FORECASTING AND COST-EFFICIENT DESIGNING RESTRICTED ENROLMENT IN CLINICAL TRIALS

PSI 2026 Conference
Belfast
15-17 June, 2026

Vlad Anisimov & Matt Austin
Data Science, Centre for Design & Analysis
Amgen: London, UK & Thousand Oaks, CA, US

AMGEN®

BACKGROUND

- Clinical trials in the modern era are characterized by their complexity and high costs as it is required to recruit hundreds/thousands patients across multiple clinical sites/countries to get a sufficient power of statistical tests.
- The efficient patient enrolment planning and forecasting is one of the main drives in efficient conducting the late-stage clinical trials.
- Citing Ken Getz, director, Tufts Center for Study of Drug Development, US, "Patient recruitment is among the greatest challenges that the clinical research enterprise faces today, and it is a major cause of drug development delays,"
- Many companies may still rely on ad-hoc simplified or deterministic models. That can lead to inefficient trial designs, delays, and increased costs.
- Predicting bounds and some important metrics like PoS (probability to complete in time) is possible only by using stochastic models for enrolment.
- Actually, about 23% of trials fail to recruit in time, > 50% in 2012; – Tufts, 2020

WHY DATA-DRIVEN STATISTICAL METHODS MATTER

- Operational performance depends on uncertain and evolving patient enrolment and site dynamics.
- Simplified deterministic planning can lead to delays, inefficiencies and missed enrolment targets.

Advanced statistical methods can help to:

- Improve enrolment forecasting and site performance
- optimize site and country selection under operational constraints
- enable data-driven interim reforecasting
- create an optimal operational design

OUTLINE

- Advanced statistical techniques for patient enrolment prediction under various restrictions
- Cost-time efficient enrolment design, sites/countries selection
- Real-time adaptive adjustments of trial operation
- Centralized statistical monitoring
- Forecasting other operational processes
- Case studies and practical implications

ENROLMENT MODELLING METHODOLOGY

Analytic data-driven methodology is developed*:

- Patients' flows in clinical sites are modelled using Poisson processes
- Variation in enrolment rates is modelled using a gamma distribution
- Sites can be initiated with random delays

The main predictive characteristics have been derived using closed-form expressions (no need in Monte Carlo simulation).

Methodology is world-wide accepted (Poisson-gamma model), published in top statistical journals and presented at major conferences (JSM, ASA, PSI,…), called now as "industry standard".

Some pharma companies and CRO: Cytel, Statsols (nQuery),.. use these models.

** Anisimov & Fedorov, Statistics in Medicine, 2007 (**189** Google Scholar citations); Anisimov; Anisimov et al., 2007-2026*

ENROLMENT MODELLING METHODOLOGY

Poisson-gamma enrolment model:

In site i, active in time interval $[u_i , b_i]$, enrolment process:

$$n_i(t) = \Pi_{\lambda_i}(t - u_i)\ I(u_i \leq t \leq b_i)$$

$\lambda_i = \gamma(\alpha,\beta)$ – gamma distributed rate, u_i – delay in site initiation, b_i - closure time.

$n_i(t)$ has a Poisson-gamma distribution (negative binomial), for $u_i \leq t \leq b_i$,

$$\Pr(\Pi_{\lambda_i}(t) = k) = \frac{\Gamma(\alpha + k)}{k!\ \Gamma(\alpha)} \frac{(t - u_i)^k \beta^\alpha}{(\beta + t - u_i)^{\alpha+k}}, k = 0,1,2,..$$

In R, function:

dnbinom(k, size= α, *prob*= β/(β+t- u_i))

This is a site-based model as sites are the main drive of the enrolment.

Prediction in country/globally – sum of individual site's processes (not a PG process).

ANALYTIC APPROXIMATION OF THE COUNTRY PROCESSES

In large countries, it can be used a normal approximation, but not in small countries.

- Country process - Poisson-gamma approximation

A new approach is developed *:

- ✓ Country process approximated by a PG process with aggregated parameters
- ✓ Works perfectly well for small and large numbers of sites
- ✓ Predictive distributions are calculated in a closed form
- ✓ Used in creating optimal design

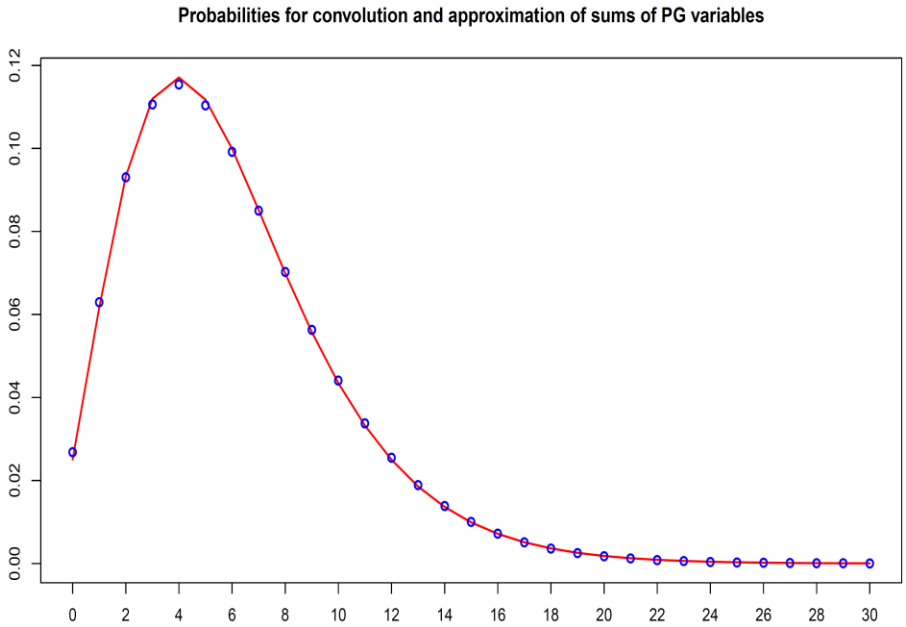
For large number of sites (>=15-20), predictive areas (mean and bounds) for exact, normal and PG approximations practically coincide.

