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 is the most common form of heart disease, prevalent among over 200 million people globally, with approximately $375,000$ deaths attributed to it annually. Then, as coronary heart disease 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 data subsumed by the ERFC containing data on $154,211$ participants across 31 prospective studies) analyzes baseline plasma fibrinogen concentration and the correlated risk of coronary heart disease, which is the focus of this paper. Fibrinogen is a protein involved in forming blood clots within the body and can be an important risk factor to developing coronary heart disease; the importance of tracking this can help many who may unknowingly have coronary heart disease (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 coronary heart disease over time.

Analysis of Methods

The paper focuses specifically on survival analysis, or a family of statistical techniques aimed at analyzing "time-to-event" data. In other words, survival analysis assesses the relationship between a given exposure and the occurrence of an outcome after a follow-up period among a cohort of individuals (Abd). This does not have to measure or refer to mortality in particular, but to any event of interest. For this analysis, the event of interest is the presence of coronary heart disease, defined as the first non-fatal myocardial infarction

1

or coronary death in those without known cardiovascular disease at an initial examination (Thompson). Then, the motivation behind implementing survival analysis stems from the objective of using baseline plasma fibrinogen concentration in approximating the relative risk of coronary heart disease.

Due to the nature of survival analysis, and the limitation of individual-based medical data, this analysis will be a rather detailed explanation behind the motivations of this paper, as well as the methods used in the paper's analysis.

One fundamental method in survival analysis, and a frequent perquisite to other methods, is called the Kaplan-Meier method; this method is the simplest way of computing survival over time. Kaplan-Meier computes the cumulative survival between exposure groups, allowing an immediate perception on the observed event with consideration of censored observations (these include enrolled patients that do not experience events of interest, such as when a patient is lost to follow-up due to an unrelated death or change of residence). In context, the Kaplan-Meier method would compute the cumulative sum of those who have developed coronary heart disease, the event of interest, over time. This would give a visual of a cumulative survivor curve over time but would not disclose much else about how other risk factors, such as fibrinogen, would play a role in the relative risk of coronary heart disease. This example illustrates well some of the limitations of the Kaplan-Meier method. These include the lack of adjustment for confounders and aim at testing causal-effect relationships rather than incremental changes in relative event risk compared to a candidate risk factor over time. These issues can be rectified with a Cox proportional hazards (PH) model, verifying this as the paper's survival analysis of choice.

Cox proportional hazards models incorporate patients' baseline characteristics such as age, clinical history, or in this case, plasma fibrinogen (measured in g/l). In addition, these models can be stratified to consider the effect of a covariate within each stratum such as age, treatment arm, or gender in this case. Stratifying PH models often accounts for heterogeneity among the strata, relaxing the assumption of proportional hazards across a whole population; this can increase precision of estimates and test for possible interactions between a covariate and strata. In this case, the covariate of interest is baseline plasma fibrinogen, and the strata is gender (male or female). It is noted in the paper that the PH models are stratified by randomized group, if applicable, however there are few instances of this in the data; then for the interpretation of these models, gender is the only notable strata included.

Using this information, PH models can be estimated for each study separately; these are aimed at analyzing the relative risk of coronary heart disease over time dependent on baseline plasma fibrinogen and stratified by sex. As a reminder, the data this study is based on comes from a subset of the Emerging Risk Factors Collaboration (ERFC) called the Fibrinogen Studies Collaboration, which houses individual records

from $154,211$ participants in 31 prospective studies. These models are set up as follows:

Each study is labeled as $s = 1 \ldots S$, where $S = 31$ to represent the 31 prospective studies in the Fibrinogen Studies Collaboration. In addition, there are strata $k = 1 \ldots K_s$ (where for most studies $K_s = 2$ for two genders) and individuals $i = 1 \ldots n_s$, where n_s is the number of participants in each respective study. These are modeled with the exposure of interest E_{si}, which is baseline plasma fibrinogen (g/l) and other covariates X_{si} that must be included as adjustment for confounding effects. Given these variables, the relative risk of coronary heart disease at time t after baseline is modelled as:

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

In this formulation, the (β_s) are the log hazard ratios per unit increase in the exposure in study s, adjusted for the confounding effects of the covariates X_{si}. This means for every 1 g/l increase in baseline fibrinogen in study s, the $exp(\beta)$ indicates the corresponding relative risk of coronary heart disease. These risk estimates, termed as log hazard ratios (HRs), can be combined over studies using random-effects meta-analysis. To break this term down, meta-analysis is used to synthesize quantitative information from related studies and produce results that summarize a whole body of research (Riley). Then, in this study, meta-analytical methods are used to combine the 31 study estimates of baseline plasma fibrinogen (g/l) on the relative risk of coronary heart disease into a summary estimate. This is helpful to analyze the whole body of research as the overall exposure of baseline fibrinogen on the relative risk of coronary heart disease.

Then, the discussion must be extended to describe a random-effects meta-analysis model. This assumes the observed estimates of treatment effect vary across studies because of real differences in each study as well as sampling variability (Riley). This is a good analysis to use because there are 31 different prospective studies to include, so considering heterogeneity in treatment effects is important. This heterogeneity could be caused by differences in study populations, follow-up length, or fibrinogen measurement discrepancies. This random-effects meta-analysis is given as follows:

$$\hat{\beta}_s = \beta_s + \epsilon_s, \;\; \text{where} \;\; \epsilon_s \sim N(0, v_s) \;\; \text{and } v_s \text{ is the variance of betas}$$

$$\beta_s = \beta + \eta_s, \;\text{where } \eta_s \sim N(0, \tau^2)$$

Again, $exp(\beta)$ is the average HR, or average relative risk of coronary heart disease dependent on 1 g/l increase in baseline plasma fibrinogen, over the 31 prospective studies. However, the equations above allow for heterogeneity in the true log HRs between studies, represented by the variance τ^2, estimated through a

standard moment estimator. The impact of heterogeneity on the imprecision of the overall log HR is given by I^2. In other words, this is the percentage of variance in the point estimates of the study-specific log HRs that is attributable to between-study variation (likely when considering differences in study population or measurement discrepancies). Values of I^2 close to 0% correspond to a lack of heterogeneity, which is far from true in this case. Heterogeneity is important to observe to correctly investigate the impact of various factors on the strength of association between baseline plasma fibrinogen and the relative risk of coronary heart disease.

Another method that has been employed in parallel analysis in the ERFC is a fixed-effects meta-analysis. This is similar to a random-effects meta-analysis but assumes no between study heterogeneity and no differences due to chance. This means the value of I^2 would be 0%, which is quite contrary to the random-effects meta-analysis, indicating variability in study estimates are due entirely to chance. Therefore, in this study, there should be more of a focus on random-effects meta-analysis due to the multitude of study populations and the resulting heterogeneity that is bound to occur, although results from both models can be observed for general comparison.

