

# Calculating Stage Duration Statistics in Multistage Diseases

Natalia L. Komarova\*, Craig J. Thalhauser

Department of Mathematics, University of California Irvine, Irvine, California, United States of America

## Abstract

Many human diseases are characterized by multiple stages of progression. While the typical sequence of disease progression can be identified, there may be large individual variations among patients. Identifying mean stage durations and their variations is critical for statistical hypothesis testing needed to determine if treatment is having a significant effect on the progression, or if a new therapy is showing a delay of progression through a multistage disease. In this paper we focus on two methods for extracting stage duration statistics from longitudinal datasets: an extension of the linear regression technique, and a counting algorithm. Both are non-iterative, non-parametric and computationally cheap methods, which makes them invaluable tools for studying the epidemiology of diseases, with a goal of identifying different patterns of progression by using bioinformatics methodologies. Here we show that the regression method performs well for calculating the mean stage durations under a wide variety of assumptions, however, its generalization to variance calculations fails under realistic assumptions about the data collection procedure. On the other hand, the counting method yields reliable estimations for both means and variances of stage durations. Applications to Alzheimer disease progression are discussed.

**Citation:** Komarova NL, Thalhauser CJ (2011) Calculating Stage Duration Statistics in Multistage Diseases. PLoS ONE 6(12): e28298. doi:10.1371/journal.pone.0028298

**Editor:** Kewei Chen, Banner Alzheimer's Institute, United States of America

**Received:** October 10, 2011; **Accepted:** November 4, 2011; **Published:** December 7, 2011

**Copyright:** © 2011 Komarova, Thalhauser. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: komarova@uci.edu

## Introduction

Many human diseases are characterized by multiple, more or less well-defined stages of progression. One well-known example is Alzheimer's disease (AD) [1], where patients' cognitive and functional deterioration follows a progression of several stages [2,3]. Other diseases that follow a certain pattern of progression, and can be characterized by several stages, include cancers (such as multiple myeloma [4,5] and Hodgkin's disease [6]), HIV [7,8], lyme disease [9,10], chronic kidney disease [11,12], Parkinson's disease [13,14], and many others.

While typical patterns of disease progression can be identified, there may be large individual variations among patients. Like any biological variable, rates of disease progression cannot be described by a fixed "number", but rather come from some (usually unknown) probability distribution. Fundamental characteristics of these distributions, such as the mean rate of progression and its variance, can be used as important tools in clinical trials and disease monitoring. For example, knowing both the mean duration and the amount of its natural "spread" is essential for evaluation of treatment efficiency. The group of treated patients must exhibit an average stage duration which is significantly (that is, more than a standard deviation) longer than that for the controls.

Quantitative measurements of the mean and the variance in stage durations of multi-stage diseases is a difficult empirical and statistical task. One way to approach it would be following many patients for the entire disease duration, to record the times when each new disease stage begins. If information of this kind could be

collected for a sufficient number of patients, then computing the means and variances of stage durations would be the matter of a trivial calculation. However, such an approach is often impossible in real clinical situations. Instead, the patient data collected in longitudinal studies are usually much more sparse and incomplete. This means that the mathematical task of extracting the mean and variance of stage durations becomes much more difficult.

A typical longitudinal patient dataset that logs the individuals' timing of disease progression consists of dates of the clinician's visits and the stages of the disease determined at the time of each visit. The medical records are often sparse, with only a few visits per patient, and the timing of the follow-up visits is sporadic. Data of this kind do not provide stage durations for individual patients. For example, if a given patient was seen at times  $t_1$ ,  $t_2$  and  $t_3$ , and was found to be at stages 1, 1, and 3 respectively, then we can say that for this patient, stage 1 lasted at least  $t_2 - t_1$  years, and stage 2 lasted at most  $t_3 - t_2$  years. In this particular example, we are able to obtain a lower bound on stage 1 duration and an upper bound on stage 2 duration. For other cases, even this is impossible. For example, if we have 2 visits at times  $t_1$  and  $t_2$  at stages 1 and 2 respectively, all we can say is that the total duration of stages 1 and 2 for this patient was at least  $t_2 - t_1$  years.

In statistics, this type of longitudinal data are referred to as "censored". Extracting information about stage duration distributions from this type of data is a nontrivial task, and large literature exists that is devoted to statistical methods in this context [15]. Statistical methods to handle such problems include the regression method [16], Kaplan-Meier (KM) statistics [17], maximum-likelihood-based methods described originally by Turnbull

[18,19] and Lagakos [20,21], and nonparametric Bayesian approaches [22,23]. In addition, a simple counting algorithm was introduced in [24] to estimate the stage duration cumulative probability distribution. All types of techniques have advantages and disadvantages. Here we will focus on two of the above methods: an extension of the linear regression technique [16,25] and the counting algorithm. These are by far the simplest methods of extracting the stage duration means and variances:

- they are non-iterative, so no “saddle-point” solutions can be observed,
- they are non-parametric, and
- they are computationally very cheap.

The importance of these considerations comes about when one studies the epidemiology of diseases, and looks for patterns of disease progression using large patient datasets. Methods of bioinformatics can be applied to sorting the datasets and identifying various subgroups of patients with different characteristics, with a further goal of correlating this with other available data such as genetic data, demographic data etc. Complex routines of this kind require estimation of progression statistics for many subsets of the patient group, to find the meaningful trends. Therefore, having reliable and very efficient methods for statistical analysis becomes essential.

The focus of the present analysis is to show that the regression method performs well for calculating the mean stage durations under a wide variety of assumptions, however, its generalization to variance calculations fails under realistic assumptions about the data collection procedure. On the other hand, the counting method yields reliable estimates for both means and variances of stage durations, where the regression method fails. While we will use AD as the main example for our calculations, this study is applicable for a wide range of multi-stage diseases.

### Example: stages of Alzheimer’s disease

We will illustrate our methodology by using the example of AD. It has been known for two decades that the rate of progression of AD varies from patient to patient [26], with illness ranging from a few years’ duration to as long as 21 years (see [27] and the references therein). Many factors have been found to be correlated with the rate of patients’ cognitive deterioration including apolipoprotein 4 genotype [28,29] and other genetic factors [30,31], brain atrophy rates [32–34], patterns of regional brain atrophy [35], ventricular enlargement [36], neuropsychological and cerebral metabolic profiles [27], vascular factors [37], and immune system factors [38].

The rate of AD progression can be estimated by using various cognitive tests, such as the Disability Rating Scale [39], the Mini-Mental State Examination (MMSE) score [40,41], and Clinical Dementia Rating (CDR) sum of boxes score [42,43]. Other measures that have been used are Global Deterioration Scale (GDS) [44] and its derivative, Functional Assessment Staging (FAST) [45]. These are reliable assessment techniques for evaluating functional deterioration in AD patients throughout the entire course of the illness [46]. They are based on a systematic examination of the functional changes occurring in patients with AD, and serve as a strong diagnostic aid for clinicians. The FAST staging technique has been compared with other scores such as MMSE, CDR sum of boxes, and others [47], and was found to correlate with other measures of progression. In a systematic review of 12 different assessment tools of AD [48], FAST staging was identified as one of the best studied techniques for reliability,

which shows good to excellent results on intrarater and interrater reliability.

Although AD is a continuous process, the patient decline usually follows a number of milestones. FAST staging technique distinguishes 7 AD stages ranging from 1 (normal adult) to 7 (severe AD). They are further characterized in [16]. The cognitive and behavioral differences among patients across different stages are very large, thus making GDS/FAST diagnostics a reliable tool. In this paper we will apply our statistical tools to a longitudinal dataset where both FAST and GDS staging systems were applied. We will focus only on the integer FAST stage values and not on the FAST substages.

## Methods

### Data considerations

Data from a longitudinal AD study typically have the following format [16]. A patient is diagnosed with AD and is seen one or more times. At each visit, the date of observation and current stage of AD is recorded according to whatever scale the clinic uses to track AD progression. Follow-up visits are often conducted at regular intervals, so that a patient’s record is a sequence of spaced visits and stage diagnoses. In the absence of knowledge about the onset date of the current stage, past analyses assume that patients’ first visit and last visit to the clinic occur uniformly within the current stage [16], so that on average each patient is halfway through that stage.

One can organize this data in many different ways. A particular structure which we find useful to impose is that of the transition class. We consider the set of all patients whose first visit is at stage  $i$  of the disease and whose last visit is at stage  $j$ , for  $i \leq j$ . We call this subset of patients a transition class. For a disease with  $K$  stages under observation, one can sort the patients into  $\frac{K(K+1)}{2}$  transition classes. Note that in clinical practice, not all patient records extend to the final stage of the disease and thus in a clinical dataset one may expect to find all transition classes with some patients in them.

