* 2009, **24**:134–143.
50. Celentano DD, Munoz A, Cohn S, Vlahov D: **Dynamics of behavioral risk factors for HIV/AIDS: a 6-year prospective study of injection drug users.** *Drug Alcohol Depend* 2001, **61**:315–322.

51. Chao C, Jacobson LP, Tashkin D, Martinez-Maza O, Roth MD, Margolick JB, Chmiel JS, Holloway MN, Zhang ZF, Detels R: **Recreational amphetamine use and risk of HIV-related non-Hodgkin lymphoma.** *Cancer Causes Control* 2009, **20**:509–516.

52. Cheung YB, Khoo KS, Karlberg J, Machin D: **Association between psychological symptoms in adults and growth in early life: longitudinal follow up study.** *BMJ* 2002, **325**:749.

53. Chien KL, Hsu HC, Sung FC, Su TC, Chen MF, Lee YT: **Hyperuricemia as a risk factor on cardiovascular events in Taiwan: the Chin-Shan Community Cardiovascular Cohort Study.** *Atherosclerosis* 2005, **183**:147–155.

54. Clays E, De Bacquer D, Leynen F, Kornitzer M, Kittel F, De Backer G: **Job stress and depression symptoms in middle-aged workers–prospective results from the Belstress study.** *Scand J Work Environ Health* 2007, **33**:252–259.

55. Conron KJ, Beardslee W, Koenen KC, Buka SL, Gortmaker SL: **A longitudinal study of maternal depression and child maltreatment in a national sample of families investigated by child protective services.** *Arch Pediatr Adolesc Med* 2009, **163**:922–930.

56. Cuddy TE, Tate RB: **Sudden unexpected cardiac death as a function of time since the detection of electrocardiographic and clinical risk factors in apparently healthy men: the Manitoba Follow-Up Study, 1948 to 2004.** *Can J Cardiol* 2006, **22**:205–211.

57. Daniels MC, Adair LS: **Growth in young Filipino children predicts schooling trajectories through high school.** *J Nutr* 2004, **134**:1439–1446.

58. de Mutsert R, Grootendorst DC, Boeschoten EW, Brandts H, van Manen JG, Krediet RT, Dekker FW: **Subjective global assessment of nutritional status is strongly associated with mortality in chronic dialysis patients.** *Am J Clin Nutr* 2009, **89**:787–793.

59. De Stavola BL, Meade TW: **Long-term effects of hemostatic variables on fatal coronary heart disease: 30-year results from the first prospective Northwick Park Heart Study (NPHS-I).** *J Thromb Haemost* 2007, **5**:461–471.

60. Di Nisio M, Barbui T, Di Gennaro L, Borrelli G, Finazzi G, Landolfi R, Leone G, Marfisi R, Porreca E, Ruggeri M, *et al*: **The haematocrit and platelet target in polycythemia vera.** *Br J Haematol* 2007, **136**:249–259.

61. Engberg J, Morral AR: **Reducing substance use improves adolescents' school attendance.** *Addiction* 2006, **101**:1741–1751.

62. Fergusson DM, Boden JM, Horwood LJ: **The developmental antecedents of illicit drug use: evidence from a 25-year longitudinal study.** *Drug Alcohol Depend* 2008, **96**:165–177.

63. Fung TT, Malik V, Rexrode KM, Manson JE, Willett WC, Hu FB: **Sweetened beverage consumption and risk of coronary heart disease in women.** *Am J Clin Nutr* 2009, **89**:1037–1042.

64. Gallo WT, Bradley EH, Dubin JA, Jones RN, Falba TA, Teng HM, Kasl SV: **The persistence of depressive symptoms in older workers who experience involuntary job loss: results from the health and retirement survey.** *J Gerontol B Psychol Sci Soc Sci* 2006, **61**:S221–228.

65. Gauderman WJ, Avol E, Gilliland F, Vora H, Thomas D, Berhane K, McConnell R, Kuenzli N, Lurmann F, Rappaport E, *et al*: **The effect of air pollution on lung development from 10 to 18 years of age.** *N Engl J Med* 2004, **351**:1057–1067.