Monte Carlo simulation (used by some companies) is not so efficient:

- large computational time, impossible to solve multivariate optimal problems, compute small critical probabilities, etc.

* *Anisimov & Austin, Communications in Statistics*, 2020

ANALYTIC APPROXIMATIONS OF THE ENROLMENT PROCESS



Quality of approximation of the enrolment process in some country by a PG process

Example: 3 sites, interim time = 300 days; initiation times: uniform grid in interval (1,300); α=1.5; β=150 for each site;

Exact distribution of the enrolled sbj – red line; Approximative PG distribution – blue dots. Perfect coincidence.

Using the discrepancy measure: K - # of sites

$$\mathrm{Dif}(K) = \max(\left|\mathrm{ppK}[i] - \mathrm{ppKPG}[i]\right|, i = 0, \ldots, L).$$

K	2	3	5	8	10	15
$\mathrm{Dif}(K)$	0.0019	0.0017	0.0011	0.00075	0.00059	0.00039

ENROLMENT FORECASTING TOOLS

The developed methodology & software-tools allow:

Forecast enrolment over time on different levels (with mean and predictive bounds)

Predict time to reach enrolment target & Probability of Success at the design and interim stages

Separate problem at the design stage – how to evaluate enrolment rates for the new trials:

Using machine learning algorithms trained on large databases of historical studies provided by consortium of pharma companies using different features

- Study indication
- Site's types
- Countries
- Number of sites
- Start-up times
- enrolment windows
- Study duration..

CLINICAL TRIAL DATA SHARING CONSORTIUM - 17 MEMBER COMPANIES

> 45% OF GLOBAL, INTERVENTIONAL TRIALS

AMGEN

NOVARTIS

ucb

MERCK

SANOFI

novo nordisk

Pfizer

Roche

AstraZeneca

REGENERON
science to medicine®

Boehringer Ingelheim

Lilly

janssen
PHARMACEUTICAL COMPANIES OF Johnson&Johnson

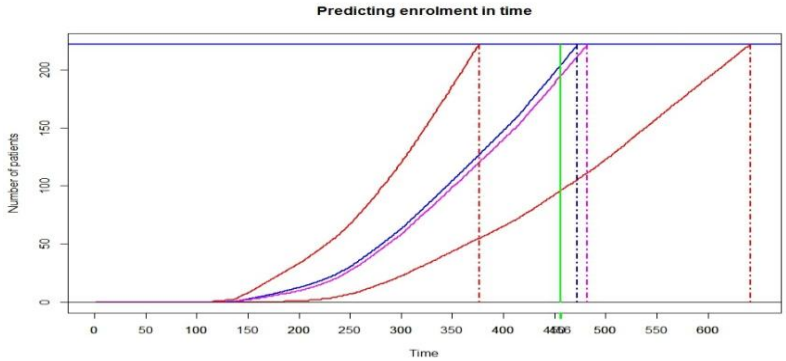
Allergan

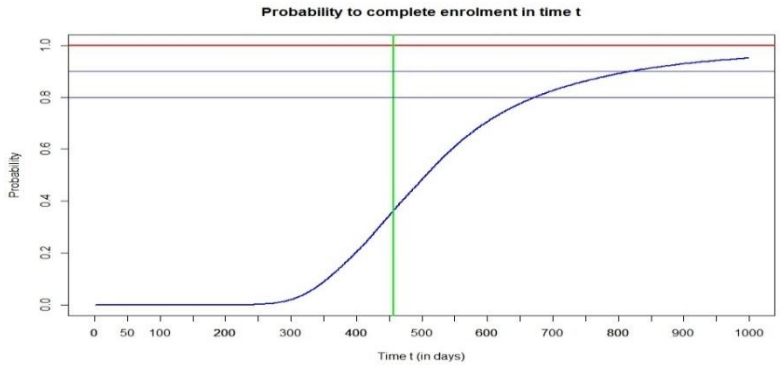
gsk
GlaxoSmithKline

abbvie

Bristol Myers Squibb™

BASELINE ENROLMENT PREDICTION





Predictive area of enrolment (mean and 90% bounds);
enrolment time: mean – 473; median – 484 (in days).
Predictive 90% bounds - (375, 646) days.
Input data: sites initiation schedule;
for each site – mean and Var of the rate.

