Original Research Paper

Survival and Mediation Analysis with Correlated Frailty

*^Usha S Govindarajulu and ^Mohsin Qadri

*Center for Biostatistics, Department of Population Health Sciences,
Icahn School of Medicine at Mount Sinai, New York, NY, USA
^Department of Epidemiology and Biostatistics,
School of Public Health, SUNY Downstate Health Sciences University, Brooklyn, NY, USA

Corresponding Author:
Dr. Usha S Govindarajulu
Center for Biostatistics,
Department of Population
Health Sciences, Icahn School of
Medicine at Mount Sinai, New
York, NY, USA
Email: usha.govindarajulu@mountsinai.org

Abstract: Many studies have focused on mediation effects, but fewer have focused on these in a time-to-event analysis and even fewer have focused on these effects in a survival model with frailty effects. The purpose of this study was to show unique modeling of correlated frailty effects with mediation and to demonstrate their use through simulations as well as on a real data application. We focused on a technique introduced by (Lange *et al.*, 2012), which involves a procedure based on Marginal Structural Models (MSM) that directly parameterize the natural direct and indirect effects of interest. Using their method of specifying the MSM structure along with a correlated frailty model allowed us to incorporate the unexplained heterogeneity still unaccounted for by these models. We were able to ascertain direct and indirect mediating effects in the simulations. In the real dataset, we were also able to demonstrate stronger frailty effect in the correlated frailty model accounting for mediation.

Keywords: Survival Analysis, Causal Mediation, Cox Model, Correlated Frailty, Heterogeniety

Introduction

Mediation is the process through which an exposure causes disease. It is up to researchers to decide that some or all, of the total effect on an outcome operates through a mediator, which is defined as an effect of the exposure and cause of the outcome (Robins and Greenland, 1992). When a mediator is hypothesized, such as in our study, the total effect can be broken down in to two parts, direct and indirect effect. The direct effect is the effect of exposure on the outcome minus the mediator. The indirect effect is the effect of exposure on the outcome that works through the mediator.

Galea (2013) states that in public health, it has been said that epidemiologists should seek to estimate parameters that have a logical correspondence with some realistic intervention that might be taken to improve population health. Similarly, causal inference suffers from the idea that one cannot estimate causal effects without some clearly defined exposure intervention (Glass *et al.*, 2013). This fueled the debate on the usefulness of decomposing effects into their natural direct and indirect components. One argument is that because natural direct and indirect effects cannot be identified using intervention-based causal models and cannot be estimated in a randomized trial and, therefore, they cannot be interpreted as effects that have a logical correspondence with some public health action (Robins and Greenland, 1992). Others argue that natural direct and indirect effects are identifiable using causal inference models that make stricter assumptions in order to infer mechanistic relations and that many causal effects cannot be estimated in a randomized intervention trial for logistical or ethical reasons, and thus, should not be ruled out as a means of providing information on actions of different mechanisms (Pearl, 2009).

Casual mediation analysis have existed in the survival analysis world. We employed a technique to allow for survival and causal mediation analysis. This technique was introduced by Lange *et al.*, 2012, which is in the realm of causal mediation. Although one of the best known methods that also works with survival analysis is that of Valeri and VanderWeele (2015), which uses a technique for causal inference of mediation to implement analysis in the presence of exposure-outcome interaction accounting for different type of outcomes; such as survival. The causal inference method allows for effect decomposition even in the presence of exposure-mediator interaction. The issue is that the outcome has to be rare to work in the Cox model format and their method is cumbersome to adapt for frailty analysis. The limitations of their methods did not allow for the Cox proportional hazard model to be employed unless the outcome is rare.

The technique introduced by Lange *et al.* (2012) involves directly modeling the natural direct and indirect effects of interest. Their method functions via a procedure based on Marginal Structural Models (MSM), which will directly parameterize the (Glass *et al.*, 2013) natural direct and indirect effects of interest. It has tended to produce

more parsimonious results than current techniques, greatly simplifies testing for the presence of a direct or an indirect effect and has the advantage that it can be implemented in standard software like SAS or R. This approach can be used for any type of outcome (binary, continuous, survival, categorical, etc.,) and any type of mediator, even when exposure-mediator interactions exists. However, its simplicity comes at the price of relying on correct specifications of models for the distribution of the mediator and exposure and accepting some loss of precision compared with more complex methods.

We have previously published our frailty model framework (Govindarajulu *et al*., 2011) and demonstrated its flexibility in handling univariate, multivariate, or correlated frailty. Frailty models have been described as essentially survival models with both fixed and random effect terms. While the fixed effects comprised the explained or observed portion of the model, the random effect term accounted for the unexplained variability of the model. In other words, the random effect, or frailty, modeled the unexplained heterogeneity in the model, which could be exhibited by the heterogeneity of hazard rates beyond recorded covariates shown in a population. The frailty term (Vaupel *et al*., 1979) was initially developed to describe heterogeneity at the individual level, but was expanded to describe heterogeneity among groups of individuals or within an individual. These were then considered levels of clustering, where an individual is one level of clustering (Duchateau, 2008) and then groups or observations within an individual were additional forms of clustering.

