Better planning through Design

Forecasting Enrollment in Clinical Trials When Site-Level Accrual Rates Vary with Time

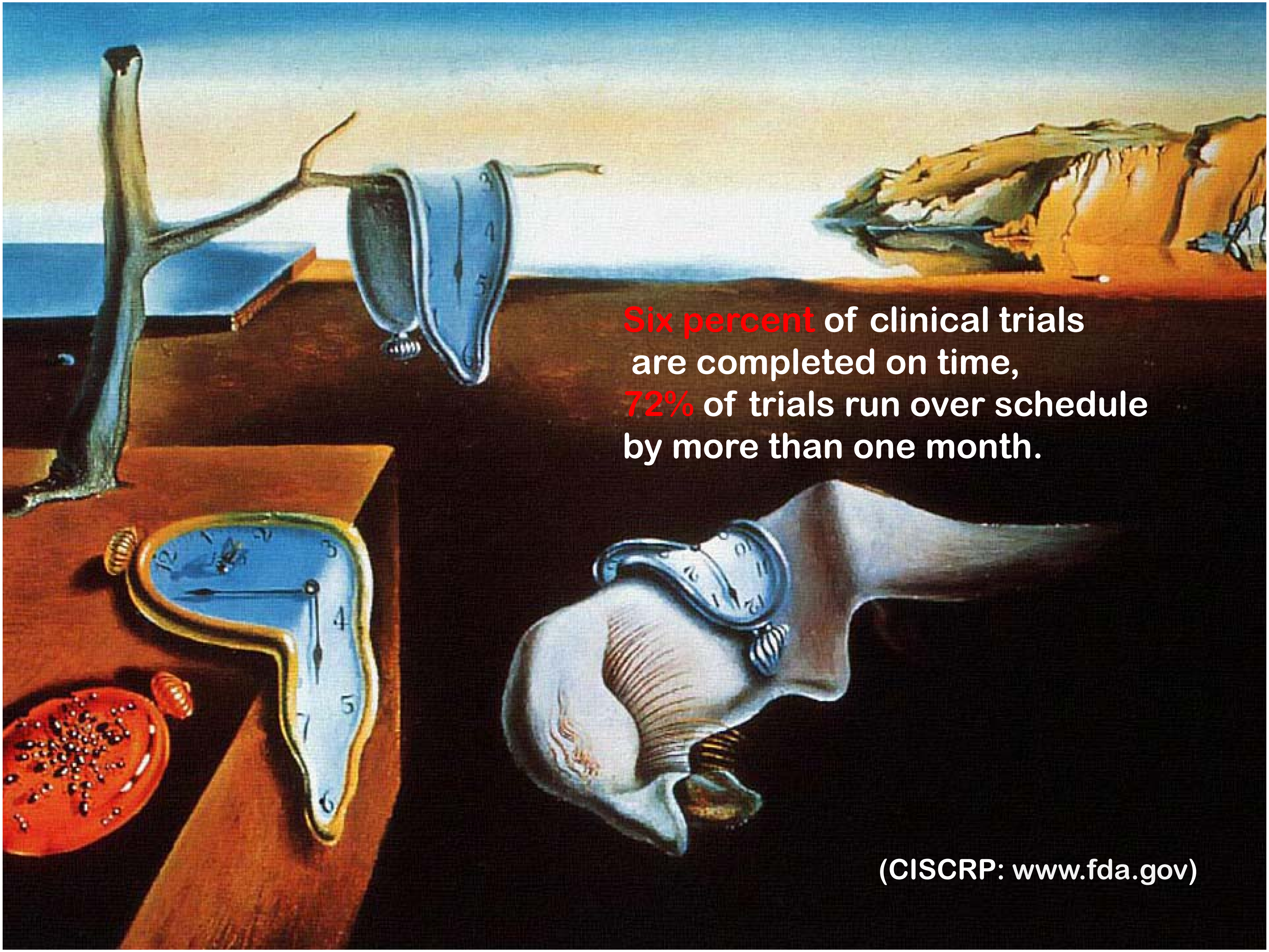
30 July 2012

Nitin Patel, Pralay Senchaudhuri, Suresh Ankolekar
Presented by Yannis Jemiai

Outline

- Motivation and background
- Poisson-Gamma: a reasonable model to forecast enrolment
- Improving predictions through stratification
- Ignoring the first inter-arrival time
- Time-variant enrolment rates
- Conclusions

Motivation and Background


Six percent of clinical trials
are completed on time,
72% of trials run over schedule
by more than one month.
(CISCRP: www.fda.gov)

Other facts and figures

- Eighty percent of total trials are delayed at least one month because of unfulfilled enrollment. (Lamberti, "State of Clinical Trials Industry", 292)

- Out of all of the research sites in the United States, less than a 1/3 contain 70% of the valuable subjects. Therefore 70% of the research sites under-perform, and somewhere between 15%-20% never even enroll a single patient. (Pierre, "Recruitment and Retention". 2006)

- Fifty percent of clinical research sites enroll one or no patients in their studies. (Pierre, "Recruitment and Retention". 2006)

Why model enrolment?

- Significant resources and strategic planning are contingent upon the timing of interim and final data analyses
- Modeling recruitment and event accrual based on current accumulated data allows early and accurate predictions of interim analysis times and study termination
- Statistical modeling also provides confidence levels for predictions

Large pharma enrolment data for trials with more than 200 subjects

Seq#	Therapeutic Area	Study Phase	# Sites	# Subjects	# Countries
1	CVID&MET	Phase 3	118	993	30
2	INFL&IM	Phase 2	36	249	1
3			62	365	9
4			71	242	8
5	Neuroscience	Phase 2	56	227	10
6			61	342	11
7			49	237	11
8			32	314	2
9			27	204	1
10		Phase 3	25	474	1
11			43	485	10
12			21	638	1
13			65	608	5
14			36	387	1
15			57	713	1
16			46	412	1
17			44	245	1

Seq#	Therapeutic Area	Study Phase	# Sites	# Subjects	# Countries
18	Vaccines	Phase 3	38	667	1
19			56	606	1
20			9	286	1
21			38	613	1
22			15	500	1
23			4	354	1
24			9	269	3
25			9	355	1
26			79	1712	1
27			23	449	1
28			11	603	1
29			9	606	1
30			35	619	1
31			25	1241	1
32			61	1053	1
33			34	1116	1
34			2	270	1
35			39	1165	4
36			21	718	1
37	WH&BR	Phase 3	21	495	3
38			34	458	1
39			62	533	2

Trials conducted in more than 5 countries highlighted in yellow

Poisson-Gamma: a reasonable model to forecast enrolment

Predicting if a trial will complete on time

- Consider a multicenter trial that starts at time zero and is planned to enrol n subjects by time T
- Suppose that at time $t_0 < T$, K centers have been opened at times $a_1, a_2, \ldots, a_K$ ($\leq t_0$).
- Let n_{0i} be the number of subjects accrued in center i, $i=1, 2, \ldots, K$ at time t_0
- We would like to estimate the probability of completing the trial on time (i.e., by time T).

A Bayesian approach

- Based on Poisson-Gamma model (Anisimov & Fedorov, 2007)
- Assume that accruals at center i follow a Poisson process with rate μ_i.
- We also assume that μ_i has a Gamma prior with parameters (α_i,β_i), and accruals at centers are mutually independent.
- It is well-known that the posterior distribution of μ_i (computed at t_0) is a Gamma distribution with parameters ($n_{0i}+ \alpha_i$, $t_0 - a_i + \beta_i$).

Predictive probability of future accruals

- The predictive probability of n_{1i} – the number of accruals at center i during (t_0, T) is Negative Binomial with parameters

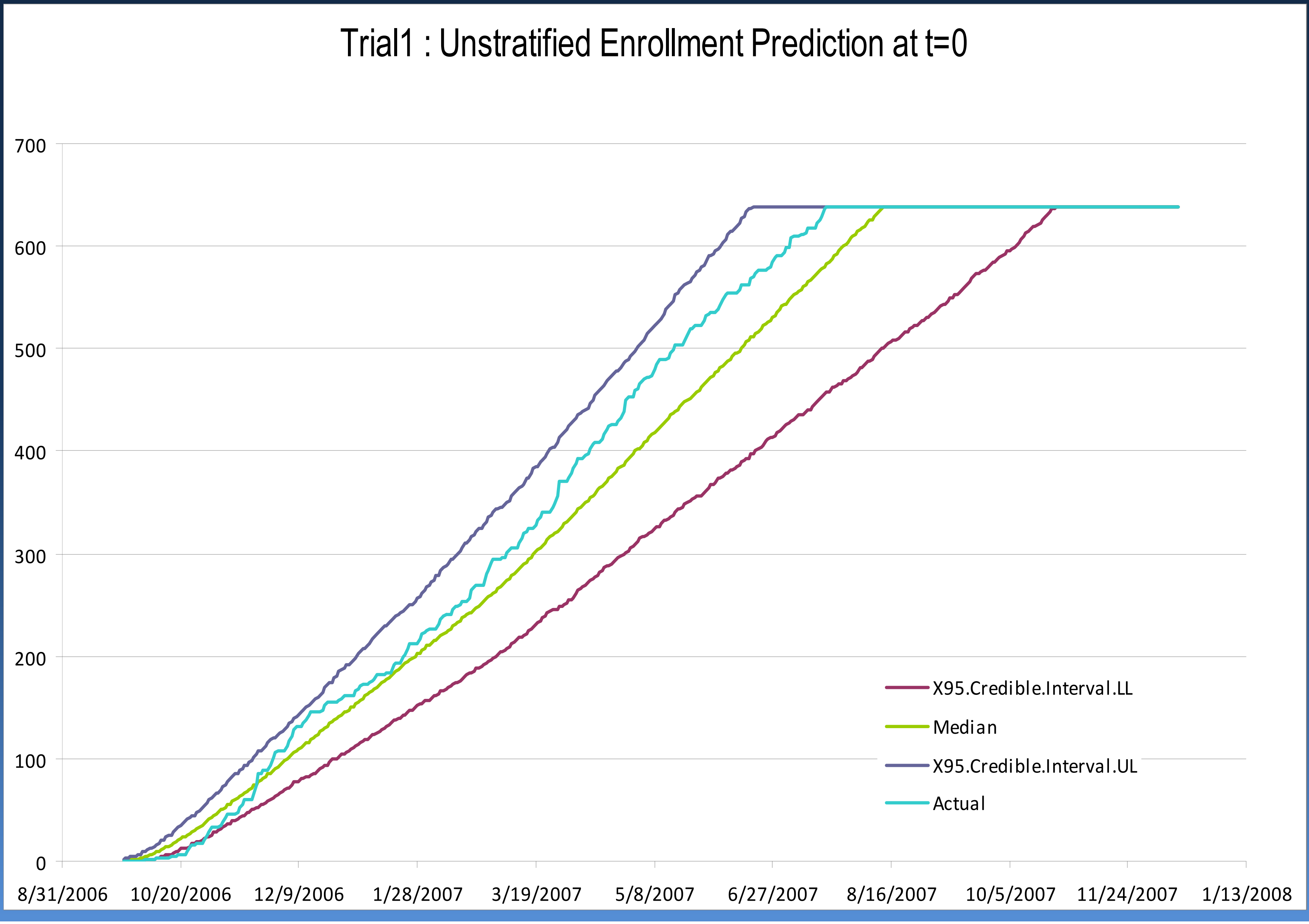
$$(n_{0i} + \alpha_i, (t_0 - a_i + \beta_i)/(T - a_i + \beta_i))$$

