Supporting Information

Corlan and Ross 10.1073/pnas.1510927112

A typical, important problem that is rarely approached in the literature is a deductive meta-analysis that is a quantitative synthesis of the results of multiple experimental studies of different types, such as a mortality, an incidence, and a prevalence study. In chemical kinetic terms, this amounts to describing a system consisting of multiple reactions. Data for tuning the system would be obtained from separate experiments involving one reaction each. Concentration measurements would be affected by substantial measurement errors.

We present below, as an illustration of the discussions in the main text, an example application of stochastic chemical kinetics for an epidemiology problem: checking the consistency of observational reports on incidence, mortality, and prevalence of atrial fibrillation (AF) and quantitative check of one hypothesis regarding the increase in AF prevalence over the last 20 y.

The Issue of AF

Permanent heart rhythm, in most humans, is either "sinus rhythm" (SR), which is the normal, regular rhythm, or AF, which is a completely irregular rhythm and a pathologic condition. There are other, much rarer rhythms that we do not consider here (we abstract them out of the model).

AF occurs increasingly with age, starting from around age 50 y. It may be due to other preexisting diseases and conditions, but it mostly occurs idiopathically—without any detectable cause. It is initially reversible, that is, patients may revert to SR, either spontaneously or with specific treatments. Then, AF may occur again. In time, it becomes irreversible.

Patients having AF have a reduction of their heart ability to pump blood, because the atria do not beat anymore, and an increased risk of stroke, because blood clots in the nonbeating atria and the resulting clots may migrate and obstruct various arteries, in particular cerebral arteries.

Because of these, and other reasons, AF is associated with increased mortality. This association was increasingly recognized over the last 20 y, resulting in growing interest in identifying and preventing AF's causes and complications.

Statement of the Problem

We use data from three sources, all on recent US populations:

- Ref. 17 reports on the incidence and prevalence of AF by gender and age group in a large population sample.
- Ref. 18 reports on AF vs. non-AF mortality by gender and age group.
- Center for Disease Control (19) general mortality data.

The prevalence of AF is a result of its cumulated incidence over time, as well as the different mortality of the patients with AF vs. patients with SR. The first question we try to solve is to check whether the prevalences reported in ref. 17 are quantitatively possible given the incidences from ref. 17 and the mortalities from ref. 18. This type of evaluation is possible only by using a kinetic model. Ref. 17 also finds that AF prevalence has increased from 1993 to 2003 and attributes this increase to a decreased reversibility of AF (possibly due to better diagnosis).

Kinetic Model of Mortality in AF

As can be seen in Fig. S1 and in table 2 of ref. 18, the mortality rate in AF is not constant in time after the occurrence of AF. Thus, if $F_{g,a}$ is a species that models patients with AF having gender g and age a, the mortality rate cannot be described as a mass-action law reaction $F_{g,a} \xrightarrow{k} D$, where D represents dead subjects. However, we note that there is a subgroup that dies in the first months after AF occurrence and another group that survives for years, dying at a lower rate, although still higher than subjects without AF.

We call these two groups the high-risk group, $F^{(H)}_{g,a}$ and the low-risk group, $F^{(L)}_{g,a}$ of AF patients having gender g and age a.

The kinetic equations for the model are thus

$$F^{(L)}_{g,a} \xrightarrow{k_{FLD(g,a)}} D \qquad [S1]$$

$$F^{(H)}_{g,a} \xrightarrow{k_{FHD(g,a)}} D. \qquad [S2]$$

This two-reaction model, for each age and gender group, is a good fit for the ref. 18 AF mortality data, as shown in Fig. S1, where the variables are the initial ratio of

$$q_{g,a} = \frac{\left[F^{(H)}_{g,a}\right](0)}{\left[F^{(L)}_{g,a}\right](0)} \qquad [S3]$$

and the two mortality rates, $k_{FHD(g,a)}$ and $k_{FLD(g,a)}$. We denote by $[X](t)$ the concentration of species X at time t. We estimated the values of these variables by fitting a sum of two exponentials on the data in each row of table 2 of ref. 18, using nonlinear least squares. The percentage values in the table, representing the data against which we performed the fit, correspond to the cumulative mortality

$$\mu_{a,g}(t) = \frac{[D](t)}{[D](t) + [F^{(H)}](t) + [F^{(L)}](t)}, \qquad [S4]$$

where time t instants are at 30 d and 1 y, 5 y, and 10 y.