Consider a patient with two or more visits to a clinic, whose first visit is at stage  $i$  and last visit is at stage  $j$ , with  $i < j$ . Then the total time that patient has been observed,  $T_p$ , is the sum of the amount of time the patient has spent in each observed stage:

$$T_p = \sum_{k=i}^j \alpha_k S_k, \quad (1)$$

where  $S_k$  is the duration of stage  $k$ . The coefficient  $\alpha_k$  with  $k=i$  or  $k=j$  is a random number distributed in  $(0,1)$ , and otherwise it satisfies:

$$\alpha_k = \begin{cases} 1, & i < k < j, \\ 0, & k < i \text{ or } k > j. \end{cases}$$

We call  $\alpha_k$  the completion coefficient of stage  $k$ . In the absence of any other information, one might assume that across all patients,  $\alpha_k$  are distributed uniformly on  $(0,1)$ . Thus, on average one might set all  $\alpha_k = \frac{1}{2}$  and thus generate, for each patient, a single equation of one or more unknowns, depending on how many stages of the disease the patient has entered. With enough such patient records, one can form a large, overdetermined system of equations of  $N$  unknowns,  $N$  being the number of stages under observation, and can solve this system in the least squares sense. This process was applied without justification in [16].

**Data generation method 1: prescribed transition class, prescribed completion coefficients.** We first generate data to

specifically conform to the assumptions used in past studies [16], namely that all patients arrive at their first and last visit at the clinic on average halfway through that stage, and that all transition classes are populated. To create the dataset, we:

1. Determine the actual timecourse of the disease from a predetermined set of distributions (one for each stage).
2. We then randomly determine the transition class to which that patient will belong.
3. Then, we draw the completion coefficient for the patient's first and last visits from a uniform (0,1) distribution.
4. Finally, given this data we can calculate the time from the patient's first visit to their last visit and create a patient record including total time and stages transited.

Theoretical records created by this method can be modified by changing the assumptions on the distribution of transition classes (uniform or skewed) and the assumptions on the completion coefficient (non-uniformly distributed, skewed to favor early or late visits).

A "dataset" of  $N_p$  patients created by this method consists of  $N_p$  records specifying the transition class (that is, the first and last stage of visits) and the total time duration between the two visits.

**Data generation method 2: prescribed initial stage, prescribed first completion coefficient, prescribed inter-office intervals.** We next formulate a method to generate test data for our analysis algorithms which is closer in mechanism to that which is used in a clinical setting:

1. As before, we specify the actual timecourse of progression for a patient through all stages of the disease.
2. Choose the stage number for the first visit. For the data sets used in this paper, the choice of initial stage is weighted towards earlier stages to ensure population of all possible transition classes.
3. Calculate the completion coefficient for the first stage. The initial arrival can be specified to occur at any time during their current stage; for example, we will employ both a uniform distribution for completion coefficients and the assumption that patients tend to first go to the clinic upon recognition of a new symptom, which occurs at or very near to the beginning of the current stage.
4. Patients return to the clinic following a loosely prescribed interval of time (for example, every two years on average). We draw the time-intervals between consecutive visits from a distribution. The distributions for inter-office visit times are given by the sum of 4 random numbers, uniformly distributed in the range [0.5,1.5] years (mean of 1 year, standard deviation of  $1/(4\sqrt{3}) \approx 0.14$ ) or [1.5,2.5] years (mean of 2 years, with the same standard deviation).
5. For each visit, the patient is evaluated and diagnosed; based on how much time has elapsed and the true timecourse of the disease, the patient may or may not have entered into a new stage.
6. Starting from the second visit, there is a fixed probability (given by 0.75) for the patient to come back for a subsequent visit, or to never come back. If a patient's visit happens after the "end" of the last stage, we disregard this visit and stop the record of the patient.

A "dataset" of  $N_p$  patients created by this method consists of  $N_p$  records specifying a list of stages at which the patient was seen, together with the inter-visit time intervals. Such datasets differ

from the ones created by method 1 because each patient may have more than 2 visits.

### Linear regression method for means and variances

We begin by combining the standard linear regression technique described previously [16] with that of the transition class. Consider the transition class of all patients starting at stage  $i$  and ending at stage  $j$  for  $i \leq j$ . Then the mean of the total observation time,  $T$ , for all patients in the class is:

$$E[T] = E\left[\sum_{k=i}^j \alpha_k S_k\right] = \sum_{k=i}^j E[\alpha_k S_k] = \sum_{k=i}^j E[\alpha_k] E[S_k],$$

where we have used the tacit assumption from prior work that the completion coefficients are independent of the stage durations. Note that if we assume that  $\alpha_k$  is uniform (0,1) for  $k=i,j$  and  $\alpha_k=1$  for  $i < k < j$ , then this system becomes:

$$E[t_{i,j}] = \frac{E[T_i]}{2} + \sum_{s=i+1}^{j-1} E[T_s] + \frac{E[T_j]}{2}, \quad 1 \leq j \leq i \leq N. \quad (2)$$

This is equivalent to the linear regression scheme described elsewhere [16]. Herein we have provided a simple justification for the prior schemes as well as explicitly delineated the assumptions needed for this method to yield accurate results.

We now turn to the calculation of the variance. As with the mean, we begin by grouping patient data by transition class. Under the same assumptions as the mean equations along with an assumption of independence of each stage duration from one another, we can calculate:

$$Var(T) = Var\left(\sum_{k=i}^j \alpha_k S_k\right) = \sum_{k=i}^j Var(\alpha_k S_k).$$

We resolve the variance as follows:

$$\begin{aligned} Var(\alpha_k S_k) &= E[(\alpha_k S_k)^2] - E^2[\alpha_k S_k] = E[\alpha_k^2] E[S_k^2] - E^2[\alpha_k] E^2[S_k] \\ &= E[\alpha_k^2] E[S_k^2] - E[\alpha_k^2] E^2[S_k] + E[\alpha_k^2] E^2[S_k] - E^2[\alpha_k] E^2[S_k] \\ &= E[\alpha_k^2] Var(S_k) + Var(\alpha_k) E^2[S_k]. \end{aligned}$$

Therefore, if we have information concerning the distribution of the completion coefficients, then we can write an equation for each transition class similar to its counterpart in the mean equations:

$$Var(T) = \sum_{k=i}^j (E[\alpha_k^2] Var(S_k) + Var(\alpha_k) E^2[S_k]).$$

Note that if  $i < k < j$ , then  $\alpha_k=1$ , and so  $Var(\alpha_k)=0$ . The variance equations are similar in nature to the means equations with a necessary adjustment to reflect the variance of the completion coefficients in the first and last stages of a transition class. For uniformly distributed completion coefficients, we have the following system of equations for the variance:

$$\begin{aligned} Var[t_{i,j}] &= \frac{Var[T_i]}{3} + \frac{E^2[T_i]}{12} + \sum_{s=i+1}^{j-1} Var[T_s] + \\ &\quad \frac{Var[T_j]}{3} + \frac{E^2[T_j]}{12}, \quad 1 \leq j \leq i \leq N. \end{aligned} \quad (3)$$

Note that another implicit assumption used in the derivation of these equations is the independence of the stage durations of one another. This assumption will be discussed later in the paper.

To solve these equations, we employ a nonlinear least squares regression algorithm implemented in MATLAB. We use a restricted variant of this general algorithm which guarantees non-negative values for both the mean and variance calculations.

**The counting method for means and variances**

The counting method first introduced in [24] is somewhat similar in design to the Kaplan-Meier statistic [17], but exhibits less bias induced by a coarse sampling methodology which is common in longitudinal studies of long-term diseases such as AD, see below.

