

Supplementary Material for "Heterogeneity's ruses: the neglected role of between-individual variability in longitudinal studies of COPD exacerbations"

Authors: Mohsen Sadatsafavi, J. Mark FitzGerald

Contents

Background	2
Methods.....	2
Details for Suissa <i>et. al.</i>	3
Details for Hurst <i>et. al.</i>	5
What happens when heterogeneity is removed?	6
Details for Figure 1.....	7
References	8

Background

The authors of the case studies have observed certain patterns in the results of their survival analysis of exacerbations in patients with COPD, and accordingly they have made conclusions on associations between exacerbations at the individual level[1,2]. The purpose of this exercise was to provide one counter example, enough to show that the observed patterns at the population level do not necessarily imply the individual-level conclusions made by the authors. We show that the heterogeneity in the rate of exacerbation can generate the same pattern of results without such conclusions holding at the individual level.

In this context, heterogeneity refers to the variability in the propensity to exacerbate across individuals. Part of such variability can be explained by variability in measurable factors such age, sex, or the degree of lung function impairment; but it is known that patients with COPD vary in their propensity to exacerbate over and above the effect of such variables[3]. We note that neither Hurst et al. nor Suissa et. al. controlled for the effect of such variables in estimating the population hazard function. Suissa et. al. did adjust for potential confounding variables in other analyses, but the administrative data used in their study does not have important variables, such as measures of lung function impairment. All in all, it is likely that in both studies there was significant (residual) heterogeneity in the population.

Methods

We use a simple simulation study with very simple features (to separate and highlight the importance of heterogeneity) to show this in action. The few parameters of the study are chosen to roughly mimic the features of the case studies. We did not aim to generate the exact results. This would have required knowing the joint distribution of several variables defining the study sample (follow-up times, event times, background rate of exacerbations, demographics characteristics, to name a few). The simulation was performed in the statistical programming environment R version 3.0.2[4].

The stylized cohort consists of longitudinal data of 100,000 individuals (to minimize sampling variability). There are no demographic characteristics or competing risk, and censoring occurs at a fixed time in follow-up. With individuals in our cohort being similar in all aspects except rate of exacerbations, we did not need to model the inclusion criteria as they do not have any implications on our results (e.g., taking the first exacerbation as the cohort entry as in Suissa et al. would have only changed the distribution of the background hazard in our population, which is already arbitrary).

Within each individual, the hazard of exacerbation was fixed over time. The background rate of exacerbations was from a gamma distribution. For each individual, time to next exacerbations were generated from an exponential distribution (corresponding to a fixed hazard) with the rate equal to the background hazard of exacerbation for that individual. This process was continued till the next exacerbation time occurred past the follow-up time. Therefore, the data for each individual with n exacerbations were represented by $(n + 1)$ time intervals adding up to follow-up time, with the first n intervals representing time to the next exacerbation, and the last interval representing time to censoring (due to end of follow-up). Note that using this scheme, the distribution of the number of exacerbations is from a Poisson distribution within each individual according to the background rate of exacerbation, and exacerbations occur completely independent of each other and with a constant hazard. The choice of the gamma distribution for modelling heterogeneity causes the distribution of between-exacerbation times across the population to follow a negative binomial distribution[5].

Details for Suissa *et. al.*

Suissa et al. have reported on the analysis of the long-term natural history of COPD with emphasis on severe exacerbations and mortality[1]. The study used 17 years longitudinal data of 73,106 individuals with COPD.

For this study, the gamma distribution had a shape and scale parameters of 0.2. This results in an average exacerbation rate of 1 per year, with a standard deviation of 2.28. Individuals were followed for 10 years. We followed the approach by the authors in estimating the population hazard function (**Figure 3A**) of the original publication. We estimated the baseline hazard and median inter-exacerbation times as a function of the previous number of exacerbations using the life table method via the *survfit* function[6]. We used *Lowess* with default values for smoothing the hazard[7]. Results are provided in the manuscript (**Figure 2**).

In addition, we estimated the relative risk (RR) of past exacerbations on future ones (similar to Table 2 of Suissa et al.) by fitting a Cox proportional hazards model (using *coxph* function[8]) with the number of previous exacerbations as a time-dependent covariate. Results are provided in the Table below.