Materials and Methods

Mediation Framework

We employed an already published technique to determine the survival and causal mediation analysis and then enhanced this method. The second technique introduced by Lange *et al*. (2012) involved a direct modeling of the natural direct and indirect effects of interest. This approach was applicable to survival outcomes with any type of mediator, even when exposure-outcome interactions exist. The method described by Lange *et al*. (2012) allowed us to model the natural direct and indirect effect yielding results simpler for reporting. We adapted the technique for a continuous outcome though they had mainly demonstrated a categorical outcome.

We then incorporated our prior published frailty model, which allows for univariate, multivariate, or correlated frailty. With this we focused on the correlated frailty aspects with mediation. We have presented a Directed Acrylic Graph (DAG) of the causal structure of variable relationships along with frailty in Fig. 1 which is modified from Lange *et al*. (2012). Since frailty is generally multiplied onto the function of covariates in the hazard function, we attempted to portray its causal effects on the covariates but even on the mediation.
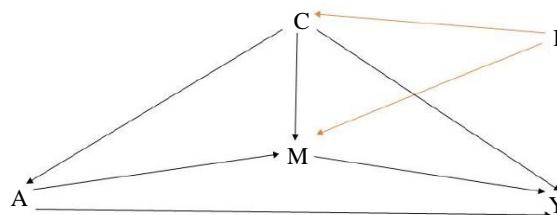

Fig. 1: DAG with frailty effect The DAG shown illustrates the causal mediation relationships between F: Frailty, A: Exposure, M: Mediator, C: Baseline confounders and Y: Outcome

We report our steps of setting up the MSM structure proposed by Lange *et al*. (2012) and with our final incorporation of the frailty:

Step 1: Start out with $g()$: being a link function, in this case, a logit link, where a is the exposure, c are the covariates represented by $\beta'X$ and m is the mediator:

$$g\left(E\left[Y_M \mid a,c\right]\right) = b_0 + b_1 a + \beta'X_i \tag{1}$$

Step 2: Continuous variable comes from a normal distribution with mean and variance from the exposure:

$$newvar \sim N\left(\mu_a, \sigma_a^2\right) \text{ with } \mu \sim \mu_a, \sigma^2 \sim \sigma_a^2 \tag{2}$$

Step 3: Using their derivation for computing the weights. Generally, there is a no unmeasured confounders assumption but we had to modify these assumptions to include the frailty effect (f): $Y_{amf} \parallel A|C|F$, $M_a \parallel A|C|F$, $Y_{amf} \parallel M|A = a$, $C = c$, and $Y_{amf} \parallel M_{a*}|C$. They eventually derive that:

$$\sum_m \sum_c E\left[Y_M \mid M = m, C = c\right] * P\left[M = m \mid C = c\right] * P\left[C = c\right]$$
$$\dots \tag{3}$$
$$= E\left[Y \mid (A = a)W\right]$$

where, W refers to the weights. These weights from Equation 3 are then utilized in the modeling in Equation 4.

Step 4: Utilizing the weights computed in Step 3 in the Cox proportional hazards model with correlated frailty, which are multiplied onto the hazard:

$$h_{ij}\left(t \mid x,z,w\right) = \left[h_0\left(t\right)\exp\left(c_0 + c_1 a + c_2 a* + \beta'x_{ij} + z_{ij}'w_{ij}\right)\right.$$
$$\left. * E\left[Y I(A = a)W\right]\right. \tag{4}$$

where, h_{ij} is the hazard function for the jth observation from the ith cluster where the cluster can be an individual or a group and w_{ij} is a vector of random effects associated with the covariates vector z_{ij} for cluster i observation j, which usually includes an intercept. The random effects account for the heterogeneity effects of the z_{ij}. When z_{ij} includes only an intercept, the model is a simple

multiplicative frailty model. The w_{ij} accounts for the correlation among individuals within a cluster. Therefore, this term represents frailty as univariate or multivariate. In univariate survival data, each cluster has only one individual with only one survival outcome. Multivariate survival data consists of a cluster of more than one individual. The clusters may be multiple survival outcomes for a single individual or one or more survival outcomes for multiple correlated individuals, such as relatives (Govindarajulu *et al*., 2011).

The correlated frailty can be considered to be an extension of the additive polygenic model often used to estimate heritability for quantitative phenotypes (Govindarajulu *et al*., 2011). The assumptions of a family-specific frailty model is that everyone within the same family shares the same risk level. Persons that are not genetically related but are related through marriage will share the same family-specific effect. For the model described in Equation 4, in the shared frailty model, the random effect, w_{ij}, is the same for all individuals (*j*) in pedigree *i*, while in the correlated frailty model, the degree of genetic relationship between individuals is specified through the frailty. Individuals that are more closely related are expected to have more similar frailty, whereas more distant relatives and unrelated individuals have less similar frailty. This model might be appropriate when individuals in a family are correlated because of unmeasured genotypic effects.