Unlike the regression method which computes all the stage durations at once, the counting method considers each stage separately. For a given stage, say, stage  $k$ , we calculate the numerical cumulative probability distribution of the stage duration,  $P_k(t)$ . The method is illustrated in figure 1. We identify all the patients whose first visit corresponds to stage  $k$  as group  $G_k$  patients. For these patients, we assume that their first visit is at  $t=0$  (the onset of stage  $k$  is therefore on average at time  $t = -\tilde{t}/2$ ). In figure 1, there are 12 patients diagnosed with stage 1 at time 0. Consider all the patients in group  $G_k$  who visit the doctor's office in some relatively short interval  $I_i = [\tau_i, \tau_{i+1}]$ . The interval lengths,  $\Delta t_i = \tau_{i+1} - \tau_i$ , are chosen to ensure that each patient within the grouping has no more than one visit in the interval. Then we can compute  $N_t^i$  to be the number of patients who, upon visiting the doctor between times  $\tau_i$  and  $\tau_i + \Delta t_i$ , transited to the next stage of the disease. Thus, for all these patients, the duration of stage  $k$  was less than  $\tau_i$ . In figure 1, among the 3 patients seen during interval  $I_1$ , two remained in stage 1 and one transited to stage 2. Therefore, we have  $N_t^1 = 1$ . Likewise, we define  $N_t^i$  as the number

of patients seen at the clinic in the time interval  $[\tau_i, \tau_i + \Delta t_i]$  who remained in stage  $k$ . For all of these patients, the duration of stage  $k$  is greater than  $\tau_i + \Delta t_i$ . For all the patients who have an office visit in the interval  $I_i$  we define  $\tilde{t}_i$  to be the mean time of these office visits, and compute  $P_k(t_i) = \frac{N_t^i}{N_t^i + N_t^{i+1}}$ . The ordered pairs  $(\tilde{t}_i, P_k(t_i))$  for  $i = 1 \dots N$  are thus a numerical approximation to the cumulative distribution function, see the bottom panel of figure 1.

We further assume that the underlying probability distribution has finite support; that is, there exist some  $t_{min}$  for which  $t \leq t_{min} \Rightarrow P(t) = 0$  and  $t_{max}$  for which  $t \geq t_{max} \Rightarrow P(t) = 1$ . We specify  $t_{min}$  and  $t_{max}$  as follows: set  $t_{min} = \text{floor}(t_1)$  and  $t_{max} = \text{ceiling}(t_M)$ . The choice of interval lengths  $\Delta t_i$  is somewhat arbitrary, and defines the discretization grid for the numerical approximation of the cumulative probability function. We choose this grid to be non-uniform: we define the time-intervals such that the number of patients,  $J$ , in each of the intervals is the same.

It is important to note here that, as it stands, the counting method is unable to analyze the final stage in a multistage disease, as there are no transitions out of the final stage. With sufficiently accurate information (for example, dates of death for patients in the study) it may be possible to include the final stage in the analysis; however, absent that information, we restrict this work to the first  $n - 1$  stages of the disease.

Given the cumulative distribution function derived above, it becomes a simple calculation to compute the mean and variance of the underlying distribution. We have:

$$E[X] = t_{max} - \sum_{i=1}^{N-1} \frac{P_k(t_i) + P_k(t_{i+1})}{2} \Delta t_i, \tag{4}$$

$$Var(X) = (t_{max} - E[X])^2 - \sum_{i=0}^{N-1} [(\tilde{t}_i - E[X])P_k(t_i) + (\tilde{t}_{i+1} - E[X])P_k(t_{i+1})] \Delta t_i. \tag{5}$$

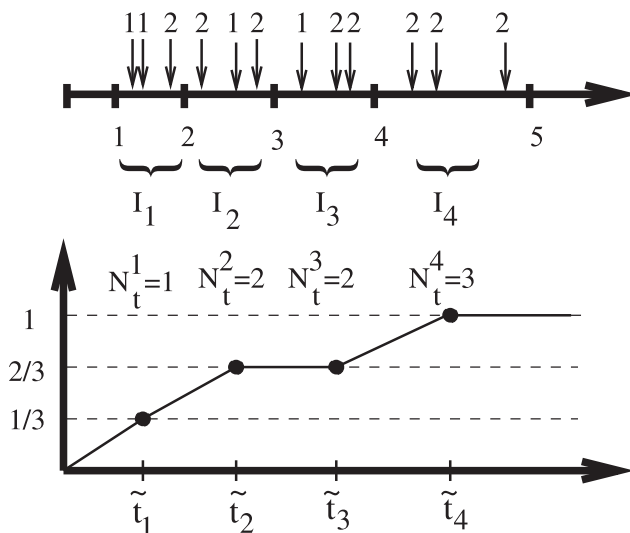
A more detailed description of the counting method, as well as an analysis of its properties, is given in Text S1.

**Underlying distributions in a 4-stage disease**

To begin testing the methods described herein for their ability to accurately calculate the mean and variance of the underlying distributions of a multi-stage disease, we first must prescribe some test distributions. We simulated a 4-stage disease by constructing three probability distributions for each of the 4 stages. Note that these probability distributions are not meant to model the real course of AD or any other disease, but are used as a technique to test our estimators. For these distribution, the mean stage durations correspond to the published means for FAST stages 4–7 of AD [49], and the variances are listed in table 1. The three distributions have different shapes: distribution A is uniform over a certain time-interval; distribution B is a triangle distribution (the sum of two uniform distributions), and distribution C is more bell-shaped and center-concentrated (the sum of four uniform distributions). Most of the results presented in the figures below pertain to distribution A, and some results to distribution B. It turns out that the main findings reported below do not change with the shape of the distributions.

**Results**

We first test the methods by using artificial datasets to assess their validity and reliability under different assumptions. The



**Figure 1. A simple illustration of the counting method.** We use the method to approximate the probability distribution of stage 1 duration of a multistage disease. The top panel shows 12 patients' visits. The horizontal axis is the time elapsed from stage 1 onset, and the numbers on top of vertical arrows indicate the stage assessment for each patient. The time-axis is split into 4 unequal intervals,  $I_1, \dots, I_4$ , with  $J = 3$  patient visits in each. The corresponding average visit times,  $\tilde{t}_i$ , and the numbers of transiting patients,  $N_t^i$ , are calculated. In the bottom panel, the approximation  $(\tilde{t}_i, P^1(t_i))$  is plotted, where  $P^1(t_i) = N_t^i/3$ .  
doi:10.1371/journal.pone.0028298.g001

**Table 1.** Test Distributions for AD simulations.

Stage	Mean	Variance A	St. Dev. A	Variance B	St. Dev. B	Variance C	St. Dev. C
4	2	3/4	0.87	3/8	0.61	3/16	0.43
5	1.5	1/3	0.58	1/6	0.41	1/12	0.29
6	2.5	4/3	1.16	2/3	0.82	1/3	0.58
7	7	169/12	3.75	169/24	2.65	169/48	1.88

doi:10.1371/journal.pone.0028298.t001

advantage of this approach is that in the case of artificial datasets, we actually know the true “answer” and can compare our estimates with the correct values. Once the properties of the statistical methods have been established, we will apply them to a real-life dataset.

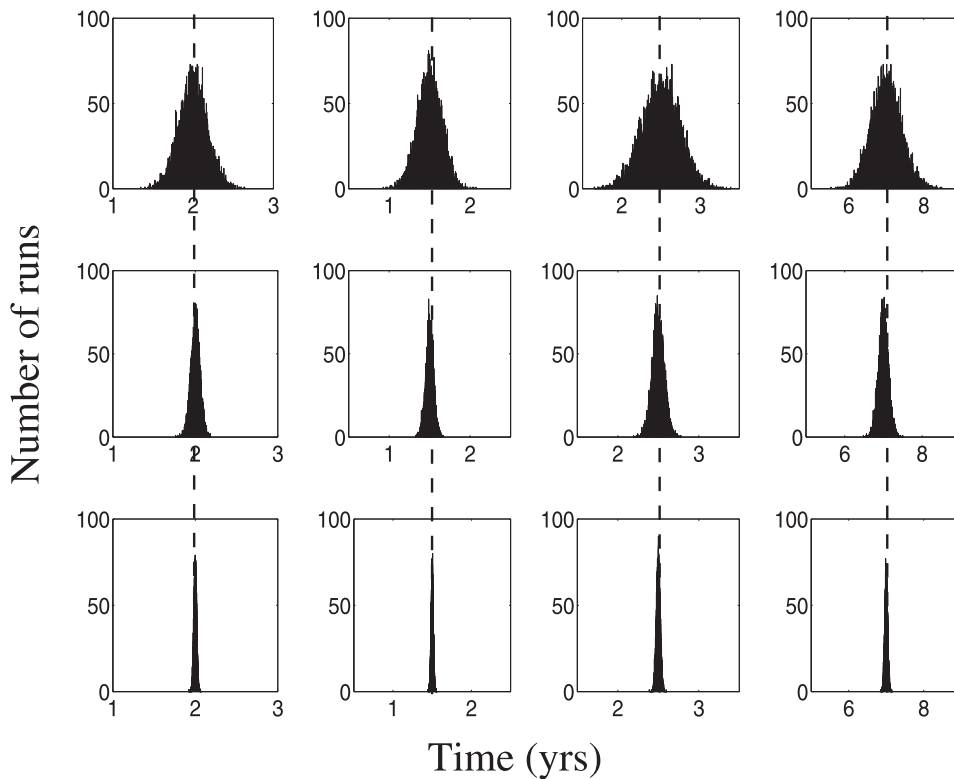
**Mean and standard deviation reconstruction via regression**

To study the accuracy of the regression-based method, we adopt the following strategy. For each total data sample size,  $N_p$ , we randomly generate a large number of datasets (namely, 10,000 for the examples presented here), and for each such dataset we calculate the stage means and standard deviations based on equations. We assume that the number of patients is given by (a)  $N_p = 1,000$ , (b)  $N_p = 10,000$ , and (c)  $N_p = 100,000$ . The calculated mean and standard deviation values are then compared with the true means and standard deviations.