Table E1: relative risk (RR) of a subsequent exacerbation associated with each successive exacerbation, as a function of the time since the first exacerbation	
Exacerbation #	RR in the presence of heterogeneity
0 (reference)	1.000
1	5.964
2	15.791
3	20.202
4	25.583
5	29.863
6	35.780
7	39.493
8	45.485
9	54.129

These RRs are associational but not causal. They can be used for predicting the time of next exacerbations, but one cannot conclude that by preventing each exacerbation, the rate of the future one is decreased. This is because of confounding by frailty: here the background rate of exacerbation affects the risk of past and future exacerbations without itself being part of the causal pathway, thus acting as a classical confounder. Note that some other observations by the authors of the original study

can be potentially due to the same phenomenon. For example, the increased hazard of death after each exacerbation can be due to the heterogeneity and positive correlation between hazard of death and exacerbation at the population level.

Details for Hurst *et. al.*

Hurst et al., followed 297 COPD patients for 904 patient-years to evaluate if COPD exacerbations tend to occur in clusters[2]. The investigators separated 1,923 pairs of successive exacerbations arisen in the cohort. They fitted a Weibull distribution to the time intervals between such successive exacerbations. In doing so, between-exacerbation intervals were considered as independent observations and the clustering of such intervals within individuals was ignored, an issue that has already raised concerns[9]. The temporal clustering of successive exacerbations was assessed by testing the shape parameter of the fitted Weibull distribution, with shape values of less than 1 indicating a declining hazard thus implying clustering of exacerbations. This parameter had a value of 0.966 (95%CI 0.948–0.985, $P < 0.001$).

In simulating data for this study, the shape parameter was set to 2.71 while the scale parameter was set to 1, corresponding to the average number of exacerbation in the study. Follow-up was for 3 years (average follow-up time in the original study). We noted that in estimating the shape of the Weibull distribution, the investigators only used between-exacerbation intervals, and not the censored intervals. This *per se* can bias the results given that censored intervals were incorrectly removed from the estimation (see [9] for explanation). We therefore improved the analysis by fitting a parametric survival model (the *survreg* function[10]) with a Weibull distribution. This framework properly takes into account censoring.

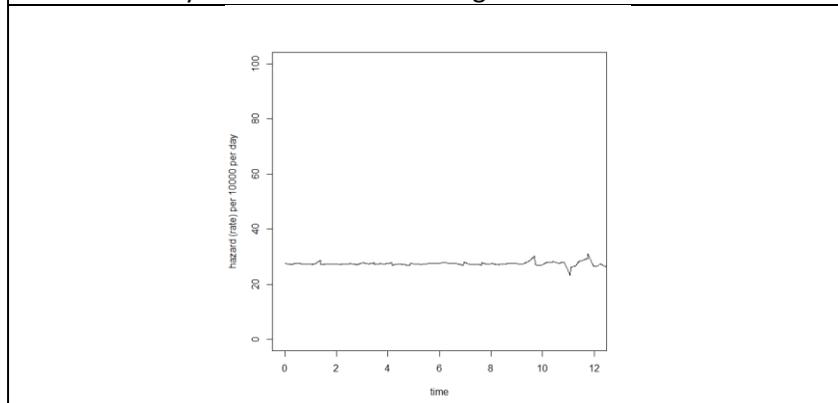
The estimated shape parameter of the Weibull distribution for the stylized cohort was 0.85, thus suggesting a strong clustering effect. This is despite the fact that within individuals, exacerbation

occurred purely at random. This is because in the presence of heterogeneity, those who exacerbate more often tend to provide more between-exacerbation periods than infrequent exacerbators. As such, there would be a preponderance of shorter exacerbation times contributed by high exacerbators, causing the population hazard to decrease over time. Further explanation of this phenomenon can be found in Keene et al.[9].

What happens when heterogeneity is removed?

