Published in final edited form as:

*J Stat Plan Inference*. 2011 May 1; 141(5): 1783–1788. doi:10.1016/j.jspi.2010.11.028.

# Fractional Brownian motion and long term clinical trial recruitment

**Qiang Zhang**[a,*] and **Dejian Lai**[b]

[a]American College of Radiology, 1818 Market Street, Suite 1600, Philadelphia, PA, 19103, U.S.A

[b]School of Public Health, University of Texas at Houston, Division of Biostatistics, Houston, TX, 77030, USA

## Abstract

Prediction of recruitment in clinical trials has been a challenging task. Many methods have been studied, including models based on Poisson process and its large sample approximation by Brownian motion (BM), however, when the independent incremental structure is violated for BM model, we could use fractional Brownian motion to model and approximate the underlying Poisson processes with random rates. In this paper, fractional Brownian motion (FBM) is considered for such conditions and compared to BM model with illustrated examples from different trials and simulations.

### Keywords

Brownian motion; Fractional Brownian motion; Recruitment; Prediction

## Introduction

Recruitment is one of the critical steps of a clinical trial. If the desired sample size can not be achieved before end of the randomization period, one could either redesign the trial with different sample size requirement, adding more centers, extend the trial duration, or simply stop the current ongoing trial. Predicting future recruitment is often done by using a straight line extrapolation, which is based on the number of subjects already been randomized. According to this method, random variation around the future path is not considered with formal inferential procedure. Lee (1983) and Williford (1987) and Carter (2004) have used different Poisson processes to model recruitment processes. By assuming a constant recruitment rate, Lee (1983) derived the minimum acceptable interim recruitment goal and constant recruitment rate needed to achieve the final sample size. Also, by allowing the current rate depending on the observed rates so far, the interim recruitment goals are derived based on contagious Poisson distribution. In addition, calculations for recruitment duration,

*Corresponding author: Tel: 1-215-574-3197 Fax: 1-215-928-0153, ed.zhang.jr@gmail.com, Postal address: American College of Radiology, 1818 Market Street, Suite 1600, Philadelphia, PA, 19103, U.S.A.

projecting final accrual at interim and length of extension are also discussed. However, by studying recruitment rates for several clinical trials, Williford (1987) concluded that a constant rate was not observed for these studies. A possible alternative negative binomial model is introduced to assess model fit. Among the trials compared, negative binomial distribution provided better fit than Poisson model for studies with sample size greater than 200. This indicates for these studies Poisson distribution may not be the appropriate model for cumulative recruitment at trial level. They introduced a Bayesian model with prior information that could be estimated during the recruitment process. With a gamma prior and Poisson distribution the posterior distribution is again a gamma distribution. The results are compared to those of Poisson model and Lee's contagious Poisson approach. It showed that the contagious method is conservative in that the expected number of patients to be recruited is greater than that of Poisson model. The Bayesian model does not depend on the monitoring frequency and is able to provide a realistic estimate of recruitment rate at time *t* given observed data through time *t*. Carter (2004) illustrated how a Poisson process can be used to decide an optimal period of time for participant recruitment by simulating Poisson arrival into a clinical trial. Other similar efforts for modeling recruitment process can be found in Bjornson-Benson *et al* (1991) and Silagy (1993), controlled clinical trials (Vol. 8, No. 45, 1987) and in Barnard (2010) for a systematic review on this topic. Lai *et al* (2001) have proposed using theory of Brownian motion to model the stochastic variation of the recruitment process by assuming independent increment structure. They have shown that when we have relatively large number of patients, BM can be used as a continuous approximation of recruitment process asymptotically with normalization. Brownian motion approach is more efficient by accounting for cumulative recruitment at trial level across time than linear modeling and better than Poisson method by modeling variance independent of expected arrival. They also suggested that if the mean is close to the variance of the recruitment process, a Poisson model could be used. And if continuous approximation is reasonable, Brownian motion can be considered regardless of the relationship between the expected arrival and variance in the unit time. This tool has the property that it can be implemented easily for use in practice. In clinical trials, we may observe mixture of Poisson process or negative binomial distribution as pointed out by Willoford (1987), so the process may have dependent increment structure, the future recruitment will depend on the observed past history, for this case, fractional Brownian motion may be a good approximation to these processes. Anisimov (2006) and Anisimov and Fedorov (2007) proposed a Poisson–gamma recruitment model. According to their approach, a different Poisson rate for each center is defined as $\lambda_i$ with a gamma distribution, so the overall rate for *N* centers is $\Lambda(t)=\sum_{i=1}^{N}\lambda_i x(u_i<t)$, with the indicator variable to account for different initiation date of each center. Thus this process has dependent increments and in general the variance is greater than the mean. A center occupancy model, parameter estimation and prediction of remaining time are also discussed in the paper. The proposed method is verified from results of analyzing many clinical trial data. One of the advantages of the method is that it takes into consideration of potential different recruitment rates for different centers which usually do not have identical initiation date. Then at trial level the recruitment process is modeled as nonhomogeneous Poisson process with possibly random rate. In this paper we generalize the BM model to FBM model using example from Lai *et al* (2001) and discuss similarities and

differences between our model and Poisson and related models, and the Poisson-gamma model proposed by Anisimov and Fedorov (2007).