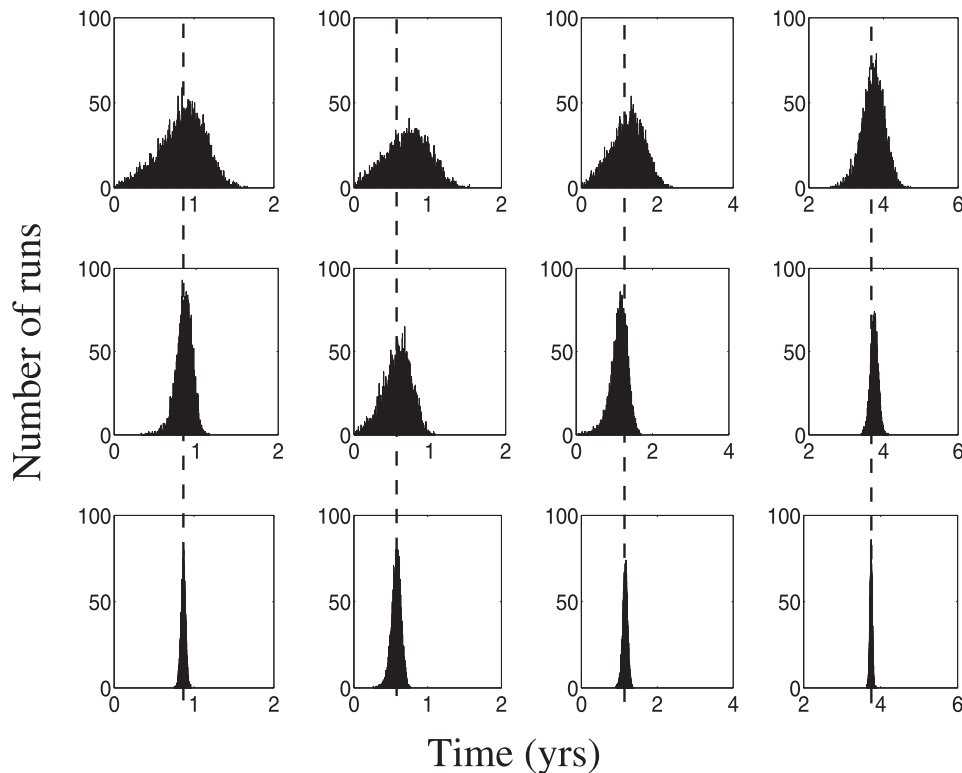
Typical histograms of calculated mean and standard deviation values are given in figure 2 for the means calculations and figure 3 for the standard deviation calculation. In both figures we used stage duration distribution A, table [?]. In each panel, the rows represent different numbers of patients considered, and the columns correspond to the four stages of the disease. The true values of the mean and the standard deviations are marked on each histogram by vertical lines, for comparison.

A convenient measure of *accuracy* of a method is as follows. The accuracy of a mean or standard deviation calculation is the length of the 95% confidence interval centered around the true value of the mean or standard deviation, respectively. From figures 2 and 3 we note the following trends:

- The histograms for the mean and standard deviation calculations have a bell-shaped form centered around the true mean. The spread of the mean and standard deviation estimates decreases with the number of patients.



**Figure 2. Regression method: the mean stage durations for data generation method 1.** Sets of 1,000 (top row), 10,000 (middle row) and 100,000 (bottom row) patients were considered. Each column presents the calculated mean for 10,000 simulations for each stage of a 4-stage disease. The true mean values are shown by dashed vertical lines.  
doi:10.1371/journal.pone.0028298.g002



**Figure 3. Regression method: the standard deviation of stage durations.** All the parameters are as in figure 1.  
doi:10.1371/journal.pone.0028298.g003

- The accuracy of the calculations increases with the patient number. For all stages, increasing the patient number by a factor of ten leads to an increase in accuracy by a factor of 3.1 for the mean calculations and by a factor of 2.5 for the variance calculations.
- At low patient numbers, the standard deviation calculations are uncertain. A fraction of variances calculated at the 1,000 patient level returned a negative value (these patients are not shown in the histograms of figure 3). Thus the standard deviation calculation procedure requires significantly more data to yield accurate and precise results than does the mean (see Text S1 for a mathematical explanation of this observation).
- The accuracy of the first and the last stage mean calculations is slightly lower than that for middle stages. The reason for that is the smaller number of transition classes which include the end stages; for example, there is no transition class which fully transits the first or the last stage. It is noteworthy that the difference in accuracy of end stages compared to the middle stages is not large (less than a factor of 2).
- We have performed calculations for three different distributions of stage durations (not shown). We observe that the accuracy of the mean calculations does not depend strongly on the underlying distribution. The accuracy of the standard deviation reconstruction is more sensitive to the relative magnitude of the distribution's variance (a smaller distribution variance increases the relative error of reconstruction).

To conclude, we note that, given data which conforms to the assumptions needed to derive the governing equations, this method is surprisingly accurate in the calculations of the mean. Even for small patient samples with a large number of disease

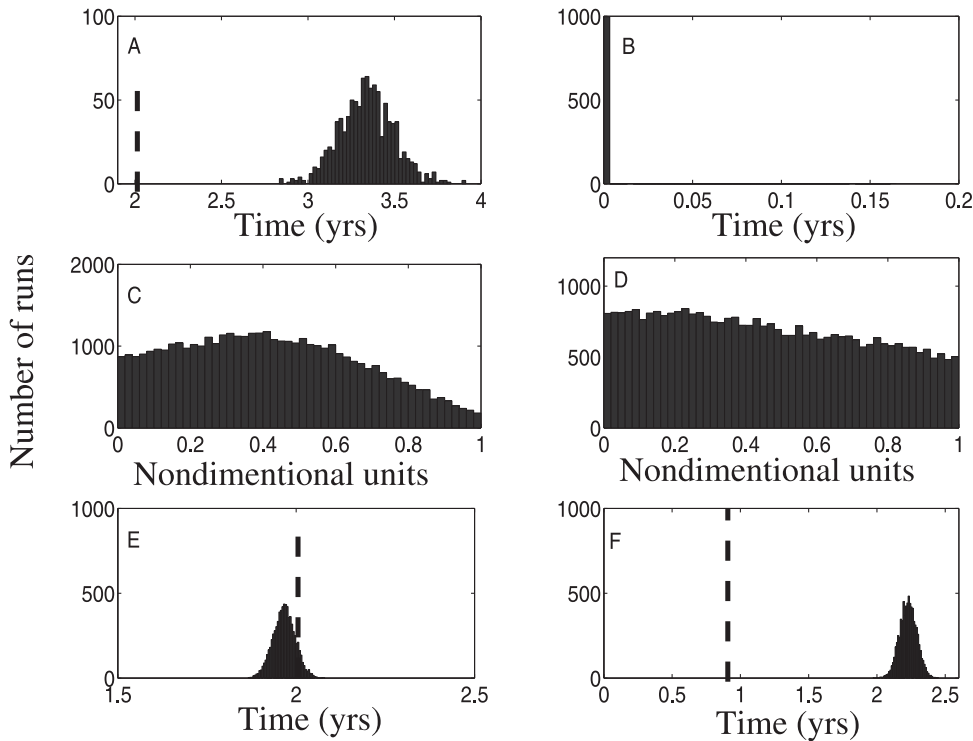
stages (which translates into a very small number of patients per transition class), we obtain accuracy in the reconstruction on the order of just a few percent error. However, the variance reconstruction proves to be much more sensitive to the number of patients in the sample size than does the mean reconstruction. With sufficient number of patients per class, it is still possible to attain accuracy to within 5% error; however, many more patients are needed in the variance reconstruction than in the mean reconstruction.