We have so far shown that the pattern of results observed in the original studies can be replicated in a stylized cohort in presence of heterogeneity without any individual-level associations among exacerbations. To fully attribute the observations to the impact of heterogeneity (and not, say, to the artefacts of data analysis), we take one additional step to show that the afore-mentioned patterns disappear in the absence of heterogeneity. Heterogeneity was removed by assigning the same rate of exacerbation to each individual which was equal to the average rate of exacerbation in the population in the respective analyses in the presence of heterogeneity. For Suissa et al., the step-wise hazard function is provided below.

Figure E1: Hazard function of successive exacerbations from Suissa et al. in the stylized cohort with homogenous rate of exacerbation



For Hurst et. al., the estimated shape parameter was 1.001. In attempt to mitigate the problem of mixing high and low exacerbators and taking the unit of the study as exacerbations not individuals, Hurst et al. repeated the analysis by dividing the sample into two subgroups (low and high exacerbators). As mentioned by others, this broad categorization hardly mitigates the frailty problem as individuals within subgroups still vary in their rate of exacerbation[9]. In addition, by dividing the data around an observed value (median), it is no longer guaranteed that under the null hypothesis of no temporal clustering, even in the absence of heterogeneity, the shape parameter will be 1. We nonetheless repeated this analysis as well, both with and without heterogeneity. In the presence of heterogeneity, the shape parameter was 1.05 for low- and 0.85 for high exacerbators. When we removed the impact of heterogeneity, the fitted shape parameter was 1.11 for low and 1.01 for high exacerbators, suggesting that at least for low exacerbators, dividing the sample around the median will change the value of the test statistic under the null hypothesis.

Details for Figure 1

This figure presents unconditional (population) hazard function for a population consisting of two equal-sized subgroups (frails and robust). While the conditional (subgroup) hazards are constant, the unconditional hazard decreases over time. This is because at time T of follow-up, the population hazard is a weighted average of hazard in subgroup 1 (denoted by h_1) and hazard in subgroup 2 (denoted by h_2), with weights being the proportion of each subgroup still event free at time T . Because of the constant hazard, the survival curve of subgroups 1 and 2 follow an exponential curve and at time T is equal to. respectively, $e^{-h_1 \cdot T}$ and $e^{-h_2 \cdot T}$. The population hazard therefore is

$$h(T) = \frac{h_1 \cdot e^{-h_1 \cdot T} + h_2 \cdot e^{-h_2 \cdot T}}{e^{-h_1 \cdot T} + e^{-h_2 \cdot T}}$$

References

- 1 Suissa S, Dell'aniello S, Ernst P. Long-term natural history of chronic obstructive pulmonary disease: severe exacerbations and mortality. *Thorax* 2012;**67**:957–63. doi:10.1136/thoraxjnl-2011-201518
- 2 Hurst JR, Donaldson GC, Quint JK, *et al.* Temporal clustering of exacerbations in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2009;**179**:369–74. doi:10.1164/rccm.200807-1067OC
- 3 Hurst JR, Vestbo J, Anzueto A, *et al.* Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med* 2010;**363**:1128–38. doi:10.1056/NEJMoa0909883
- 4 R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: : R Foundation for Statistical Computing 2011. <http://www.r-project.org/> (accessed 16 Nov2009).
- 5 Keene ON, Calverley PMA, Jones PW, *et al.* Statistical analysis of exacerbation rates in COPD: TRISTAN and ISOLDE revisited. *Eur Respir J* 2008;**32**:17–24. doi:10.1183/09031936.00161507
- 6 R: Create survival curves. <http://stat.ethz.ch/R-manual/R-devel/library/survival/html/survfit.html> (accessed 21 Dec2013).
- 7 R: Scatter Plot Smoothing. <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/lowess.html> (accessed 21 Dec2013).
- 8 R: Fit Proportional Hazards Regression Model. <http://stat.ethz.ch/R-manual/R-patched/library/survival/html/coxph.html> (accessed 21 Dec2013).
- 9 Keene O, Roger J. Temporal clustering of exacerbations in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010;**182**:984–985; author reply 985. doi:10.1164/ajrccm.182.7.984
- 10 R: Regression for a Parametric Survival Model. <http://stat.ethz.ch/R-manual/R-devel/library/survival/html/survreg.html> (accessed 21 Dec2013).