Probability to reach target 222 patients.
Planned time 15 months.
Probability of Success = 0.365

Similar predictions can be created in countries and globally using a developed methodology for approximation of country enrolment processes by Poisson-gamma processes.

PREDICTING RESTRICTED ENROLMENT

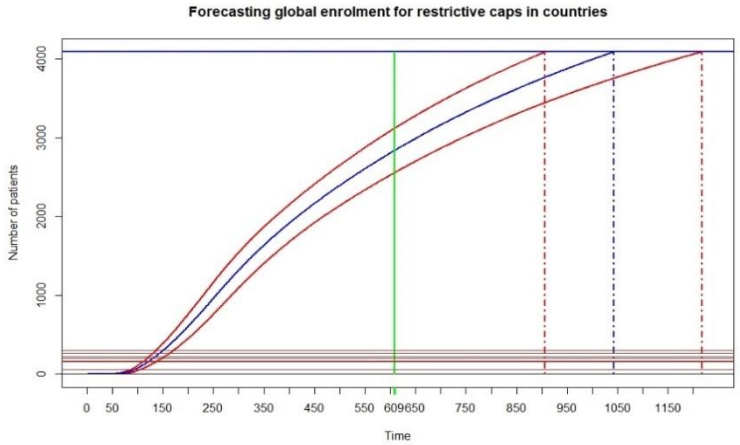
Upper restrictions in countries - Caps. Country I_s , L(s) – cap.

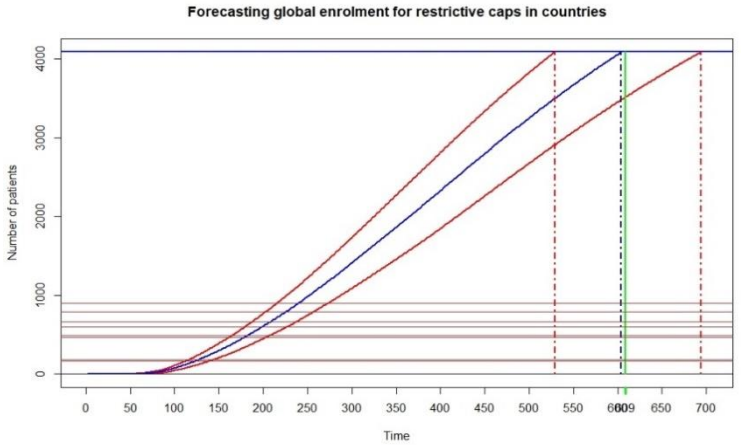
- Country process is approximated by a PG restricted process $n^{L(s)}(I_s, t)$.
- Global process – convolution of country restricted processes (analytic approach).

  Bounds and PoS are calculated using a distribution of the global process – convolve() in R – distributional approach based on a discrete Fourier Transform.
- For a large # of countries we can use a normal approximation using formulae for mean and Var of aggregated country processes (much faster, used in optimization):

$$
\begin{aligned}
\mathbf{E}[n^{L(s)}(I_s,t)] &= E(I_s,t)\mathbf{P}(\mathrm{PG}(A(I_s,t)+1, B(I_s,t)) \le L(s)-2) \\
&+ L(s)\Big(1-\mathbf{P}(\mathrm{PG}(A(I_s,t), B(I_s,t)) \le L(s)-1)\Big) \\
\mathbf{E}[(n^{L(s)}(I_s,t))^2] &= (E^2(I_s,t)+S^2(I_s,t))\mathbf{P}(PG(A(I_s,t)+2, B(I_s,t))) \le L(s)-3) \\
&+ E(I_s,t)\mathbf{P}(PG(A(I_s,t)+1, B(I_s,t)) \le L(s)-2) \\
&+ L^2(s)\Big(1-\mathbf{P}(PG(A(I_s,t), B(I_s,t)) \le L(s)-1)\Big)
\end{aligned}
$$

PREDICTING RESTRICTED ENROLMENT





Upper bounds (Caps) in countries – enrolment stops in a country where Cap is reached.

Newly developed analytic methodology *

Left: Initial Caps. Right: Caps with 200% increase. Horizontal brown lines – the values of Caps in countries.

Non-efficient (rather low) Caps may substantially increase enrolment time. Here on 14 months.

**Anisimov & Austin, chapter in book,* Springer*, 2022*

OPTIMAL ENROLMENT DESIGN

- The key question:
  - Which countries and how many sites to select that:
    enrol the fastest with min cost to get a desired PoS?
- The developed tools allow:

  Realize a new paradigm: Plan enrolment with confidence, e.g.

  "design a trial with 80% chance to complete enrolment in time"
  - Evaluate PoS;
  - If study is likely to go late, update design and calculate the minimal number of additional sites needed to reach a desired PoS.