### Regression method: non-independent completion coefficients

We next explore the ability of the linear regression algorithm to compute the mean and variance of data created by data generation method 2. We maintain the assumption that patients initially arrive at the clinic at a random time during their current stage, and are seen thereafter at time-intervals determined by a preset distribution. We simulate data sets of 10,000 patients with disease course determined by distribution set B, with a mean inter-office interval of 1 year. The results for 1,000 independent runs are presented for the first stage of the disease in figure 4.

As seen in panels A and B of figure 4, the regression method is able to accurately recreate neither the means (panel A) nor the standard deviation (panel B), in sharp contrast to the results from data generation method 1. Note that the variance values generated by the algorithm were mostly negative, and they are replaced by zero values in panel B of figure 4. We next attempt to determine why the method has failed. Panels C and D of figure 4 show histograms of the completion coefficients for one particular realization of a patient dataset, for transition classes 1→2 and 2→3. We can see that in neither of these cases are the completion coefficients uniformly distributed across the interval [0,1].





**Figure 4. Failure of the regression method to accurately calculate standard deviations in data generation method 2.** (A,B). Accuracy histogram of mean (panel A) and standard deviation (panel B) reconstruction for the first stage of a 4-stage disease using a data set of 10,000 patients created via method 2 with an inter-office interval mean of 1 year. Entry coefficients were assumed to occur uniformly. (C,D). Actual distribution of entry-stage completion coefficients for the transition classes 1→2 (panel C) and 2→3 (panel D) for the data in panels A and B. (E,F) Same as in panels (A,B) but the entry coefficients were assumed to be 1 (patients first visit the clinic at the beginning of their stages). The true values of the mean and the standard deviations are marked on histograms A, E and F by vertical lines.  
doi:10.1371/journal.pone.0028298.g004

Moreover, despite the fact that both of these transitions represent  $i \rightarrow i + 1$  transitions, the distribution of completion coefficients are not the same, indicating that the assumption that completion coefficients are independent of the underlying stage distribution is no longer valid. In fact, the distribution for the 1→2 class appears to have a peak around the value 0.4, whereas the 2→3 class appears to be monotonically decreasing across the domain. Given that this fundamental assumption of the regression method no longer holds, it is clear that the method cannot give reliable estimates of stage means or standard deviations.

We next look to determine if the regression method might still be of some use for data of type 2. Instead of arriving randomly throughout their current stage, in this new experiment the patients are assumed to come to their first clinician’s visit at the beginning of the stage. This is reflected both in the way we generate the datasets, and in the way we implement the regression algorithm: all transitions of type  $i \rightarrow j$ ,  $i < j$  will have an entry coefficient of 1. This assumption still gives no information about the exit coefficients nor the single coefficient of  $i \rightarrow i$  transitions, so we will leave the initial estimate of a value of  $\frac{1}{2}$  for the mean and  $\frac{1}{3}$  for the variance in place. We plot the results of mean and standard deviation calculation on a 10,000 patient dataset in panels E and F of figure 4 respectively. We see here that correct knowledge of the entry coefficient is sufficient to recover most of the accuracy of the means calculations, but there is still significant error in the standard deviation calculation.

Thus, we must conclude that, short of extensive knowledge of all the completion coefficients, the regression technique is in general only suitable for calculation of the mean stage durations if

- the data is collected in a manner similar to method 1, or
- the data is collected in a manner similar to our method 2, where patients arrive to their first visit near the onset of the current stage.

The standard deviation calculations by regression technique are only reliable in the case (a) above.

### Mean and standard deviation reconstruction via the counting method

We adopt a similar strategy for exploring the accuracy of the counting method as we did for the regression model. We generate datasets based upon our three underlying distribution sets and perform a counting analysis on those sets. This process is repeated 7,500 times so that accuracy statistics may be generated. We perform these simulations twice, first on a dataset of 1,000 patients, and then on a set of 10,000 patients. In order for the counting method to be valid, the patients’ first visit must coincide with the current stage onset (see the Methods section). Therefore, to test the counting method we use data generated from procedure 2, which adds another layer of complexity to the analysis. Data type 2 involves sampling patients at time intervals drawn from a given distribution. For this test, we consider two such sampling distributions, one with a mean inter-office visit interval of 1 year and the other with a mean of 2 years. In other words, in the former set patients return on average each year for a visit to the clinic, whereas in the latter set patients come on average every 2 years. The latter sampling methodology represents a much coarser study, which is not uncommon in AD longitudinal studies. We will refer

to the sampling distribution with 1 year means by a “finer sampling” and the other one as the “coarser sampling” case.

We plot representative histograms for the results for distribution A for sampling intervals of both 1 and 2 years. Figures 5 and 6 show the histograms for mean and standard deviation calculations for the coarser sampling distribution, which is a more difficult case. The results for the finer sampling case are presented in Text S1.

The results shown in figures 5 and 6 are striking. Unlike the regression method, the counting method works very well for type 2 data, both for mean and standard deviation calculation. We do note that a small bias exists in both the mean and standard deviation calculations for data sampled at 2 year intervals for 1,000 patients; however, that bias is eliminated with a larger dataset (see figure 5 and compare row 1 versus row 2, the effect is most apparent in the left and right columns). This is expected from our analysis of the counting method, see Text S1. Results for the counting method can be compared with those for the traditional KM statistics (Text S1), which suffer from a strong bias in the context of nonuniform sampling distributions. Unlike the case of the counting method, this bias is not eliminated by larger sampling sizes. Thus the counting method can provide accurate and precise estimates of both stage means and stage standard deviations for a multistage disease observed in even a coarsely sampled longitudinal study.

We further observe the following trends in the calculations by the counting method:

- Unlike the regression method, there is no loss of accuracy (defined in the previous subsection) in the calculation of the first stage parameters. This is a feature of treating each stage

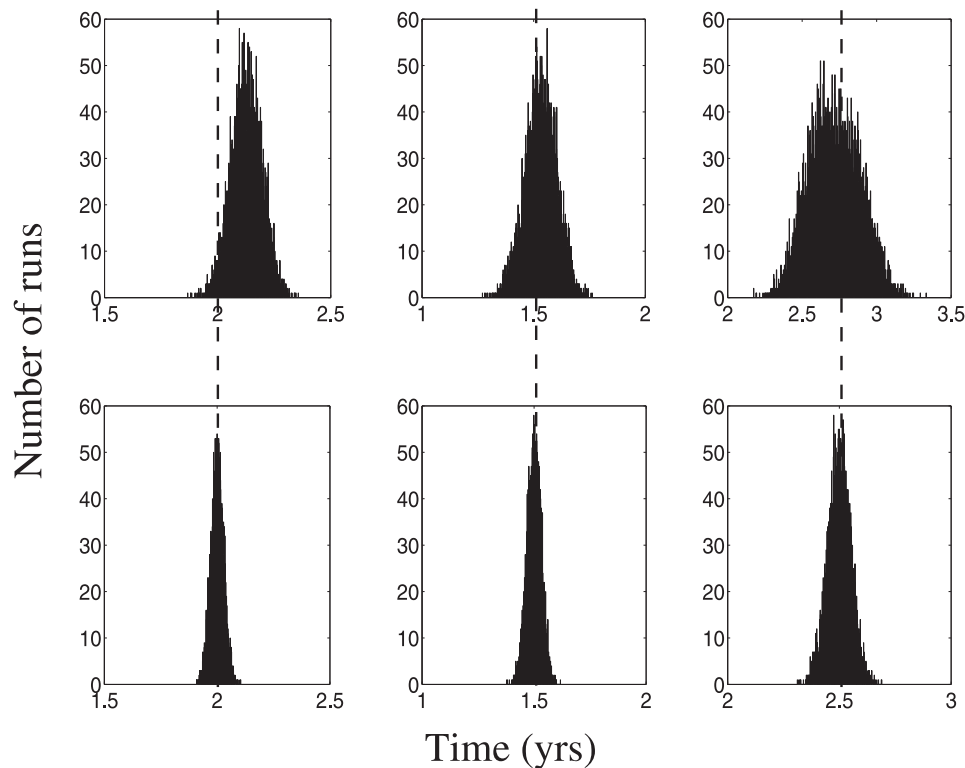
separately rather than grouping all patients together for a single calculation.

- However, due to the fact that each stage is treated separately, the increase in accuracy after increasing the number of patients by a factor of 10 is less than that observed for the regression method.