- Computing the probability of accruing n subjects by time T involves figuring out the convolution of K negative binomials
- In general no closed solution. It's easiest to obtain a Monte Carlo estimate through simulation.

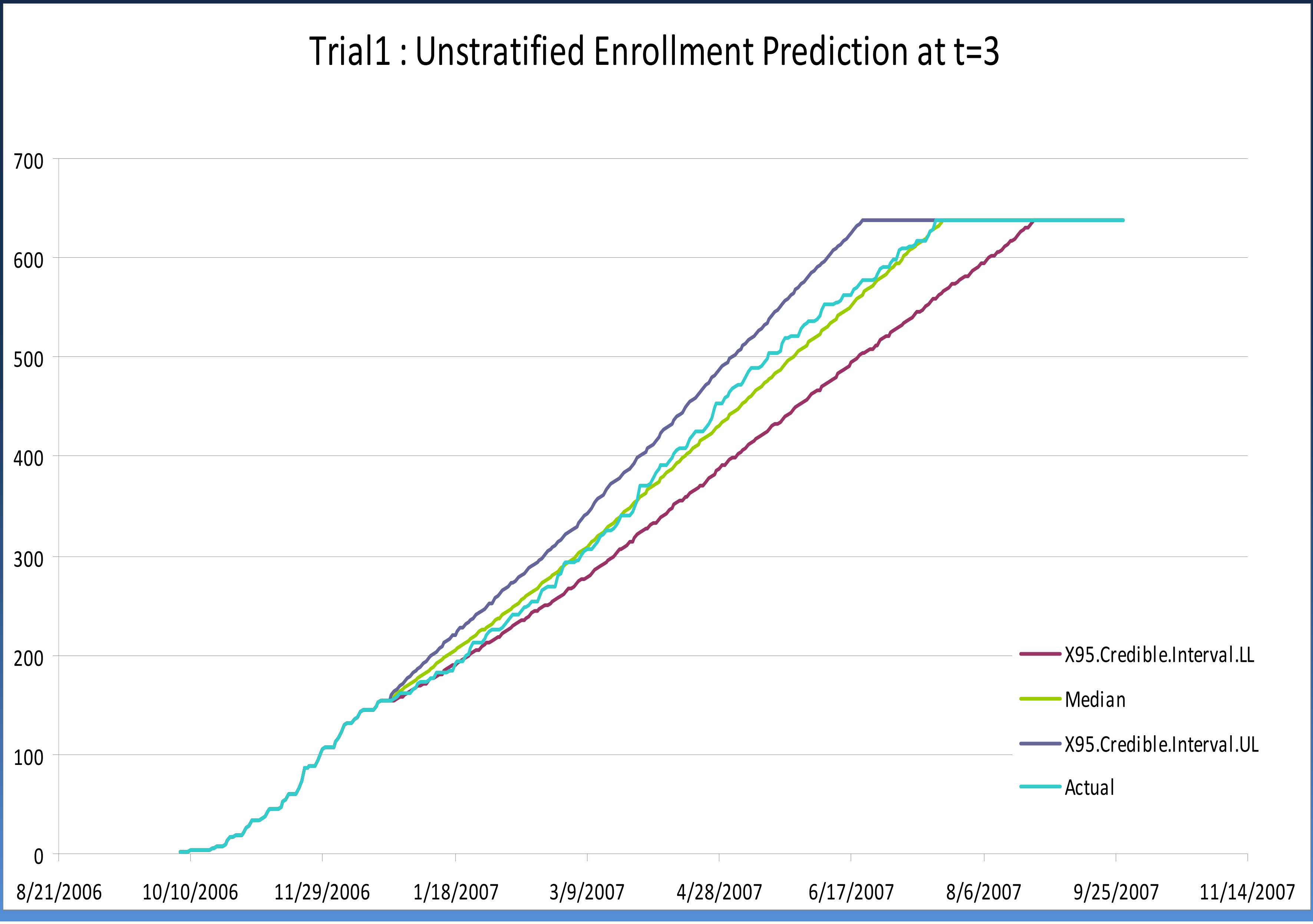
Simulation assumptions

- In the simplest version of the model, we assume that μ_i has a Gamma prior with parameters (α,β)
- We simulate 500 runs to obtain median and credible intervals for end time T
- To make comparisons between models fair, we take the truth as our prior, i.e. get α and β from the data
- Calculate mean and standard deviation of the enrolment rate from data and use to obtain α and β
- Site Initiation Visit (SIV) times are assumed to be known and not predicted

