Potential Bias in the Indirect Methods for Extracting Summary Statistics in Literature-based Meta-Analyses: an empirical evaluation

Background: In literature-based meta-analyses of cancer prognostic studies, methods for extracting summary statistics from published reports have been extensively employed. However, no assessment of the magnitude of bias produced by these methods or comparison of their influence on fixed vs. random effects models have been published previously. Therefore, the purpose of this study is to empirically assess the degree of bias produced by the methods used for extracting summary statistics and examine potential effects on fixed and random effects models. **Methods:** Using published data from cancer prognostic studies, systematic differences between reported statistics and those obtained indirectly using log-rank test p-values and total number of events were tested using paired *t* tests and the log-rank test of survival-agreement plots. The degree of disagreement between estimates was quantified using an information-based disagreement measure, which was also used to examine levels of disagreement between expressions obtained from fixed and random effects models. **Results:** Thirty-four studies provided a total of 65 estimates of lnHR and its variance. There was a significant difference between the means of the indirect lnHRs and the reported values (mean difference = -0.272, *t* = -4.652, p-value <0.0001), as well as between the means of the two estimates of variances (mean difference = -0.115, *t* = -4.5556, p-value <0.0001). Survival agreement plots illustrated a bias towards under-estimation by the indirect method for both lnHR (log-rank p-value = 0.031) and its variance (log-rank p-value = 0.0432). The magnitude of disagreement between estimates of lnHR based on the information-based measure was 0.298 (95% CI: 0.234 – 0.361) and, for the variances it was 0.406 (95% CI: 0.339 – 0.470). As the disagreement between variances was higher than that between lnHR estimates, this increased the level of disagreement between lnHRs weighted by the inverse of their variances in fixed effect models. In addition, results indicated that random effects meta-analyses could be more prone to bias than fixed effects meta-analyses

as, in addition to bias in estimates of lnHRs and their variances, levels of disagreement as high as 0.487 (95% CI: 0.416 – 0.552) and 0.568 (95% CI: 0.496 – 0.635) were produced due to between-studies variance calculations. **Conclusions:** Extracting summary statistics from published studies could introduce bias in literature-based meta-analyses and undermine the validity of the evidence. These findings emphasise the importance of reporting sufficient statistical information in research articles and warrant further research into the influence of potential bias on random effects models.

Author:

Dima Yousef Abdallah

Affiliation:

School of Public Health, Physiotherapy, and Population Science. University College Dublin. Dublin, Ireland

Corresponding Author:

Dima Youdef Abdallah

Email: dima.abdallah@gmail.com

Mobile: 00353 87 964 7164

9 Introduction

PeerJ PrePrints | https://peerj.com/preprints/142v1/ | v1 received: 5 Dec 2013, published: 5 Dec 2013, doi: 10.7287/peerj.preprints.142v1

One of the main objectives of research in oncology is the identification of prognostic factors associated with cancer outcomes. Yet, in spite of the growing research in this field over the last two decades, the prognostic value of many factors remains unclear and their clinical utility not established, partly due to the inconsistent and, sometimes, conflicting results reported in the literature (Simon & Altman, 1994; Bossard et al., 2003). A concern for these studies is the small sample size and inadequate statistical power (Altman et al., 1995; Pajak et al., 2000). Thus, to improve power and precision in results, pooling of estimates in the form of meta-analysis has been widely applied in cancer survival research (Hirooka et al., 2009). Pooling results through the collection of individual patient data (IPD) is considered the 'gold standard' approach for meta-analyses, as it provides reliable results and circumvents several forms of bias such as outcome selection bias (Stewart et al., 1993; Williamson et al., 2005). However, collecting IPD is not always possible and can be time-consuming and resource intensive (Stewart et al., 1993, Williamson et al., 2002). Therefore, most published meta-analyses use summary data extracted from published reports. In these instances, a series of methods have been described by Parmar et al. (1998) to extract summary statistics from individual studies.

The summary statistic most appropriately used for the analysis of time-to-event outcomes is the natural logarithm of the hazard ratio (lnHR) and its variance (Michiels et al. 2005). The HR has the same interpretation as the relative risk and odds ratio (Clark et al., 2003), with the added benefit of incorporating censoring and time to event (Michiels et al., 2005). In meta-analyses of time-to-event outcomes, the overall pooled lnHR is expressed as a weighted average of the individual lnHRs, with the weights representing the inverse of the variances of lnHRs (Parmar, Torri & Stewart, 1998). Thus, calculation of the overall lnHR requires extraction of individual estimates and their variances from each study, which might not be always available. The three major methods described by Parmar et al. (1998), the direct, indirect, and survival curve methods, allow for extracting these statistics when they are not reported. The direct method is based on calculating lnHR and its variance from the 'observed' and 'expected' number of events and the Mantel-Haenzel variance of lnHR. A number of indirect methods were described, which mainly use the reported p-value of the log-rank test and the number of events to estimate lnHR and its variance. The survival curve method relies on extracting data from the published survival curves by splitting the time axis into several intervals and calculating lnHR for each interval under the assumption of uniform censoring. The summary lnHR is then calculated as a weighted average of lnHRs across all intervals. Based on empirical analyses, the authors conclude that the direct

method gives the most accurate estimates, and a simple average of the indirect methods would also perform well. While no evidence of systematic bias was found for the survival curve method, it was deemed the least precise. In particular, due to the assumption of uniform censoring, the survival curve method tended to under-estimate the variance of the lnHR (Parmar, Torri & Stewart, 1998). Consequently, Williamson et al. (2002) proposed a modified survival curve approach that would yield improved estimates by incorporating the reported numbers at risk and, thus, providing information on censoring pattern.

Since its publication, the paper by Parmar et al. (1998) received 1097 citations according to Google Scholar. Most citations were from meta-analyses that have applied one or more of the methods described in the paper to extract summary statistics. However, few studies evaluated the performance of these methods. Tudur et al. (2001) compared estimates obtained from these methods with those obtained from IPD for seven randomised controlled trials (RCTs). The authors concluded that the indirect methods performed well, while the survival curve methods tended to under-estimate effect size and were the least consistent, especially under low event rates. Hirooka et al. (2009) used simulation to compare results from these methods for RCTs with several combinations of sample sizes, HRs, and survival and censoring rates. The direct method was found to yield the most reliable results. The indirect method using total number of events and log-rank test p-values was generally also accurate but tended to under-estimate effect sizes when there were large effects with large sample size. Similar to the results reported by Tudur et al. (2001), the survival curve methods were the least reliable and tended to under-estimate effect size when event rates were low. As the study was based on simulated data, it was assumed that the true values of the lnHRs were known with no uncertainty and, therefore, no comparisons between the methods in relation to variance estimates were done.

