

Minireview

Pooling Biomarker Data from Different Studies of Disease Risk, with a Focus on Endogenous Hormones

Timothy J. Key, Paul N. Appleby, Naomi E. Allen, and Gillian K. Reeves

Abstract

Large numbers of observations are needed to provide adequate power in epidemiologic studies of biomarkers and cancer risk. However, there are currently few large mature studies with adequate numbers of cases with biospecimens available. Therefore, pooling biomarker measures from different studies is a valuable approach, enabling investigators to make robust estimates of risk and to examine associations in subgroups of the population. The ideal situation is to have standardized methods in all studies so that the biomarker data can be pooled in their original units. However, even when the studies do not have standardized methods, as with existing studies on hormones and cancer, a simple approach using study-specific quantiles or percentage increases can provide substantial information on the relationship of the biomarker with cancer risk. *Cancer Epidemiol Biomarkers Prev*; 19(4); 960–5. ©2010 AACR.

Introduction

Studies of biomarkers and cancer risk require large numbers of observations to provide adequate power, particularly to examine consistency across subgroups. Because there are currently few large mature studies with adequate numbers of cases with biospecimens available, the only way to provide sufficient numbers for analysis is to pool the data from individual studies for collaborative reanalysis. Pooled analyses provide several advantages over meta-analyses of published results: data from all the studies can be analyzed with the same statistical approach, using the same categories and adjustments, and heterogeneity of the main association can be assessed according to characteristics of the studies and of the individuals, such as the type of assay used and the subtype of tumors.

Ideally, biomarker data would be combined using the raw data from the original studies to provide hazard ratios per unit change in the concentration of the biomarker. For example, in the Prospective Studies Collaboration, data from 61 studies worldwide were combined and vascular mortality was reported per mmol/L difference in cholesterol concentration (1). In this example, cholesterol assays have very well established methods

and provide measurements that are both accurate and precise, making it reasonable to pool the results from different studies. In contrast, other biomarkers are measured using less well-standardized assays and accurate measurements may be difficult. A well-known example is estradiol in postmenopausal women; concentrations are very low (around 30 pmol/L) and although research laboratories have selected or developed assays sensitive enough to measure these low concentrations, the methods vary and this leads to large differences in measured concentrations (2). This potential variation in biomarker measurements has important implications for the choice of appropriate methods for pooling biomarker data across studies.

Pooling Biomarker Data: Assessing whether Variation in Measures Is due to Methods or due to True Variation between Populations

Figure 1 shows circulating estradiol concentrations (medians and interquartile ranges) for postmenopausal women who were controls in the Endogenous Hormones and Breast Cancer Collaborative Group (EHBCCG; ref. 3) and the European Prospective Investigation into Cancer and Nutrition (EPIC; ref. 4). In the EHBCCG, studies were conducted in different places at different times, with no standardization of methods for blood collection and storage, or type of assay used. The figure shows that median estradiol concentrations in controls varied from 21.7 pmol/l in the ORDET (Study of Hormones and Diet in the Etiology of Breast Tumors) study in Italy to 101 pmol/l in the New York University Women's Health Study in the United States, a 4.7-fold difference,

Authors' Affiliation: Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, Oxford University, Richard Doll Building, Roosevelt Drive, Oxford, United Kingdom

Corresponding Author: Timothy J. Key, Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, Oxford University, Richard Doll Building, Roosevelt Drive, Oxford OX3 7LF, United Kingdom. Phone: 44-1865-289-600; Fax: 44-1865-289-610. E-mail: tim.key@ceu.ox.ac.uk

doi: 10.1158/1055-9965.EPI-10-0061

©2010 American Association for Cancer Research.