## Methodology

Let $X(t)$ be the number of patients on study since the start of the clinical trial until time $t$. From which we can derive $X(0)=0$. Let $f(t)$ denote the expected value of $X(t)$ at time $t$, we assume $B(t)$, a one dimensional Brownian motion, is a continuous approximation for $X(t) - f(t)$. For $t_2 > t_1$, the increment $B(t_2)-B(t_1)$ has the following properties.

- **i.** $B(t_1)$ and $B(t_2)-B(t_1)$ are normal and independent;
- **ii.** $E(B(t_1))=0$ and $E(B(t_2)-B(t_1))=0$;
- **iii.** $\text{Var}(B(t_1))= \sigma^2 t_1$ and $\text{Var}(B(t_2)-B(t_1))= \sigma^2(t_2 - t_1)$.

Brownian motion is a special case of Fractional Brownian motion with Hurst parameter, $H=1/2$. Whereas for fractional Brownian motion, $B_H(t)$ with increment $B_H(t_2) - B_H(t_1)$ has the following properties:

(i) $B_H(t_2) - B_H(t_1)$ is normal; (ii) $E(B_H(t_2) - B_H(t_1))=0$; and (iii) $E(B_H(t_2) - B_H(t_1))^2 = \sigma^2 c|t_2 - t_1|^{2H}$,

where $c = (1/2H[\Gamma(H + 1/2)]^2\})^{2H}$ is the constant of proportionality, and $H$ is the Hurst constant with range $0<H<1$ (Mandelbrot & Van Ness, 1968). The covariance between $B_H(t_2)$ and $B_H(t_1)$ is $Cov(B_H(t_2), B_H(t_1))=\sigma^2(c/2)(t_2^{2H}+t_1^{2H} - |t_2 - t_1|^{2H})$. So the conditional probability $P(X(t_k) > a \mid X(t_n) = r(t_n), \ldots, X(t_1) = r(t_1))$, for $k>n$ and $t_1 \le t_2 \ldots \le t_n \le t_k$ under fractional Brownian motion is derived as follows:

$$\begin{aligned} &P(X(t_k)>a|X(t_n)=r(t_n),\ldots,X(t_1)=r(t_1)) \\ &\quad =P(X(t_k) - f(t_k)>a - f(t_k)|X(t_n)=r(t_n),\ldots,X(t_1)=r(t_1)) \\ &\quad =P(B(t_k)>a - f(t_k)|X(t_n)=r(t_n),\ldots,X(t_1)=r(t_1)) \\ &\quad =1 - \Phi\left(\frac{a - f(t_k) - \mu_{H_n}}{\sigma_{H_n}}\right) \end{aligned}$$

where $\mu_{H_n}=\sigma^2\Sigma_{12}\Sigma_{22}^{-1}\mu_n$, $\mu_n = (B_H(t_n), B_H(t_{n-1}), \ldots, B_H(t_1))$ and $\Phi(\cdot)$ is CDF for standard normal distribution. $\Sigma_{12}$ and $\Sigma_{21}$ are the covariance matrices between $B_H(t_k)$ and vector $(B_H(t_n), B_H(t_{n-1}), \ldots, B_H(t_1))$. The variance of the conditional distribution is $\sigma^2_{H_n}=\sigma^2(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$, $\Sigma_{11}$ is the variance covariance matrix of $B_H(t_k)$ and $\Sigma_{22}$ is the variance covariance matrix of $B_H(t_n), B_H(t_{n-1}), \ldots, B_H(t_1)$. We can use method of moments estimator to estimate the variance of the increments, $\text{Var}(B(t + 1) - B(t)) = \sigma^2 c|t + 1 - t|^{2H}$ as $\tilde{\sigma}^2=\sum_{t=1}^{25}(X(t+1) - f(t+1) - (X(t) - f(t)))^2/25$ and calculate $\sigma^2$ with estimated $c$ and $H$ according to above (Lai, et al, 2001). $c$ is the constant of proportionality described earlier and $f(t)$ is the expected growth curve with zero intercept. The 95% confidence interval $[a_{t_k}, b_{t_k}]$ of the future path $X(t_k)$ of the recruitment process given existing observations $X(t_n), X(t_{n-1}), \ldots, X(t_1)$ is:

$$a_{t_k}=\mu_{H_n}-1.96SD=f(t_k)+\mu_{H_n}-1.96\sigma_{H_n},$$

$$b_{t_k}=\mu_{H_n}+1.96SD=f(t_k)+\mu_{H_n}+1.96\sigma_{H_n}.$$

There are different methods to estimate the Hurst parameter, including the maximum likelihood estimation and rescaled adjusted range analysis (Mandelbrot and Wallis, 1969, Davies & Harte, 1987; Mesa & Poveda, 1993; Molz & Boman, 1993; Beran, 1994a, Lai, 2004). For this analysis, the perFit procedure in fArma package of R is used to calculate Hurst parameter. The function `perFit` estimates the Hurst exponent from the periodogram. In the finite variance case, the periodogram is an estimator of the spectral density of the time series. A series with long range dependence will show a spectral density with a lower law behavior in the frequency. Thus, we expect that a log-log plot of the periodogram versus frequency will display a straight line, and the slope can be computed as $1-2H$ (J. Geweke and S. Porter-Hudak 1983). As pointed out by Beran (1994a) and Lai (2004), for irregularly observed short time series, such as recruitment process for a typical clinical trial with a few hundreds of patients, exact and direct maximum likelihood methods can be used to estimate Hurst parameter. Simulation results in Lai (2004) showed that with sample sizes of 20 – 100, the theoretical formulas based on large sample arguments can be used when the number of observations is relatively small.

In clinical trials, the inter-correlation among patient recruitment, patients' survival, and loss to follow up is usually not verified before using Brownian motion method, while the extra parameter $H$ in fractional Brownian motion can measure this long term dependence. For example, if $H > 1/2$ the correlation is positive, if $H = 1/2$, then the fractional Brownian motion becomes the classical Brownian motion, whereas $H < 1/2$ implies negative correlation. So, we suggest estimate Hurst parameter when modeling recruitment process and it is illustrated in the following examples.

## Results

We will first model the recruitment process of the CARE study (Sacks, 1995) using both Brownian and fractional Brownian motion methods. As described previously by Lai *et al* (2001), the data from first 26 weeks can be fitted using least square regression without intercept, the model is $f(t) = 39.54t$. The estimated standard deviation for Brownian motion is 13.94 using the method of moments and 14.42 under fractional Brownian motion. Hurst parameter ($H_n$) is 0.5713 using the above mentioned procedure and fArma package in R for the 26 weeks data from CARE study. So, based on these parameters, the conditional probability of having 3800 patients at week 100 is:

$P(X(t_2)>a|X(t_1))=1-\Phi(\frac{a-f(t_1)-\mu_{H_n}}{\sigma_{H_n}})=1-\Phi(\frac{3800-f(100)-4}{162})=0.835$, $a$ is the recruitment goal at week 100, $t_1$ $t_2$ represent weeks 26 and 100 for this example. However, under Brownian motion it is 0.91, of course in this case we have overlooked or underestimated the Hurst parameter. To better understand the influence of Hurst parameter,

we calculated probabilities for having 4000 patients in the study from 90 to 100 weeks given 1035 patients at week 26, Figure 1 shows the results. We can see that when $H$ increases from 0.5 to 0.7, the conditional power also increases for achieving the recruitment goal of 4000. For the observed $H$=0.5713, the conditional probability is about 39.8% at 100 weeks, which is slightly higher than that based on Brownian motion (37%). To further explore this relationship, we calculated conditional probabilities of reaching different recruitment goals using different Hurst parameters, Figure 2. shows that for $H$<0.5, the probabilities are greater than that of $H$=0.5 for recruitment goal about less than 4000, but decrease quickly to less than that of $H$=0.5 when the goal is greater than 4000. On the other hand, for $H$>0.5, the probabilities are lower and then become greater than that of $H$=0.5 when recruitment goal is higher than about 4000. This implies the importance of estimating Hurst parameter and considering dependent increments structure when predicting patients' accrual. We also plotted original trajectory, linear prediction along with 95%CIs based on Brownian motion and Fractional Brownian motions. We can see from Figure 3. 95%CIs using various values of $H$ can lead to different coverage probabilities, from very narrow ($H$=0.2) to very wide confidence intervals ($H$=0.8). These imply again the importance of estimating Hurst parameter and modeling potential dependent increment structure of the recruitment process.

In addition, we have modeled recruitment data from six oncology trials conducted by Radiation Therapy Oncology Group (RTOG). These examples include Head and Neck, Lung, Prostate, and Breast cancers. Since typically during the first six month, study sites will obtain IRB approval and the recruitment is minimal, so these data are not included in these models. For these analyses we have used data up to one third of the sample size to model and predict the probability of reaching recruitment goals at the study end. As we can see from Table 1 that sample sizes range from 377 to 626, 1535, 1997 which are typical sizes for phase III oncology trials. The number of centers range from 84 to 216. In Figure 4 we plotted the trajectories across time for each trial. For two of the trials, Hurst parameters have value of less than 0.5, conditional probabilities are lower under BM model than those under FBM model and $H$ is greater than 0.5 for the rest of four trials, conditional probabilities are higher under BM than those under FBM except the case of trial No. 3. This is similar to scenario (e) in Figure 2, where the analysis has not reached turning point of the curve. In general, the higher the conditional probability is the narrower the 95%CI except for the case of trial No. 3. If we model recruitment in the middle of the process, for trial No. 3, the conditional probability becomes 0.999 and 95%CI: 1925–2350 under BM and the probability is 0.9 and 95%CI: 2048–2193 under FBM. The trend is now similar to other trials.