The above suggestions can be observed in Figure 1. This visualization displays the study-specific HRs for the relationship of baseline fibrinogen with the risk of coronary heart disease in the 31 prospective studies from the Fibrinogen Studies Collaboration. The overall random-effects meta-analysis is displayed in comparison to the overall fixed-effects meta-analysis at the bottom of the plot. This visual comparison reinforces the idea that a random-effects model is best suited to represent the considered studies, especially given the variation in study-specific HRs clearly displayed between studies in the visualization. In addition, the figure gives a random-effects prediction interval for the true HR in a new study, estimated by $exp\left(\hat{\beta} \pm t_{s-2}\sqrt{var(\hat{\beta}) + \tau^2}\right)$, where t is the 2.5-percentile of the t-distribution, and S is the number of studies. This interval is wide due to the presence of obvious heterogeneity, but it remains above 1, indicating a positive correlation between baseline fibrinogen and relative risk of coronary heart disease.
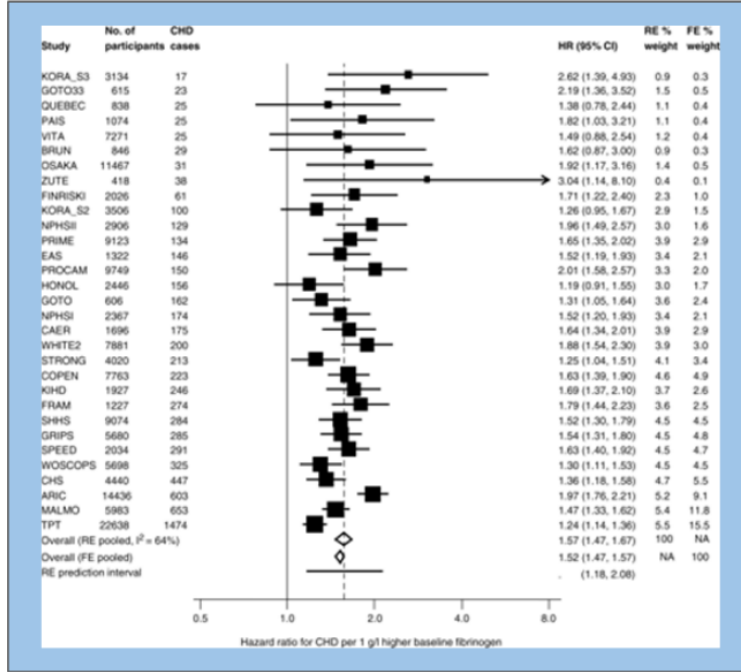
Figure 1: Study-specific HRs and 95% CI (log scale) for the relationship of baseline fibrinogen with coronary heart disease in 31 studies and meta-analysis.

The results of the random-effects meta-analysis and fixed-effects meta-analysis are displayed in the table below (Figure 2). Again, these are the combined HRs for the relationship between baseline fibrinogen (g/l) and risk of coronary heart disease, stratified by age in each study separately. The numbers in the HR column represent the relative risk of coronary heart disease per 1 g/l higher baseline fibrinogen concentration (and intuitively the log HR is simply the log of these values). The between-study variance, or heterogeneity in the true log HRs between studies, is given by τ^2, where the p-value corresponds to the test for heterogeneity. Since the p-value is less than 0.05, or a typical level of significance, there is statistically significant proof for heterogeneity in the analysis, which solidifies the reason for random-effects meta-analysis over fixed-effects meta-analysis. The I^2 value, or impact of heterogeneity on the imprecision of the overall log HR, is given in the last column.

Method	HR (95% CI)	Log HR, $\hat{\beta}$ (SE)	Heterogeneity		
			Between-study variance $\hat{\tau}^2$	P-value	I^2 (95% CI)
Untransformed fibrinogen: log HRs per 1 g/l increase					
Random-effects meta-analysis	1.57 (1.47–1.67)	0.450 (0.033)	0.018	<0.0001	64% (48, 76)
Fixed-effects meta-analysis	1.52 (1.47–1.57)	0.419 (0.018)	NA	<0.0001	NA

Figure 2: Combined HRs for the relationship between baseline fibrinogen (g/l) and coronary heart disease risk.