66. Glotzer TV, Daoud EG, Wyse DG, Singer DE, Ezekowitz MD, Hilker C, Miller C, Qi D, Ziegler PD: **The relationship between daily atrial tachyarrhythmia burden from implantable device diagnostics and stroke risk: the TRENDS study.** *Circ Arrhythm Electrophysiol* 2009, **2**:474–480.

67. Gunderson EP, Jacobs DR Jr, Chiang V, Lewis CE, Tsai A, Quesenberry CP Jr, Sidney S: **Childbearing is associated with higher incidence of the metabolic syndrome among women of reproductive age controlling for measurements before pregnancy: the CARDIA study.** *Am J Obstet Gynecol* 2009, **201**(177):e171–179.

68. Haag MD, Bos MJ, Hofman A, Koudstaal PJ, Breteler MM, Stricker BH: **Cyclooxygenase selectivity of nonsteroidal anti-inflammatory drugs and risk of stroke.** *Arch Intern Med* 2008, **168**:1219–1224.

69. Hart CL, Hole DJ, Davey Smith G: **Are two really better than one? Empirical examination of repeat blood pressure measurements and stroke risk in the Renfrew/Paisley and collaborative studies.** *Stroke* 2001, **32**:2697–2699.

70. Hogg RS, Bangsberg DR, Lima VD, Alexander C, Bonner S, Yip B, Wood E, Dong WW, Montaner JS, Harrigan PR: **Emergence of drug resistance is associated with an increased risk of death among patients first starting HAART.** *PLoS Med* 2006, **3**:e356.

71. Jacobs EJ, Thun MJ, Connell CJ, Rodriguez C, Henley SJ, Feigelson HS, Patel AV, Flanders WD, Calle EE: **Aspirin and other nonsteroidal anti-inflammatory drugs and breast cancer incidence in a large U.S. cohort.** *Cancer Epidemiol Biomarkers Prev* 2005, **14**:261–264.

72. Jamrozik E, Knuiman MW, James A, Divitini M, Musk AW: **Risk factors for adult-onset asthma: a 14-year longitudinal study.** *Respirology* 2009, **14**:814–821.

73. Jimenez M, Krall EA, Garcia RI, Vokonas PS, Dietrich T: **Periodontitis and incidence of cerebrovascular disease in men.** *Ann Neurol* 2009, **66**:505–512.

74. Juhaeri, Stevens J, Chambless LE, Nieto FJ, Jones D, Schreiner P, Arnett D, Cai J: **Associations of weight loss and changes in fat distribution with the remission of hypertension in a bi-ethnic cohort: the Atherosclerosis Risk in Communities Study.** *Prev Med* 2003, **36**:330–339.

75. Karlamangla A, Zhou K, Reuben D, Greendale G, Moore A: **Longitudinal trajectories of heavy drinking in adults in the United States of America.** *Addiction* 2006, **101**:91–99.

76. Keller MC, Neale MC, Kendler KS: **Association of different adverse life events with distinct patterns of depressive symptoms.** *Am J Psychiatry* 2007, **164**:1521–1529. quiz 1622.

77. Kersting RC: **Impact of social support, diversity, and poverty on nursing home utilization in a nationally representative sample of older Americans.** *Soc Work Health Care* 2001, **33**:67–87.

78. Lacson E Jr, Wang W, Lazarus JM, Hakim RM: **Change in vascular access and mortality in maintenance hemodialysis patients.** *Am J Kidney Dis* 2009, **54**:912–921.

79. Lamarca R, Ferrer M, Andersen PK, Liestol K, Keiding N, Alonso J: **A changing relationship between disability and survival in the elderly population: differences by age.** *J Clin Epidemiol* 2003, **56**:1192–1201.

80. Lawson DW, Mace R: **Sibling configuration and childhood growth in contemporary British families.** *Int J Epidemiol* 2008, **37**:1408–1421.

81. Lee DH, Ha MH, Kam S, Chun B, Lee J, Song K, Boo Y, Steffen L, Jacobs DR Jr: **A strong secular trend in serum gamma-glutamyltransferase from 1996 to 2003 among South Korean men.** *Am J Epidemiol* 2006, **163**:57–65.