## Discussion

In long term studies such as CARE trial, the recruitment process can be influenced by many factors, thus the independent incremental structure assumption for standard Brownian motion may be violated. When this happens, we can quantify the dependent increment structure using the Hurst parameter. So the current observation will not only depend on the previous data but the entire past history. For these situations Brownian motion can be generalized to fractional Brownian motion to model recruitment process. Lee's contagious Poisson model and Williford's Bayesian model provides the advantages of using all data up

to date too. But the model assumptions need to be verified as to if there is over-dispersion in the data. The Poisson-gamma model proposed by Anisimov and Fedorov (2007) includes center specific recruitment rates and starting dates, so it captures the non-homogeneous structure of the process. Also, in Anisimov and Fedorov (2006), under the Poisson-gamma model, the authors proved asymptotic approximation of the recruitment process by a combined linear process with normal slope and Brownian motion, of the form $AF(t) = a * X *t + b *W(t)$, where $X$ is a Gaussian random variable that is independent of the Brownian motion $W(t)$. This process has dependent increment and is not a fractional Brownian motion. It's second moment is $E((AF(t_2) - AF(t_1))^2 = A* (t_2 - t_1)^2 + B* (t_2 - t_1)$. For a fractional Brownian motion, *FBM(t)*, we have $E(FBM(t_2) - FBM(t_1))^2 = C* (t_2 - t_1)^{2H}$. This can be viewed as an approximation to $A* (t_2 - t_1)^2 + B* (t_2 - t_1)$ for a proper value of $H$. Therefore, the FBM model can be used as an approximation to *AF(t)* process when the number of patients in the trial is large enough. The difference between their model and the FBM model is that for our methods the mean recruitment is modeled separately, but the BM process is included in their model as a covariate. This would lead to improved model fit and reflect fluctuations around the linear trend of recruitment process. Also for the FBM method considered here, the number of currently active clinical centers is not included in the model. So it can be improved and applications can be extended to cases when we have different number of active centers at various analysis times. Simulations can be carried out to show empirical evidence of fit for each method for different sample sizes and number of centers. Alternatively, a FBM may be used to combine with the linear process instead of BM and see if there is any improvement of model fit to all scenarios. Lai *et al* (2001) compared differences between Brownian motion method to that of linear and Poisson approach and they showed that the BM model did not underestimate the actual variation in the process. Their results also indicated that when the number of patients is relatively small, such as at center levels, we may use Poisson process to model the recruitment, However, for clinical trials with larger number of patients, Brownian motion may be more suitable for proper time intervals (i.e. weekly). From the 95%CI, we can see that the confidence intervals based on fractional Brownian motion is different from that of Brownian motion. Comparing to the BM method, the proposed method takes into account of correlation of accumulated data across time, possible dependent increment structure and separated estimation of mean and variance. As illustrated in the six oncology trials, only one trial has a Hurst parameter value close to 0.5. So, in practice, we suggest check the distribution assumptions and select the FBM method if the assumptions for BM do not hold given FBM/BM models are good approximations for the recruitment trajectory. Again, model fit for FBM/BM, Poisson model and Poisson-gamma model in clinical trials with relatively small, medium and large sample sizes can also be compared through simulations. These results will help us to identify situations when all models perform well and for cases some model will have better fit to the data. Finally BM and FBM methods are tools that could be used together with Poisson and Poisson-gamma models for better monitoring the recruitment process in clinical trials with various numbers of patients and centers.

## Acknowledgments

Project supported by RTOG grant U10 CA21661 and CCOP grant U10 CA37422 from the NCI.

## References


Anisimov, V.; Fedorov, VV. Design of multicentre clinical trials with random enrolment, in book "Advances in Statistical Methods for the Health Sciences.". In: Balakrishnan, N.; Auget, J-L.; Mesbah, M.; Molenberghs, G., editors. Series: Statistics for Industry and Technology. Vol. Ch. 25. Birkhäuser; 2006. p. 387-400.

Anisimov V, Fedorov VV. Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. Statistics in Medicine. 2007; 26:4958–4975. [PubMed: 17639505]

Anisimov V. Recruitment modeling and predicting in clinical trials. Pharmaceutical Outsourcing. 2009; 10(1):44–48.

Barnard KD, Dent L, Cook A. A systematic review of models to predict recruitment to multicentre clinical trials. BMC Medical Research Methodology. 2010; 10:63. [PubMed: 20604946]

Beran, J. Statistics for Long-Memory Processes. New York: Chapman and Hill; 1994a.