Simulation Framework

We utilized a simulation framework akin to methods previously used to simulate survival data (Govindarajulu and Malloy, 2015; Govindarajulu *et al*., 2007; 2009) adapted to correlated frailty models. We created 50 families. For each family, each of which ranged in size from 1 to 5, we simulated family id, father id and mother id, each generated from sampling without replacement between a particular range of integer values for each. We then randomly sampled covariates with replacement, allowing age to range from 25 to 90 years old and systolic blood pressure (SBP) from 116 to 156 units. Gender and smoking status were binary so we sampled those from a uniform distribution.

Bender *et al*. (2005) discussed the use of different distributions for generating survival times. We allowed for a baseline Weibull hazard (Bender *et al*., 2005; Klein, 1997), so that:

$$h_0(t) = \theta v t^{\theta-1} \qquad (5)$$

depending on parameters θ and v, are chosen to generate realistic survival data. Survival times are found from this generating distribution by using the relationship between the hazard function, survival function, $S(t)$ and cumulative distribution function (CDF), $F(t)$, of the survival time random variable, T. Given the baseline hazard function, $h_0(t)$, the CDF is found through the relationship:

$$F(t) = 1 - S(t) = 1 - e^{-H(t)} \qquad (6)$$

where, $H(t)$ is the cumulative hazard function. This in turn is given by:

$$H(t) = e^{g(x,y,w)} \int_0^t h_0(u) du = e^{g(x,y,w)} v t^{\theta}, \qquad (7)$$

where, $g(x,y,w) = \beta_T x + (1|id)$ is the true log hazard ratio defined in equation for $H(t)$ above. Using the probability integral transformation, for a given t, the CDF, F, has a Uniform distribution on the interval from 0 to 1. Generating $U1$ from a Uniform $(0,1)$ distribution and solving for t gives the simulated survival time:

$$t = \left(\frac{-e^{-g(x,y,w)} \log(1-U1)}{v} \right)^{1/\theta} \qquad (8)$$

for given values of x, y, w, θ and v. We also incorporated a competing risk into the simulations. Competing risk times were generated in a similar manner based on an exponential distribution with scale parameter γ, giving:

$$t_{cr} = \frac{\log(1-U2)}{\gamma} \text{ where } U2 \sim Uniform(0,1). \qquad (9)$$

The observed survival time was taken to be the minimum of t_0, t_{cr} and a pre-specified end-of-study time, τ. An observation was considered censored if t_0 was larger than the minimum of t_{cr} and τ. The user-specified parameters, θ, v,γ and τ were set to generate realistic survival time distributions. We used $\tau = 20$ years of follow-up time and $\theta = 5$ which is considered typical used value. The parameters, v and γ, were chosen to control the amount of censoring in the simulated data. We had 71 cases on average in the dataset where the average number of cases increased across the scenarios from 47 to 91. The final simulated datasets contain: all covariates as well as the clustering variables for id, family id, father and mother ids, survival time, event indicator, center number, and physician indicator.

Real Dataset Framework

We examined age-at-death data from the original cohort of the Framingham Heart Study, the longest running cohort study in the United States. There were a total of n = 5205 individuals with age-at-death or last contact data, coming from 2601 nuclear and extended families. Of the 5205 individuals, 4653 had died, and 552 were censored at age of last contact for this analysis. Families ranged in size from having a single individual to twenty-five individuals with age data. Most families (96%) contributed five or fewer members, and 38% of families contributed two individuals. Among the 5205 individuals, there are 1307 first degree relative pairs (sib and parent-offspring pairs), 61 second degree relative

pairs (avuncular and half-sib pairs), and 11 third degree relative pairs (cousins).

All computations were performed in *R* for both the simulations and real data analysis. In building the correlated frailty models, we utilized the `kinship` package to fit the models. The `kinship` package has a `coxme` function fits mixed-effect Cox models, including those with correlated frailties. In addition, the kinship package uses sparse block diagonal matrices to compute and represent kinship matrices.

Results

From the real dataset application from the Framingham Heart Study, we utilized the dataset described in the Methods section with n = 5205 individuals. We were then able to model univariate at an individual level as well as multivariate correlated frailty models with the family and sib-pair information. These models have been described in the Methods section.

The simulations were based off the covariates from this dataset as described in the Methods, which were then emulated around the covariates from the real dataset previously described. The simulation results show that the natural direct effect of age is statistically significant with a Hazard Ratio (HR) around 2.7 ($p<0.0001$), while for 2 out of the 20 simulations, the natural indirect mediating effect was significant in the correlated frailty model without weighting (Table 1). In the correlated frailty model with the weighting (Table 2), it was statistically significant for 5 out of 20 of the simulations with a reduced HR rate as compared to the direct effect. The other covariates, gender, smoke and SBP were generally statistically significant in the model.