82. Lee DS, Evans JC, Robins SJ, Wilson PW, Albano I, Fox CS, Wang TJ, Benjamin EJ, D'Agostino RB, Vasan RS: **Gamma glutamyl transferase and metabolic syndrome, cardiovascular disease, and mortality risk: the Framingham Heart Study.** *Arterioscler Thromb Vasc Biol* 2007, **27**:127–133.

83. Li G, Higdon R, Kukull WA, Peskind E, Van Valen Moore K, Tsuang D, van Belle G, McCormick W, Bowen JD, Teri L, *et al*: **Statin therapy and risk of dementia in the elderly: a community-based prospective cohort study.** *Neurology* 2004, **63**:1624–1628.

84. Li LW, Conwell Y: **Effects of changes in depressive symptoms and cognitive functioning on physical disability in home care elders.** *J Gerontol A Biol Sci Med Sci* 2009, **64**:230–236.

85. Limburg PJ, Anderson KE, Johnson TW, Jacobs DR Jr, Lazovich D, Hong CP, Nicodemus KK, Folsom AR: **Diabetes mellitus and subsite-specific colorectal cancer risks in the Iowa Women's Health Study.** *Cancer Epidemiol Biomarkers Prev* 2005, **14**:133–137.

86. Luchenski S, Quesnel-Vallee A, Lynch J: **Differences between women's and men's socioeconomic inequalities in health: longitudinal analysis of the Canadian population, 1994–2003.** *J Epidemiol Community Health* 2008, **62**:1036–1044.

87. Melamed ML, Eustace JA, Plantinga L, Jaar BG, Fink NE, Coresh J, Klag MJ, Powe NR: **Changes in serum calcium, phosphate, and PTH and the risk of death in incident dialysis patients: a longitudinal study.** *Kidney Int* 2006, **70**:351–357.

88. Menotti A, Lanti M, Kromhout D, Blackburn H, Jacobs D, Nissinen A, Dontas A, Kafatos A, Nedeljkovic S, Adachi H: **Homogeneity in the relationship of serum cholesterol to coronary deaths across different cultures: 40-year follow-up of the Seven Countries Study.** *Eur J Cardiovasc Prev Rehabil* 2008, **15**:719–725.

89. Michaelsson K, Olofsson H, Jensevik K, Larsson S, Mallmin H, Berglund L, Vessby B, Melhus H: **Leisure physical activity and the risk of fracture in men.** *PLoS Med* 2007, **4**:e199.

90. Michaud DS, Liu Y, Meyer M, Giovannucci E, Joshipura K: **Periodontal disease, tooth loss, and cancer risk in male health professionals: a prospective cohort study.** *Lancet Oncol* 2008, **9**:550–558.

91. Mirzaei M, Taylor R, Morrell S, Leeder SR: **Predictors of blood pressure in a cohort of school-aged children.** *Eur J Cardiovasc Prev Rehabil* 2007, **14**:624–629.

92. Mishra GD, McNaughton SA, Bramwell GD, Wadsworth ME: **Longitudinal changes in dietary patterns during adult life.** *Br J Nutr* 2006, **96**:735–744.

93. Monda KL, Adair LS, Zhai F, Popkin BM: **Longitudinal relationships between occupational and domestic physical activity patterns and body weight in China.** *Eur J Clin Nutr* 2008, **62**:1318–1325.

94. Moss SE, Klein R, Klein BE: **Long-term incidence of dry eye in an older population.** *Optom Vis Sci* 2008, **85**:668–674.

95. Nabi H, Consoli SM, Chastang JF, Chiron M, Lafont S, Lagarde E: **Type A behavior pattern, risky driving behaviors, and serious road traffic accidents: a prospective study of the GAZEL cohort.** *Am J Epidemiol* 2005, **161**:864–870.

96. Nakano T, Tatemichi M, Miura Y, Sugita M, Kitahara K: **Long-term physiologic changes of intraocular pressure: a 10-year longitudinal analysis in young and middle-aged Japanese men.** *Ophthalmology* 2005, **112**:609–616.