Then, it can be seen the random-effects meta-analysis combined HR exp(beta) is estimated at 1.57 per 1 g/l higher baseline fibrinogen concentration with substantial heterogeneity across studies, portrayed by an I^2 value of 64%. This means the relative risk for coronary heart disease increases by 1.57 for every 1 g/l increase in baseline fibrinogen concentration by this model. In contrast, the fixed-effects meta-analysis model estimates 1.52 per 1 g/l higher baseline fibrinogen concentration. However, since there is obvious heterogeneity in the analysis, this model is not as precise when considering between-study variation. The model described above is a two-step procedure, where the first part analyzes each study independently through PH models, and the variables of interest (β), denoting the log HRs, are combined in the random-effects meta-analysis. Hypothetically, these steps could be consolidated into a one-step method which could provide different insights to the variance of betas (v_s) in the first step. This would be represented by the following:

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

$$\beta_s = \beta + \eta_s, \text{ where } \eta_s \sim N(0, \tau^2)$$

A one-step method, such as the one suggested above, is computationally infeasible in a dataset the size of the ERFC, however both the one-step and two-step methods produce similar results.

Revisiting the random-effects meta-analysis described above, the use of a study-specific standard deviation score (or other non-linear risk relationships) might reduce the heterogeneity of risk association between studies. This is because the distributions seemed to be skewed slightly positively, and the standard deviation varied considerably between studies. Previously, there was an assumption that the log HR increases linearly with exposure, but there should always be caution when accepting assumptions. Then, an assessment of non-linear risk relationships should be considered that could possibly indicate a threshold or plateau for

risk.

An assessment of non-linear risk relationships could be approached as follows:

(1) Divide the distributions of the exposure into arbitrary groups such as fifths, which can be defined within or across studies.

(2) HRs in each quantile group are estimated by using Cox PH regression in each study separately, which reflects the original two-step analysis; however, these log HRs are not independent given they are relative to the same reference group.

(3) Due to the dependence defined in (2), a multivariate version of random-effects meta-analysis must be conducted to allow for pooling of log HRs with consideration of inter-correlations within and between studies.

(4) Pooled log HRs found in (3) are plotted against mean exposure level in each quantile group and assessed through sensitivity analyses.

As an additional note to (4), sensitivity analyses provide a more reliable way to assess the effect of these non-linear changes on the observed results; if the results are similar to the original, it can be assured that a linear risk relationship is appropriate.

The table below (Figure 3) shows the results from the sensitivity analyses aimed at determining the shape of the risk relationship between baseline fibrinogen and coronary heart disease. The below analyses compare untransformed fibrinogen, log fibrinogen, study-specific standard deviation fibrinogen score, and study-specific standard deviation log fibrinogen score. These results are displayed similarly to the ones in the first random-effects meta-analysis study, expressed as the relative risk in coronary heart disease per 1-standard deviation higher baseline fibrinogen (g/l). Since the values of HR, log HR, and heterogeneity are similar both to the original analysis and to each other, a linear risk relationship can again be assumed.

Method	HR (95% CI)	Log HR, $\hat{\beta}$ (SE)	Heterogeneity		
			Between-study variance $\hat{\tau}^2$	P-value	I^2 (95% CI)
Transformed fibrinogen: log HRs per SD increase—random-effects meta-analysis					
Untransformed fibrinogen	1.34 (1.29–1.40)	0.294 (0.022)	0.008	<0.0001	64% (48, 76)
Log fibrinogen	1.38 (1.32–1.45)	0.325 (0.025)	0.010	<0.0001	65% (48, 76)
Study-specific SD score fibrinogen	1.34 (1.29–1.40)	0.292 (0.021)	0.007	<0.0001	63% (45, 75)
Study-specific SD score log fibrinogen	1.37 (1.31–1.44)	0.316 (0.024)	0.009	<0.0001	64% (47, 76)

Figure 3: Combined HRs for the relationship between baseline fibrinogen (g/l) and coronary heart disease risk.

Special attention must be brought to the confounding effects of covariates, which was included as X_{si} in the original PH model. Confounders are a threat to the validity and precision of models and can be especially detrimental in the case of meta-analysis. In this case, a confounder would cause comparison groups to differ with respect to their risk of coronary heart disease beyond the exposure of baseline plasma fibrinogen. This would be a common cause of exposure and outcome such as age. Age is the most important confounder in many epidemiological applications, since age is associated with many physiological changes, so naturally this is the confounder that attracts the most attention for the study at hand.