The previous two evaluation studies were based on RCTs. Although the methods described by Parmar et al. (1998) were originally intended for RTCs, they have been applied in several meta-analyses of prognostic factor studies in cancer. Prognostic factor studies are observational studies that are susceptible to several sources of bias (Egger, Schneider & Smith, 1998), and it is not known how the methods perform in these situations. Any potential biases introduced by the indirect and survival curve methods would threaten the validity of the results from these meta-analyses. Furthermore, it is important to quantify the magnitude of bias and to probe whether random effects meta-analyses, where weights assigned to studies are based on

within-study as well as between-studies variances (Borenstein et al., 2009; Borenstein et al., 2010), are more prone to bias than meta-analyses based on fixed effect models. Therefore, using published data from cancer prognostic studies, the purpose of this study is to empirically assess the degree of bias produced by the methods used for extracting summary statistics and examine potential effects on fixed and random effects models. A survey of the literature revealed that data on the 'expected number of events' required for application of the direct method were not available in any of the studies. Thus, it was not possible to evaluate the direct method. Further, due to the low reliability of the survival curve methods and their sensitivity to figure quality and inter-reader variability (Williamson et al., 2002), the focus of this paper has been on evaluating bias produced by the indirect methods only.

Methods

Data used for the analyses in this paper were obtained from studies that have assessed the prognostic role of microRNAs in cancer survival in humans. MicroRNAs are small non-coding RNAs that regulate many cellular processes such as cellular differentiation, cell cycle progression, and apoptosis. Sine the discovery of their role in chronic lymphocytic leukemia in 2002 (Calin et al., 2002), numerous clinical studies investigated the potential prognostic value of microRNAs in cancer by comparing survival among groups of patients with high vs. low levels of microRNAs. Consequently, several meta-analyses that have pooled results from these prognostic studies have been published (Fu et al., 2011;Ma et al., 2012; He et al., 2013; Lin et al., 2013; Yang et al., 2013). When summary statistics were not reported in individual studies, these meta-analyses have applied several of the methods proposed by Parmar et al. (1998). Validity of the results from these meta-analyses partly relies on the accuracy of estimates obtained from individual studies, and thus, this study ascertains whether there is evidence of potential bias in the estimates obtained through the indirect methods.

Literature Search and Eligibility Criteria

A search of the PubMed database for English language studies that have assessed the prognostic role of microRNAs in cancer survival was conducted in 06/07/2013. The search strategy consisted of combinations of the following terms: ("microRNA" OR "miRNA" OR "miR") AND ("cancer" OR "malignancy" OR "tumour" OR "tumor" OR "lymphoma" OR "leukaemia" OR "leukemia") AND ("prognosis" OR "prognostic" OR "survival" OR "recurrence" OR

103 "metastasis" OR "metastases" OR "outcome") AND ("hazard rate" OR "hazard ratio"
104 OR "kaplan meier" OR "Cox").

105 For the indirect calculation of lnHR and its variance, the number of events and log-rank p-value
106 are required (Parmar, Torri & Stewart, 1998). Therefore, to be included in the evaluation of the
107 indirect methods, studies were eligible if they provided the following information: (1) HR or
108 lnHR or coefficient derived from Cox regression, along with 95% confidence interval, variance,
109 standard error, or p value, (2) the total number of events, (3) the number of patients in high and
110 low levels of microRNAs, and (4) the p-value for the log rank test. Only original research articles
111 in humans were included.

112 **Data Extraction**

113 For studies eligible for evaluation of the indirect methods, the following data were extracted:
114 surname of first author, year of publication, microRNAs and outcomes investigated, total sample
115 size, total number of events, number of patients in each microRNA group, p-value for log rank
116 test, unadjusted HR, 95% CI for HR or variance, and its p-value.

117 **Statistical Analysis**

118 For studies that did not report the variance, the confidence intervals were used to approximate the
119 variance of the reported lnHR using the following expression (Tierney et al., 2007):

$$Var(\ln(HR_i)) = \left[\frac{\ln(upperCI) - \ln(lowerCI)}{2 \times z}\right]^2$$

120 Where upper CI and lower CI are the upper and lower confidence intervals of the reported HR,
121 respectively, and *z* is the *z* score for the upper limit of the confidence interval. When 95%
122 confidence intervalss are used, *z* would be equal to 1.96.

123 Variances estimated from the confidence intervals were compared with indirect variances
124 calculated using the following equations (Parmar, Torri & Stewart, 1998):

$$V_{ri} = \frac{O_i R_{ri} R_{ci}}{(R_{ri} + R_{ci})^2}$$

$$Var(\ln(HR_i)) = 1 / V_{ri}$$

125 Where, for study i:

V_{ri} = variance of the log-rank statistic

O_i = total observed number of events

R_{ri} = number of patients in the high risk group

R_{ci} = number of patients in low risk group

The logarithms of the reported HRs were compared with indirect lnHRs calculated using the total number of events, the log-rank p-value, and the variance of the log-rank statistic through the following equations (Parmar, Torri & Stewart, 1998):

$$O_{ri} - E_{ri} = \frac{\sqrt{(O_i R_{ri} R_{ci}}}{(R_{ri} + R_{ci})} \times \Phi^{-1}(1 - \frac{p_i}{2})$$

$$\ln(HR_i) = (\frac{O_{ri} - E_{ri}}{V_{ri}})$$

Where, for study *i*:

O_{ri} = observed number of events in high risk group

E_{ri} = expected number of events in high risk group

P_i = the reported (two-sided) log-rank p-value

Φ = the cumulative distribution function of the Normal distribution

Log-rank test p-values that were reported as less than a particular level (e.g. $p < 0.05$) were rounded as $p = 0.05$ (Hirooka et al., 2009). In addition, to ensure consistency in the interpretation of hazard ratios, the low risk group was set as the reference category for all studies. Thus, for studies that reported HR for low-risk vs. high-risk groups for a certain microRNA level, the inverse of the reported HRs were used.