OPTIMAL COST-EFFICIENT DESIGN

Find an optimal geographic site's allocation $\{N_1 \ldots N_J\}$ among J different countries given enrolment rates and country's costs that:

minimizes

$$\text{TrialCost}\ (N_1 \ldots N_J) = (\text{Patients}) + (\text{Sites}) + (\text{Countries})$$

given

a) $\Pr(\ \text{enrolment Time}\ (N_1 \ldots N_J) \leq T_{plan}\ ) \geq P_{plan}\ ,$

b) $L_j \leq N_j \leq U_j,\ j = 1,..,J$

- To complete enrolment earlier will cost more – optimal decision making
- Analytic tools are developed using different criteria of optimality for various enrolment scenarios (non-linear constraint optimization*, and also genetic evolution algorithms and metaheuristic optimization**).

**Anisimov & Austin, chapter in book Springer, 2022;*
***Schepps, Wong, Austin, Anisimov, 2021, 2024*

OPTIMAL ENROLMENT DESIGN

Table 23.1 Optimal centre's allocation

Country	Low	Upp	Rate	Cost	Opt.alloc.	Opt.alloc.	Opt.alloc.	Opt.alloc.	Opt.alloc.
Probab.					0.5	0.6	0.7	0.8	0.9
Country1	0	7	0.42	15600	0	0	0	0	0
Country2	0	4	0.43	14250	0	0	0	0	1
Country3	2	5	0.22	13550	2	4	5	5	5
Country4	0	4	0.55	14200	3	4	4	4	4
Country5	0	6	0.3	13800	6	6	6	6	6
Country6	1	7	0.57	14300	1	1	2	4	6
Country7	1	5	0.21	13400	5	5	5	5	5
Country8	1	7	0.25	14250	1	1	1	1	1
Country9	2	5	0.16	12300	5	5	5	5	5
Country10	0	7	0.19	13800	1	1	0	0	0
Country11	2	7	0.18	14600	2	2	2	2	2
Country12	2	7	0.62	16380	2	2	2	2	2
Country13	0	4	0.45	13400	4	4	4	4	4
Country14	0	5	0.23	11200	5	5	5	5	5
Country15	0	5	0.3	14000	1	1	1	1	1
Country16	2	7	0.39	14100	2	2	2	2	2
Total	14	92	–	–	40	43	45	46	49
Opt cost	–	–	–	–	3,643,470	3,902,851	4,135,948	4,415,110	4,879,621

16 countries.

Restrictions: Low and Upper number of sites.

Rate – monthly enrolment rates

Cost – cost per patient

Optimal site's allocation for different PoS: 0.5, .., 0.9

Anisimov & Austin, chapter in book, Springer, 2022, 511-540.

DATA-DRIVEN RE-PROJECTION AND ADAPTIVE ADJUSTMENT

Interim re-forecasting of the remaining enrolment:

Data in each active site i:
(enrolment window v_i,
# of patients recruited k_i)

Schedule for new sites to be initiated and planned enrolment rates

Methodology:

The global parameters of the enrolment rates in active sites (a,b) are estimated

The posterior rates are adjusted using site's data and Bayesian technique

- The posterior rate in site i has a gamma distribution with parameters (a+k_i,b+v_i)

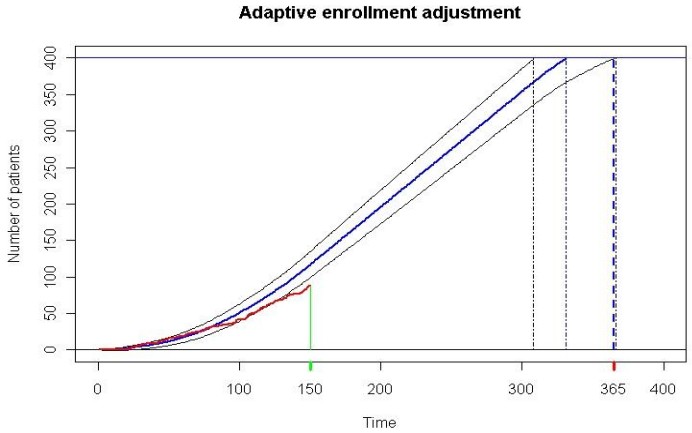
Future enrolment is predicted using posterior rates in active sites and the expected rates in the new sites accounting also for initiation times

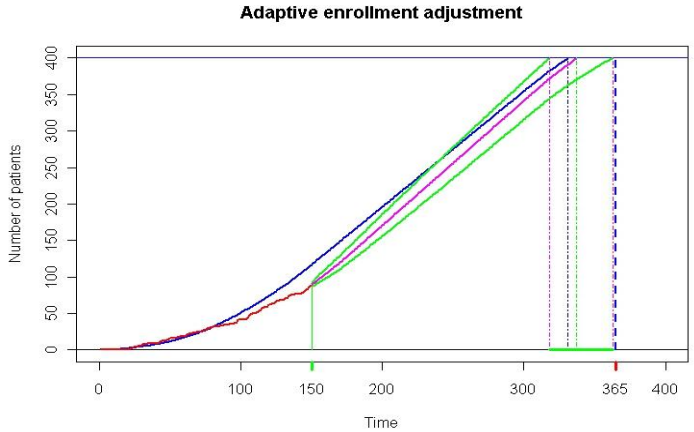
Optimal adaptive adjustment:

if PoS is not so large (unlikely to complete in time), calculate the optimal number of additional sites required to reach a desired PoS.