Real dataset results showed potentially a significant natural direct effect and a nonsignificant natural indirect effect of age in all three models shown for correlated frailty: univariate, multivariate and multivariate with weighting. The natural direct effect appears to become more significant in the multivariate correlated frailty models while the indirect effect stays the same, including in our version of the correlated frailty model adjusted for mediation. Meanwhile the other covariates stayed statistically significant in each model.

The AIC results varied across simulations and do not appear correlated with a significant indirect mediating effect (i.e., lower AIC ≈ significant indirect mediating effect). In the real dataset analysis, it does not seem like AIC improved from just shared frailty only to a multivariate frailty model. We also included variance of frailty results separately for both individual frailty, σ_p^2 and shared frailty, σ_f^2, between families. The frailty variances for both effects are generally very small in the simulations on the order of 10^{-4} and 10^{-5} (Table 1) and then perhaps only show perhaps minor increases if at all in Table 2 where the weighting is included. In the real data analyses, the initial correlated frailty shows 23% of variance accounted for by the frailty (Table 3) but then Table 4 shows small effect mainly for individual frailty of about 0.03% while the shared frailty is about 14%, which stays about the same in Table 5 with the weighted correlated multivariate frailty model. However, the individual frailty variance went up to > 11% so at least from an individual level, accounting for causal mediation induces a stronger frailty effect.

Table 1: Simulations of correlated frailty

| Simulation # | Correlated frailty | | | | | |
	Variable	Coef	S.E.	P-value	Frailty variances	AIC
1	Age	0.9983	0.0790	<0.0001	$\sigma_p^2 = 0.000025$	690.2971
	Varstar	-0.0007	0.0002	0.0089	$\sigma_f^2 = 0.000000095$	
	Gender*	-0.7080	0.2040	0.0005		
	Smoke*	1.1386	0.2180	<0.0001		
	sbp	0.0035	0.0090	<0.0001		
2	Age	0.9151	0.0700	<0.0001	$\sigma_p^2 = 0.000064$	683.6775
	Varstar	-0.0002	0.0003	0.4374	$\sigma_f^2 = 0.0000083$	
	Gender*	-0.5231	0.1956	0.0074		
	Smoke*	0.8753	0.2145	<0.0001		
	Sbp	0.0210	0.0086	0.0146		
3	Age	0.9657	0.0785	<0.0001	$\sigma_p^2 = 0.00017$	697.5636
	Varstar	-0.0006	0.0003	0.0363	$\sigma_f^2 = 0.0000032$	
	Gender*	-0.6150	0.2121	0.00374		
	Smoke*	1.2079	0.2302	<0.0001		
	sbp	0.0144	0.0086	0.0958		

Table 1: Continue

4	Age	1.0511	0.0834	<0.0001	$\sigma_p^2 = 0.0000087$	703.8998
	Varstar	-0.0004	0.0003	0.2907	$\sigma_f^2 = 0.0000083$	
	Gender*	-0.8965	0.2172	<0.0001		
	Smoke*	0.8872	0.2338	0.0001		
	Sbp	0.0231	0.0093	0.0124		
5	Age	1.0824	0.0879	<0.0001	$\sigma_p^2 = 0.0217260$	711.2872
	Varstar	0.00003	0.0002	0.9004	$\sigma_f^2 = 0.000103$	
	Gender*	-0.8869	0.2138	<0.0001		
	Smoke*	1.0238	0.2276	<0.0001		
	Sbp	0.0176	0.0091	0.0543		
6	Age	1.0352	0.0832	<0.0001	$\sigma_p^2 = 0.000273$	701.9531
	Varstar	0.00008	0.0003	0.8135	$\sigma_f^2 = 0.0000014$	
	Gender*	-0.86834	0.2131	<0.0001		
	Smoke*	0.9583	0.2223	<0.0001		
	Sbp	0.0157	0.0087	0.07239		
7	Age	0.9910	0.0777	<0.0001	$\sigma_p^2 = 0.0002456$	701.4104
	Varstar	-0.00009	0.0002	0.6924	$\sigma_f^2 = 0.0000279$	
	Gender*	-0.6559	0.2123	0.0020		
	Smoke*	1.3287	0.2306	<0.0001		
	Sbp	0.0227	0.0090	0.0119		
8	Age	1.1355	0.0916	<0.0001	$\sigma_p^2 = 0.000188$	734.4097
	Varstar	-0.0002	0.0003	0.5375	$\sigma_f^2 = 0.0000064$	
	Gender*	-0.7939	0.2149	0.00022		
	Smoke*	1.0359	0.2261	<0.0001		
	Sbp	0.0047	0.0086	0.5827		
9	Age	0.9785	0.0793	<0.0001	$\sigma_p^2 = 0.0002080$	688.1612
	Varstar	-0.0002	0.0003	0.4605	$\sigma_f^2 = 0.0000235$	
	Gender*	-1.0565	0.2264	<0.0001		
	Smoke*	1.0328	0.2164	<0.0001		
	Sbp	0.0092	0.0081	0.2551		
10	Age	0.9392	0.0753	<0.0001	$\sigma_p^2 = 0.0380070$	669.2119
	Varstar	0.0004	0.0003	0.1452	$\sigma_f^2 = 0.000143$	
	Gender*	-0.3910	0.2074	0.0594		
	Smoke*	0.8503	0.2065	<0.0001		
	Sbp	0.0098	0.0090	0.2763		
11	Age	1.1539	0.0942	<0.0001	$\sigma_p^2 = 0.0015984$	722.5457
	Varstar	0.0002	0.0003	0.5667	$\sigma_f^2 = 0.0000106$	
	Gender*	-0.8363	0.2152	0.0002		
	Smoke*	0.8274	0.2152	0.0001		
	Sbp	-0.0013	0.0091	0.8863		
12	Age	1.0943	0.0866	<0.0001	$\sigma_p^2 = 0.000307$	726.0126
	Varstar	-0.0002	0.0003	0.3980	$\sigma_f^2 = 0.000000174$	
	Gender*	-0.7558	0.2340	0.0003		
	Smoke*	1.1868	0.2340	<0.0001		
	sbp	0.0351	0.0094	0.0002		
13	Age	0.9785	0.0769	<0.0001	$\sigma_p^2 = 0.0000078$	701.3641
	Varstar	-0.0002	0.0003	0.4337	$\sigma_f^2 = 0.0000079$	
	Gender*	-0.4781	0.2107	0.0233		
	Smoke*	0.5366	0.2032	0.0083		
	Sbp	-0.0061	0.0092	0.5083		