To assess agreement between measurements, the indirectly calculated estimates were plotted against the reported ones and the divergence of points from lines of equality, where all points would lie if there was perfect agreement between the two measurements, was examined (Bland & Altman, 1986). Component Plus Residual (CPR) plots were used to assess deviations from linearity in the relationship between measurements (Vittinghoff et al., 2012). Systematic differences between measurements were tested using paired *t* tests and the log-rank test of survival-agreement plots (Luiz et al., 2003; Llorca & Delgado-Rodríguez, 2005). The degree of

disagreement between the reported and estimated measurements was quantified using an information-based disagreement measure described by Cost-Santos et al. (2010), which is calculated as follows:

$$d(A,B) = \frac{1}{n}\sum_{i=1}^{n}\log_2\left(\frac{|a_i - b_i|}{\max\{a_i, b_i\}} + 1\right)$$

where,

d(A,B) = information-based measure of disagreement between reported and indirectly estimated measurements, $0 \le d(A,B) \ge 1$ and $d(A,B) = d(B,A)$

n = number of studies

a_i = reported lnHR or its variance calculated from reported 95% CI

b_i = indirectly estimated lnHR or its variance

For this measure, higher values correspond to higher disagreement. Further, the measure has differential weighting, which means that differences between high values of lnHR or its variance contribute less to the disagreement measure than equal differences between low values (Costa-Santos et al., 2010).

To compare the potential influence of bias on fixed and random effects meta-analyses, the disagreement measure between reported and indirectly estimated statistics was calculated for each of the following expressions: $\ln HR_i/Var(\ln HR_i)$, $(\ln HR_i/Var(\ln HR_i))^2$, $(\ln HR_i)^2/Var(\ln HR_i)$, and $(1/Var(\ln HR_i))^2$, where the first of these expressions is used in the fixed effect model, while the second, third, and fourth expressions are used for the calculation of between-studies variance in random effects models (DerSimonian & Laird, 1986; Borenstein et al., 2009).

For the paired *t* test and the log-rank test of the survival-agreement plots, two-sided p values <0.05 were considered significant. For the information-based measure of disagreement, non-parametric bootstrapping was employed, where 95% confidence intervals were obtained from the 2.5th and 97.5th percentiles of the 1000 bootstrapped samples (Costa-Santos et al., 2010). Analyses were performed using Excel for Mac (Version 14.3.6, Microsoft Corporation) and R for Mac (Version 3.0.0, R Foundation for Statistical Computing, Vienna, Austria), with the following packages: 'car' (Fox & Weisberg, 2011), 'boot' (Canty and Ripley, 2013), and 'survival' (Therneau, 2013).

Results

Literature Search Results

The electronic search identified 358 articles. Of these, 10 were reviews and meta-analyses and 181 were excluded based abstract relevance. Of the remaining 167 articles that underwent full text review, 5 were not prognostic studies. Thus, a total of 162 papers used survival analyses techniques to investigate the prognostic role of one or more microRNAs in cancer. The HR was provided in 136 (84%) of all papers, while only 4 (2.5%) studies reported the variance or standard error of lnHR. The 95% confidence intervals for HR, on the other hand, were given in 123 (76%) studies. None of the studies reported the expected number of events required for the application of the direct method. The total number of observed events required for the application of one of the indirect methods was reported by around 82 (50.6%) studies. Based on the literature search results, the indirect method of estimating lnHR and its variance using log-rank p-value and total number of events was applicable in 68 studies.

As the log-rank test does not adjust for confounders except in cases of stratified analysis, estimates derived from the indirect method were compared to reported HRs that were unadjusted for confounders. Of the 68 studies suitable for the application of the indirect method, 8 did not report HR, and 26 did not report unadjusted HRs. Therefore, 34 studies providing a total of 65 estimates of lnHR and its variance were eligible for evaluation of the indirect method. As none of the 34 studies reported the variance of the lnHR, variances were calculated using the 95% confidence intervals and were compared with variances calculated using the indirect method. Table 1 summarises the medians (and ranges) of statistics extracted from the 65 analyses. List of references for the 34 studies is provided in Supplemental File 1, and the data used for analyses are provided in Supplemental File 2.

Table 1: Statistics from 34 studies (consisting of 65 analyses)

Evaluation of the Performance of the Indirect Method

Figures 1 and 2 show plots of the indirectly estimated lnHRs against reported lnHRs and indirectly estimated variances against variances calculated from reported 95% confidence intervals. It can be seen that at low values of lnHR and its variance, numerous points lie about and close to the line of equality, especially in the case of lnHR. However, both plots show a

tendency for indirect estimates to under-estimate effect sizes especially at higher values. CPR plots shown in Figures 3 and 4 suggest marked deviations from linearity between reported and indirect estimates for both lnHRs and their variances. Thus, linear regression analyses were not applied to assess bias.

Figure 1: Plot of indirect against reported values for lnHR. Straight line represents line of equality

Figure 2: Plot of indirect against reported values for variance. Straight line represents line of equality

Figure 3: CPR plot lnHR and (b) variance

Figure 4: CPR plot for variance

A significant difference between the means of the indirect lnHRs and the reported values was found (mean difference = -0.272, t = -4.652, p-value <0.0001), as well as between the means of the two estimates of variances (mean difference = -0.115, t = -4.5556, p-value <0.0001). Figures 5 and 6 display the survival agreement plots, where the solid lines represent absolute differences when the indirect estimates are less than the reported ones, and the dashed lines represent absolute differences when indirect estimates are larger than the reported ones. As the solid lines lie above the dashed ones, the plots suggest a bias towards under-estimation by the indirect method (Llorca & Delgado-Rodríguez, 2005). Further, the log-rank p-values reveal significant differences between the lines (χ^2 = 4.7, p = 0.031) and (χ^2 = 4.1, p = 0.0432) for the curves of lnHR and its variance, respectively.

Figure 5: Survival agreement plots for absolute differences between reported and indirect lnHRs. (solid lines: indirect < reported, dashed lines: indirect > reported)

Figure 6: Survival agreement plot for absolute differences between variances estimated from 95% CIs and indirect variances. (solid lines: indirect < reported, dashed lines: indirect > reported)

The magnitudes of disagreements between reported and indirect estimates of lnHRs and their variances were quantified using the information-based measurement of disagreement (Costa-Santos et al, 2010). As shown in Table 2, the disagreement between estimates of lnHR is equal to 0.298 (95% CI: 0.234 – 0.361), while between variances, it is equal to 0.406 (95% CI: 0.339 – 0.470). In addition, disagreements were measured for four expressions used in the calculation of the pooled lnHR in fixed and random effects meta-analyses. Table 2 shows how bias in the estimates of lnHRs and their variances influences calculations used for pooling results in meta-analyses. Weighting of lnHRs by the inverse of their variances in fixed effect models results in a disagreement level higher than that for lnHRs because of the magnitude of disagreement between estimates of variances. In addition, results suggest that random effects meta-analyses could be more prone to bias than fixed effects meta-analyses as, in addition to bias in estimates of lnHRs and their variances, levels of disagreement as high as 0.487 (95% CI: 0.416 – 0.552) and 0.568 (95% CI: 0.496 – 0.635) are produced as a result of between-studies variance calculations.