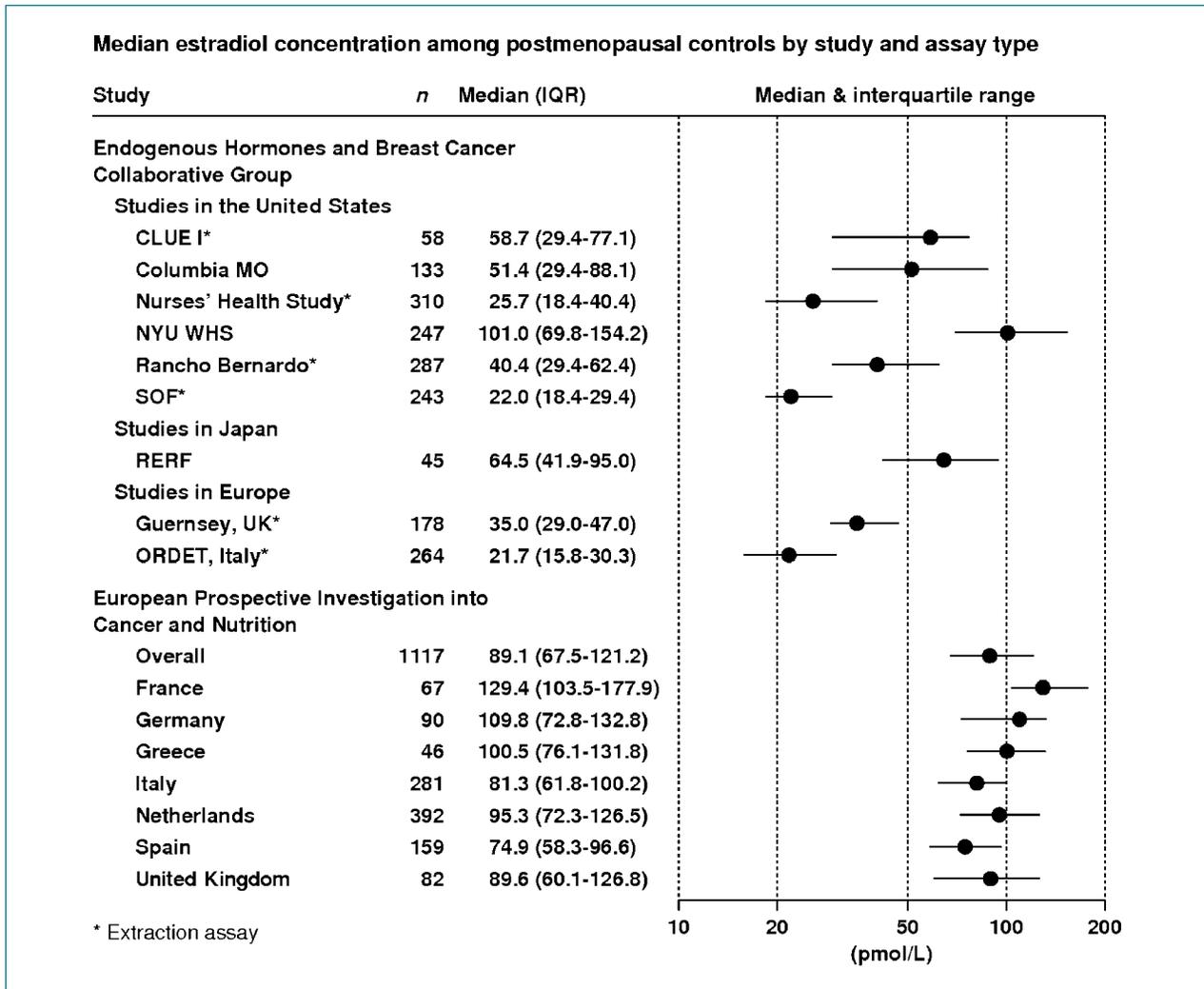


Figure 1. Median circulating estradiol concentrations [interquartile ranges (IQR)] for postmenopausal control women from studies in the EHBCCG and the EPIC (3, 4). CLUE, Give Us a CLUE to Cancer; NYU WHS, New York University Women's Health Study; RERF, Radiation Effects Research Foundation.

and concentrations varied almost as much (4.6-fold) between the six studies within the United States. In EPIC, with standardized methods of blood collection and all assays conducted in the same laboratory (but not in the same batch), there was relatively little difference in median circulating concentrations between women in the seven participating European countries (1.5-fold variation), but estradiol concentrations in all seven countries in EPIC were much higher than those in two other studies in European countries in the EHBCCG (Italy and the United Kingdom). The large differences between studies even within one country, together with the smaller differences between countries in the standardized EPIC study, strongly suggest that most of the difference in median estradiol concentrations between studies is due to the assay and other methodologic characteristics (e.g., blood collection procedures, time between blood collection and freezing of the samples, and temperature of storage), rather

than due to true differences in hormone levels between populations. Some of the difference between studies seems to be due to whether the estradiol assay incorporated an extraction step because nonextraction methods generally produce higher estimates of estradiol than extraction methods (Fig. 1, studies using extraction assays marked with an asterisk; ref. 5).