Table 1: Continue

14	Age	1.1284	0.0922	<0.0001	$\sigma_p^2 = 0.00000780$	726.2838
	Varstar	-0.0001	0.0003	0.7428	$\sigma_f^2 = 0.00000532$	
	Gender*	-0.7533	0.2175	0.0005		
	Smoke*	1.1410	0.2218	<0.0001		
	Sbp	0.0117	0.0086	0. 1724		
15	Age	0.0305	0.1045	<0.0001	$\sigma_p^2 = 0.00079382$	741.9662
	Varstar	0.0001	0.0002	0.7015	$\sigma_f^2 = 0.00006725$	
	Gender*	-0.8628	0.2223	0.0001		
	Smoke*	1.7906	0.2715	<0.0001		
	Sbp	0.0305	0.0095	0.0013		
16	Age	0.0229	0.0793	<0.0001	$\sigma_p^2 = 0.0002575$	708.0735
	Varstar	-0.0004	0.0003	0.1207	$\sigma_f^2 = 0.0000271$	
	Gender*	-0.4277	0.2091	0.0409		
	Smoke*	0.8698	0.2160	<0.0001		
	Sbp	0.0136	0.0091	0.1332		
17	Age	0.0136	0.0764	<0.0001	$\sigma_p^2 = 0.00000714$	693.8137
	Varstar	<0.0001	0.0003	0.8443	$\sigma_f^2 = 0.00000562$	
	Gender*	0.4597	0.2091	0.0279		
	Smoke*	1.2046	0.2246	<0.0001		
	Sbp	0.0229	0.0089	0.0101		
18	Age	0.0018	0.0786	<0.0001	$\sigma_p^2 = 0.0000077$	695.5984
	Varstar	-0.000006	0.0003	0.9806	$\sigma_f^2 = 0.0000073$	
	Gender*	-0.7682	0.2142	0.0003		
	Smoke*	0.5202	0.2112	0.0138		
	Sbp	0.0018	0.0096	0.8480		
19	Age	0.9980	0.0808	<0.0001	$\sigma_p^2 = 0.084670$	695.0876
	Varstar	-0.0003	0.0003	0.2203	$\sigma_f^2 = 0.0005842$	
	Gender*	-0.7139	0.2160	0.0009		
	Smoke*	1.1982	0.2230	<0.0001		
	Sbp	0.0358	0.0091	<0.0001		
20	Age	0.9476	0.0754	<0.0001	$\sigma_p^2 = 0.0001673$	690.7863
	Varstar	-0.0001	0.0003	0.5821	$\sigma_f^2 = 0.00002150$	
	Gender*	-0.8319	0.2112	0.0001		
	Smoke*	1.0914	0.2193	<0.0001		
	sbp	0.0128	0.0084	0.1270		

