Fibrinogen's Role in the Risk of Coronary Heart Disease

Abigail Mabe

Introduction

Cardiovascular diseases are the leading cause of death globally. Coronary Heart Disease (CHD) is the most common form of heart disease, defined in the following analysis as the first non-fatal myocardial infarction or coronary heart in those without known cardiovascular disease at an initial examination (Thompson). This disease is prevalent among over 200 million people globally, with approximately 375,000 deaths attributed to CHD annually. Then, as CHD is such a widespread and pressing disease, there are many global collaborations aimed at addressing possible risk factors, with The Emerging Risk Factors Collaboration (ERFC) being one of the largest. The ERFC hosts individual records from more than 1.2 million participants in 116 prospective studies of major cardiovascular disease outcomes. Of this large database, the Fibrinogen Studies collaboration (a subset of the ERFC containing data on 154,211 participants across 31 prospective studies) analyzes baseline plasma fibrinogen concentration and the correlated risk of CHD, which is what this paper focuses on. Fibrinogen is a protein involved in forming blood clots within the body and can be an important risk factor to developing CHD; the importance of tracking this can help many who may unknowingly have CHD (which is about 1 in 5 people) and can bring awareness to get them proper help. The following study focuses on the association between baseline fibrinogen and the correlated risk of CHD over time.

Methods

The main analysis of this paper considers a Cox Proportional Hazards (PH) model, estimated for each of the 31 studies separately. The PH model displays the evolution of risk for CHD over time based on differing levels of baseline fibrinogen in patients, independently analyzed for each study; this model is considered the "first step" of the analysis. The models are stratified by sex and randomized group (if the respective studies include randomization). Then, for each study s with strata k (where individuals i have exposure interest E_{si}, effected by covariates X_{si}, which will be mentioned again later), the hazard at time t after baseline is modelled by:

$$\log(h_{ski}(t|E_{si}, X_{si})) = \log(h_{0sk}(t)) + \beta_s E_{si} + \gamma_s X_{si}$$

In the above equation, the parameters of interest are the log hazard ratios (HRs, denoted β_s per unit increase in exposure in study s, adjusted for X_{si}. The variance of the estimated β_s, written as v_s, is utilized in the "second step", as outlined below.

The "second step" in this analysis is a random-effects meta-analysis model, which combines within-study information and between-study variance (denoted τ^2) on the relationship between exposure and risk, as written below:

$$\hat{\beta}_s = \beta_s + \epsilon_s, \text{ where } \epsilon_s \sim N(0, v_s)$$
$$\beta_s = \beta + n_s, \text{ where } n_s \sim N(0, \tau^2)$$

A fixed-effects model was created parallel to the random-effects model above to facilitate further analysis.

Due to the high level of heterogeneity found in the above models, various confounder adjustments on baseline measurements were analyzed. Examples of these confounders include age, sex, smoking, total cholesterol, systolic blood pressure, and BMI, as well as the interactions between them. In addition, joint effects and effect modifiers were introduced in the original PH model to incorporate both within and between-study effects for the confounders stated above. This part of the analysis ultimately estimates the interaction between the respective potential effect modifier and fibrinogen to assess the correlated risk of CHD. For effect modifiers such as age, the model creates an overall interaction term based on within-study information, δ_w, by adding the term $\delta_s E_{si} X_{si}$ to the original "first step". For effect modifiers such as type of population recruited, which are assessed at the study level and rely on between-study comparisons, a between-study interaction term (δ_B) is created through the "second step" model. By adding $\delta_B \ X_s$ to the second equation in that model, the interaction term allows for continual analysis of the balance between within-study and between-study information.

In addition to the main analysis above, additional analyses are conducted to explore the linearity of risk exposure and log HR. To do so, the cox PH regression for each study is separately analyzed. The set of the log HRs from these analyses are pooled across studies using a multivariate version of random-effects meta-analysis and plotted against mean exposure level in each quantile group. The purpose of this is to view the linearity of the model while allowing for inter-correlations within and between studies.

Finally, throughout the analysis, there has been an assumption that regression coefficients in the PH model do not change with time since the baseline measurements of fibrinogen. The authors combine interaction terms between exposure and time over studies and use random-effects meta-analysis along with chi-squared tests to assess the truth of the above assumption.

Results

The random-effects model produced a combined hazard ratio (HR $exp(\beta)$) of 1.57 per 1 g/l higher baseline fibrinogen, with a 64% heterogeneity (I^2) impact on the imprecision of the overall log HR. In contrast, the fixed-effects model produced an HR point estimate of 1.52 per 1 g/l higher baseline fibrinogen. Both results relate to the mean HR across all studies, rather than reporting each study independently. These results indicate a strong association between baseline fibrinogen levels and the evolution of CHD risk over time, given the hazard ratios are both significantly larger than 1. It also indicates a slightly stronger association between baseline fibrinogen and the risk of CHD within the random-effects model, although the high heterogeneity needs to be investigated further.