Bjornson-Benson WM, Stibolt TB, Manske KA, Zavela KJ, Youtsey DJ, Buist AS. Monitoring recruitment effectiveness and cost in a clinical trial. Controlled Clinical Trials. 1991; 14(Suppl): 52S–67S. [PubMed: 8500313]

Carter RE. Application of stochastic processes to participant recruitment in clinical trials. Controlled Clinical Trials. 2004; 25:429–436. [PubMed: 15465613]

Davies RB, Harte DS. Tests for Hurst effect. Biometrika. 1987; 74:95–101.

Geweke J, Porter-Hudak S. The estimation and application of long memory time series models. J. Time Ser. Anal. 1983; 4:221–238.

Lai D, Moye LA, Davis BR, Brown LE, Sacks FM. Brownian Motion and Long-Term Clinical Trial Recruitment. Journal of Statistical Planning and Inference. 2001; 93:2239–246.

Lai D. Estimating the Hurst effect and its application in monitoring clinical trials. Computational Statistics & Data Analysis. 2004; 45:549–562.

Lee YJ. Interim recruitment goals in clinical trials. J. Chronic Diseases. 1983; 36:379–389. [PubMed: 6853664]

Mesa OJ, Poveda G. The Hurst effect: The scale of fluctuation approach. Water Resources Research. 1993; 29:3995–4002.

Mendelbrot BB, Van Ness JW. Fractional Brownian motions, fractional noise and applications. SIAM Review. 1968; 10:422–437.

Mendelbrot BB, Wallis JR. Robustness of the rescaled range R/S in the measurements of non-cyclic long-run statistical dependence. Water Resources Research. 1969; 5: 967–988.

Molz FJ, Boman GK. A fractal-based stochastic interpolation scheme in subsurface hydrology. Water Resources Research. 1993; 20:3769–3774.

Sacks FM, Rouleau JL, Moye LA, Pfeffer MA, Warnica W, Arnold M, Nash D, Corder C, Brown LE, Davis D, Hawkins CM, Braunwald E, for the CARE Investigators. Baseline characteristics in the Cholesterol and Recurrent Events (CARE) trial of secondary prevention in patients with average serum cholesterol level. Amer. J. Cardiol. 1995; 75:621–623. [PubMed: 7887392]

Silagy CA, Campion K, McNeil JJ, Worsan B, Donnan GA, Tonkin AM. Comparison of recruitment strategies for a large-scale clinical trial in the elderly. J. Clinical Epidemiol. 1993; 44:1105–1114. [PubMed: 1941003]

Williford WO, Bingham SF, Weiss DG, Collins JF, Rains CK, Krol WF. The constant intake rate" assumption in interim recruitment goal methodology for multicenter clinical trials. J. Clinical Epidemiol. 1987; 40:297–307.