### Application to an AD dataset

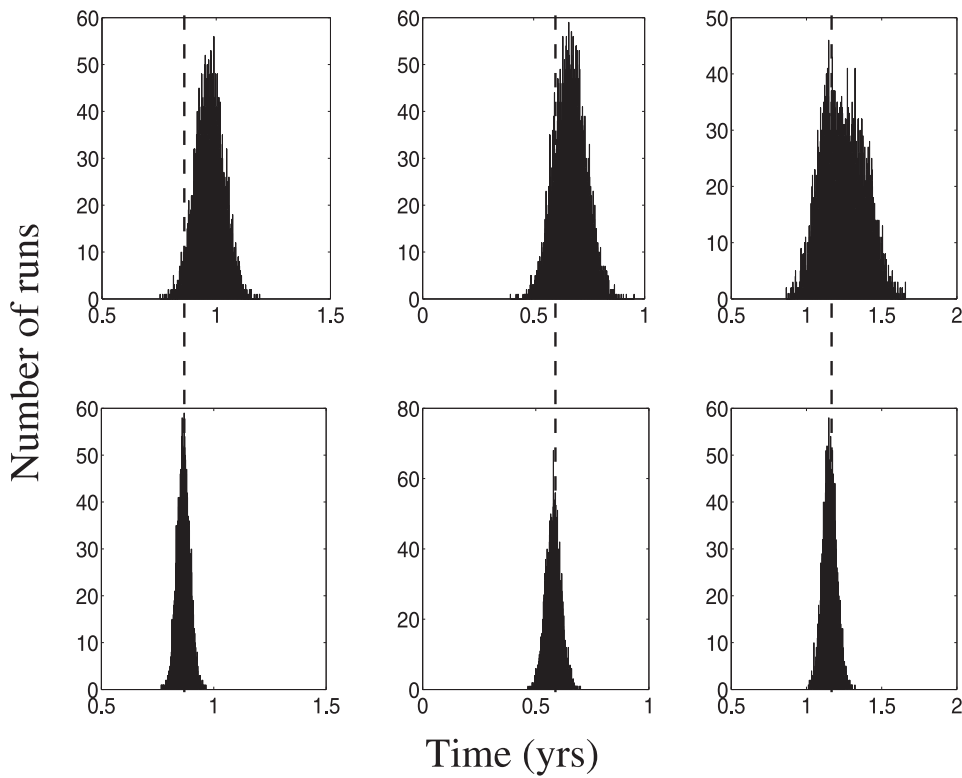
In order to demonstrate the applicability of the two methods, we use these techniques to analyze a longitudinal dataset of AD patients, which is an outcome of a study performed between the years 1983 and 2006 [25]. The following information is contained in the dataset: the date of each patient’s visit to the clinic, current GDS and FAST stage, and some demographic information on each patient (such as gender, age and years of education). The total number of AD patients in the dataset is 1321, of which 648 have repeated records (that is, they were seen more than once). The latter group is the one we considered in this study. The mean number of records per patient is  $2.6 \pm 0.9$ . The patients’ age at the first visit to the clinic is  $73.1 \pm 8.7$  years. 66% of the patients are female, and 34% male. The average length of education received by the patients is  $13.1 \pm 3.4$  years.

In figure 7 we present the inter-visit time distribution, which shows how long the patients waited before their next visit to the doctor. Two observations are important. (1) The distribution has a strong peak around 2 years, and then a weaker mode around 4 years, which tells us that the sampling times are strongly biased. This is because the patients were instructed to schedule their subsequent visits after 2 years, see [25]. This results in a highly



**Figure 5. Counting method: the mean stage durations for data generation method 2 (the coarse case).** Sets of 1,000 (top row) or 10,000 (bottom row) patients were considered. Each column presents the calculated mean for 7,500 simulations for stages 1–3 of a 4-stage disease. The true mean values are shown by dashed vertical lines. doi:10.1371/journal.pone.0028298.g005

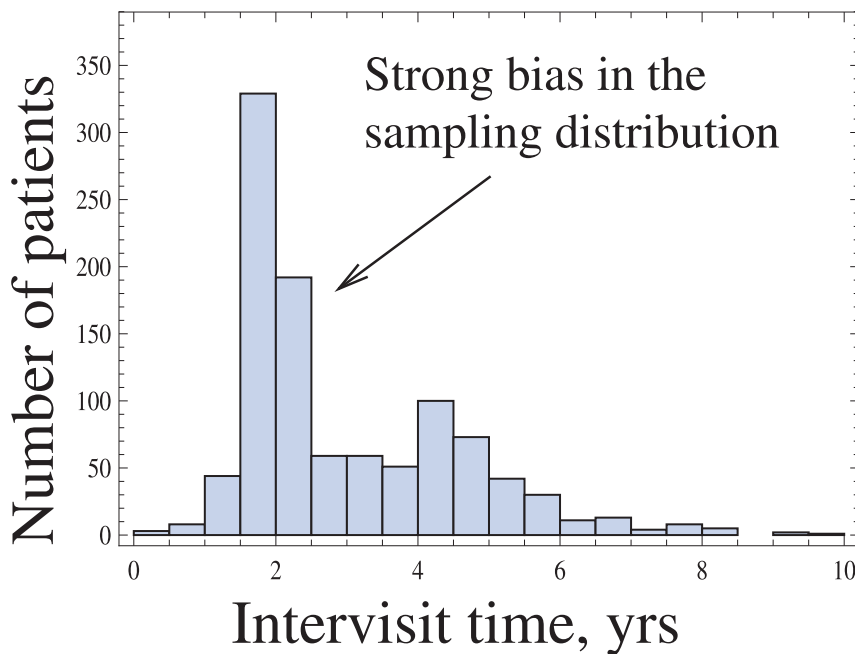




**Figure 6. Counting method: the standard deviation of stage durations.** All the parameters are as in figure 4.  
doi:10.1371/journal.pone.0028298.g006

biased distribution of sampling times in the dataset. For this reason, a traditional method such as Kaplan Meier would be strongly compromised, see Text S1. (2) The average inter-visit time in the dataset is  $3.03 \pm 1.59$ , which is comparable with the

approximate average stage duration for GDS/FAST stages 4–6. For these reasons we conclude that the dataset at hand is most in line with the data collecting method 2, under the coarse sampling conditions.



**Figure 7. Real-life AD dataset: a strong bias in the sampling distribution.**  
doi:10.1371/journal.pone.0028298.g007

An example of results for GDS and FAST stage calculations is presented in figure 8, see [24] for more information. We plot the cumulative probability distribution estimated by the counting method for AD stages 4–6. The mean stage durations and their standard deviations calculated by this method are provided on top of the graphs. We further applied the regression method to calculate the mean stage durations. The results were 2.39 yrs, 1.49, and 2.49 yrs for stages 4, 5 and 6, respectively, see [25]. We did not attempt to calculate the standard deviations of stage durations by the regression method because we have established the method’s poor reliability in this context.

First we note that the graphs of the cumulative probability distribution approximations (figure 8) have very non-uniform number of points along the time axes. The striking increase in the density of points near the 2-yr mark corresponds to the peak of the inter-visit time distribution, figure 7. Looking at the results for the mean values calculated by the two methods, we can see that they are within the standard deviation of each other, and also within the standard deviation of the previously reported values, which are 2 yrs, 1.5 yrs and 2.5 yrs for stages 4, 5, and 6, respectively [49]. We further notice that the values calculated by the counting method are somewhat higher than those obtained by the regression method. This is because for a small number of patients (see figure 5, top row), there is a certain bias in the estimates for the mean. This bias can be corrected by increasing the sampling size.