Since 64% heterogeneity has such a high impact on the overall log HR results, additional analysis on confounders are required for greater precision. As stated above, these confounders include age, sex, smoking, total cholesterol, systolic blood pressure, and BMI (as well as the interactions between them) in different studies. It was found that 29% of the effect is explained by the observed values of these confounders, which leads us to believe there can be biases induced by within-study confounding factors.

By further analyzing these confounding issues by observing the joint effects and effect modifiers, interactions between BMI and age are clear, as they both have small P-values (< 0.05). These interactions are estimated as a potential effect modifier on fibrinogen as additional contributions to estimating the risk of CHD. However, BMI and age are the only two potential effect modifiers that are significant to interacting with fibrinogen in the model. Another notable observation from this part of the analysis is the balance between within-study and between-study information. The standard error of the interaction of within-study information is smaller (0.061) than between-study information (0.092), which leads to the conclusion that it is more reliable to utilize within-study pooled interaction estimates.

Considering the additional analyses mentioned briefly above, the plot displaying pooled log HRs over mean baseline fibrinogen (g/l) displays the result that a log-linear model for risk is satisfactory. In addition, there is no evidence of departures of regression coefficients in the PH model for fibrinogen, and no evidence of heterogeneity between studies in this regard.

Discussion of Normative Concerns

The study outlined above, although it tries to limit biased results by thorough methodology, has a potential for biases that could possibly lead to both unreliable results and ethical dilemmas. First and foremost, there are fundamental issues considering the source of data utilized for the study. The Emerging Risk Factors Collaboration (ERFC) gathers its data from various publications, literature searches, and correspondence with authors of relevant reports (Thompson). Since this data is not coming directly from medical directories or other completely reliable medical sources, the ERFC could display publication and reporting biases, especially considering this is data gathered in predominantly Western populations. This can lead to arbitrary discrimination within the analysis itself, or in the implementation of it globally. This can be dangerous in many regards considering possible misestimation in the risk of CHD within non-western populations, especially given CHD is the most common form of heart disease.

Another issue with the ERFC concerns data privacy. The ERFC does not censor individuals at the time of cardiovascular investigations or intervention, or at the diagnosis of angina because the incidence of such occurrences is not recorded reliably enough in sufficient studies (Thompson). This can also lead to potential biases within the model, as well as potential issues in regulation of sensitive information. If sensitive information was somehow breached or accessed from the ERFC, theft or manipulation of personal health records could occur, as well as unauthorized hacking of personal data, data mining/selling, or resulting arbitrary discrimination. These are all prominent examples of why healthcare data is well protected, and why it should not be in large databases such as the ERFC, due to a higher possibility of breach.

In addition to these concerns with the ERFC itself, there are a few potential biases that could arise from the methodology itself. The first of these is an issue with the variability of the standard deviation between studies. Due to the considerably large variation, issues rise when considering within-study confounding factors. It is hard to discriminate between what is a between-study confounding factor and a within-study confounding factor, especially when considering the large amount of data. In addition, some of the studies were randomized, where others were not; this could lead to biases within the respective studies themselves, and once again lead to arbitrary discrimination against (or for) certain populations. There is a trade-off to consider between precision and bias to use between-study information in addition to within-study information; too much between-study information can lead to greater bias within the model while increasing precision, so a balance needs to be met. Otherwise, there might be breaches of fairness in the realm of bias and arbitrary discrimination, as well as impartiality towards all respective study populations.

Possible underestimation, caused by measurement error and within-person fibrinogen variation, can lead to downward bias of an estimated association between fibrinogen and CHD risk. If the model estimates a lower risk association for the patient than what their actual risk association is, in relation to baseline fibrinogen, it puts the patient at a much higher risk of CHD-related issues. This is a concerning point, especially when considering 20% of people are not aware of their CHD symptoms as-is. The study tries to correct for possible measurement error/underestimation, and results in a corrected HR of 1.96 per 1 g/l increase in fibrinogen v. 1.38 uncorrected, further displaying the issue highlighted above.

Overall, the study displays useful analysis in estimating the associated risk of CHD based on fibrinogen levels in patients. Although its methods are extensive, and results seem reliable, further exploration uncovers possible biases that could lead to fundamental issues and moral dilemmas with the study. However, the authors of the paper are finding ways to limit these biases with various methods of confounder adjustments, citing data collection follow-ups (within the ERFC itself), and exploring ways that aim at limiting measurement errors as a whole.

References

Thompson, Simon, et al. "Statistical Methods for the Time-to-Event Analysis of Individual Participant Data from Multiple Epidemiological Studies." OUP Academic, Oxford University Press, 3 May 2010, academic.oup.com/ije/article/39/5/1345/801905#91888123.