The ~5-fold variation between studies in median estradiol concentrations in postmenopausal women may be an extreme case, but substantial variation also occurs for other hormones. For the other hormones in postmenopausal women analyzed by the EHBCCG, variations in median concentrations between studies were over 3-fold for estrone and between 2 and 3-fold for androstenedione, DHEA sulfate, testosterone, and sex hormone binding globulin (SHBG). In the Endogenous Hormones and Prostate Cancer Collaborative Group, a collaborative analysis of 18 prospective studies with hormone

measures in men (6, 7), variations between studies were between 1.5- and 2-fold for dihydrotestosterone and androstenedione; between 2- and 3-fold for testosterone, androstenediol glucuronide, and DHEA sulfate; and over 3-fold for estradiol and SHBG.

Hormones may vary relatively little between populations because they are subject to homeostasis, whereas biomarkers of nutrient intake can vary more widely because they are dependent on dietary habits. Figure 2 shows the medians and interquartile ranges for two hormones, for SHBG, and for three nutritional biomarkers, measured in controls in nested case-control analyses of prostate cancer risk in men in six countries in EPIC (8-12). For all these measures, blood was collected, processed, and stored according to a standardized protocol, and all assays for each biomarker were conducted in the same laboratory. The biomarker concentrations are plotted with values scaled to the overall median concentration to enable easy comparison. For testosterone, SHBG, and insulin-like growth factor I, the variation in medians between European countries is <1.5-fold. Docosahexaenoic acid also differs relatively little between countries (1.6-fold), but for genistein and lycopene, which are present at very high levels in soya products and tomatoes, respectively, the difference between countries is substantially greater (4.8- and 2.1-fold, respectively), and the interquartile ranges within each country are also relatively wide. In this multicountry study with standardized methods, the large differences in some biomarker levels between countries are due to true differences in diet.

Laboratories routinely include quality control samples in each assay batch to estimate the within-assay and between-assay variation. This information should be reported in the methods for pooled analyses and could be used to categorize studies according to this measure of assay precision. If all contributing studies used the same samples for quality control, it would be possible to use the quality control measures to adjust values from each study to a common scale (13). Data of this type have not been available in the pooled analyses we have done to date, but this would be a desirable design feature for future studies. Examples of standardized quality control schemes are the United Kingdom National External Quality Assessment Service, which provides collaborating laboratories with the same quality control samples so that the differences in measures between laboratories can be quantified (14, 15), and the Centers for Disease Control project on standardizing steroid hormone measurements (16, 17).

One factor that could cause real differences in biomarker concentrations between populations is ethnic group and its associated life-style and genetic characteristics. As with the general discussion above, uncertainties in the comparability of assays can make it impossible to determine whether differences in biomarker levels in different ethnic groups from separate studies are real or are due to methodologic differences. On the other hand, if

some of the studies include more than one ethnic group with biomarker measurements made using the same methods, ethnic group can be examined as a potential modifier of any associations observed.

Methods for Combined Analysis of Pooled Biomarker Data from Different Studies

For studies in which blood collection, storage, and assay methods have been standardized, and or in pooled analyses of biomarkers that do not vary much by type of assay (for example cholesterol), we can be reasonably confident that differences between cohorts are due to true differences between populations; ideally, the comparability of the measures from the contributing studies can be confirmed by the use of common quality control samples or by remeasurement in one laboratory of a sample of specimens from each study. Standard methods of statistical analysis are appropriate; relative risks can be estimated in relation to the actual concentration of the biomarker, and analyses in categories should use the same cut-points across all contributing centers. This approach maximizes the ability to examine associations across the whole range of biomarker levels covered by the contributing individual studies.