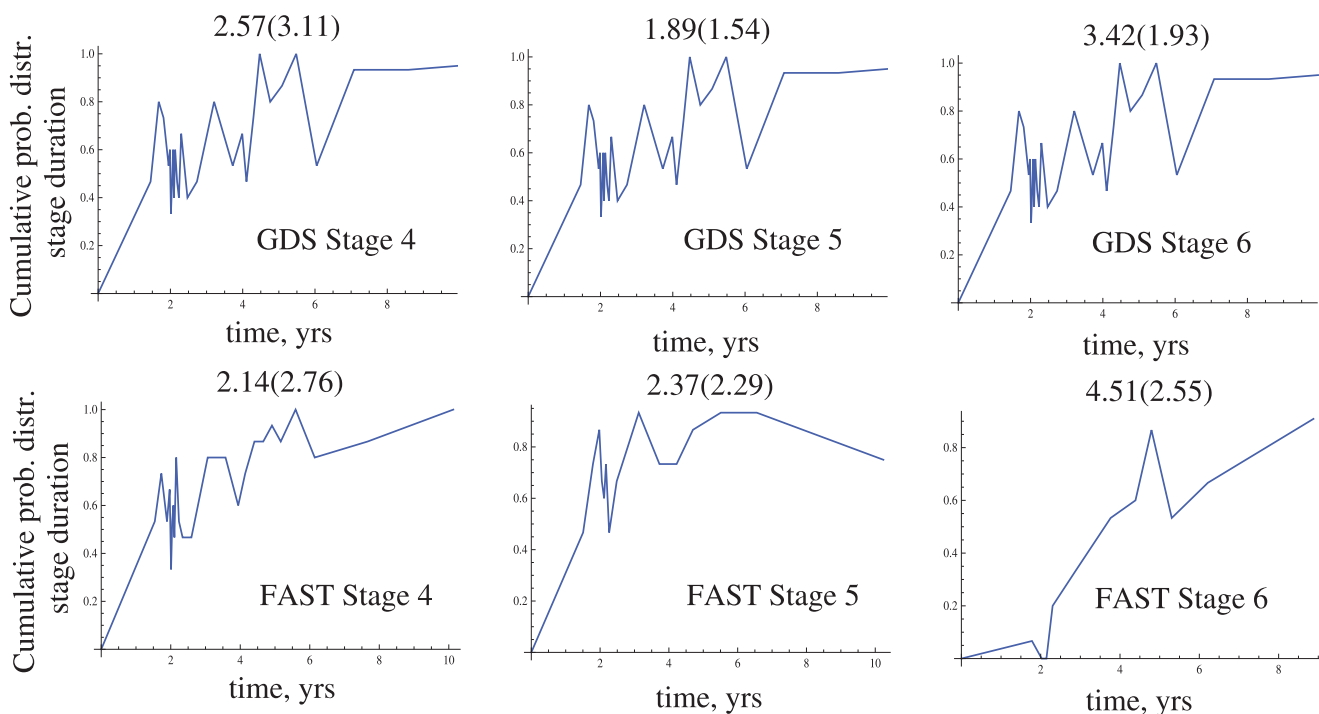
Finally, we observe that the values of the standard deviations are very large and are comparable with the mean values. It is true that these values is a slight overestimation because of the aforementioned bias which comes with small sampling sizes. However, in [24], these values were compared with the ones obtained by a different method (the method of Lagakos [20]), and the results

turned out to be very similar, confirming that the values obtained by the counting method were valid.

Although before Ref. [24], no values for the standard deviations of GDS/FAST stage durations of AD had been published, the magnitude of the calculated values is consistent with the general notion of AD being a heterogeneous disease [50]. In [51], inhomogeneity is observed with respect to the rates of ventricle enlargement, which are related to rates of cognitive decline. Many papers report a wide spread of progression rates of AD patients and find different correlates of progression speed. In [52], the presence of aphasia in AD patients is correlated with a more rapid course of the disease. Ref. [27] discovers an association between relatively severe frontal lobe involvement and a rapid clinical course of AD, measured by using the dementia rating scale and estimating the symptom duration time. In [53], it is found that the average rates of decline vary with respect to three types of measures: a cognitive measure (Alzheimers disease Assessment Scale-Cognitive Subscale), a functional measure (Physical Self-Maintenance Scale), and a global measure (CDR sum of boxes). Although no direct estimate of the variation is presented, these results clearly show that AD progression rates are heterogeneous in many respects.

### Discussion

In this paper we have presented two different methods which may be applied to the analysis of multi-stage diseases, where the goal is to reconstruct the individual stage distribution parameters (mean and standard deviation). This information is critical for statistical hypothesis testing needed to determine if an experiment is having a significant effect on the progression through one or more stages, e.g. if a new therapy is showing a delay of progression through a multistage disease. The present study is purely statistical



**Figure 8. Application of the counting method to estimating the cumulative probability distributions of AD GDS/FAST durations for stages 4, 5, and 6 from an AD dataset [25].** The parameter  $\tilde{t}=0.3$  yrs was used. The mean and the standard deviation estimated by this method are given on top of each figure.

doi:10.1371/journal.pone.0028298.g008

and no elements of modeling are used. In other words, our focus is to develop analytical tools for existing and future patient datasets in multistage disease with the goal to infer (in the most efficient and precise way) information about stage durations and their variations. To this end, we tested two statistical methods and demonstrated particular situations in which each method might be applied properly.

The counting method requires more stringent restrictions on the entry-stage completion coefficients, namely that the onset of a stage be known as precisely as possible. It, however, makes no restriction on the exit coefficients. It is also not possible to reconstruct the parameters of the final stage in the sequence unless detailed exit data is obtained. Despite these restrictions, this method gives relatively accurate estimates of both means and variances even under coarsely sampled data. Thus this method is most useful for measurements of stage variances, which are critical for any hypothesis test needed to evaluate treatments, assuming a sufficiently large patient dataset. This method was used in [24] to study the statistics of AD progression in a longitudinal patient set. In particular, it was discovered that GDS and FAST stage durations of AD patients are characterized by large variances, confirming the notion of AD being a highly heterogeneous disease [27,50–53].

Conversely, the regression method is most applicable in situations where, firstly, the general distribution of the completion coefficients is known and, secondly, where the data are insufficient to properly perform the counting method. However, in those situations, based on our results on the means, calculation may be trusted at lower numbers of patients if (a) enough information is available concerning the completion coefficients, and (b) different stage durations are independent of one another. To see that, recall that when we derived the equations for the regression method, we used the assumption of independence of different stages. Care must be taken when considering the validity of this assumption, as there are many possible avenues by which it can be broken. For example, in the AD community there are hypotheses [27,39–41,54] which concern the presence of different subclasses of progression speed. That is, there may be sub-populations of patients in the dataset which progress through each and every stage at rates drawn from a distribution with a lower mean than those of other patients. In this case, the stages would only be conditionally independent from one another, and the equations will not represent the variances accurately. On the other hand, given that a patient is in the fast or slow subgroup, the stage distributions become independent.

## References

1. Querfurth HW, LaFerla FM (2010) Alzheimer's disease. *N Engl J Med* 362: 329–344.
2. Forstl H, Kurz A (1999) Clinical features of Alzheimer's disease. *Eur Arch Psychiatry Clin Neurosci* 249: 288–290.
3. Shay KA, Duke LW, Conboy T, Harrell LE, Callaway R, et al. (1991) The clinical validity of the Mattis Dementia Rating Scale in staging Alzheimer's dementia. *J Geriatr Psychiatry Neurol* 4: 18–25.
4. Durie BG, Salmon SE (1975) A clinical staging system for multiple myeloma. Correlation of measured myeloma cell mass with presenting clinical features, response to treatment, and survival. *Cancer* 36: 842–854.
5. Greipp PR, San Miguel J, Durie BG, Crowley JJ, Barlogie B, et al. (2005) International staging system for multiple myeloma. *J Clin Oncol* 23: 3412–3420.
6. Lister TA, Crowther D, Sutcliffe SB, Glatstein E, Canellos GP, et al. (1989) Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. *J Clin Oncol* 7: 1630–1636.
7. Royce RA, Luckmann RS, Fusaro RE, Winkelstein W (1991) The natural history of HIV-1 infection: staging classifications of disease. *AIDS* 5: 355–364.
8. Vandendruene M, Colebunders R, Goeman J, Alary M, Farber CM, et al. (1993) Evaluation of two staging systems for HIV infection for use in developing countries. *AIDS* 7: 1613–1615.
9. Duray PH, Steere AC (1988) Clinical pathologic correlations of Lyme disease by stage. *Ann N Y Acad Sci* 539: 65–79.
10. Steere AC (1989) Lyme disease. *N Engl J Med* 321: 586–596.
11. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, et al. (2003) National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Ann Intern Med* 139: 137–147.
12. Winearls CG, Glasscock RJ (2009) Dissecting and refining the staging of chronic kidney disease. *Kidney Int* 75: 1009–1014.
13. Braak H, Del Tredici K, Rub U, de Vos RA, Jansen Steur EN, et al. (2003) Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging* 24: 197–211.
14. Braak H, Ghebremedhin E, Rub U, Bratzke H, Del Tredici K (2004) Stages in the development of Parkinson's disease-related pathology. *Cell Tissue Res* 318: 121–134.
15. Singer J, Willett J (2003) *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.
16. Reisberg B, Ferris SH, Franssen EH, Shulman E, Monteiro I, et al. (1996) Mortality and temporal course of probable Alzheimer's disease: a 5-year prospective study. *Int Psychogeriatr* 8: 291–311.
17. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *J Amer Statist Assn* 53: 457–481.
18. Turnbull B (1974) Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association* 69: 169–173.