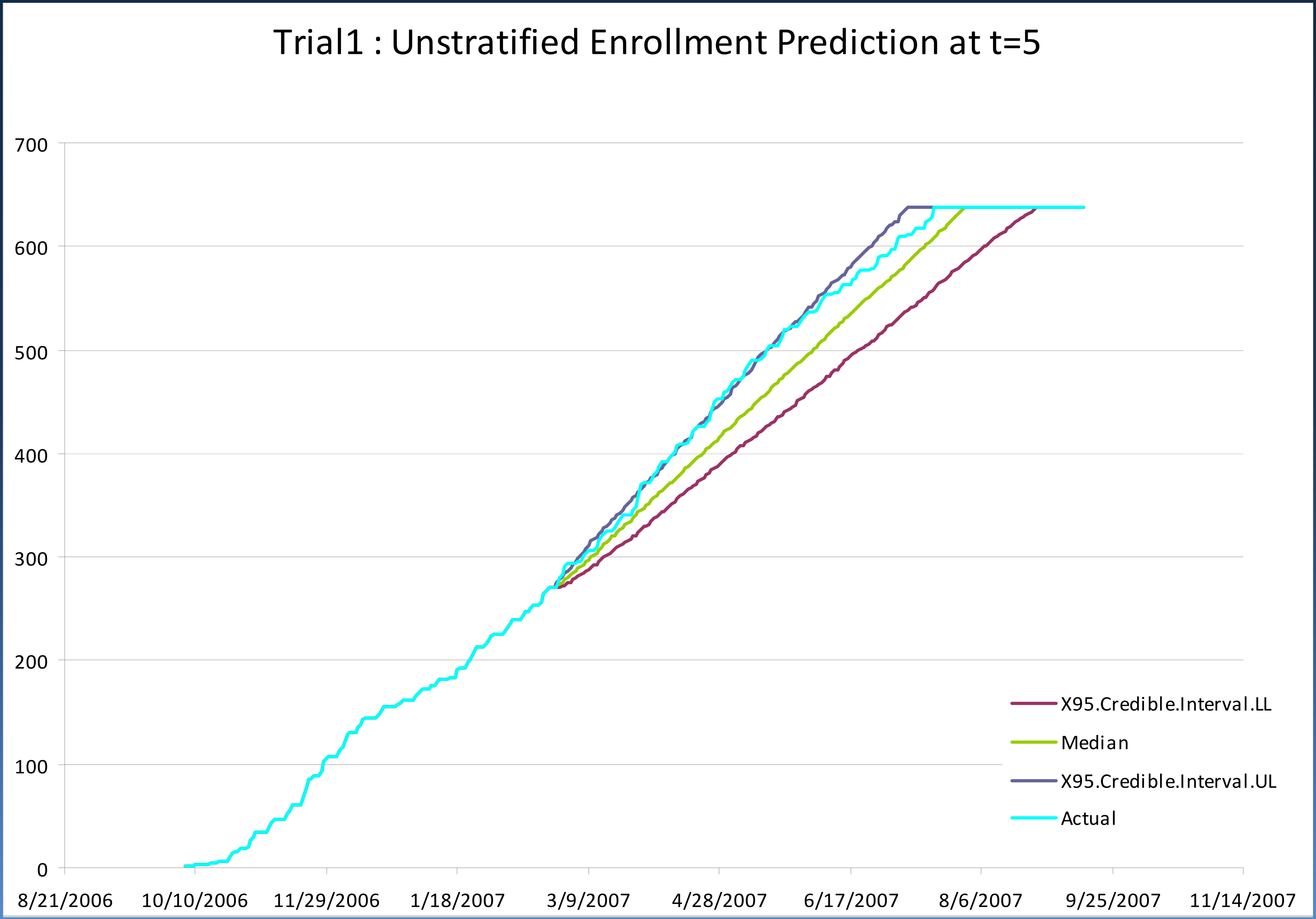
Prediction at time t_0 = 0 months



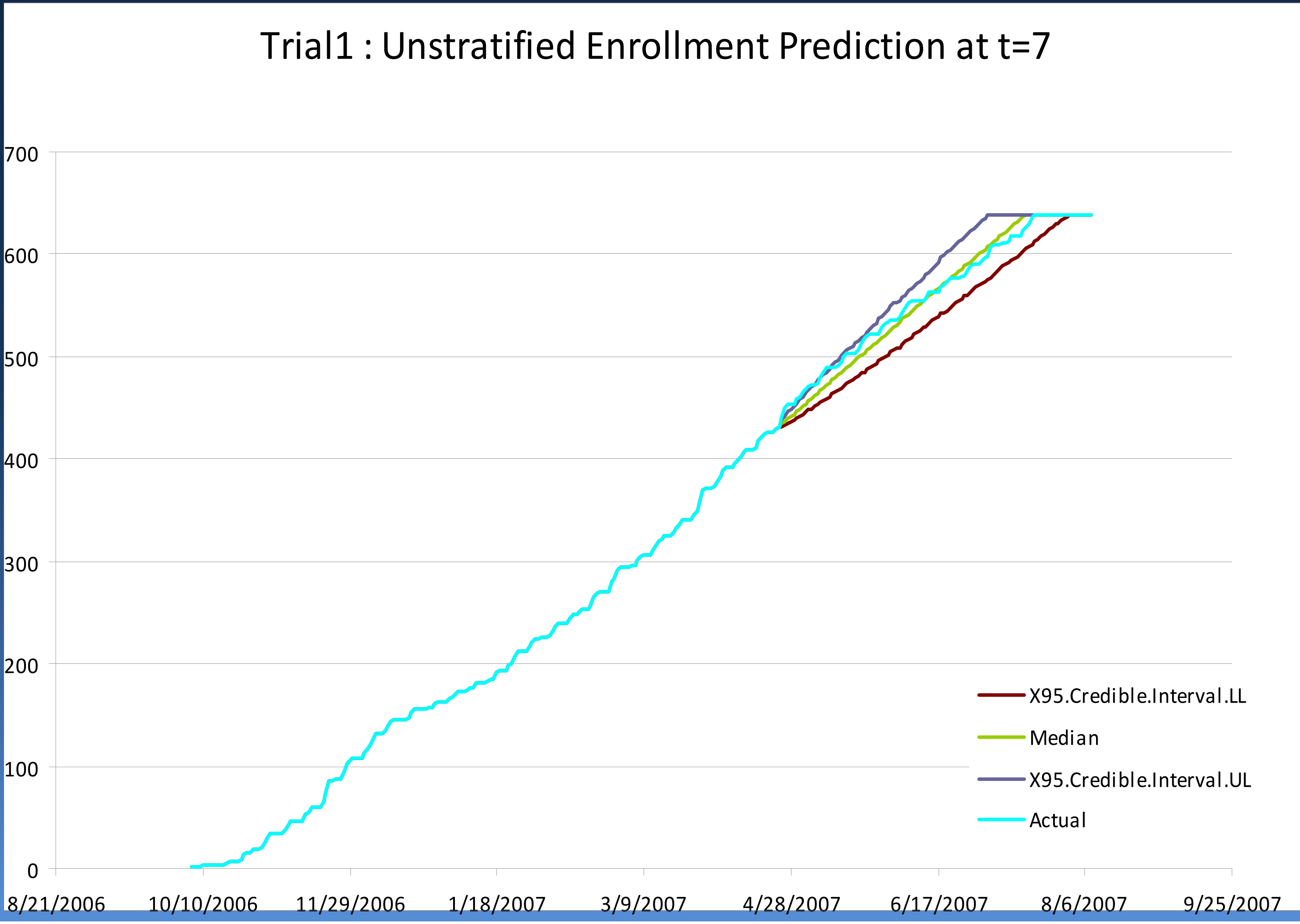
Prediction at time t_0 = 3 months



Prediction at time t_0 = 5 months



Prediction at time t_0 = 7 months

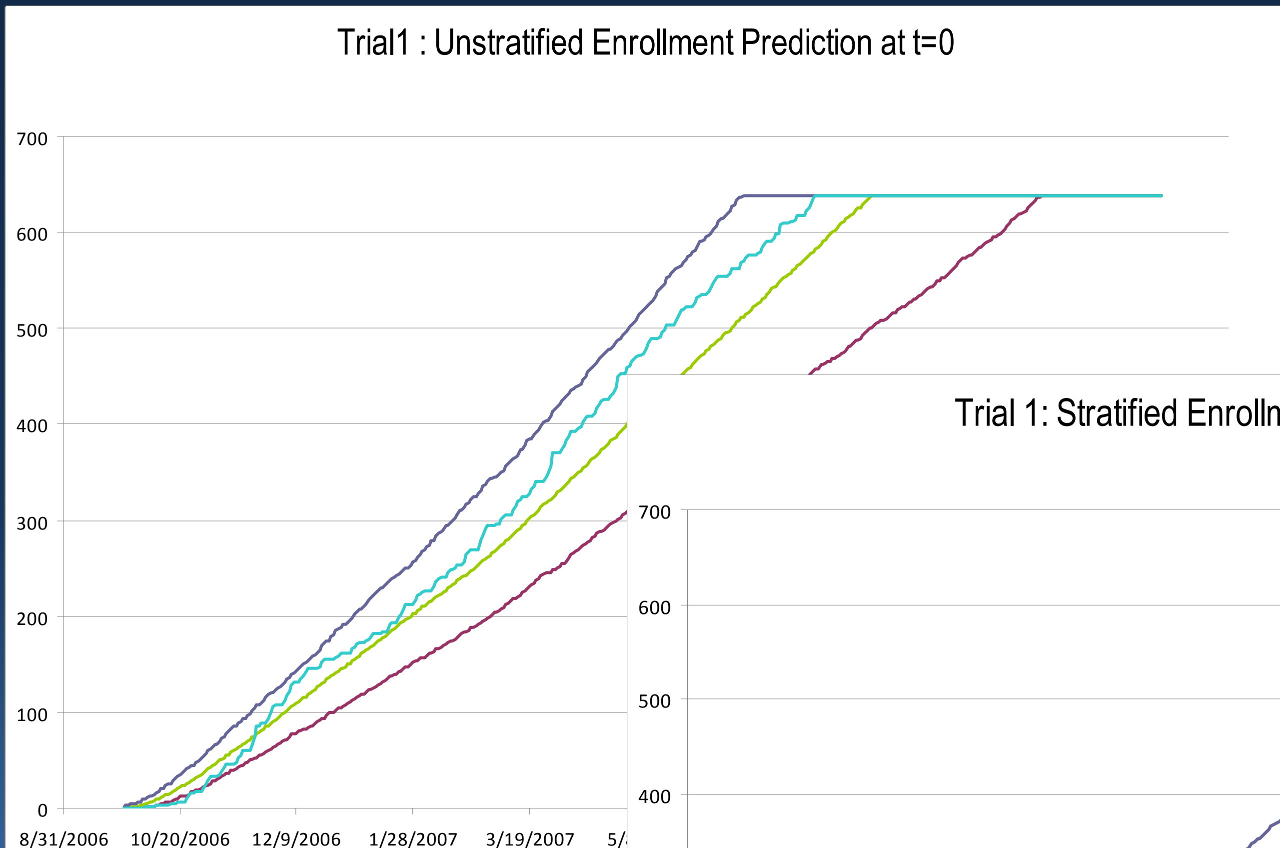


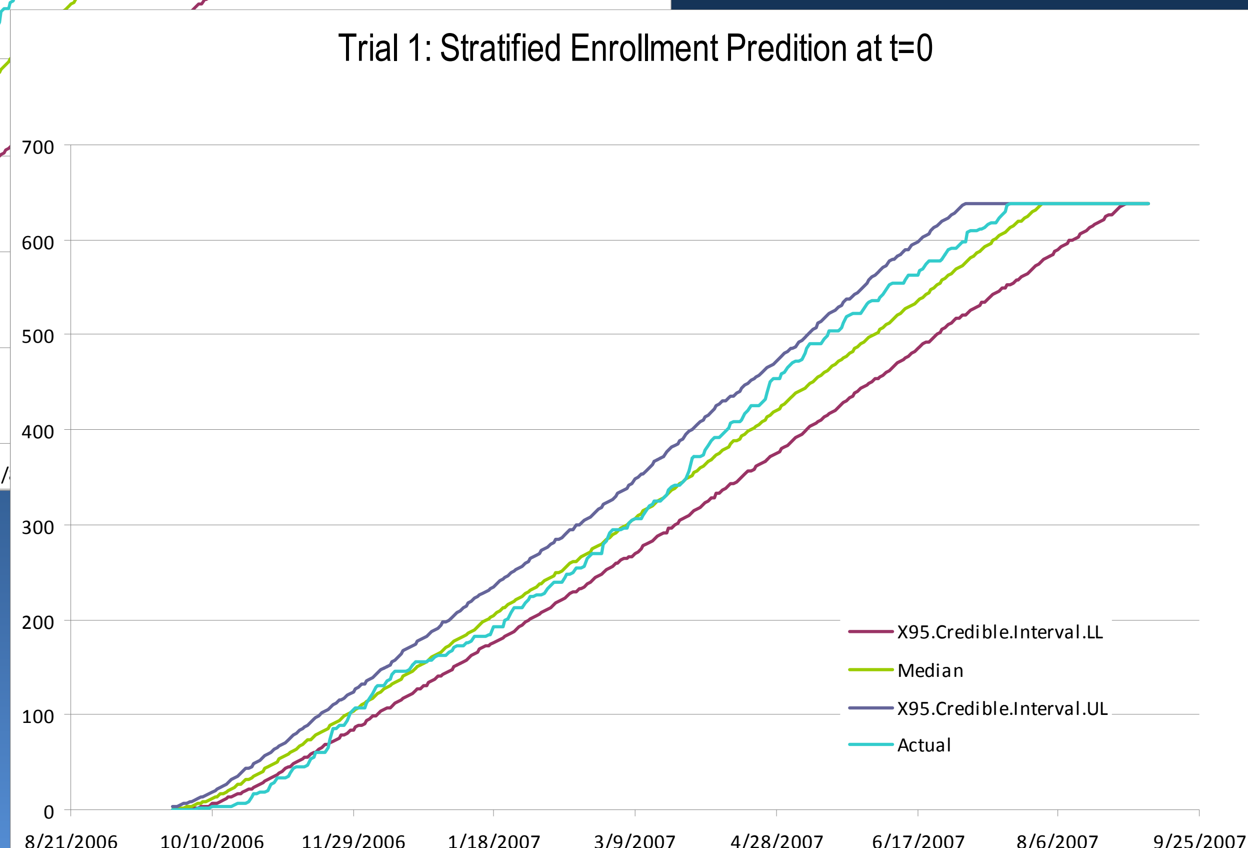
Improving predictions through stratification

Stratifying sites