For linear terms, age-at-baseline in a PH model is typically the first adjustment for a possible age confounding factor. An age-at-baseline model is equivalent to including current age as a time-dependent variable, which allows for age to be accounted for in both the individual study PH models and the random-effects meta-analysis model. More complex age adjustments, such as stratification by age categories at baseline (as 5-year bands), as well as polynomial terms and interactions with other covariates (such as sex) can also be used in PH models and corresponding random-effects meta-analysis; this is done to ensure a linear age-at-baseline model is a reasonable adjustment. The relative risk of coronary heart disease per 1 g/l increase in baseline fibrinogen, adjusted for baseline confounding variables age (including adjustments to age), sex, smoking, total cholesterol, systolic blood pressure, and body mass index are displayed in the following table (Figure 4).

With adjustment for	HR (95% CI)	Log HR $\hat{\beta}$ (SE)	χ_1^2	Heterogeneity		
				Between-study variance $\hat{\tau}^2$	P-value	I^2 (95% CI)
Age	1.57 (1.47–1.67)	0.450 (0.033)	181	0.018	<0.0001	64% (48, 76)
Age as 5-year age bands	1.57 (1.47–1.68)	0.451 (0.033)	183	0.018	<0.0001	64% (48, 76)
Stratification by 5-year age bands	1.57 (1.47–1.68)	0.451 (0.033)	182	0.018	<0.0001	64% (48, 76)
Age sex × age	1.57 (1.47–1.68)	0.450 (0.034)	180	0.018	<0.0001	65% (48, 76)
Age age^2	1.56 (1.46–1.67)	0.447 (0.033)	179	0.018	<0.0001	64% (48, 76)
Age age^2 sex × age sex × age^2	1.57 (1.47–1.67)	0.448 (0.034)	177	0.019	<0.0001	65% (49, 76)
Age smoking tchol sbp bmi[a]	1.38 (1.31–1.45)	0.320 (0.026)	156	0.006	0.028	35% (0, 58)

Figure 4: Combined HRs for the relationship between baseline fibrinogen (g/l) and coronary heart disease risk; random-effects meta-analysis adjusting for baseline confounding variables.

As viewed above, since the combined HR for coronary heart disease per 1 g/l increase in baseline fibrinogen, adjusted for age, is the same as adjusting for any of the more complex age models, the linear age-at-baseline PH model is adequate. However, a decrease in both the overall HR for fibrinogen and extent of heterogeneity can be observed when 4 additional covariates are introduced. This is indicated by the I^2

value of 35%, and the HR value of 1.38. In this case, the relative risk of coronary heart disease per 1 g/l increase in baseline fibrinogen falls by 29%; this is calculated by subtracting the difference of the new HR value of 1.38 from the original random-effect meta-analysis HR value of 1.57. In context, this means 29% of the relative risk of coronary heart disease dependent on baseline fibrinogen is explained by the linear age-at-baseline, smoking, total cholesterol, systolic blood pressure, and body mass index confounders. A less important aspect to the above table, but one that should still be noted due to its unfamiliarity, is the presence of the X_1^2 value; this is the Wald statistic. A Wald test can be used for models with dichotomous and for continuous variables. A higher Wald X_1^2 value indicates better predictive power for a set of variables; a variable is said to be "significant" if that variable adds some incremental value to the model. Then, as the X_1^2 value decreases with the addition of confounders in the model shown above, there is a slight decrease in the strength of evidence for an association between them.