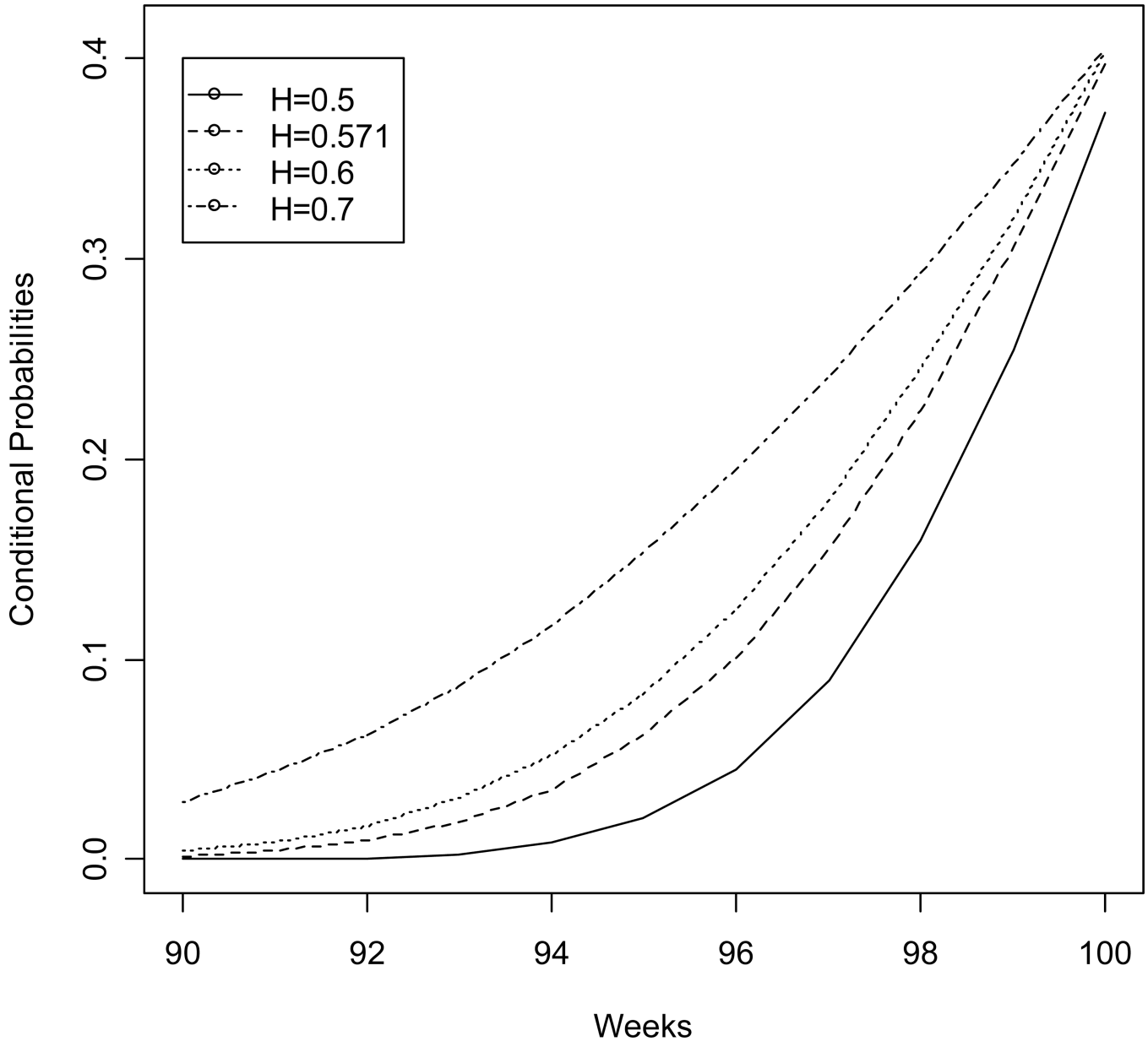


**Figure 1.**
Conditional probabilities of having 4000 patients in the study from 90 to 100 weeks based on FBM with different Hurst parameters.