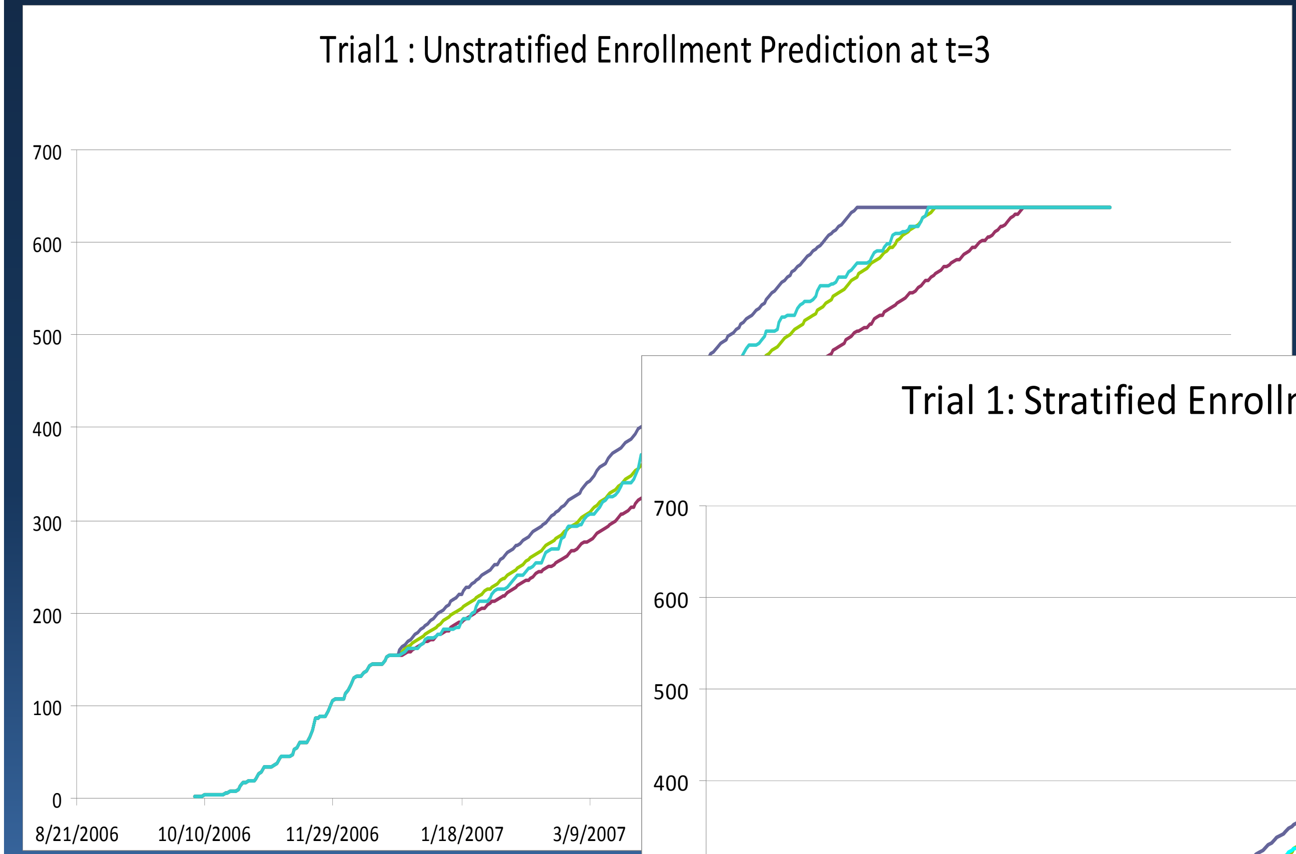
- Classify sites according to enrolment rates into high ($\mu_i > 0.4$), medium ($0.2 < \mu_i < 0.4$), and low ($\mu_i < 0.2$) enrolling categories
- Put different Gamma priors for sites belonging to each of the three categories and combine predictions
- Simulations presented were performed for single country trials