Table 2: Information-based Measurements of Disagreement

Discussion and Conclusions

As the number of meta-analyses of time-to-event outcomes assessing the prognostic role of microRNAs in cancer is rising, it is important to establish the degree of bias produced by methods of extracting summary statistics applied in these meta-analyses. Results from this paper suggest that estimates of lnHR and its variance calculated through the indirect methods were systematically different from the reported estimates, as the paired *t* tests and the log-rank tests of the survival agreement plots revealed significant under-estimation of effect sizes. In line with results by Hirooka et al. (2009), the tendency for the indirect methods to under-estimate measurements was particularly obvious at higher values of lnHR and its variance. In addition, bias in these estimates could potentially influence results from both fixed effects and random effect meta-analyses. Because effects sizes are weighted by the inverse of their variances in fixed effect models, under-estimation of variances would result in higher weights given to these effect sizes in meta-analysis (Williamson et al., 2002). Further, based on the measurements of disagreement, the random effects model is expected to be prone to a higher degree of bias as a result of between-studies variance calculations.

Parmar et al. (1998) described a hierarchy in the methods, in which reported lnHRs and their variances should be used if they are available, followed by the direct method, an average of the indirect methods, and finally, the survival curve methods. Applicability of any of these methods relies on the data available in published reports. Results from the literature search indicated that 84% of studies reported HRs, 76% studies reported HRs with their 95% confidence intervals, and only 2.5% reported HRs with variance. These results compare favourably with those presented by other authors. A survey of RCTs published during the years 2004 and 2005 in two oncology journals indicated that 50% of 129 articles reported HRs with their confidence intervals (Hirooka et al., 2009), while another study in 2005 showed that only 3% of 131 chemotherapy trials reported lnHR (Michiels et al., 2005). As the articles surveyed for the purposes of the current study were published between 2008 and 2013, these results could signify an improvement in reporting practices among cancer survival studies and less need to rely on the other methods for extracting summary statistics when conducting meta-analyses.

As none of the studies surveyed in this paper reported the expected number of events, it was not possible to apply the direct method. Hirooka et al. (2009) also found that the direct method was applicable in only 1% of 129 articles reviewed. On the other hand, median survival times have been found to be reported in more than half of research articles (Michiels et al., 2005; Hirooka et al., 2009). Nevertheless, their use in the analysis of time-to-event outcomes is not recommended, as they have been shown to produce markedly imprecise estimates (Michiels et al., 2005). Taken together, these results indicate that when summary statistics are not reported, meta-analysts would have to use the indirect or survival curve methods to extract statistics due to the rare applicability of the direct method and the inappropriateness of using alternative survival measures such as median survival times. Although the indirect methods are more reliable than the survival curves methods, findings from analyses in this paper suggest that they could be prone to systematic bias.

This study has some limitations that need to be taken into account when interpreting results. The effect of rounding log-rank p-values was not examined. Tudur et al. (2001) reported that their results were robust to rounding errors. Hirooka et al. (2009), on the other hand, found effect sizes to be under-estimated when rounded p-values were used in the indirect method. Another limitation is that variances calculated indirectly using p-values and event numbers were

compared to those calculated using confidence intervals rather than to directly reported variances. Thus, the degree of bias might be different when direct variances are used for comparison. Furthermore, methods used to assess bias have their own limitations. Paired *t* tests and survival agreement plots allow for the detection of fixed bias, where differences between measures are consistent, but not for the detection of proportional bias, where differences increase or decrease in proportion to the values of the measurement (Ludbrook, 2002; Luiz & Szklo, 2005; Ludbrook, 2010). Proportional bias is detected using linear regression techniques (Ludbrook, 2002; Ludbrook 2010), however, as the linearity assumption was not met, it was not possible to apply linear regression. Disagreements were quantified using the measure proposed by Costa-Santos et al. (2010). Although this measure provides a quantitative estimate of bias and is useful for comparing disagreement among groups, it does not detect proportional bias and has not been widely adopted. Due to the limitations in all of these methods, more than one strategy was employed as recommended by Luiz and Szklo (2005). Nevertheless, findings presented in this study might not be generalizable to other settings as they are based on a subset of studies in a particular field of research.

In conclusion, extracting summary statistics from published studies could introduce bias in literature-based meta-analyses and undermine the validity of the evidence. These findings emphasise the importance of reporting sufficient statistical information in research articles and warrant further research into the influence of potential bias on random effects models.

References:

Altman DG, De Stavola BL, Love SB, Stepniewska KA. 1995. Review of survival analyses published in cancer journals. *British Journal of Cancer* 72(2):511-8.

Bland JM, Altman D. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476), 307-310.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. 2009. Random-effects model. In: *Introduction to meta-analysis*. West Sussex: Wiley, 69-75.

Borenstein M, Hedges LV, Higgins J, Rothstein HR. 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods* 1(2):97-111.

Bossard N, Descotes F, Bremond AG, Bobin Y, De Saint Hilaire P, Golfier F, Awada A, Mathevet PM, Berrerd L, Barbier Y, Estève J. 2003. Keeping data continuous when analyzing the prognostic impact of a tumor marker: an example with cathepsin D in breast cancer. *Breast Cancer Research and Treatment* 82(1):47-59.

Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, Bullrich F, Croce CM. 2002. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences* 99(24):15524-15529.

Canty A, Ripley B. 2013. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-9.

Clark TG, Bradburn M J, Love SB, Altman DG. 2003. Survival analysis part I: basic concepts and first analyses. *British journal of cancer* 89(2):232-238.

Costa-Santos C, Antunes L, Souto A, Bernardes J. 2010. Assessment of disagreement: a new information-based approach. *Annals of epidemiology* 20(7):555-561.

DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. *Controlled clinical trials* 7(3):177-188.

Egger, M., Schneider, M., & Davey Smith, G. 1998. Spurious precision? Meta-analysis of observational studies. *British Medical Journal* 316(7125):140-144.

Fox J, Weisberg S. 2011. An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage.

Fu X, Han Y, Wu Y, Zhu X, Lu X, Mao F, Wang X, He X, Zhao Y, Zhao Y. 2011. Prognostic role of microRNA-21 in various carcinomas: a systematic review and meta-analysis. *European Journal of Clinical Investigation* 41(11):1245-1253.

He J, Zhang F, Wu Y, Zhang W, Zhu X, He X, Zhao Y, Zhang W, Zhao Y. 2013. Prognostic role of microRNA-155 in various carcinomas: Results from a meta-analysis. *Disease markers* 34(6):379-386.

Hirooka, T., Hamada, C., & Yoshimura, I. 2009. A Note on Estimating Treatment Effect for Time-to-event Data in a Literature-based Meta-analysis. *Methods of information in medicine* 48(2):104-112.