ENROLMENT RE-FORECASTING & OPTIMAL ADJUSTMENT

Scenario: (artificial) 400 pts, 70 sites, time = 365 days. Sites are initiated in 5-month period, half of sites will be closed in two months before the end of enrolment.





Initial design: to complete with 90% confidence.
Predictive area: mean and confidence bounds.
Interim analysis after 150 days: 88 pts recruited.
enrolment is going slower than predicted.

Optimal interim adjustment:
to complete with 90% confidence:
22 new sites to add (according to a given schedule).
Adjusted enrolment is going on time.

CENTRALIZED STATISTICAL MONITORING

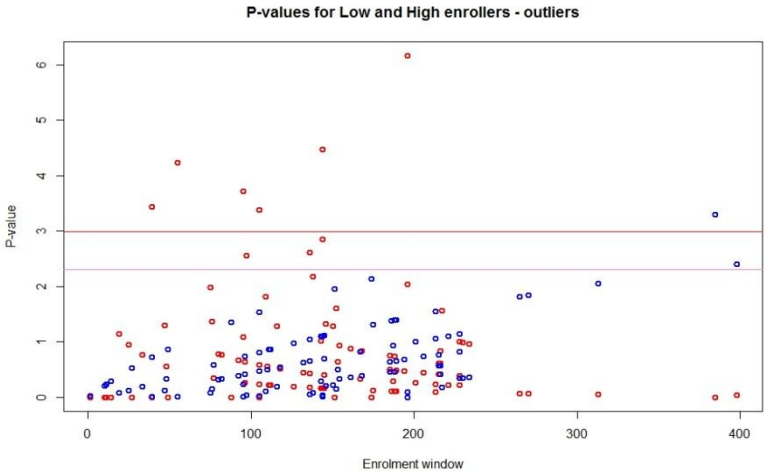
- Modelling enrolment is an underlying methodology for data-driven statistical monitoring of trial enrolment performance.
- Goal – early detecting unusual data patterns for making informed decisions
- The newly developed methodology* allows to detect
  - low and high enrolling sites, countries, regions;
  - low active and dormant sites

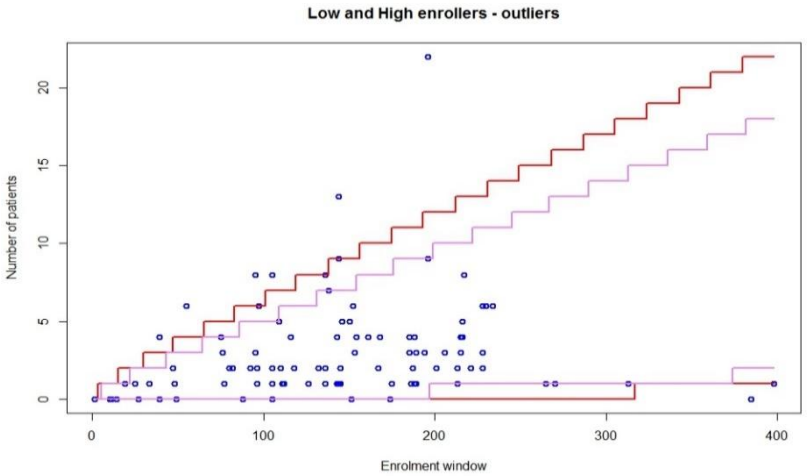
It can be used different types of detection and visualization:

- P-values
  - ✓ Upper P-values to detect high enrolling sites; Low P-values – low enrolling;
- 2D classification

** Anisimov & Austin, Communications in Statistics, 2020*

DETECTION OF LOW & HIGH RECRUITING SITES





Each dot corresponds to a site with data:
(enrolment window, # of patients recruited).

Upper P-values – red dots;

Low P-values – blue dots (using –log(…) scale)

Horizontal lines – 5% and 10% -P-values

Step-wise separation lines – quantiles of a PG model:

Upper 95% and Low 5% quantiles – red lines;

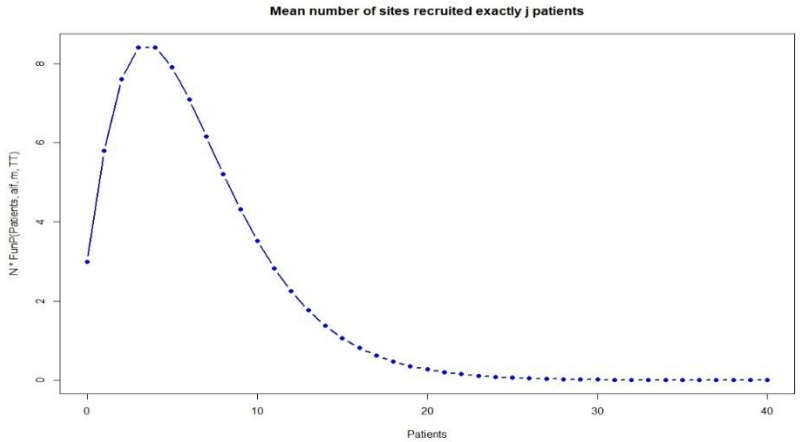
Upper 90% and Low 10% quantiles – violet lines;