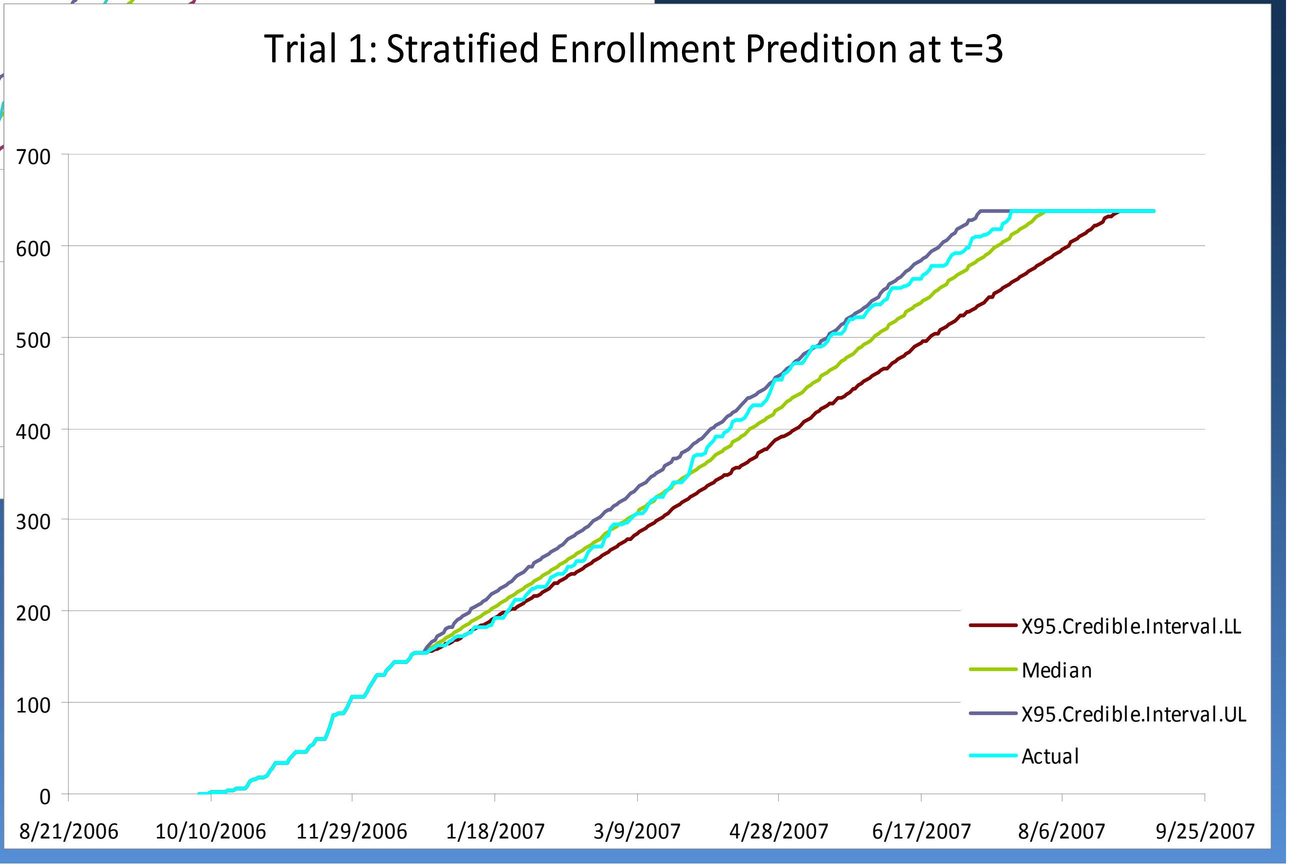
Prediction at time t_0 = 0 months



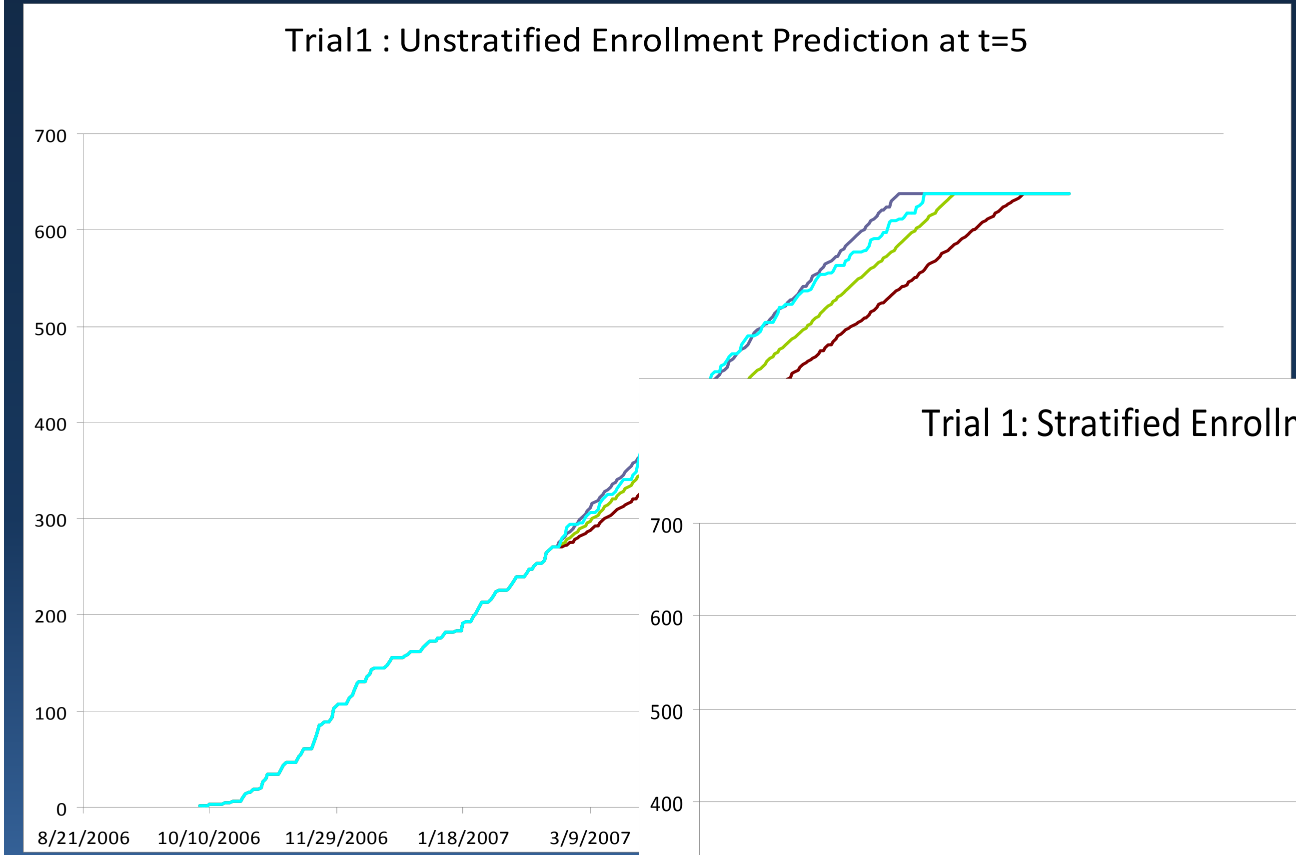


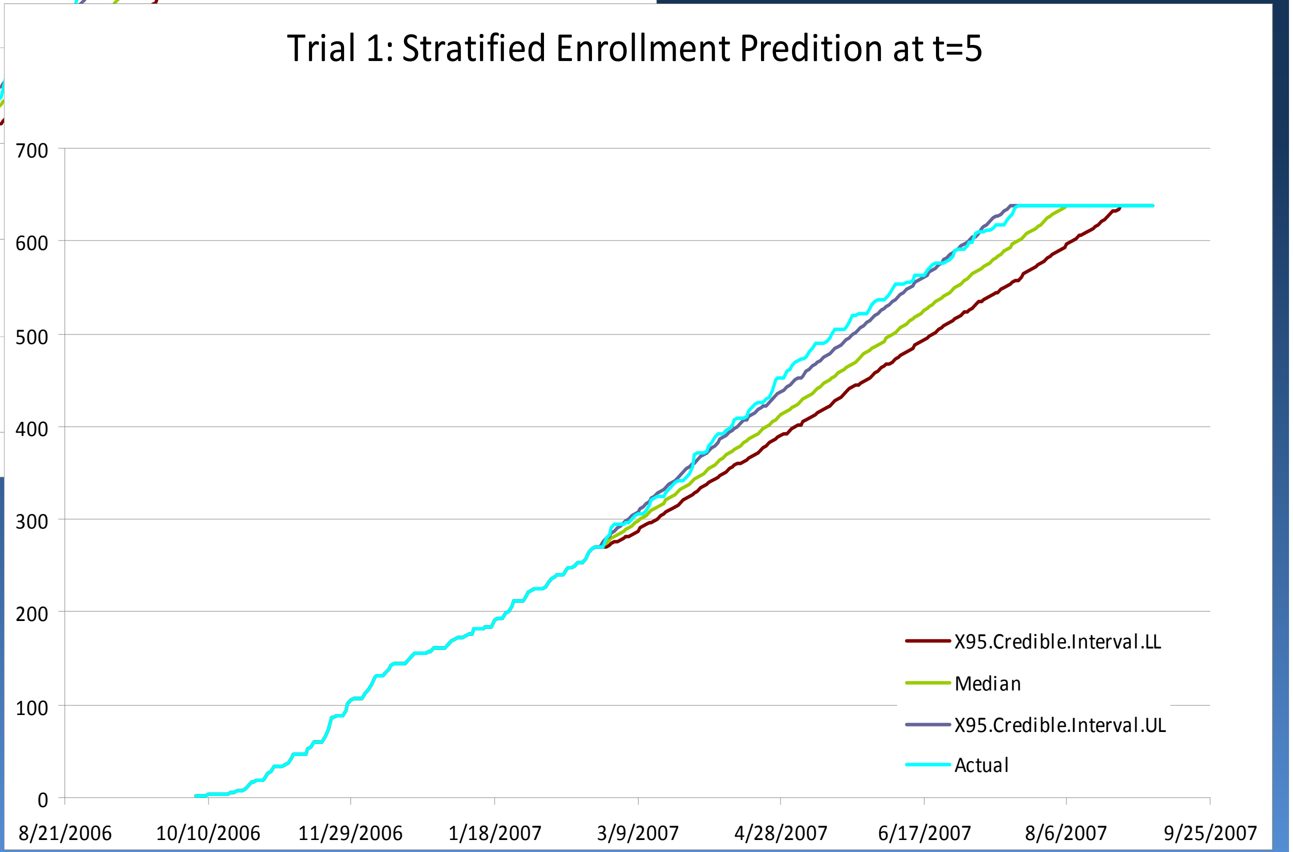
Prediction at time t_0 = 3 months



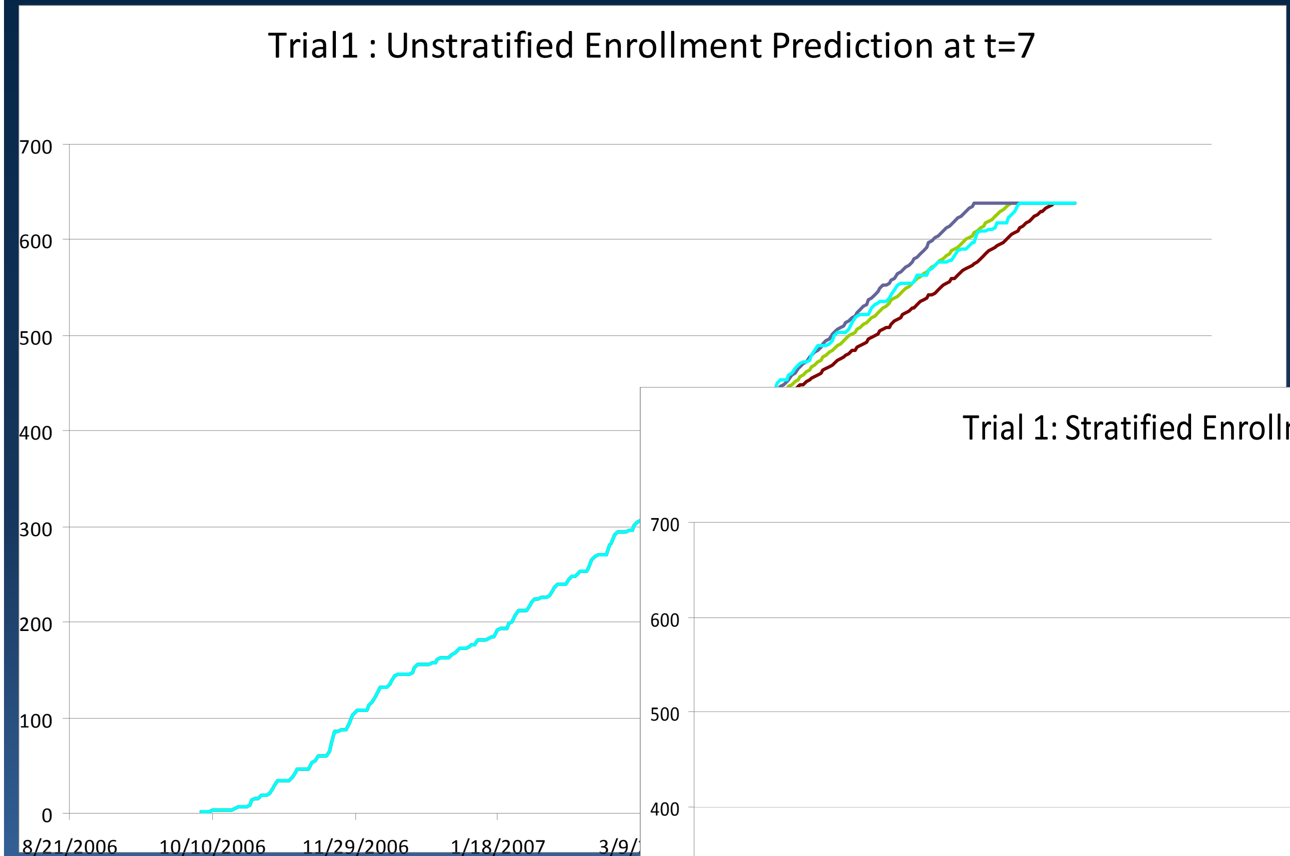