Lin S, Pan L, Guo S, Wu J, Jin L, Wang JC, Wang S. 2013. Prognostic role of microRNA-181a/b in hematological malignancies: a meta-analysis. *PloS one* 8(3):e59532.

Llorca J, Delgado-Rodríguez M. 2005. Survival analytical techniques were used to assess agreement of a quantitative variable. *Journal of clinical epidemiology* 58(3):314-315.

Ludbrook, J. 2002. Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clinical and Experimental Pharmacology and Physiology* 29(7):527-536.

Ludbrook, J. 2010. Linear regression analysis for comparing two measurers or methods of measurement: but which regression? *Clinical and Experimental Pharmacology and Physiology* 37(7):692-699.

Luiz RR, Costa AJL, Kale PL, Werneck, GL. 2003. Assessment of agreement of a quantitative variable: a new graphical approach. *Journal of clinical epidemiology* 56(10):963-967.

Luiz RR, Szklo M. 2005. More than one statistical strategy to assess agreement of quantitative measurements may usefully be reported. *Journal of clinical epidemiology* 58(3):215-216.

Ma XL, Liu L, Liu XX, Li Y, Deng L, Xiao ZL, Liu YT, Shi HS, Wei YQ. 2012. Prognostic role of microRNA-21 in non-small cell lung cancer: a meta-analysis. *Asian Pacific Journal of Cancer Prevention* 13(5):2329-2334.

Michiels S, Piedbois P, Burdett S, Syz N, Stewart L, Pignon JP. 2005. Meta-analysis when only the median survival times are known: a comparison with individual patient data results. *International journal of technology assessment in health care* 21(01):119-125.

Pajak TF, Clark GM, Sargent DJ, McShane LM, Hammond ME. 2000. Statistical issues in tumor marker studies. *Archives of Pathology and Laboratory Medicine* 124(7):1011-5.

Parmar MK, Torri V, Stewart L. 1998. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine* 17(24):2815-2834.

Simon R, Altman DG. 1994. Statistical aspects of prognostic factor studies in oncology. *British Journal of Cancer* 69(6):979-85.

Stewart LA, Parmar MKB. 1993. Meta-analysis of the literature or of individual patient data: is there a difference? *The Lancet* 341(8842):418-422.

Therneau T. 2013. _A Package for Survival Analysis in S_. R package version 2.37-4.

Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. 2007. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 8(1):16.

Tudur C, Williamson PR, Khan S, Best LY. 2001. The value of the aggregate data approach in meta-analysis with time-to-event outcomes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 164:357–370.

Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. 2012. Linear Regression. In: *Regression Methods in Biostatistics*. New York: Springer, 109-111.

Williamson PR, Smith CT, Hutton JL, Marson AG. 2002. Aggregate data meta-analysis with time-to-event outcomes. *Statistics in Medicine* 21(22):3337-3351.

383 Williamson PR, Gamble C, Altman DG, Hutton JL. 2005. Outcome selection bias in
384 meta-analysis. *Statistical Methods in Medical Research* 14(5):515-524.

385 Yang M, Shen H, Qiu C, Ni Y, Wang L, Dong W, Liao Y, Du J. 2013. High expression of miR-21
386 and miR-155 predicts recurrence and unfavourable survival in non-small cell lung cancer.
387 *European Journal of Cancer* 49:604-615.

Table 1(on next page)

Statistics from 34 studies

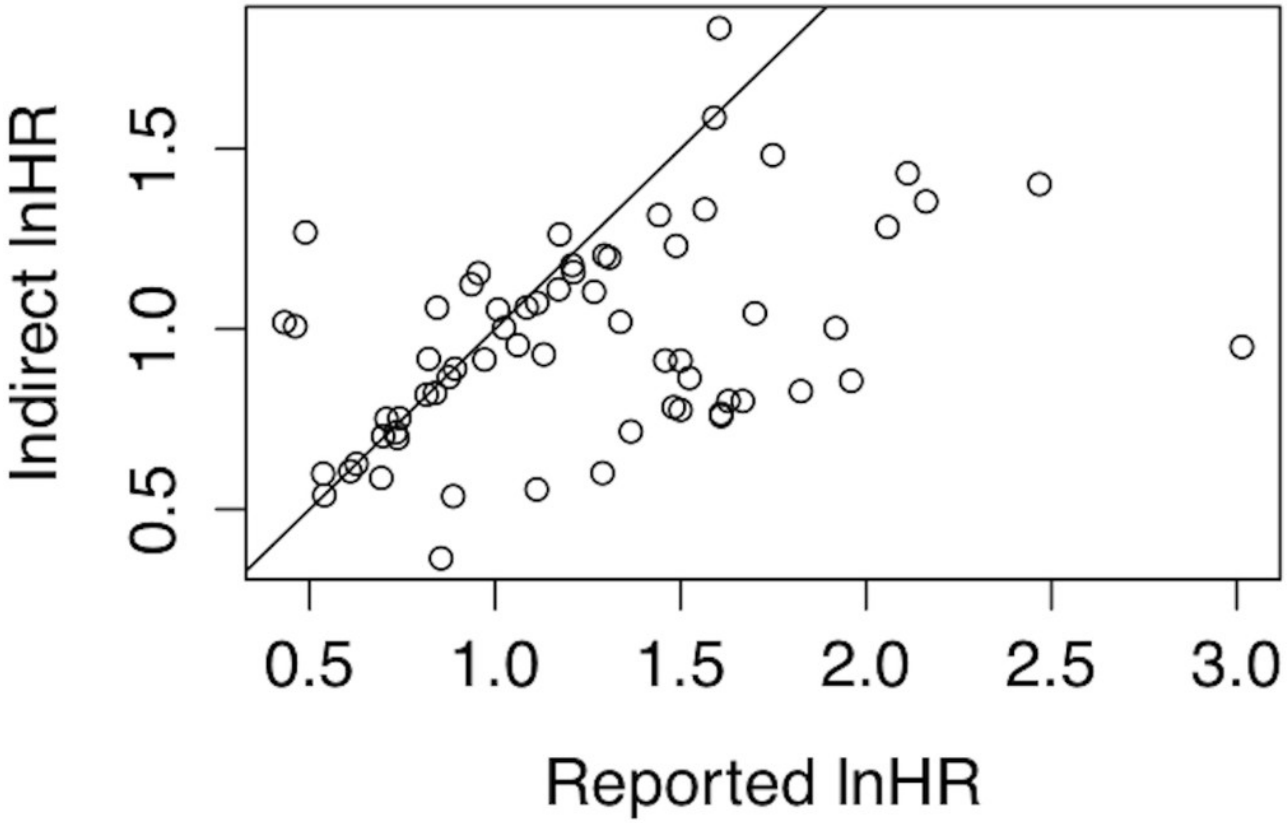
Table 1: Statistics from 34 studies (consisting of 65 analyses)