For studies in which methods have not been standardized and there are substantial differences between contributing studies that are thought to be largely due to methodologic differences, statistical analyses cannot directly use the measured concentrations of the biomarker of interest. For example, in the EHBCCG and the Endogenous Hormones and Prostate Cancer Collaborative Group, we assumed that the true mean and range of hormone concentrations in each study is similar and we analyzed the data using study-specific quintiles, summing the study-specific odds ratios to provide an overall risk estimate. A possible alternative would be to standardize the data for each study (possibly after log transformation) to have a mean of zero and a SD of one, and then analyze the data using an increment of one SD. This method makes similar assumptions to using study-specific quintiles, but is slightly more complex in that the results are expressed in units of SDs, which may be less easy to interpret.

In addition to categorical analyses, it is useful to summarize the results with a continuous measure of risk. In individual studies, or in multicenter studies with standardized methods, this is best done by estimating the relative risk per unit increment in the concentration of the biomarker of interest (on an appropriate scale). This is not possible when the measures from different studies vary due to methodologic differences. A simple alternative is to estimate the relative risk in each contributing study for a difference in the biomarker from the median of the lowest fifth to the median of the highest fifth—that is, from the 10th to the 90th centile. This gives a linear estimate that should be similar to the relative risk in the

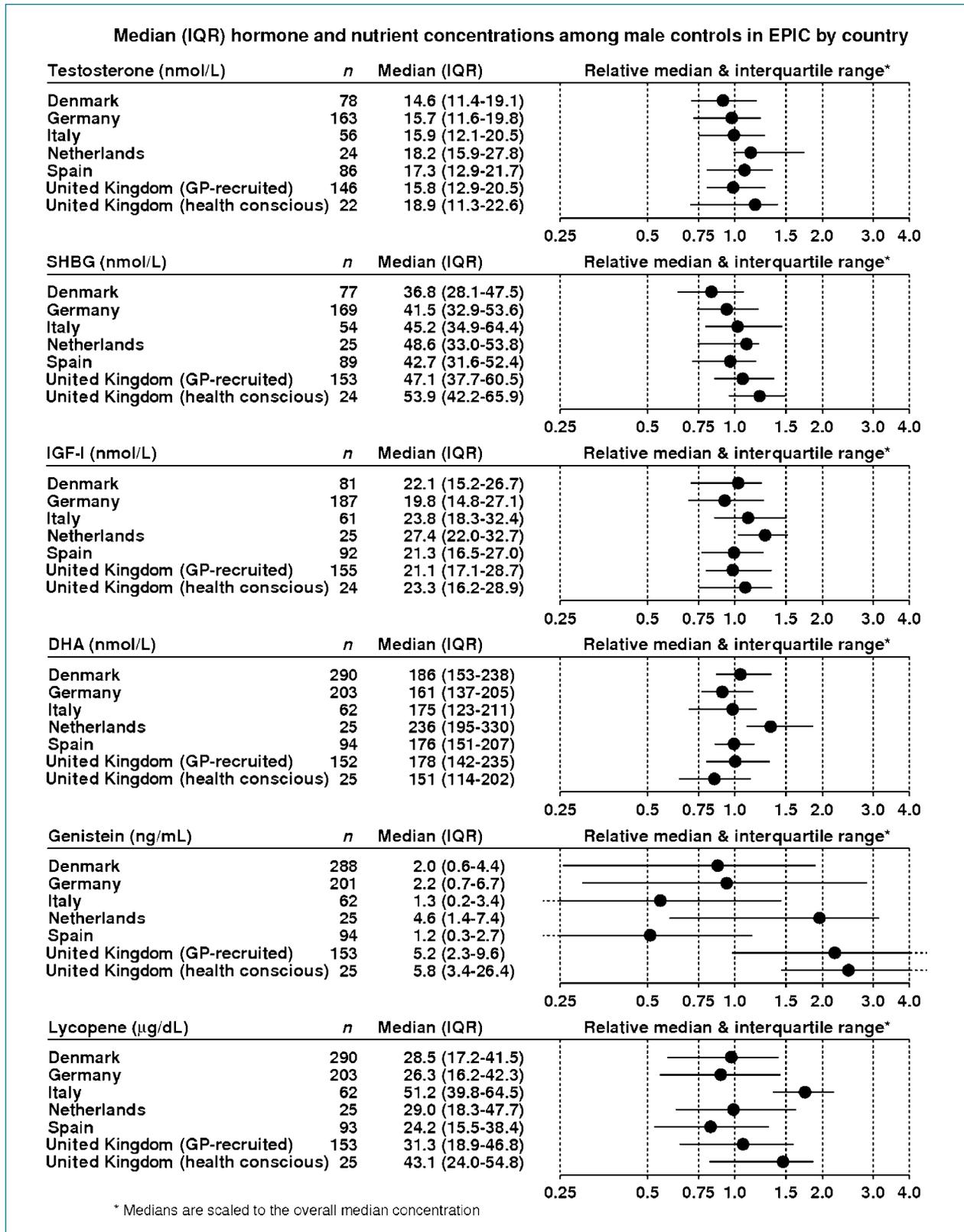


Figure 2. Median circulating biomarker concentrations [interquartile ranges (IQR)] and relative medians (interquartile ranges) for control men in nested case-control studies of prostate cancer from the EPIC (8-12). IGF-I, insulin-like growth factor; DHA, docosahexaenoic acid.

highest versus the lowest fifth, and this is the approach we used in analyses of sex hormones, insulin-like growth factor I, and prostate cancer risk (6, 7); however, a limitation of this approach is that it gives no information on the magnitude of the difference in biomarker concentrations between the highest and lowest fifths. An alternative is to estimate the relative risk for a biologically plausible percentage increase of the biomarker, such as a doubling (100% increase). This is again a simple approach, which can be readily transferred to biomarker concentrations in an individual study and is the one we have used in analyses of sex hormones and breast cancer risk in postmenopausal women (3); a disadvantage is that this method assumes a log-linear relationship.

Matching