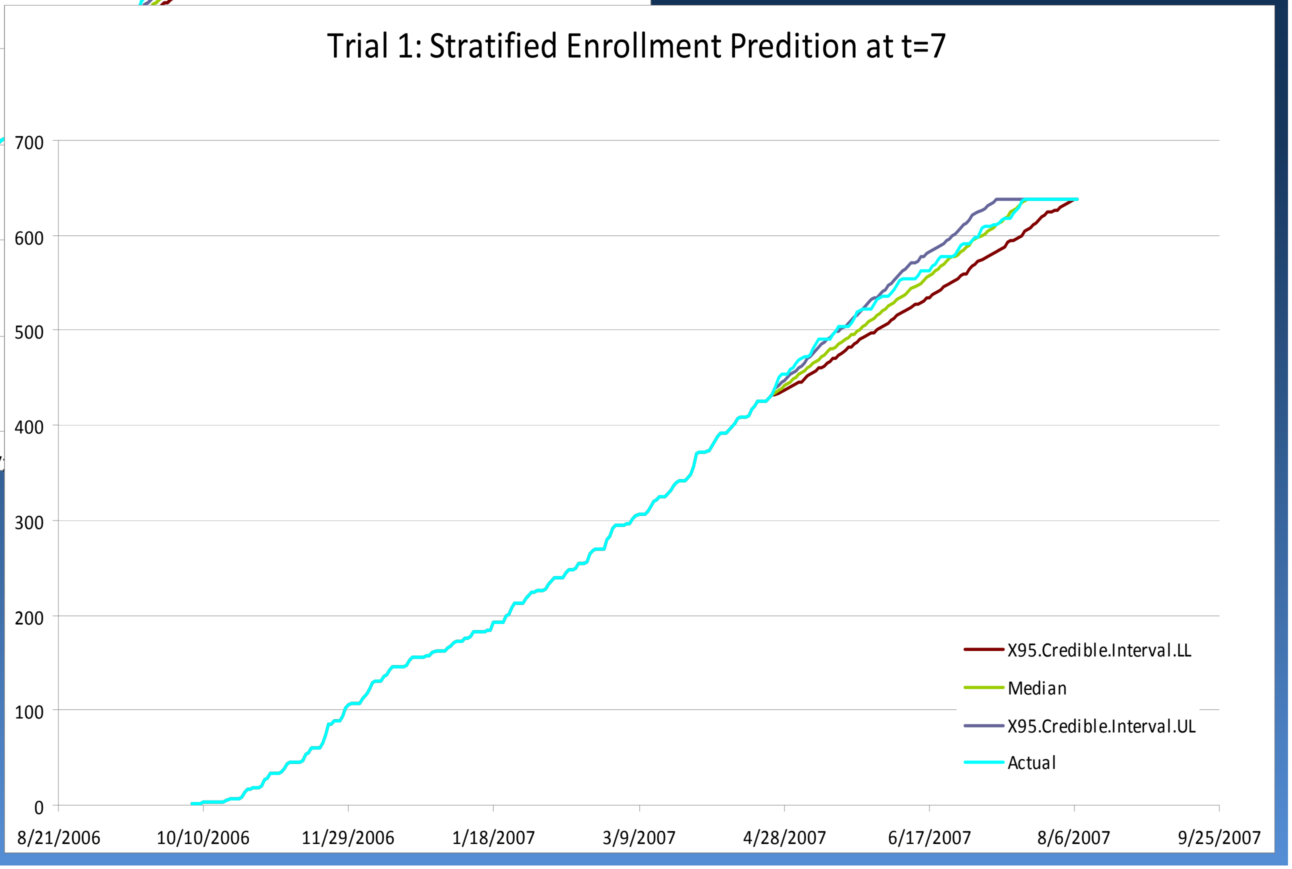
Prediction at time t_0 = 5 months





Prediction at time t_0 = 7 months



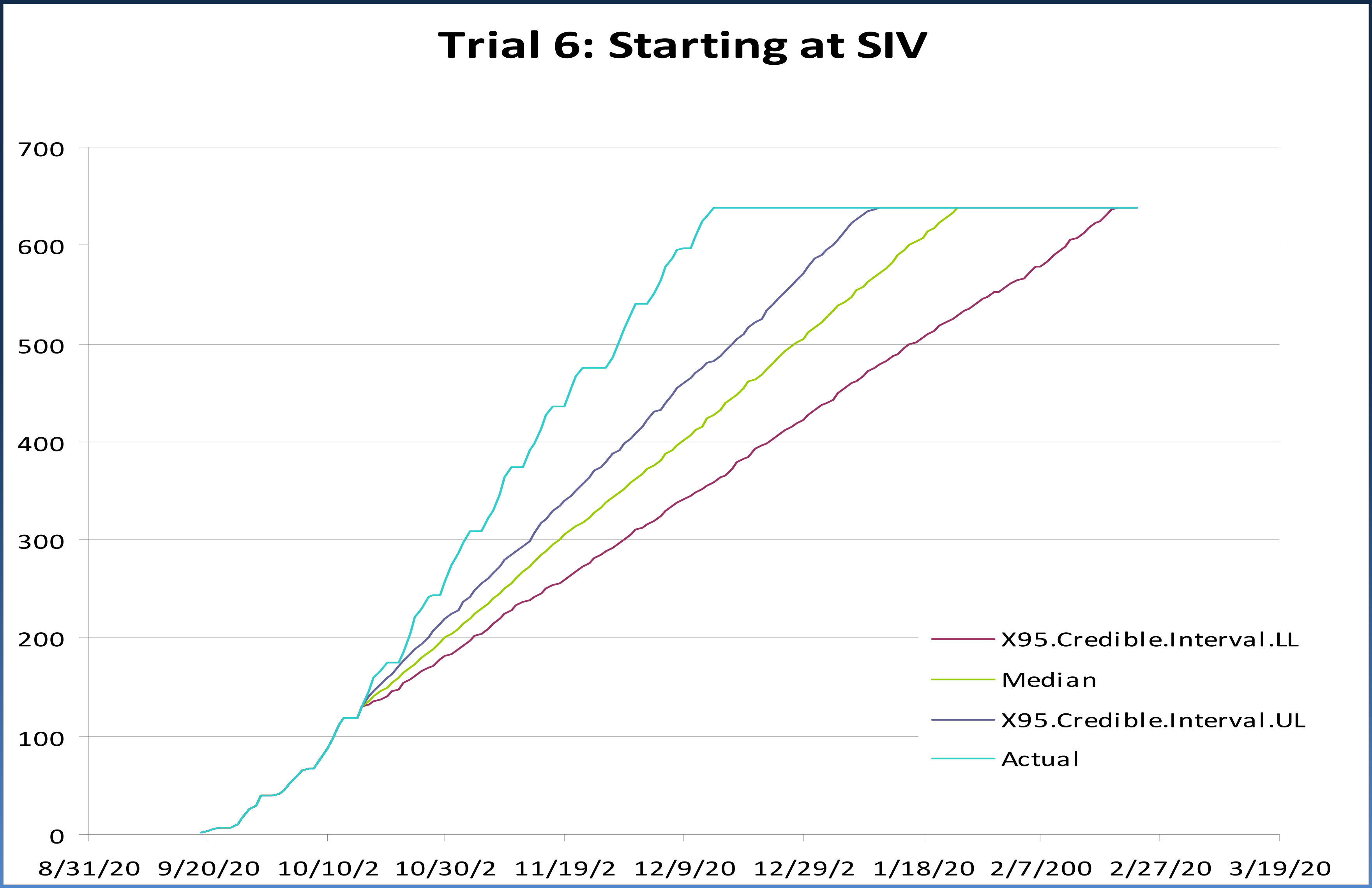


Prediction times for a few other trials (in days)