Statistics	Median (Range)
Sample Size	96 (29 - 470)
Total number of events	42 (11 - 186)
p-value for log-rank test	0.007 (0.0001 - 0.058)
HR	3.24 (1.54 - 20.36)
lnHR	1.174 (0.432 – 3.014)
Variance of lnHR	0.154 (0.017 – 1.070)

Figure 1

Plot for Indirect Against Reported Values of lnHR

Straight line represents line of equality



PeerJ PrePrints | https://peerj.com/preprints/142v1/ | v1 received: 5 Dec 2013, published: 5 Dec 2013, doi: 10.7287/peerj.preprints.142v1

Figure 2

Plot of Indirect Against Reported Values for Variance

Straight line represents line of equality

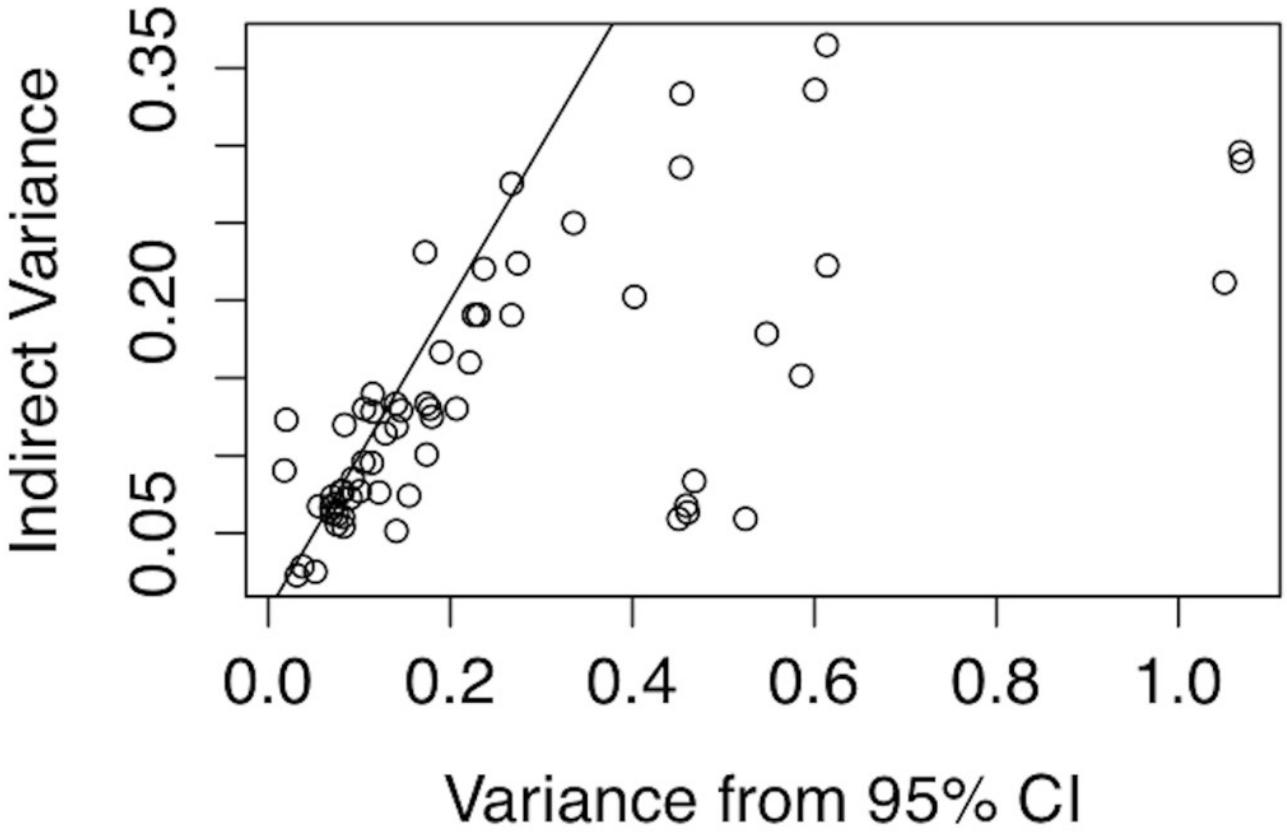


Figure 3

CPR Plot for lnHR

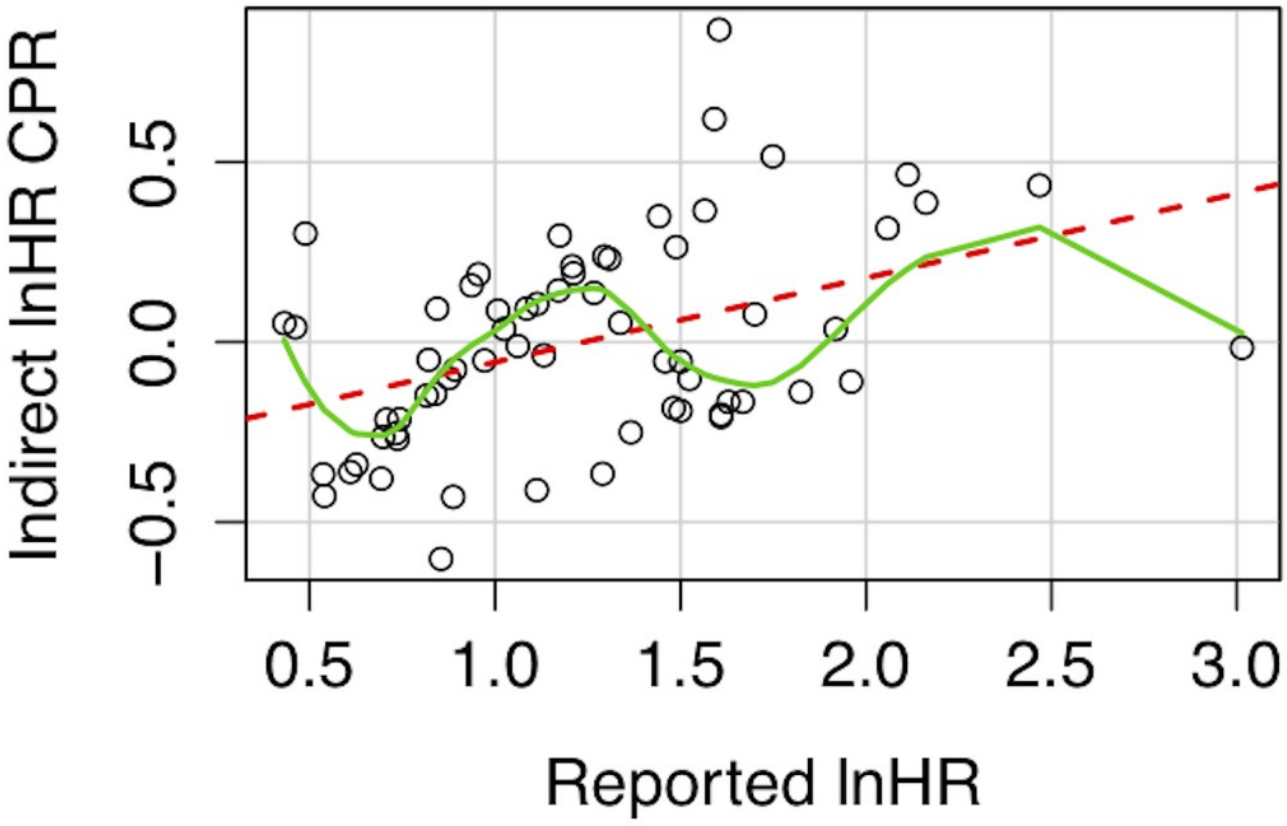


Figure 4

CPR Plot for Variance

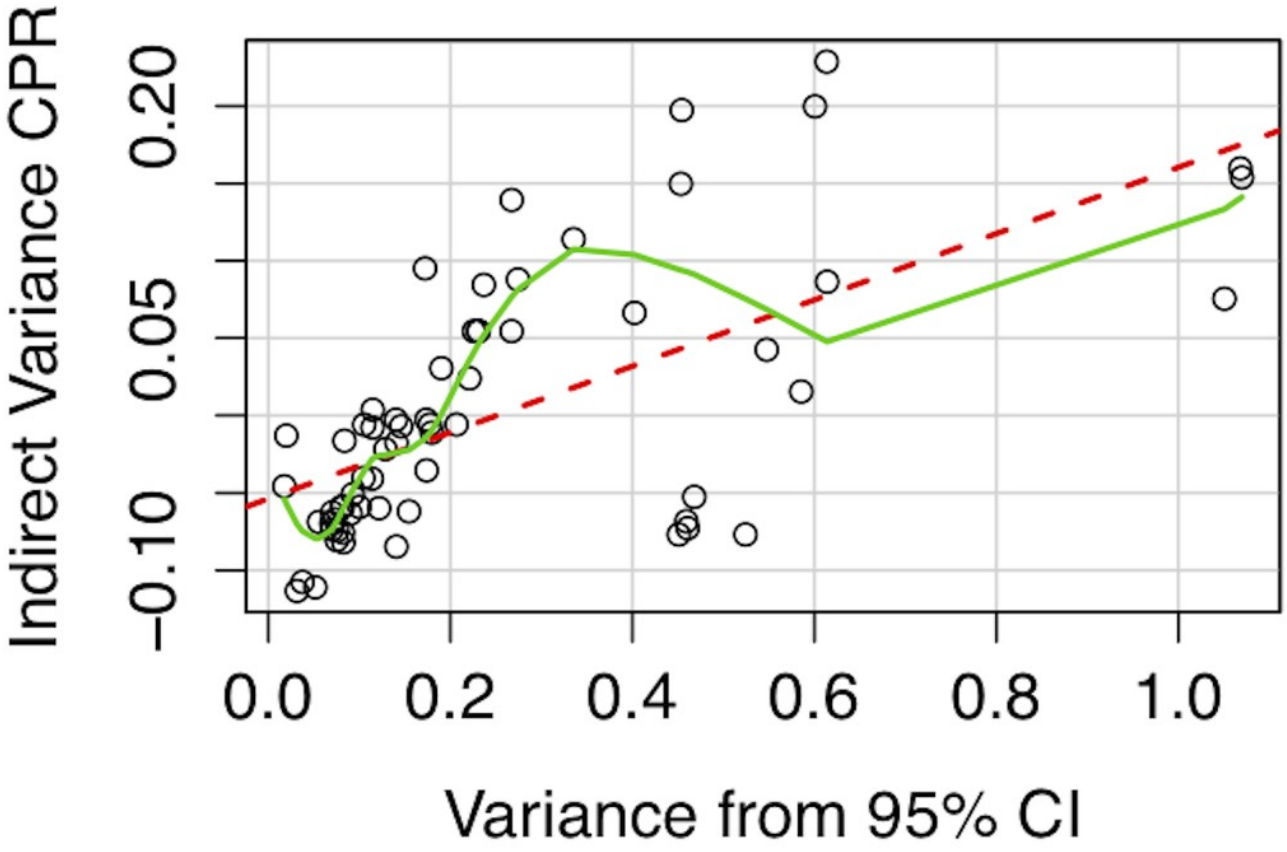


Figure 5

Survival Agreement Plot for Absolute Differences between Reported and Indirect lnHRs

(solid lines: indirect < reported, dashed lines: indirect > reported)