97. Nowicki MJ, Vigen C, Mack WJ, Seaberg E, Landay A, Anastos K, Young M, Minkoff H, Greenblatt R, Levine AM: **Association of cells with natural killer (NK) and NKT immunophenotype with incident cancers in HIV-infected women.** *AIDS Res Hum Retrovir* 2008, **24**:163–168.

98. Ormel J, Oldehinkel AJ, Vollebergh W: **Vulnerability before, during, and after a major depressive episode: a 3-wave population-based study.** *Arch Gen Psychiatry* 2004, **61**:990–996.

99. Rabbitt P, Lunn M, Wong D, Cobain M: **Sudden declines in intelligence in old age predict death and dropout from longitudinal studies.** *J Gerontol B Psychol Sci Soc Sci* 2008, **63**:P205–P211.

100. Randolph JF Jr, Sowers M, Bondarenko I, Gold EB, Greendale GA, Bromberger JT, Brockwell SE, Matthews KA: **The relationship of longitudinal change in reproductive hormones and vasomotor symptoms during the menopausal transition.** *J Clin Endocrinol Metab* 2005, **90**:6106–6112.

101. Rousseau MC, Abrahamowicz M, Villa LL, Costa MC, Rohan TE, Franco EL: **Predictors of cervical coinfection with multiple human papillomavirus types.** *Cancer Epidemiol Biomarkers Prev* 2003, **12**:1029–1037.

102. Ryu S, Chang Y, Woo HY, Lee KB, Kim SG, Kim DI, Kim WS, Suh BS, Jeong C, Yoon K: **Time-dependent association between metabolic syndrome and risk of CKD in Korean men without hypertension or diabetes.** *Am J Kidney Dis* 2009, **53**:59–69.

103. Seid M, Varni JW, Cummings L, Schonlau M: **The impact of realized access to care on health-related quality of life: a two-year prospective cohort study of children in the California State Children's Health Insurance Program.** *J Pediatr* 2006, **149**:354–361.

104. Silfverdal SA, Ehlin A, Montgomery SM: **Protection against clinical pertussis induced by whole-cell pertussis vaccination is related to primo-immunisation intervals.** *Vaccine* 2007, **25**:7510–7515.

105. Spence SH, Najman JM, Bor W, O'Callaghan MJ, Williams GM: **Maternal anxiety and depression, poverty and marital relationship factors during early childhood as predictors of anxiety and depressive symptoms in adolescence.** *J Child Psychol Psychiatry* 2002, **43**:457–469.

106. Stewart R, Xue QL, Masaki K, Petrovitch H, Ross GW, White LR, Launer LJ: **Change in blood pressure and incident dementia: a 32-year prospective study.** *Hypertension* 2009, **54**:233–240.

107. Strasak AM, Kelleher CC, Klenk J, Brant LJ, Ruttmann E, Rapp K, Concin H, Diem G, Pfeiffer KP, Ulmer H: **Longitudinal change in serum gamma-glutamyltransferase and cardiovascular disease mortality: a prospective population-based study in 76,113 Austrian adults.** *Arterioscler Thromb Vasc Biol* 2008, **28**:1857–1865.

108. Strawbridge WJ, Cohen RD, Shema SJ: **Comparative strength of association between religious attendance and survival.** *Int J Psychiatry Med* 2000, **30**:299–308.

109. Sung M, Erkanli A, Angold A, Costello EJ: **Effects of age at first substance use and psychiatric comorbidity on the development of substance use disorders.** *Drug Alcohol Depend* 2004, **75**:287–299.

110. Tehard B, Lahmann PH, Riboli E, Clavel-Chapelon F: **Anthropometry, breast cancer and menopausal status: use of repeated measurements over 10 years of follow-up-results of the French E3N women's cohort study.** *Int J Cancer* 2004, **111**:264–269.

111. Vikan T, Johnsen SH, Schirmer H, Njolstad I, Svartberg J: **Endogenous testosterone and the prospective association with carotid atherosclerosis in men: the Tromso study.** *Eur J Epidemiol* 2009, **24**:289–295.

112. Wang NY, Young JH, Meoni LA, Ford DE, Erlinger TP, Klag MJ: **Blood pressure change and risk of hypertension associated with parental hypertension: the Johns Hopkins Precursors Study.** *Arch Intern Med* 2008, **168**:643–648.