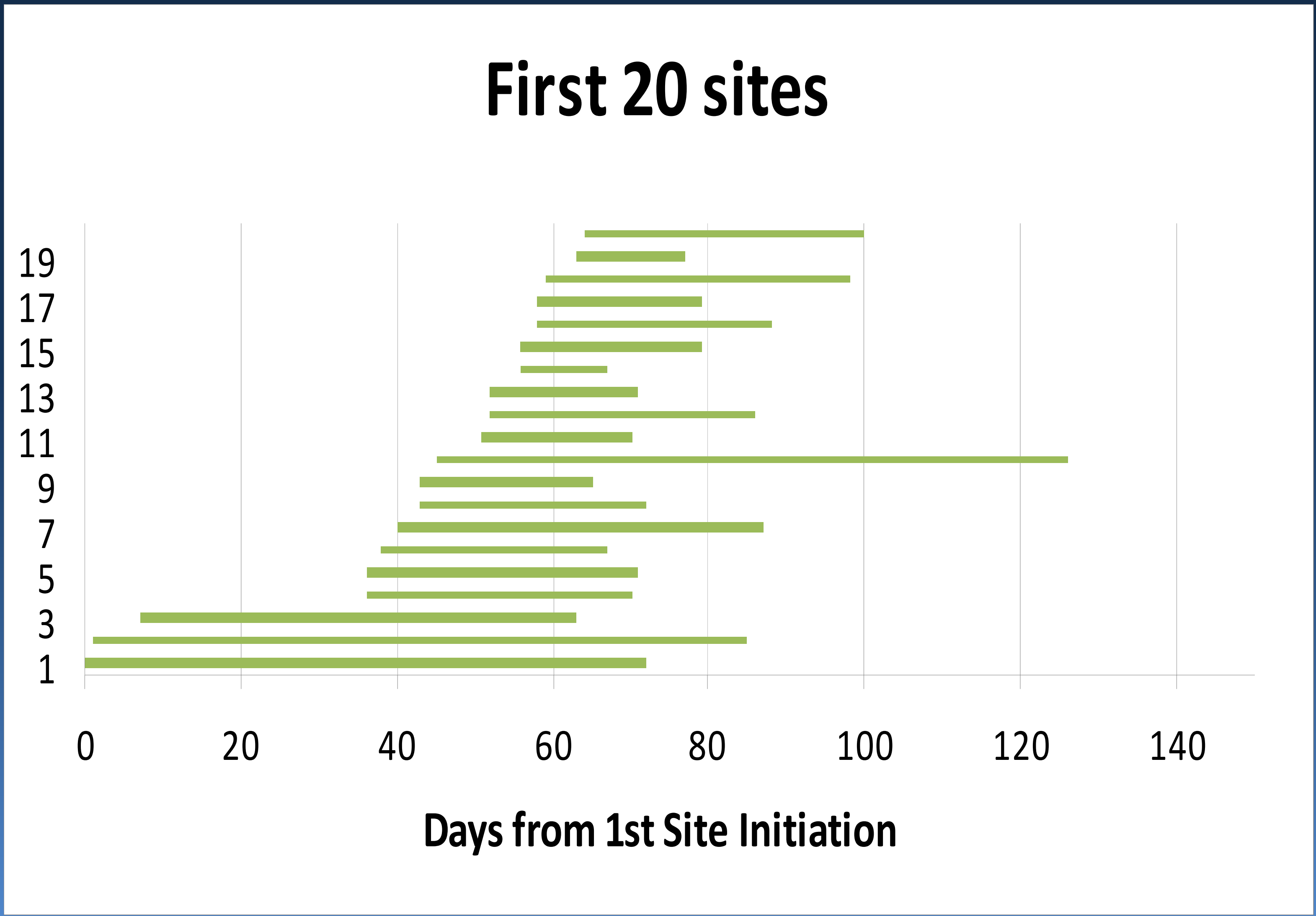
		P2.5	Median	P97.5	Actual
Trial 1	Stratified	281	308	338	297
	Unstratified	266	321	394	297
Trial 2	Stratified	147	164	186	153
	Unstratified	122	135	150	153
Trial 3	Stratified	131	145	163	134
	Unstratified	117	142	176	134
Trial 4	Stratified	309	317	326	304
	Unstratified	309	325	347	304