Table 2: Simulations of correlated frailty with weighting

Simulation #	Correlated frailty					
	Variable	Coef	s.e.	p-value	Frailty variances	AIC
1	Age	1.0443	0.0848	0.0000	$\sigma_p^2 = 0.50474413$	764.8039
	Varstar	-0.0012	0.0004	0.0019	$\sigma_f^2 = 0.01474529$	
	Gender*	-0.8697	0.2263	0.0001		
	Smoke*	1.1130	0.2303	<0.0001		
	sbp	0.0086	0.0093	0.3585		
2	Age	0.9257	0.0793	0.0000	$\sigma_p^2 = 0.000064$	630.341
	Varstar	0.0001	0.0005	0.8911	$\sigma_f^2 = 0.0000032$	
	Gender*	-0.4415	0.2255	0.0503		
	Smoke*	1.3194	0.2545	<0.0001		
	sbp	0.0199	0.0089	0.0252		

Table 2: Continue

3	Age	1.0573	0.0946	0.0000	$\sigma_p^2 = 0.0000083$	647.5097
	Varstar	-0.0006	0.0004	0.1309	$\sigma_f^2 = 0.0000063$	
	Gender*	-1.1489	0.2627	<0.0001		
	Smoke*	1.0906	0.2521	<0.0001		
	sbp	0.0147	0.0103	0.1511		
4	Age	1.0473	0.0844	0.0000	$\sigma_p^2 = 0.0001139$	683.6146
	Varstar	-0.0003	0.0004	0.3802	$\sigma_f^2 = 0.00000182$	
	Gender*	-0.9168	0.22	<0.0001		
	Smoke*	0.8437	0.2335	0.0003		
	sbp	0.0222	0.0093	0.0165		
5	Age	1.037	0.0872	0.0000	$\sigma_p^2 = 0.00001748$	669.258
	Varstar	<0.0001	0.0003	0.7432	$\sigma_f^2 = 0.000000241$	
	Gender*	-0.8007	0.219	0.0003		
	Smoke*	0.9671	0.2273	<0.0001		
	sbp	0.0243	0.0096	0.0115		
6	Age	1.0402	0.0832	0.0000	$\sigma_p^2 = 0.00000776$	711.4358
	Varstar	<0.0001	0.0003	0.9600	$\sigma_f^2 = 0.00000235$	
	Gender*	-0.8668	0.2119	<0.0001		
	Smoke*	0.9521	0.2207	<0.0001		
	sbp	0.0147	0.0086	0.0893		
7	Age	0.9699	0.0755	0.0000	$\sigma_p^2 = 0.000279177$	711.5901
	Varstar	-0.0002	0.0002	0.4398	$\sigma_f^2 = 0.000000188$	
	Gender*	-0.6828	0.2057	0.0009		
	Smoke*	1.3237	0.2258	<0.0001		
	sbp	0.0231	0.0089	0.0092		
8	Age	1.17	0.0964	0.0000	$\sigma_p^2 = 0.00017414$	799.0312
	Varstar	-0.0004	0.0004	0.3556	$\sigma_f^2 = 0.000000415$	
	Gender*	-0.5384	0.233	0.0208		
	Smoke*	0.8658	0.2384	0.0003		
	sbp	0.0026	0.0095	0.7833		
9	Age	0.9608	0.0831	0.0000	$\sigma_p^2 = 0.1477929$	649.9346
	Varstar	0.0003	0.0005	0.4603	$\sigma_f^2 = 0.0001321$	
	Gender*	-1.1595	0.2615	<0.0001		
	Smoke*	1.2282	0.2583	<0.0001		
	sbp	0.0195	0.0088	0.0272		
10	Age	0.9483	0.0764	0.0000	$\sigma_p^2 = 0.00014645$	667.4898
	Varstar	0.0004	0.0003	0.1888	$\sigma_f^2 = 0.000000139$	
	Gender*	-0.4705	0.2103	0.0253		
	Smoke*	0.9037	0.2076	<0.0001		
	sbp	0.0119	0.0091	0.1891		
11	Age	1.1322	0.0954	0.0000	$\sigma_p^2 = 0.000122383$	681.8475
	Varstar	0.0004	0.0004	0.3981	$\sigma_f^2 = 0.000004319$	
	Gender*	-0.7409	0.2308	0.0013		
	Smoke*	0.7973	0.2255	0.0004		
	sbp	-0.0023	0.0093	0.8073		
12	Age	1.1023	0.0881	0.0000	$\sigma_p^2 = 0.107791650$	720.6712
	Varstar	-0.0003	0.0003	0.4169	$\sigma_f^2 = 0.000436692$	
	Gender*	-0.745	0.213	0.0005		
	Smoke*	1.2564	0.2375	<0.0001		
	sbp	0.0384	0.0095	0.0001		