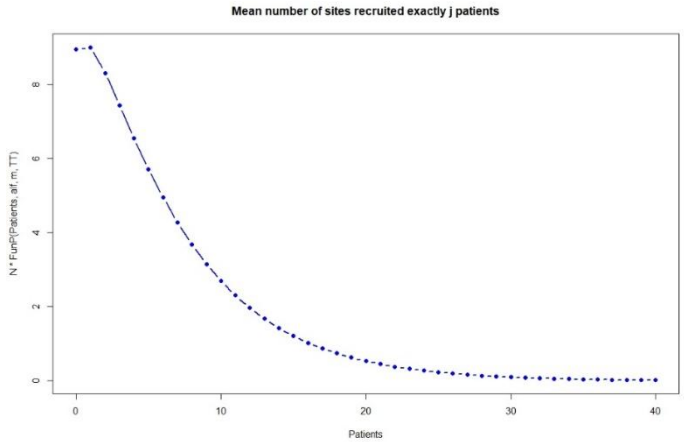
Upper area – high recruiting sites, lower area – low recruiting.

EVALUATING GLOBAL SITE'S PRODUCTIVITY

- The advanced enrolment model also allows to explain a large heterogeneity of sites productivity
- Global site's productivity:
  - for studies with low enrolment rates, e.g. oncology, it can be about 20-30% of sites enrolled 'zero' patients;
  - 20-25% of the most productive sites recruit ~ 50% of patients
  - ~ 50% of the most productive sites recruit ~ 80% of patients;
- This is NOT a result of inefficient planning, but a Statistical Law confirmed by applying a Poisson-gamma model and by the empirical analysis of many trial's databases.

ENROLLMENT PROFILE OF THE STUDY





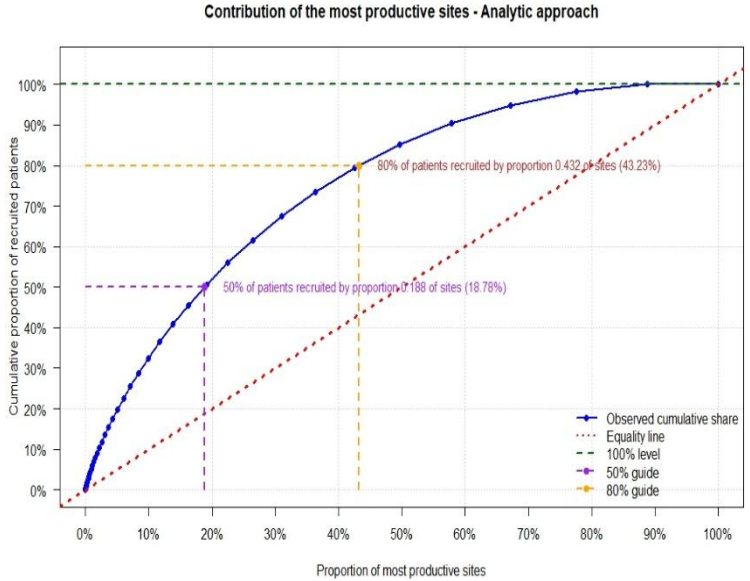
Mean Number of sites recruited exactly j patients

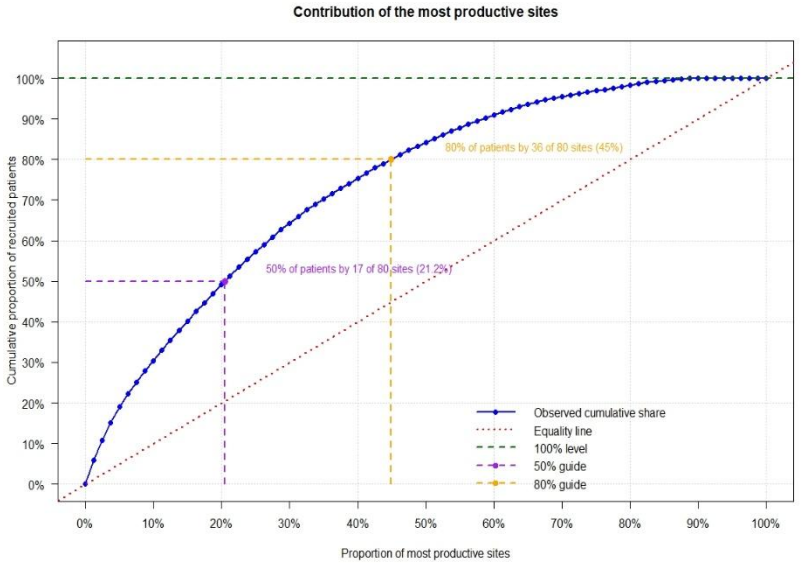
n <- 500 - sample size; N <- 80 - number of sites; enrolment time - 365 days

alpha <- 2.8; ~ 3 sites are “empty” (on average)

alpha <- 1.2; ~ 9 sites are “empty” (on average)