Stochastic Mortality Coefficients

The data in ref. 18, table 2 represent observed frequencies in finite and relatively small samples of 31–268 individuals. Consistent with these observations are population frequencies with the probability distribution $B(\mu_{a,g}(t), [D](t) + [F^{(H)}](t) + [F^{(L)}](t))$, where $B(p,n)$ is the binomial distribution.

We generated 1,000 binomially distributed random samples representing the possible population mortalities corresponding to each data point and, by repeating the fit for each set, obtained a sample of the population distribution of $q_{g,a}$, $k_{FHD(g,a)}$, and $k_{FLD(g,a)}$ given the observations reported in ref. 18, table 2.

Integration of Mortality, Incidence, and Prevalence Data

To model the incidence and prevalence data from ref. 17 we considered a stationary population to which 100,000 newborns in species $S_{g,0}$ (sinus rhythm of gender g and age 0) are added each year.

The non-mass-action law reaction

$$S_{g,a} \xrightarrow{\delta=1y} S_{g,a+1} \qquad [S5]$$

represents the age progression from age group a to age group $a+1$, where the "$\delta = 1$ y" coefficient means that individuals in species $S_{g,a}$ that were in species $S_{g,a}$ for 1 y all move into age group $S_{g,a+1}$ at the end of the year.

Besides [**S5**], for each age and gender group, we introduced the mass-action law reactions

$$S_{g,a} \xrightarrow{k_{SFH(g,a)}} F^{(H)}_{g,a} \xrightarrow{k_{FHD(g,a)}} D \qquad \textbf{[S6]}$$

$$S_{g,a} \xrightarrow{k_{SFL(g,a)}} F^{(L)}_{g,a} \xrightarrow{k_{FLD(g,a)}} D \qquad \textbf{[S7]}$$

$$F^{(H)}_{g,a} \xrightarrow{k_{FS}} S_{g,a} \qquad \textbf{[S8]}$$

$$F^{(L)}_{g,a} \xrightarrow{k_{FS}} S_{g,a} \qquad \textbf{[S9]}$$

$$S_{g,a} \xrightarrow{k_{SD(g,a)}} D, \qquad \textbf{[S10]}$$

where k_{FS} represents the rate of the reversal of AF to SR. Death rates $k_{SD(g,a)}$ were taken from ref. 19 as they describe the general population.

The age progression of subjects with AF fibrillation is, in practice, that they are theoretical species resulting from the biexponential approximation. At stationarity, these prevalences over suitable age and gender groups were compared with prevalences from ref. 17. Fig. S2 presents the results of 2,000 scenarios, 1,000 for $k_{FS} = 0.19$ and 1,000 for $k_{FS} = 0.37$. Each 1,000 scenarios result from the Monte Carlo simulation of the binomial distribution of mortalities in the populations that are possible given the rates observed in ref. 18. The simulated prevalences largely fall within the range of the measured values, using the incidences and the reversal rates reported in the supplementary information of ref. 17. However, these reversal rates do not explain the total differences in prevalence between 1993 and 2007. Also, progression of tprevalence with age was slower in the simulated results compared with observation. This could be due to the fact that we considered a constant reversal rate (constant k_{FS} irrespective of the age and of the duration of fibrillation), as in ref. 17, but in clinical practice this rate is reduced after a few months of continuous AF. The dispersion of values represented by the vertical extent of the whisker plots in Fig. S2 reflects the uncertainty due to small samples sizes of mortality estimations. It can be seen that this factor alone is not the cause of the observed differences between simulations and experimental data.