Ignoring the first inter-arrival time

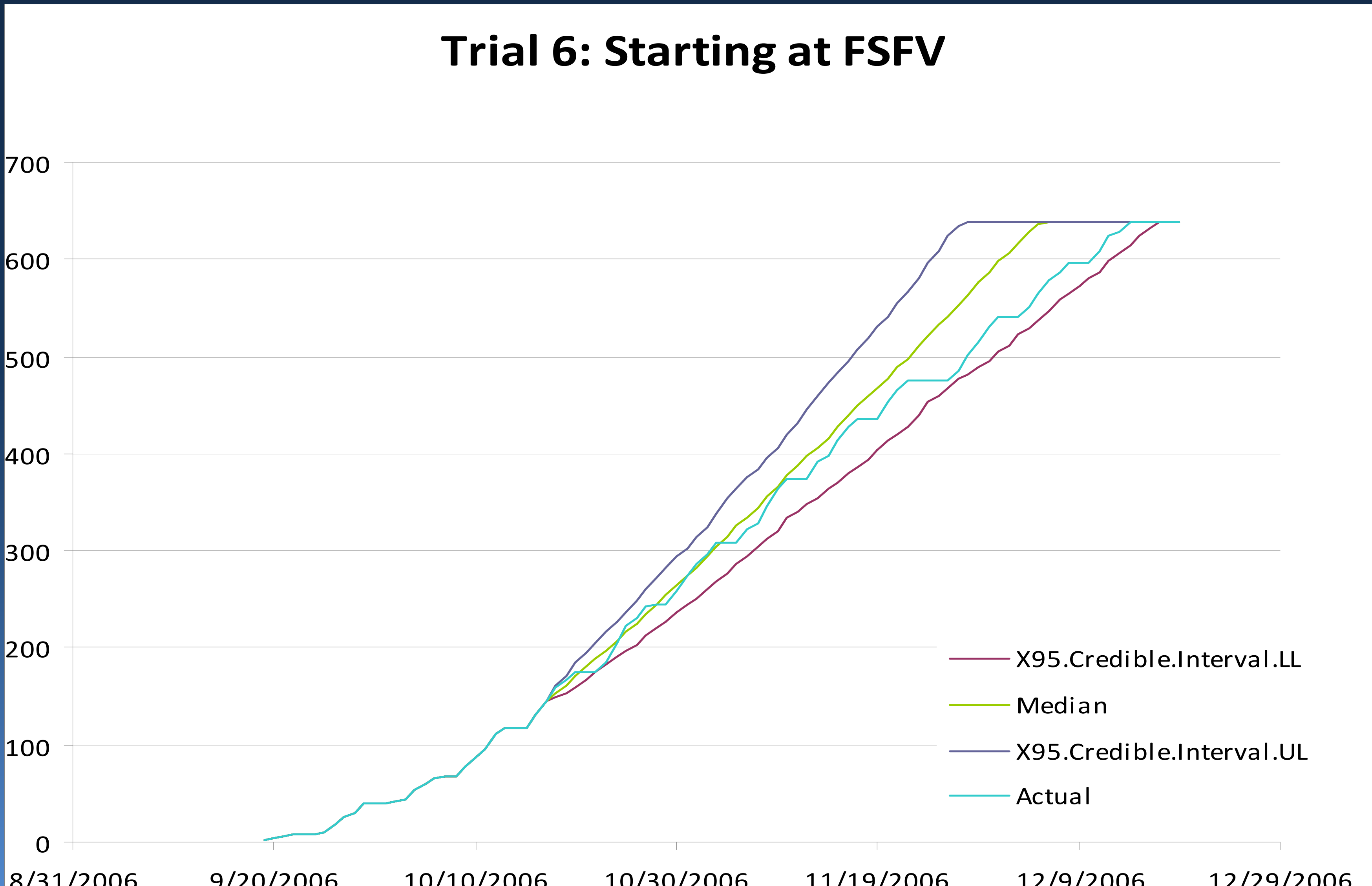
Model breakdown #1



Time from Site Initiation Visit to First Subject First Visit

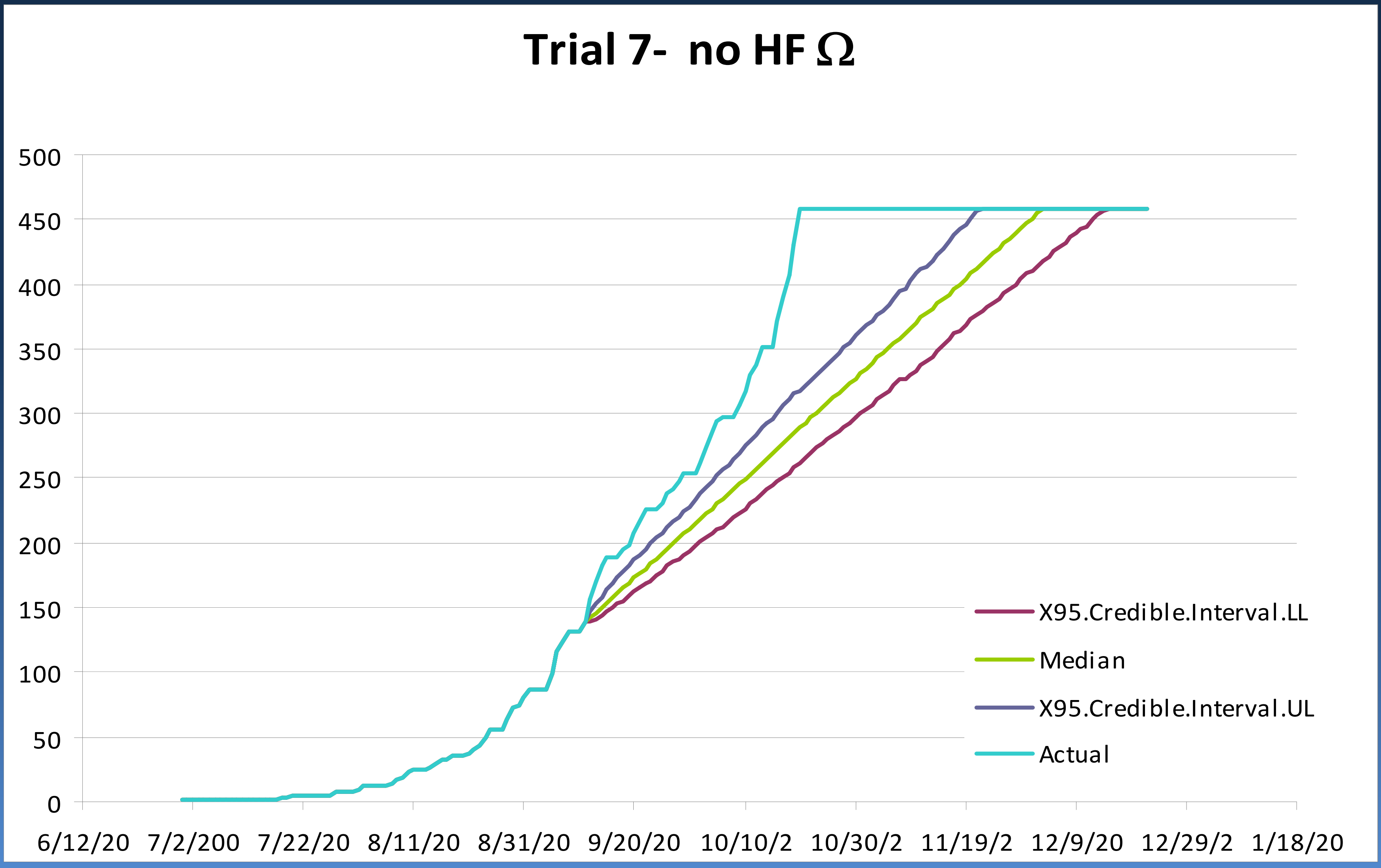


Prediction starting at FSFV



Time-variant enrolment rates

Model breakdown #2



Time varying parameters

- Harvey and Fernandes (1989) proposed a model in the econometrics literature that modifies the Poisson-Gamma to allow the underlying mean of the process to change over time
- Parameters of Gamma change from time $t-1$ to t

$$a_t = \omega a_{t-1} \text{ and } b_t = \omega b_{t-1}$$

- Mean of the Poisson remains the same
- Variance is inflated by a factor $1/\omega$ with $0 < \omega < 1$
- ω is chosen to control the amount of drift

  (Could be varied with t or modeled in some way as a function of past observations either as a numeric function or as parameter that is updated in a Bayesian manner)

Model for drift in Gamma parameters

- Posterior at time t-1 (Gamma):

$$G(\mu_{t-1};\alpha_{t-1},\beta_{t-1}) = \frac{\mu_{t-1}^{\alpha_{t-1}-1} e^{-\beta_{t-1}}}{\Gamma(\alpha_{t-1})\beta_{t-1}^{\alpha_{t-1}}} \qquad \alpha_{t-1},\beta_{t-1} > 0$$

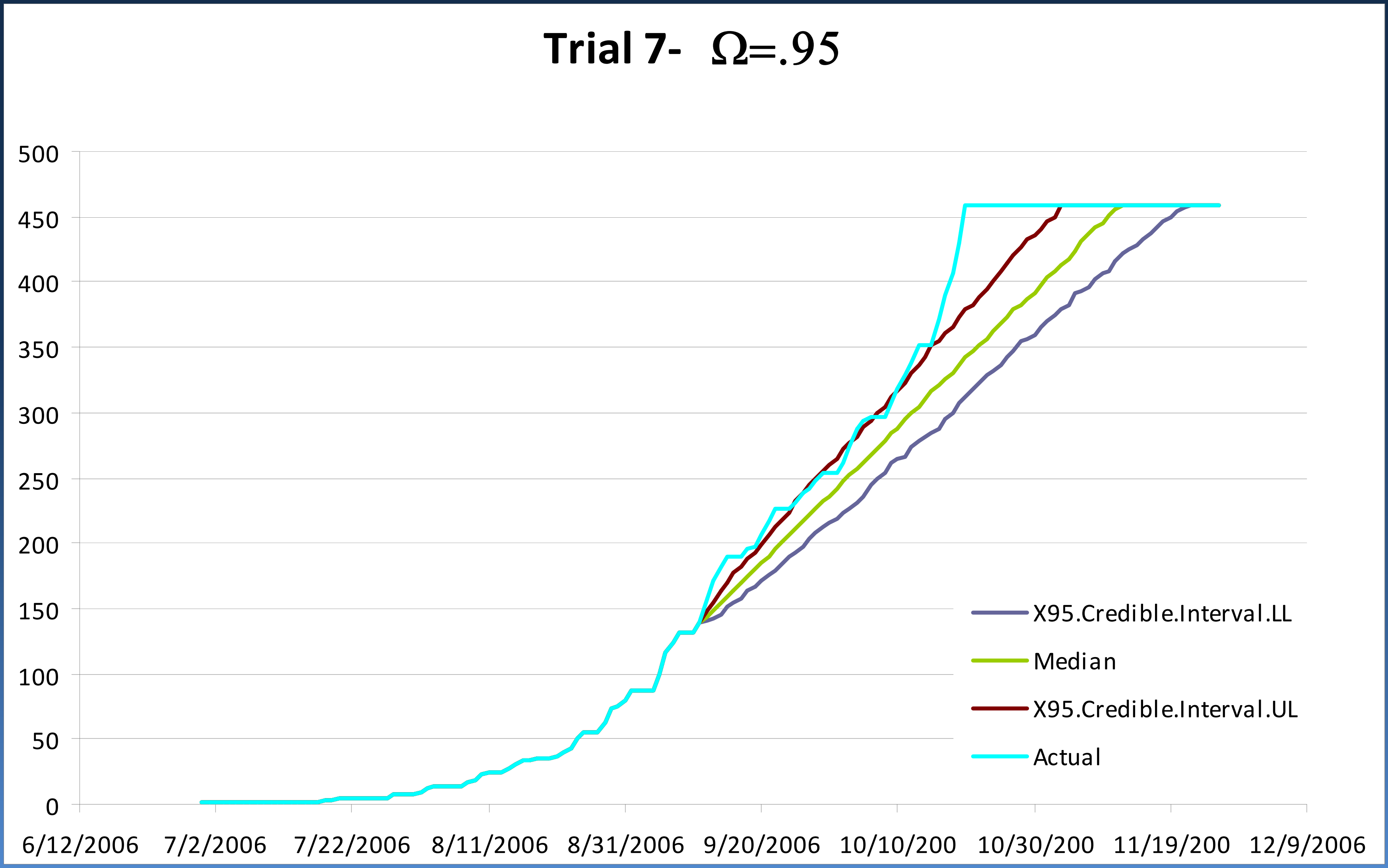
- Prior at time t (before observing n_t):

$$G(\mu_t;\alpha_{t|t-1},\beta_{t|t-1}) = \frac{\mu_{t-1}^{\alpha_{t|t-1}-1} e^{-\beta_{t|t-1}}}{\Gamma(\alpha_{t|t-1})\beta_{t|t-1}^{\alpha_{t|t-1}}} \qquad \alpha_{t|t-1} = \omega\alpha_{t-1}, \beta_{t|t-1} = \omega\beta_{t-1} \qquad 0 < \omega < 1$$

- Posterior at time t-1 (after observing n_t in period t):

$$G(\mu_t;\alpha_t,\beta_t) = \frac{\mu_t^{\alpha_t-1} e^{-\beta_t}}{\Gamma(\alpha_t)\beta_t^{\alpha_t}} \qquad \alpha_t = \alpha_{t|t-1} + y_t, \beta_t = \beta_{t|t-1} + 1$$

Application of HF model to Trial 7



Conclusions

Further extensions

- Incorporating important covariates into the model to monitor and predict appropriate representation of subpopulations and balance in randomization strata
- Predicting changes in enrolment patterns, acceleration of slowdown of enrolment at specific sites
- We have already incorporated event modeling into our software to predict timing of interim analyses and trial end in survival studies

Some practical conclusions and suggestions

- <u>Exclude or limit</u> countries likely to have low recruitment in trial design
- Prioritize <u>early initiation</u> for countries likely to have high recruitment
- Drop countries that have <u>not initiated sites within a certain time</u> after several countries are recruiting at a good rate

- Identify <u>attributes of low enrolling sites</u> by analyzing past data
- <u>Limit number of potentially low enrolling sites</u> in trial design. Prioritize <u>early initiation</u> for sites likely to have high recruitment. Create “standby” list of sites.
- <u>Monitor sites using a statistical model</u> (similar in spirit to quality control charts in manufacturing). Drop sites that are recruiting below minimum performance levels. Add sites from standby list.
- <u>Save cost of drug by redistributing supplies</u> at dropped sites to other sites in the same country
- Evaluate effect of using <u>fewer sites with greater resources allocated per site</u> (e.g. advertising budget, training, clinical research associate time)

Take home messages

Cytel

- Modeling, simulating and forecasting enrolment and the arrival of events can help a sponsor or CRO get a grip on the uncertainty inherent in trial timelines
- This quantitative exercise combines with the "art of patient recruitment and retention" to run more efficient and successful studies
- It can help track issues such as poor-performing countries or sites, and guide decisions such as when to open new sites, or how and when to resupply sites with drug
- Cost savings can be substantial if impact on NPV, drug supply chain, and site monitoring is taken into account and acted upon

References

- Anisimov, V., and Fedorov, V. Modelling, prediction, and adaptive adjustment of recruitment in multicenter trials. *Statistics in Medicine. 2007, 26: 4958-4975*
- Harvey, A.C., and Fernandes, C. Time Series Models for Count or Qualitative Observations. *The Journal of Business and Economic Statistics.* 1989, 7:4:407-422.
- Lamberti, MJ. State of Clinical Trials Industry. Thomson Centerwatch. 2006, 292.
- Pierre, C. Recruitment and Retention. 2006.