CONTRIBUTION OF THE MOST PRODUCTIVE SITES



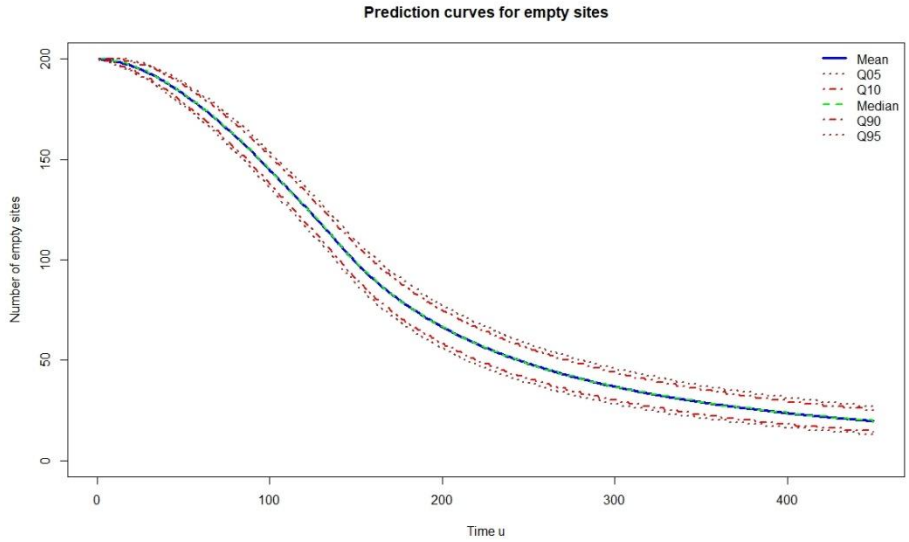


alpha <- 1.2 - shape, Analytic,
50% of sbj – by 18% of sites
80% of sbj – by 43% of sites

alpha <- 1.2, Simulation,
50% of sbj – by 21% of sites
80% of sbj – by 45% of sites

FORECASTING SITES OF A PARTICULAR TYPE

Predicting dynamics of the number of “empty” sites during the study



Artificial scenario: N = 200; shape alpha in [1,3]; mean monthly rate m in [0.3,0.45]; sites start in interval [1,140] days, study duration 15 months.

At the end, on average 20 “empty” sites left with 95% bounds (13,27).

DATA-DRIVEN FORECASTING SITE'S ENROLMENT PERFORMANCE

Country	ID	Patients	Window	Prob.zero	Prob≤2	Bound
Country1	X1	1	189	0.737	0.904	2
Country2	X2	6	97	0.302	0.316	5
Country3	X3	4	215	0.569	0.733	3
Country4	X4	3	95	0.471	0.583	4
Country5	X5	4	185	0.537	0.687	3
Country6	X6	8	217	0.393	0.464	4
Country6	X7	8	105	0.237	0.314	6
Country7	X8	4	168	0.516	0.656	3
Country8	X9	4	215	0.569	0.733	3
Country9	X10	0	385	0.886	0.986	1
Country10	X11	3	153	0.559	0.713	3
Country2	X12	0	151	0.793	0.937	2
Country11	X13	3	206	0.617	0.791	2
Country11	X14	1	145	0.698	0.867	2
Country12	X15	3	185	0.596	0.763	3

Anisimov, in Proc. of ICoMS 2023

Real study with 16 countries.

Predicting enrolment in sites using interim data and Bayesian adjustment.

Table of predictive probabilities for sites:

1) "Prob.zero" - not to recruit any patient in 30-day period
2) "Prob≤2" recruit at most 2 patients in 30-day period.
3) "Bound" – upper 90% bound, 60-days.

Low recruiting critical sites: X10, X12,

Prob Zero > 0.79

Tool to alert study teams.

Technique is analytic. Monte Carlo simulation would require ~ 10^12 runs.

MODELLING OTHER OPERATIONAL PROCESSES

- Many operational processes are driven by patient enrolment

  each arriving patient generates some follow-up process with random life-time:
  - follow-up visits, randomisation, other events, associated costs (visits, supply,…)
- The evolution of follow-up processes together with enrolment – a hierarchic evolving process
- Closed-form solutions are derived for many practical scenarios:
  - predicting event counts in event-driven trials accounting for cure, dropout and ongoing enrolment *
  - predicting follow-up processes, visits, other events, costs, etc. **
  - forecasting screening/randomization pipeline ***

* Anisimov, Pharma. Statistics, 2011, chapter, 2020; Anisimov et al., 2021, 2022

** Anisimov, Communications in Statistics - Simulation & Computation, 2016

*** Anisimov, 2025; Anisimov & Gortzak, 2026; PSI Poster P016

MODELLING EVENT COUNTS IN EVENT-DRIVEN TRIALS

Patients are recruited and followed-up until clinical events (oncology, CRV).