Most nested case-control studies of biomarkers and cancer have used a matched design, in which controls are matched on characteristics such as date of and age at blood collection. Usually, the matched sets of blood samples from each case and corresponding control(s) are sent to the laboratory in batches to ensure that they will be assayed on the same day, thus eliminating between-batch variation from the case-control comparisons. In this situation, it is preferable to maintain the original matching in future pooled analyses unless there is a particular reason for breaking the matching.

Other Aspects of the Analysis of Pooled Data on Biomarkers

Most aspects of the analysis of pooled biomarker data are the same as those that apply to observational epidemiology in general. An area that requires particular attention is to examine the data for evidence of heterogeneity between studies due to technical differences, such as the type of blood sample, conditions of blood processing, storage time and temperature, and type of assay. In practice, this may be complex because studies are likely to differ from each other in several ways and it is hard to be confident which differences are most relevant. There may be an *a priori* reason to expect heterogeneity due to technical factors. For example, in our analyses of sex hormones and breast cancer in postmenopausal women, we divided the studies according to whether the assays for

the major sex hormones had, or had not, incorporated an extraction step. These analyses did not suggest large differences in associations with breast cancer risk according to the assay protocol (3).

Measurement error in biomarkers and within-person variation lead to regression dilution bias and various methods have been developed to correct for this bias, using repeat measures on a sample of participants (18). Thus far, there has been relatively little work on correcting pooled analyses of hormone data for measurement error, but this is an important topic and will require more studies with repeat measures of hormones.

Conclusions

Pooling biomarker results from different studies is important, mainly because it increases the sample size, enabling investigators to make robust estimates of risk and to examine associations in major subgroups of the population. The ideal situation is to have standardized methods in all studies so that the biomarker data can be pooled in their original units and, after appropriate correction for measurement error, the results can be expressed as the change in risk for a specified change in biomarker concentration. However, even when the studies do not have standardized methods, as with existing studies on hormones and cancer, simple approaches using study-specific quantiles or percentage increases can provide valuable information on the relationship of the biomarker with cancer risk (3, 6, 7).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Sabina Rinaldi and Rudolf Kaaks for comments on the manuscript and Elio Riboli and collaborators in EPIC for use in the figures of data adapted from several EPIC publications.

Grant Support

Cancer Research UK.

Received 01/15/2010; revised 02/12/2010; accepted 02/15/2010; published OnlineFirst 03/16/2010.