Table 2: Continue

13	Age	0.9408	0.0786	0.0000	$\sigma_p^2 = 0.00089252$	647.8483
	Varstar	-0.0007	0.0004	0.0958	$\sigma_f^2 = 0.000007128$	
	Gender*	-0.5983	0.2284	0.0088		
	Smoke*	0.3572	0.2135	0.0943		
	sbp	-0.0094	0.0095	0.3227		
14	Age	1.0359	0.096	0.0000	$\sigma_p^2 = 0.000007683$	629.254
	Varstar	-0.0005	0.0005	0.3531	$\sigma_f^2 = 0.000005401$	
	Gender*	-0.3324	0.2403	0.1667		
	Smoke*	1.1852	0.2518	<0.0001		
	sbp	0.013	0.0098	0.1840		
15	Age	1.3264	0.1226	0.0000	$\sigma_p^2 = 0.00011563$	681.8803
	Varstar	0.0006	0.0004	0.1565	$\sigma_f^2 = 0.000002952$	
	Gender*	-0.8028	0.2377	0.0001		
	Smoke*	1.8526	0.2958	<0.0001		
	sbp	0.0365	0.0106	0.0006		
16	Age	1.1747	0.0967	0.0000	$\sigma_p^2 = 0.000026318$	766.0266
	Varstar	-0.0008	0.0004	0.0899	$\sigma_f^2 = 0.000001882$	
	Gender*	0.3668	0.2243	0.1020		
	Smoke*	1.3909	0.2478	<0.0001		
	sbp	0.0241	0.0107	0.0241		
17	Age	1.0823	8.3322	0.8966	$\sigma_p^2 = 0.000006433$	1917391.9
	Varstar	0.0012	0.2663	0.9963	$\sigma_f^2 = 0.0000064934$	
	Gender*	-0.9187	37.855	0.9806		
	Smoke*	1.1533	2.615	0.6592		
	sbp	0.0192	0.009	0.0341		
18	Age	0.9893	0.0885	0.0000	$\sigma_p^2 = 0.0085066912$	673.2487
	Varstar	-0.0008	0.0004	0.0648	$\sigma_f^2 = 0.0005803667$	
	Gender*	-0.726	0.2379	0.0023		
	Smoke*	0.8941	0.2684	0.0009		
	sbp	0.0059	0.0121	0.6253		
19	Age	0.9992	0.0841	0.0000	$\sigma_p^2 = 0.1356772283$	744.3109
	Varstar	-0.0007	0.0004	0.0769	$\sigma_f^2 = 0.0001395592$	
	Gender*	-0.6931	0.2339	0.0030		
	Smoke*	1.0534	0.2324	<0.0001		
	Sbp	0.03	0.009	0.0008		
20	Age	1.0066	0.0891	0.0000	$\sigma_p^2 = 0.0003916249$	647.2443
	Varstar	0.0006	0.0005	0.2027	$\sigma_f^2 = 0.0000016955$	
	Gender*	-0.8721	0.2449	0.0004		
	Smoke*	1.3018	0.2492	<0.0001		
	sbp	0.0114	0.0093	0.2193		

*For gender reference level is 0, varstar is indirect effect, sbp: Systolic blood pressure

Table 3: Real data Univariate correlated frailty

Variable	Coef	Exp(coef)	Se(coef)	Z	p
age1	-0.0115	0.9886	0.0072	-1.60	0.1100
Varstar	0.0005	1.0005	0.0005	1.02	0.1000
Sex	-0.6744	0.5095	0.0989	-6.82	<0.0001
smoke1	0.9902	2.6918	0.096	10.31	<0.0001
sbp1	0.0135	1.0136	0.0028	4.89	<0.0001
n = 2578, number of events = 490	AIC: 59.09	Frailty variance: 0.2350691			

Table 4: Real data multivariate correlated frailty

Variable	Coef	Exp(coef)	Se(coef)	Z	p
age1	-0.0127	0.9874	0.0073	-1.74	0.0810
varstar	0.0006	1.0006	0.0005	1.18	0.2400
sex	-0.6893	0.5019	0.0993	-6.95	<0.0001
smoke1	0.9778	2.6587	0.0966	10.12	<0.0001
sbp1	0.0135	1.0136	0.0028	4.85	<0.0001
AIC: 184.32	Frailty variances:	$\sigma_p^2 = 0.0003995222$			
		$\sigma_f^2 = 0.1421022498$			

Table 5: Real data multivariate correlated frailty with weighting from mediation

Variable	Coef	Exp(coef)	Se(coef)	Z	p
age1	-0.0124	0.9877	0.0073	-1.70	0.0890
varstar	0.0005	1.0005	0.0005	1.07	0.2800
sex	-0.6838	0.5047	0.09998	-6.84	<0.0001
smoke1	0.9844	2.6763	0.0976	10.09	<0.0001
sbp1	0.0138	1.0139	0.0028	4.97	<0.0001
AIC: 185.77	Frailty variances:	$\sigma_p^2 = 0.1107320$			
		$\sigma_f^2 = 0.1430282$			

Discussion

The purpose of this study was to show unique modeling of these frailty effects with mediation and to demonstrate their use through simulations as well as on a real data application. We focused on a technique employed introduced by Lange, Vandsteelandt and Baeker[1] which involves a procedure based on Marginal Structural Models (MSM) that directly parameterize the natural direct and indirect effects of interest. Using these MSM models along with a correlated frailty model allowed us to incorporate the unexplained heterogeneity still unaccounted for by these models. We demonstrated this in both simulations and a real dataset with a correlated frailty model including weighting using the methodology of Lange *et al*. (2012). We were able to ascertain direct and indirect mediating effects in the simulations. In the real dataset we are able to ascertain also stronger frailty effect in the correlated frailty model accounting for mediation.