Summary

We developed a kinetic model of AF epidemiology as an illustration of the use of kinetic methods for deductive meta-analysis of data sources of different types. Our model demonstrates that the incidence, prevalence, and mortality data from refs. 17–19 are consistent with each other. It also confirms that AF diagnostic reversal, whether due to encoding practices, actual AF reversal, or other reasons, plays a part in the increase of AF prevalence over the last two decades, but also indicates that it may not be the only factor.

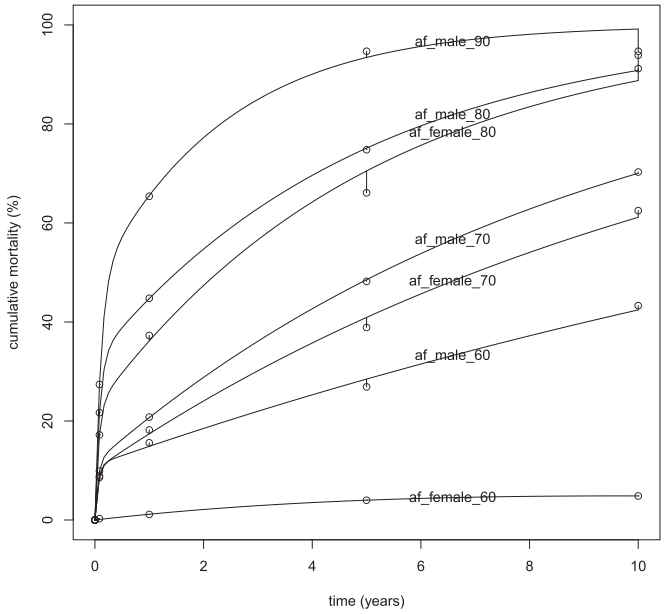

Fig. S1. Fit of the kinetic model (Eqs. **1** and **2**)—the lines, one for each gender and age group—to data from table 2 of ref. 18 (circles). It shows the extent to which a set of biexponential models, with a high-risk and a low-risk component for age and gender subgroup, fits the observed data and thus justifies why we modeled each age and gender group with the two species.

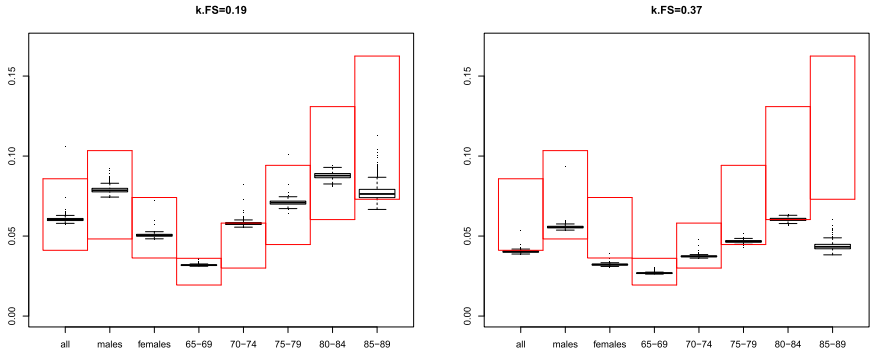

Fig. S2. (*Left* and *Right*) Simulated and literature results on the prevalence of AF. The *y* axis shows AF prevalence. The *x* axis shows various gender and age subgroups. The large, red rectangles represent measured AF prevalences, taken from ref. 17, supplementary table 5. The top level of the range corresponds to year 2007 and the bottom level to year 1993. The whisker plots represent the distributions of the prevalences obtained for each scenario resulting from the Monte Carlo simulation of mortality rates for high- and low-risk AF. *Left* shows the results obtained with a uniform AF reversal rate of $k_{FS} = 0.19$, whereas *Right* shows a value of $k_{FS} = 0.37$ (as reported in ref. 17). Ideally, the simulated prevalences (black whisker plots) at *Left* would be expected to occur toward the top, 2007 level of the measured prevalence range (top of the red box) whereas those at *Right* would occur toward the bottom, 1993 level. Systematic differences from the expected behavior need to be explained by further refinement of the model, but it can be seen that the uncertainty of the mortality coefficients that is attributable to small sample sizes cannot account for these differences.