References

- Lewington S, Whitlock G, Clarke R, et al. Prospective Studies Collaboration. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet* 2007;370:1829–39.
- Santen RJ, Lee JS, Wang S, et al. Potential role of ultra-sensitive estradiol assays in estimating the risk of breast cancer and fractures. *Steroids* 2008;73:1318–21.
- Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst* 2002;94:606–16.
- Kaaks R, Rinaldi S, Key TJ, et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. *Endocr Relat Cancer* 2005;12:1071–82.
- Lee JS, Ettinger B, Stanczyk FZ, et al. Comparison of methods to measure low serum estradiol levels in postmenopausal women. *J Clin Endocrinol Metab* 2006;91:3791–7.
- Endogenous Hormones and Prostate Cancer Collaborative Group.

- Endogenous sex hormones and prostate cancer: a collaborative analysis of 18 prospective studies. *J Natl Cancer Inst* 2008;100:170–83.
7. Roddam AW, Allen NE, Appleby P, et al. Insulin-like growth factors, their binding proteins, and prostate cancer risk: analysis of individual patient data from 12 prospective studies. *Ann Intern Med* 2008;149:461–71, W83-8.
 8. Travis RC, Key TJ, Allen NE, et al. Serum androgens and prostate cancer among 643 cases and 643 controls in the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer* 2007;121:1331–8.
 9. Allen NE, Key TJ, Appleby PN, et al. Serum insulin-like growth factor (IGF)-I and IGF-binding protein-3 concentrations and prostate cancer risk: results from the European Prospective Investigation into Cancer and Nutrition. *Cancer Epidemiol Biomarkers Prev* 2007;16:1121–7.
 10. Crowe FL, Allen NE, Appleby PN, et al. Fatty acid composition of plasma phospholipids and risk of prostate cancer in a case-control analysis nested within the European Prospective Investigation into Cancer and Nutrition. *Am J Clin Nutr* 2008;88:1353–63.
 11. Travis RC, Spencer EA, Allen NE, et al. Plasma phyto-oestrogens and prostate cancer in the European Prospective Investigation into Cancer and Nutrition. *Br J Cancer* 2009;100:1817–23.
 12. Key TJ, Appleby PN, Allen NE, et al. Plasma carotenoids, retinol, and tocopherols and the risk of prostate cancer in the European Prospective Investigation into Cancer and Nutrition study. *Am J Clin Nutr* 2007;86:672–81.
 13. Stanczyk FZ, Lee JS, Santen RJ. Standardization of steroid hormone assays: why, how, and when? *Cancer Epidemiol Biomarkers Prev* 2007;16:1713–9.
 14. Middle JG, Kane JW. Oestradiol assays: fitness for purpose? *Ann Clin Biochem* 2009;46:441–56.
 15. United Kingdom National External Quality Assessment Service (UK NEQAS). <http://www.ukneqas.org.uk/>.
 16. Vesper HW, Botelho JC, Shacklady C, Smith A, Myers GL. CDC project on standardizing steroid hormone measurements. *Steroids* 2008;73:1286–92.
 17. Centers for Disease Control and Prevention (CDC) project on standardizing steroid hormone measurements. http://www.cdc.gov/nceh/dls/hormone_standardization.htm.
 18. Wood AM, White I, Thompson SG, et al, Fibrinogen Studies Collaboration. Regression dilution methods for meta-analysis: assessing long-term variability in plasma fibrinogen among 27,247 adults in 15 prospective studies. *Int J Epidemiol* 2006;35:1570–8.

BLOOD CANCER DISCOVERY

Pooling Biomarker Data from Different Studies of Disease Risk, with a Focus on Endogenous Hormones

Timothy J. Key, Paul N. Appleby, Naomi E. Allen, et al.

Cancer Epidemiol Biomarkers Prev 2010;19:960-965. Published OnlineFirst March 16, 2010.

Updated version Access the most recent version of this article at:
doi: [10.1158/1055-9965.EPI-10-0061](https://doi.org/10.1158/1055-9965.EPI-10-0061)

Cited articles This article cites 16 articles, 5 of which you can access for free at:
<http://cebp.aacrjournals.org/content/19/4/960.full#ref-list-1>

Citing articles This article has been cited by 7 HighWire-hosted articles. Access the articles at:
<http://cebp.aacrjournals.org/content/19/4/960.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cebp.aacrjournals.org/content/19/4/960>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.