In meta-analysis of individual participant data, there should be many considerations with possible interactions between individual effect modifiers such as age and the exposure of interest, or baseline fibrinogen. These interactions are best assessed using within-study information, so another two-step procedure is recommended where the interaction should be estimated in each study separately first, as follows:

For a single potential effect modifier X_{si}, the model in study s is

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

This model is very similar to the original PH model, with the inclusion of the interaction terms δ_s, which can be combined using random-effects meta-analysis to give the overall interaction Sw based on within-study information. However, some effect modifiers are exclusively assessed at the study level; these could include population recruited or the methods of fibrinogen measurement. Then, any information on interactions relies on between-study comparisons utilizing random-effects meta-regression. Meta-regression is an extension of meta-analysis, serving to appropriately combine multiple subsets of studies when a single summary measure does not seem sufficient to capture clinical diversity (How to Perform a Meta-Regression). Then, using the estimates of betas from the original PH model, the original random-effects meta-analysis model can be extended to include a study-level covariate X_s as follows:

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

$$\beta_s = \beta + \delta_B X_s + \eta_s, \text{ where } \eta_s \sim N(0, \tau^2)$$

9

Then, in this model, δ_B is the between-study interaction term. Some variables, notably sex and ethnic groups, have potential interactions with baseline fibrinogen for both within and between-study information. However, it should be noted that in general, between-study interactions are often prone to potential sources of between-study confounding. This poses a trade-off between increased precision and possible bias when choosing whether to use between-study information in addition to within-study information.

In the context of this study, there are many examples of interaction analyses for fibrinogen, some of which include age, systolic blood pressure, total cholesterol, and sex for both between and within-study interactions. Interactions of both age and body mass index with baseline fibrinogen are clear, as they display statistical significance in the form of a p-value that communicates the strength of the interaction. These p-values are small enough to reject the null hypothesis that there is no interaction effect between the respective effect modifier and baseline fibrinogen when assessing the relative risk of coronary heart disease. These p-values are derived from the standard error of the summary interaction effects between fibrinogen and the respective effect modifiers. In addition, with the sex interaction term analyzed for both between and within-study interactions, it is sensible to rely solely on within-study information given both the standard error of the interaction term for Sw is smaller than Sb, and the caution of potential bias for between-study estimates (which could be due to several factors, like population diversity, between studies).

Finally, there is an assumption that the regression coefficients in the original PH model do not change with time since baseline measurement. In other words, this is the assumption that there are no departures from estimating the relative risk of coronary heart disease dependent on baseline fibrinogen, hence the word proportional in a proportional hazards model. Covariates (such as the ones explored above) could plausibly alter the relative risk of coronary heart disease over time, but the general analysis is still focused on the primary effect of baseline fibrinogen. Then, this assumption can be estimated based on Schoenfeld residuals. The mean of these values should be 0 if the regression coefficients in the original PH model do not change with time, which makes it a good estimator of this assumption. Then, to test the assumption over all studies, interaction terms between baseline fibrinogen should be combined across studies.

Using random-effects meta-analysis one more time, the model is given by:

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

$$\xi_s = \xi + \eta_s, \text{ where } \eta_s \sim N(0, \tau^2)$$

Where β_s are separate fixed effects (fixed effects are used as a method for controlling all confounders

since the focus is on baseline fibrinogen rather than covariates), and the variable of interest is ξ, which can be tested using the Schoenfeld residual test. Then, since the results over each study combine to less than the expectation, there is no evidence of departures from PH for fibrinogen, and no evidence of heterogeneity between studies in this regard. This is reinforced by the p-value for the Schoenfeld residual test, which is 0.9, and too large to reject the null hypothesis that the regression coefficients in the original PH model do not change with time. Hence, the above assumption that there are no departures from estimating the relative risk of coronary heart disease dependent on baseline fibrinogen is confirmed.

Analysis of Normative Consideration

The largest concern with the analysis outlined above is with the collection of data within the Emerging Risk Factors Collaboration (ERFC), which is where the Fibrinogen Studies Collaboration resides. 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 is an example of what is called paternalism. Paternalism is an act of regulating or restricting the independence of a subordinate for their "own good." In this case, the ERFC is using patients' private cardiovascular and medical information to promote future studies given the lack of reliability with this information in current resources. The ERFC could make the claim they are collecting this data to work towards studies like the one described in the paper, ultimately aimed at helping people assess their own medical risks. In this claim, it is for the patients' "own good" that the ERFC is collecting their personal information. However, paternalism is most often used to defend a claim (such as the one previously described) that informed consent is not necessary, and therefore limits user autonomy even in their ability to keep their private health information censored.