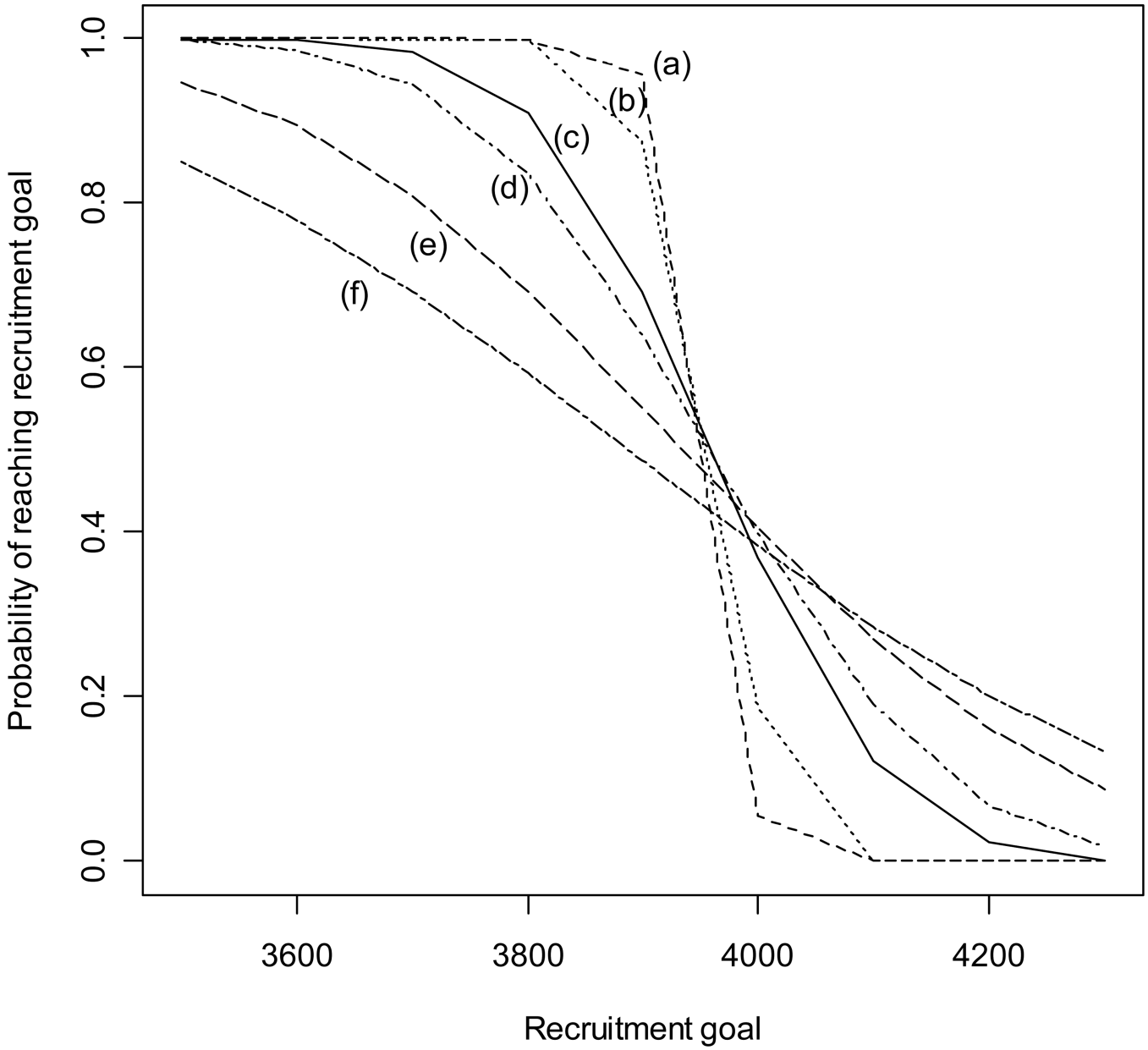


**Figure 2.**
Conditional probabilities of reaching recruitment goal based on different Hurst parameters: (a) $H$=0.2, (b) $H$=0.3, (c) $H$=0.5, (d) $H$=0.5713, (e) $H$=0.7, (f) $H$=0.8.

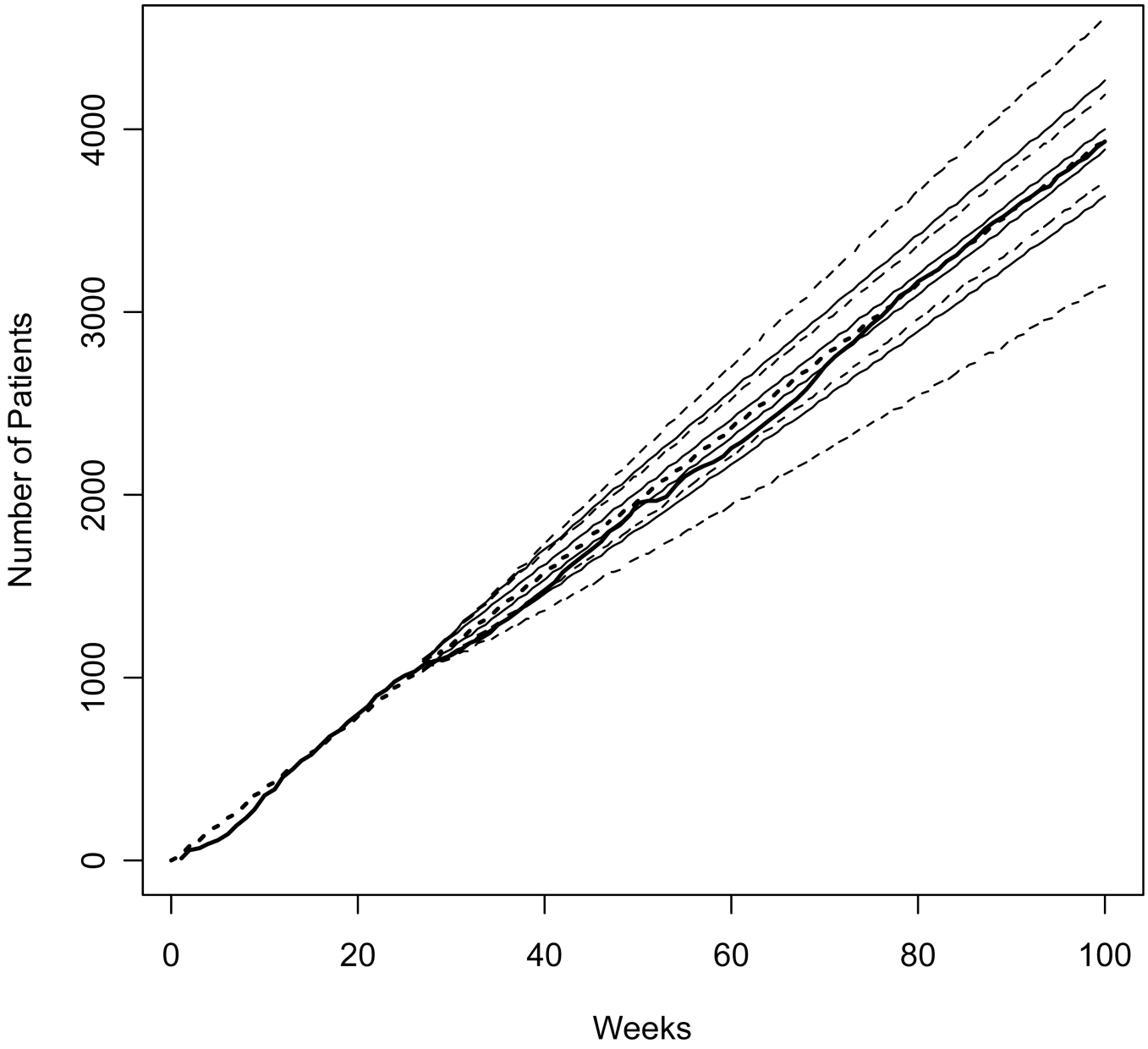


**Figure 3.**
Observed (Solid) and expected (Dotted) paths of CARE study with 95%CIs based on different Hurst parameters: (starting from inside) H=0.2, 0.5, 0.5713 and 0.8.