There is a broad and sustained interest in mediation analysis from many areas of public health and other fields. In particular, for time-to-event studies, employing survival analysis for modeling has been a standard and with the advent of frailty models, they could additionally incorporate unexplained heterogeneity, which generally is thought to be due to unmeasured confounding. While mediation analysis in the counterfactual methodology helps to define causal relationships between outcome, exposure and predictors, the idea of unexplained heterogeneity seems lost in this world, mainly relying without that possibility. In fact, Valeri and Vanderweele (2013) made these assumptions of no confounding as an identifiability assumption for their models incorporating mediation (Valeri and VanderWeele, 2015; Valente *et al*., 2017; Valeri and Vanderweele, 2013).

In terms of future work, this research can be extended to other datasets which can be modeled with correlated frailty and which also have potential mediating effects. Modeling different datasets with different genetic associations and some factor with mediating influence would be interesting to see how they perform with this new modeling. What is especially important is ascertaining the degree of associations amongst measures as we had attempted in the beginning of this article with a DAG to explain the counterfactual diagram. It might be interesting also to explore the weighting mechanisms behind the MSM structure.

Acknowledgement

Authors Contributions

Govindarajulu: Involved with all aspects of paper.
Qadri: Assisted in literature review and introduction.

Ethics

We have no ethical issues to report.

References

Bender, R., T. Augustin and M. Blettner, 2005. Generating survival times to simulate Cox proportional hazards models. Stat. Med., 24: 1713-1723. DOI: 10.1002/sim.2059

Duchateau, L.J.P., 2008. The Frailty Model. 1st Edn., Springer Science and Business Media, ISBN-10: 038772835X, pp: 316.

Galea, S., 2013. An argument for a consequentialist epidemiology. Am. J. Epidemiol., 178: 1185-1191. DOI: 10.1093/aje/kwt172

Glass, T.A., S.N. Goodman, M.A. Hernan and J.M. Samet, 2013. Causal inference in public health. Ann. Rev. Public. Health, 34: 61-75. DOI: 10.1146/annurev-publhealth-031811-124606

Govindarajulu, U.S. and E.J. Malloy, 2015. Evaluating treatment effect in multicenter trials with small centers using survival modeling. Int. J. Stat. Med. Res., 4: 8-25.

Govindarajulu, U.S., D. Spiegelman, S.W. Thurston, B. Ganguli and E.A. Eisen, 2007. Comparing smoothing techniques in cox models for exposure-response relationships. Stat. Med., 26: 3735-3752. DOI: 10.1002/sim.2848

Govindarajulu, U.S., E.J. Malloy, B. Ganguli, D. Spiegelman and E.A. Eisen, 2009. The comparison of alternative smoothing methods for fitting non-linear exposure-response relationships with cox models in a simulation study. Int. J. Biostat.

Govindarajulu, U.S., H. Lin, K.L. Lunetta and R.B. D'Agostino, 2011. Sr. Frailty models: Applications to biomedical and genetic studies. Stat. Med., 30: 2754-2764. DOI: 10.1002/sim.4277

Klein, J.P., 1997. Survival Analysis: Techniques for Censored and Truncated Data. 2nd Edn., Springer, ISBN-10: 0387948295, pp: 502.

Lange, T., S. Vansteelandt and M. Bekaert, 2012. A simple unified approach for estimating natural direct and indirect effects. Am. J. Epidemiol., 176: 190-195. DOI: 10.1093/aje/kwr525

Pearl, J., 2009. Causality: Models, Reasoning and Inference. 2nd Edn., Cambridge University Press, ISBN-10: 052189560X, pp: 484.

Robins, J.M. and S. Greenland, 1992. Identifiability and exchangeability for direct and indirect effects. Epidemiology, 3: 143-155. DOI: 10.1097/00001648-199203000-00013

Valente, M.J., W.E. Pelham, H. Smyth and D.P. MacKinnon, 2017. Confounding in statistical mediation analysis: What it is and how to address it. J. Couns. Psychol., 64: 659-671. DOI: 10.1037/cou0000242

Valeri, L. and T.J. VanderWeele, 2013. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol. Meth., 18: 137-150. DOI: 10.1037/a0031034

Valeri, L. and T.J. VanderWeele, 2015. SAS macro for causal mediation analysis with survival data. Epidemiology, 26: e23-24. DOI: 10.1097/EDE.0000000000000253

Vaupel, J.W., K.G. Manton and E. Stallard, 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography, 16: 439-454. DOI: 10.2307/2061224