However, this claim of paternalism is easily overturned when thinking about the harm principle. The harm principle determines autonomy should extend up until its use results in the objective harm of another agent. In this case, there is great potential to harm patients whose personal health data remains uncensored. Some of these harms could include the acquisition of this uncensored information by other companies or insurance agencies. An insurance agency that has information on private health data such as cardiovascular investigations could easily raise rates or even deny insurance for a patient, based on information they did not consent to give. This is inherently an outcome that very negatively affects the patient and inflicts great harm unto them. In addition, the ERFC is a very large collection of databases, and could easily be susceptible

to breaches, giving way to more risk for the patients whose health data remains uncensored. The larger the database, the more potential there is to cause harm and violate the harm principle; this solidifies the idea that lucrative data acquisitions in the healthcare sector, such as the collection of uncensored cardiovascular information, is not justified.

In addition to the harm principle, the use of sensitive data outside the scope of giving informed consent treats patients as means to an end rather than the end. This is in clear violation of Immanuel Kant's categorical imperative, which encourages the following: treat people as an end, and never as a means to an end.

There would be different considerations if fully informed consent was given by patients. However, this is likely not the case as patients do not have a choice to be censored if they are participants in the ERFC at the time of their cardiovascular investigation or diagnosis of angina. Although the argument could be made that participants in the ERFC chose to have this information disclosed (as their information is already present in one of the many studies housed by the ERFC), there is no evidence of non-coercive fully informed consent. Coercion could come in the form of denied medical care, especially in the context of cardiovascular disease, if a patient refused to participate in a study housed by the ERFC. Then, fully informed consent is no longer valid, and the harm principle applies once more, outweighing any claims of paternalism or fully informed consent.

Therefore, there is no justification for a lack of data privacy that could ultimately lead to the harm of countless patients, as well as the possibility of introducing potential bias into the many models the ERFC harbors.

Conclusion

Time-to-event analyses are important in the context of analyzing relative risk in relation to both primary effectors and effect modifiers. These studies, such as the one above analyzing the relative risk of coronary heart disease dependent on baseline fibrinogen, can give insights to both patients and providers on what they should be looking out for. Large analysis efforts such as the ones made by the Emerging Risk Factors Collaboration leads to many beneficial outcomes in identifying risk factors for many different diseases, however their use of uncensored information can ultimately lead to more harm for participants than good. Then, if methods in data acquisition can be justified (and this must come as a pre-requisite), analyses following time-to-event data should continue to be pursued and utilized by both patients and providers.

References

Abd ElHafeez, Samar, et al. "Methods to Analyze Time-to-Event Data: The Cox Regression Analysis." Oxidative Medicine and Cellular Longevity, U.S. National Library of Medicine, 30 Nov. 2021, www.ncbi.nlm.nih.gov/pmc/articles/PMC8651375/.

"How to Perform a Meta-Regression: Columbia Public Health." Columbia University Mailman School of Public Health, 13 Mar. 2023, www.publichealth.columbia.edu/research/population-health-methods/meta-regression.

Riley, Richard D, et al. "Interpretation of Random Effects Meta-Analyses." The BMJ, British Medical Journal Publishing Group, 10 Feb. 2011, www.bmj.com/content/342/bmj.d549.

Thompson, Simon, et al. "Statistical methods for the time-to-event analysis of individual participant data from multiple epidemiological studies." International Journal of Epidemiology, vol. 39, no. 5, 3 May 2010, pp. 1345–1359, https://doi.org/10.1093/ije/dyq063.