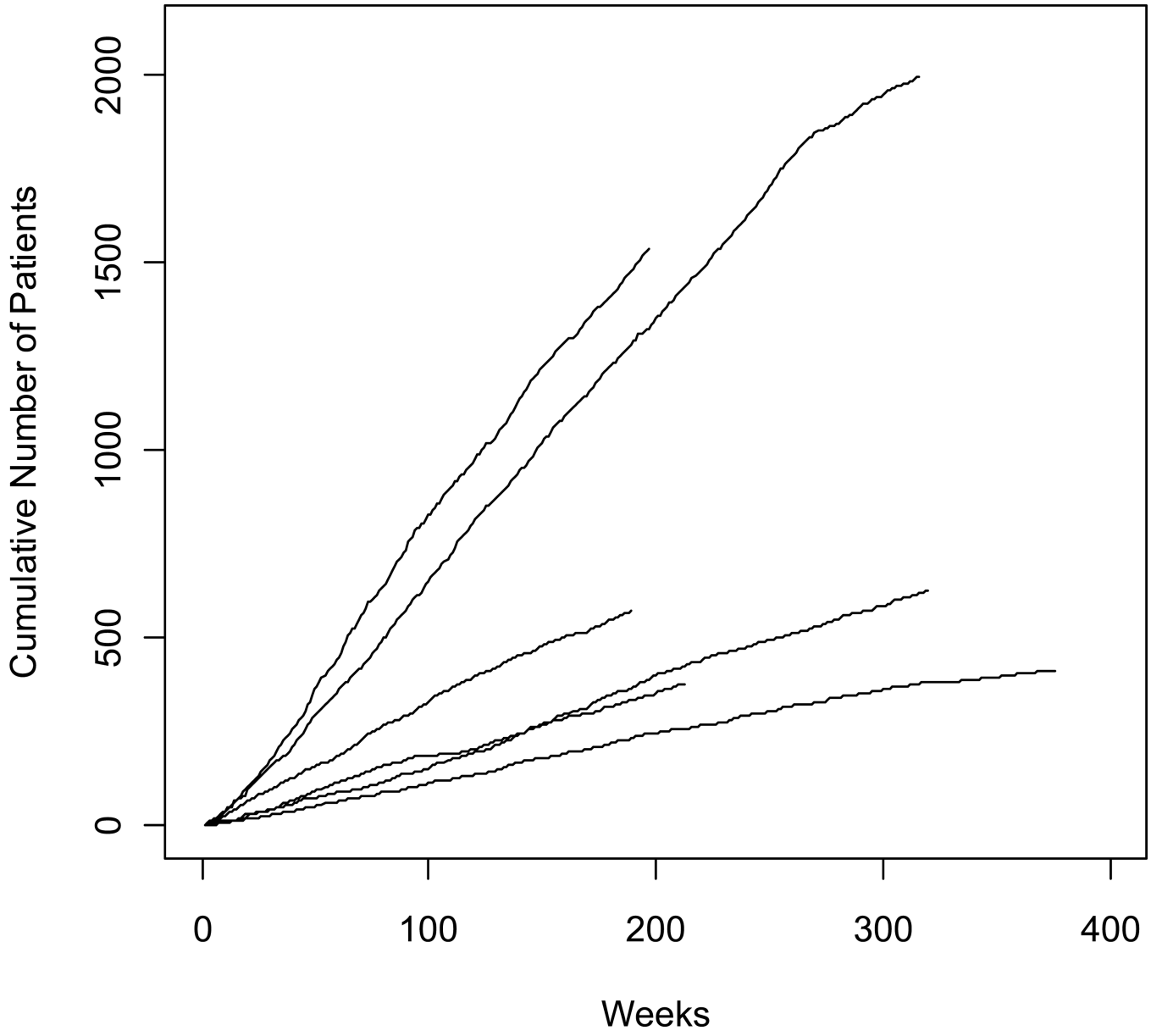


**Figure 4.**
Number of patients by weeks for each of the six example trials.

**Table 1**

Hurst parameters and conditional probabilities for different trials.

| Trials | N | Centers | H | P (BM) | 95%CI | P (FBM) | 95%CI |
|---|---|---|---|---|---|---|---|
| 1 | 416 | 149 | 0.41 | 0.65 | (389–457) | 0.72 | (401–443) |
| 2 | 1997 | 216 | 0.70 | 0.49 | (1914–2078) | 0.60 | (1805–2249) |
| 3 | 573 | 153 | 0.53 | 0.90 | (560–633) | 0.86 | (554–638) |
| 4 | 626 | 194 | 0.35 | 0.70 | (598–674) | 0.83 | (618–650) |
| 5 | 377 | 84 | 0.63 | 0.94 | (370–439) | 0.84 | (345–473) |
| 6 | 1535 | 152 | 0.70 | 0.89 | (1350–1524) | 0.85 | (1282–1712) |