Because of these considerations, the regression method is most useful in searching for partitions within a larger dataset. For example, several previous reports in the AD literature hypothesize the presence of fast progressors and slow progressors within the general AD population [26,27,34]. That is, there are patients whose AD stage durations are in general shorter than those of their counterparts across all stages of the disease. A sorting routine, e.g. based on a genetic algorithm, can be used to partition these patients into fast progressing versus slow progressing subgroups. The regression method is the better option to use in such sorting routines within a defined dataset [25].

In this paper we demonstrated an example of how our methodology can be applied to studying the stage duration distributions in AD. In the data set we used, GDS and FAST staging systems were applied to assess the patients' decline. Recent clinical trials involving candidate treatments for AD have targeted cohorts that already exhibit mild dementia (roughly equivalent to FAST stage 4). It is generally hypothesized [55] that the failure of those trials is attributable in large part to enrolling only demented participants; clinicopathological evidence suggests that even mild clinical symptoms appear only after AD pathology has advanced to the point at which neurons are damaged or destroyed. Therefore, the strongest current focus of AD research is to identify and treat the early symptomatic and even pre-symptomatic stages of the disease (before FAST stage 4), and prevent the development of significant pathology before ostensibly irreparable brain damage has occurred [56]. Because these early stages may last for several years and may have rather poorly-defined borders, future trials are likely to require more precise measures of cognitive decline, with greater sensitivity to small cognitive/behavioral changes. Methods developed in this paper are relevant to the development of new treatments for AD in the context of other rating systems with greater sensitivity to detect small changes in cognition early in the disease. This is the subject of future work.

## Supporting Information

**Text S1** Details of the methodology developed in the paper. (PDF)

## Author Contributions

Conceived and designed the experiments: NK CT. Performed the experiments: CT NK. Analyzed the data: NK CT. Contributed reagents/materials/analysis tools: NK CT. Wrote the paper: NK CT.

19. Turnbull B (1976) The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society B* 38: 290–295.
20. De Gruttola V, Lagakos SW (1989) Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* 45: 1–11.
21. Gomez G, Lagakos SW (1994) Estimation of the infection time and latency distribution of AIDS with doubly censored data. *Biometrics* 50: 204–212.
22. Ferguson T (1973) A bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1: 209–230.
23. Susarla V, Van Ryzin J (1976) Nonparametric bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association* 71: 897–902.
24. Komarova NL, Thalhauser CJ (2011) High degree of heterogeneity in Alzheimers disease progression patterns. *PLoS Comp Biol*;Epub ahead of print.
25. Thalhauser CJ, Komarova NL (2011) Alzheimer's disease: rapid and slow progression. *J R Soc Interface*: Epub ahead of print.
26. Chui HC (1987) The significance of clinically defined subgroups of Alzheimer's disease. *J Neural Transm Suppl* 24: 57–68.
27. Mann UM, Mohr E, Gearing M, Chase TN (1992) Heterogeneity in Alzheimer's disease: progression rate segregated by distinct neuropsychological and cerebral metabolic profiles. *J Neurol Neurosurg Psychiatr* 55: 956–959.
28. Craft S, Teri L, Edland SD, Kukull WA, Schellenberg G, et al. (1998) Accelerated decline in apolipoprotein E-epsilon4 homozygotes with Alzheimer's disease. *Neurology* 51: 149–153.
29. Ito K, Corrigan B, Zhao Q, French J, Miller R, et al. (2011) Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database. *Alzheimers Dement* 7: 151–160.
30. Farrer LA, Cupples LA, van Duijn CM, Connor-Lacke L, Kiely DK, et al. (1995) Rate of progression of Alzheimer's disease is associated with genetic risk. *Arch Neurol* 52: 918–923.
31. Murphy GM, Claassen JD, DeVoss JJ, Pascoe N, Taylor J, et al. (2001) Rate of cognitive decline in AD is accelerated by the interleukin-1 alpha -889 \*1 allele. *Neurology* 56: 1595–1597.
32. Jack CR, Shiung MM, Gunter JL, O'Brien PC, Weigand SD, et al. (2004) Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology* 62: 591–600.
33. Ridha BH, Barnes J, Bartlett JW, Godbolt A, Pepple T, et al. (2006) Tracking atrophy progression in familial Alzheimer's disease: a serial MRI study. *Lancet Neurol* 5: 828–834.
34. Shuimer JD, Vrenken H, Blankenstein MA, Fox NC, Scheltens P, et al. (2008) Whole-brain atrophy rate in Alzheimer disease: identifying fast progressors. *Neurology* 70: 1836–1841.
35. McEvoy LK, Fennema-Notestine C, Roddey JC, Hagler DJ, Holland D, et al. (2009) Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. *Radiology* 251: 195–205.
36. Nestor SM, Rupsingh R, Borrie M, Smith M, Accomazzi V, et al. (2008) Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. *Brain* 131: 2443–2454.
37. Mielke MM, Rosenberg PB, Tschanz J, Cook L, Corcoran C, et al. (2007) Vascular factors predict rate of progression in Alzheimer disease. *Neurology* 69: 1850–1858.
38. Prolo P, Chiappelli F, Angeli A, Dovio A, Perotti P, et al. (2007) Physiologic modulation of natural killer cell activity as an index of Alzheimer's disease progression. *Bioinformatics* 1: 363–366.
39. Bhargava D, Weiner MF, Hyman LS, Diaz-Arrastia R, Lipton AM (2006) Vascular disease and risk factors, rate of progression, and survival in Alzheimer's disease. *J Geriatr Psychiatry Neurol* 19: 78–82.
40. Nagahama Y, Nabatame H, Okina T, Yamauchi H, Narita M, et al. (2003) Cerebral correlates of the progression rate of the cognitive decline in probable Alzheimer's disease. *Eur Neurol* 50: 1–9.
41. Doody RS, Massman P, Dunn JK (2001) A method for estimating progression rates in Alzheimer disease. *Arch Neurol* 58: 449–454.
42. O'Bryant SE, Waring SC, Cullum CM, Hall J, Lacritz L, et al. (2008) Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's research consortium study. *Arch Neurol* 65: 1091–1095.
43. O'Bryant SE, Lacritz LH, Hall J, Waring SC, Chan W, et al. (2010) Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. *Arch Neurol* 67: 746–749.
44. Reisberg B, Ferris SH, de Leon MJ, Crook T (1982) The Global Deterioration Scale for assessment of primary degenerative dementia. *Am J Psychiatry* 139: 1136–1139.
45. Reisberg B (1988) Functional assessment staging (FAST). *Psychopharmacol Bull* 24: 653–659.
46. Sclan SG, Reisberg B (1992) Functional assessment staging (FAST) in Alzheimer's disease: reliability, validity, and ordinality. *Int Psychogeriatr* 4 Suppl 1: 55–69.
47. Na HR, Kim SY, Chang YH, Park MH, Cho ST, et al. (2010) Functional assessment staging (FAST) in Korean patients with Alzheimer's disease. *J Alzheimers Dis* 22: 151–158.
48. Olde Rikkert MG, Tona KD, Janssen L, Burns A, Lobo A, et al. (2011) Validity, Reliability, and Feasibility of Clinical Staging Scales in Dementia: A Systematic Review. *Am J Alzheimers Dis Other Dement* 26: 357–365.
49. Reisberg B (1986) Dementia: a systematic approach to identifying reversible causes. *Geriatrics* 41: 30–46.
50. Mayeux R, Stern Y, Spanton S (1985) Heterogeneity in dementia of the Alzheimer type: evidence of subgroups. *Neurology* 35: 453–461.
51. Friedland RP, Koss E, Haxby JV, Grady CL, Luxenberg J, et al. (1988) NIH conference. Alzheimer disease: clinical and biological heterogeneity. *Ann Intern Med* 109: 298–311.
52. Knesevich JW, Toro FR, Morris JC, LaBarge E (1985) Aphasia, family history, and the longitudinal course of senile dementia of the Alzheimer type. *Psychiatry Res* 14: 255–263.
53. Doody RS, Pavlik V, Massman P, Rountree S, Darby E, et al. (2010) Predicting progression of Alzheimer's disease. *Alzheimers Res Ther* 2: 2.
54. Capitani E, Cazzaniga R, Francescani A, Spinnler H (2004) Cognitive deterioration in Alzheimer's disease: is the early course predictive of the later stages? *Neurol Sci* 25: 198–204.
55. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, et al. (2011) Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging- Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 7: 280–292.
56. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, et al. (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 7: 270–279.