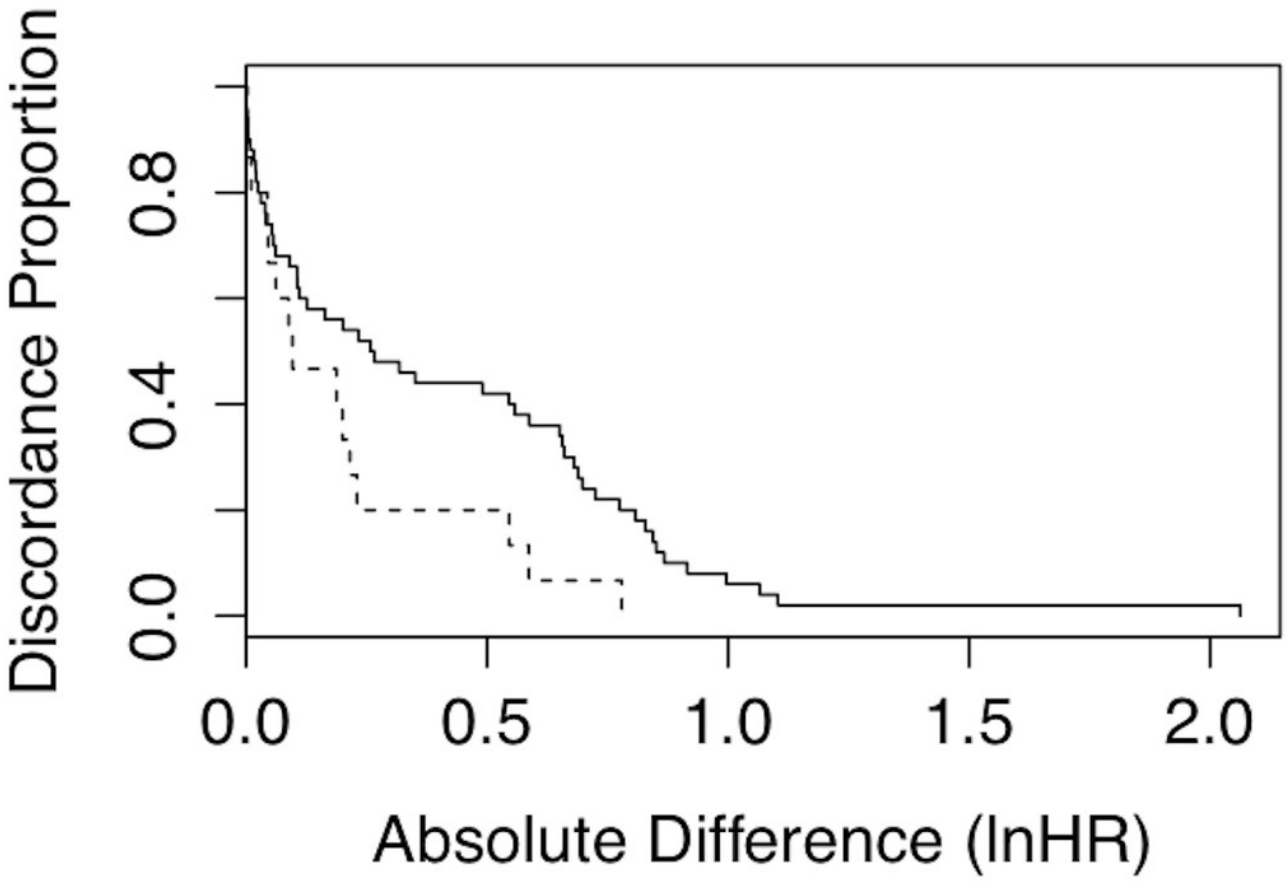


Figure 6

Survival Agreement Plot for Absolute Differences between Variances Estimated from 95% CIs and Indirect Variances.

(solid lines: indirect < reported, dashed lines: indirect > reported)

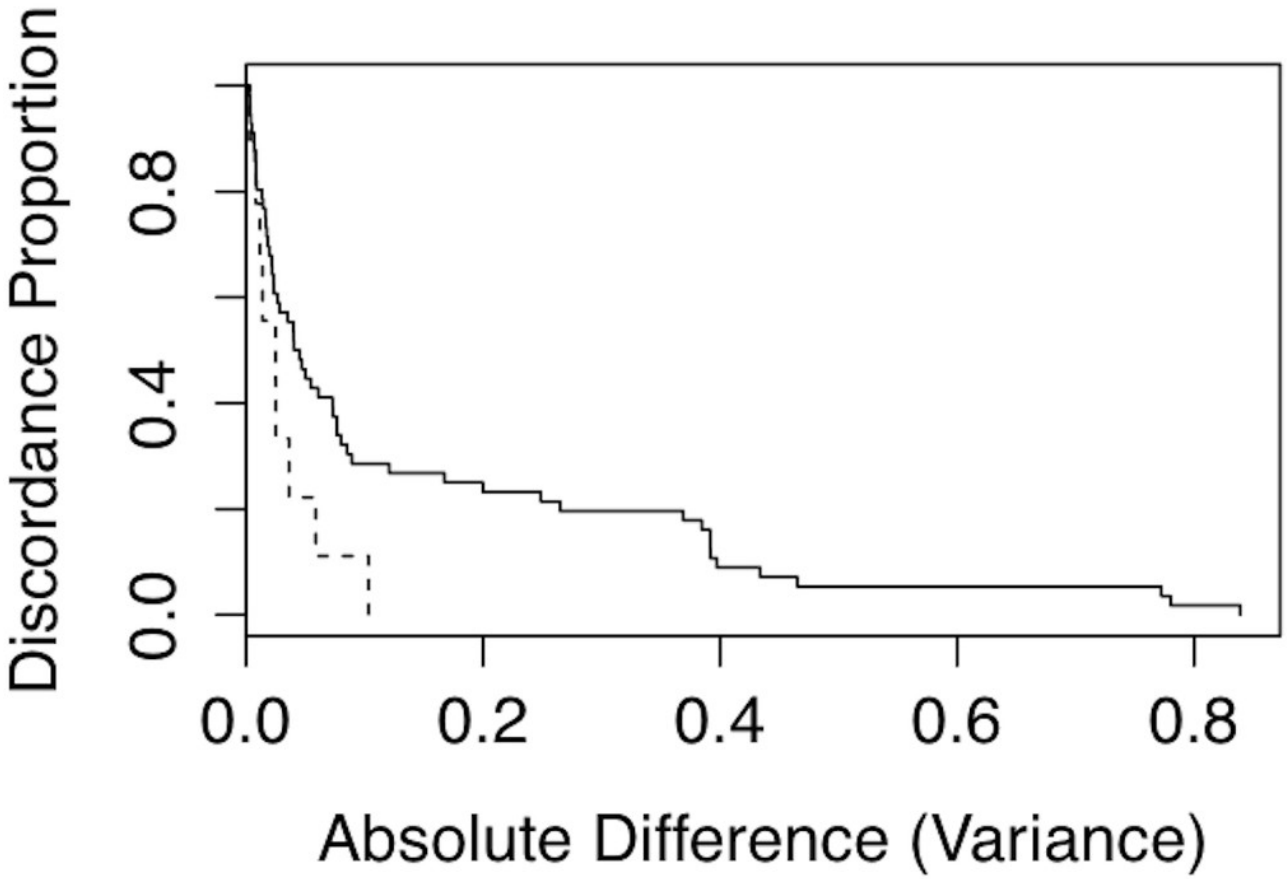


Table 2(on next page)

Information-based measurements of disagreement

1 Table 2: Information-based measurements of disagreement

Statistic	Disagreement	Bootstrap 95% CI
Used in Fixed and Random Effects Models		
lnHR	0.298	0.234 – 0.361
Variance	0.406	0.339 – 0.470
Used in Fixed Effect Models		
lnHR/Var(lnHR)	0.329	0.276 – 0.388
Used in Random Effects Models		
$(lnHR/Var(lnHR))^2$	0.487	0.416 – 0.552
$(lnHR)^2/Var(lnHR)$	0.304	0.242 – 0.367
$(1/Var(lnHR))^2$	0.568	0.496 – 0.635