- Assumptions:
  - patient enrolment follows a stochastic Poisson-gamma model
  - each patient can experience the main event or dropout
  - each patient can be cured with some probability
- Models for event appearance with cure & dropout
  - Exponential, Weibull, Log-normal
- Analytic methodology* is developed for predicting over time:
  - the event counts together with ongoing enrolment; time to reach milestones
- Predictive characteristics are derived in a closed-form*

* Anisimov, 2011, 2020; Anisimov et al., 2021, chapter, 2022

* R-package “EventPrediction” is developed and implemented in-house

FORECASTING SCREENING/RANDOMISATION PIPELINE

- In randomized trials, patients typically undergo one or more screening stages before being randomised.
- To reduce unnecessary oversampling by the patients randomised from those in the screening pipeline yet, screening new patients should be stopped earlier, prior to reaching randomization target.
- **Key idea:** Using interim data, predict the distribution of the number of patients randomized out of those in screening stages yet and estimate the probability of reaching the planned randomisation target.
- **Decision rule:** If this probability is sufficiently high (e.g., 95%–99%), stop screening new patients earlier*

* Anisimov, 2025; Anisimov & Gortzak, 2026; PSI Poster P016, Anisimov & Henderson

CONCLUSIONS

- Advanced data-driven statistical techniques can make clinical trial operations more efficient, predictable and adaptive.
- They support better trial design, improved enrolment forecasting and more effective operational monitoring.
- These techniques help to reduce delays, manage cost, improve trial execution and implemented in our tools on R&D level.
- The work *”Forecasting and cost-efficient designing restricted enrolment in clinical trials"* joint with Matt Austin, Amgen, received the 2025 Award for Statistical Excellence in the Pharmaceutical Industry from the RSS and PSI (UK).

REFERENCES

1. Anisimov V., Fedorov V., Modelling, prediction and adaptive adjustment of recruitment in multicentre trials, *Statistics in Medicine*, 26, 2007, 4958–4975.
2. Anisimov V., Downing D., Fedorov V., Recruitment in multicentre trials: prediction and adjustment, in book: mODa 8 – *Advances in Model-Oriented Design and Analysis*, Physica-Verlag, 2007, 1–8.
3. Anisimov V., Using mixed Poisson models in patient recruitment in multicentre clinical trials, *Proc. the World Congress on Engineering*, II, 1046–1049, 2008.
4. Anisimov V., Predictive modelling of recruitment and drug supply in multicentre clinical trials, *Proc. of JSM*, USA, 2009, 1248-1259.
5. Anisimov V., Predictive event modelling in multicentre clinical trials with waiting time to response, *Pharmaceutical Statistics*, 10, iss. 6, 2011, 517-522.
6. Anisimov V., Statistical modelling of clinical trials (recruitment and randomization), *Communications in Statistics – Theory and Methods,* 40: iss. 19-20, 2011, 3684-3699.
7. Anisimov V., Predictive hierarchic modelling of operational characteristics in clinical trials, *Communications in Statistics - Simulation and Computation*, 45, iss. 05, 2016, 1477 – 1488.
8. Anisimov V., Discussion on the paper ”Real-time prediction of clinical trial enrollment and event counts: a review” by DF Heitjan et al., *Contemporary Clinical Trials*, 46, 2016, 7–10.

REFERENCES

9. Anisimov V., Modern analytic techniques for predictive modeling of clinical trial operations, in book *"Quantitative Methods in Pharmaceutical Research and Development*", O. Marchenko, N. Katenka (Eds), Springer, 2020, 361-408.
10. Anisimov V., Austin M., Centralized statistical monitoring of clinical trial enrollment performance, *Communications in Statistics – Case Studies and Data Analysis*, v.6, iss. 4, 2020, 392-410.
11. Schepps M., Wong WK., Austin M., Anisimov V., Optimizing patient enrollment in clinical trials with metaheuristics, *Biostatistics*, 2021.
12. Anisimov V., Gormley S., Baverstock R., Kineza C., Advanced models for predicting event occurrence in event-driven clinical trials accounting for patient dropout, cure and ongoing recruitment, *arXiv:2108.09196*, 20 Aug 2021.
13. Anisimov V., Austin M., Forecasting and optimizing patient enrolment in clinical trials under various restrictions, chapter 23, in book: "*Stochastic Processes, Statistical Methods, and Engineering Mathematics",* Springer, Proc. in Mathematics & Statistics 408, 2022, 511-540.
14. Anisimov V., Gormley S., Baverstock R., Kineza C., Predicting event counts in event-driven clinical trials accounting for cure and ongoing recruitment, chapter 9, in book "*Data Analysis and Related Applications 2*", Volume 10 - Big Data, Artificial Intelligence and Data Analysis SET, Jacques Janssen (edt.), Wiley & ISTE, Aug 2022, 121-142.
15. Anisimov V., An analytic methodology for forecasting patient enrolment performance in multicentre clinical trials, "*2023 6th International Conference on Mathematics and Statistics*", Leipzig, Germany, July 14–16, 2023, pp. 91-97.

REFERENCES

16. Schepps M., Wong WK., Austin M., Anisimov V., Optimizing Patient Recruitment in Global Clinical Trials using Nature-Inspired Metaheuristics, *Statistics in Biopharmaceutical Research,* Mar 2024, 1-15.
17. Anisimov V., Modelling/forecasting patient recruitment in clinical trials using Poisson-gamma model with time-dependent rates, Proc. 2024 6th International Communication Engineering and Cloud Computing Conference (CECCC), Oct 2024, pp. 79-86
18. Anisimov V., Oliver L., Patient recruitment forecasting in clinical trials using time-dependent Poisson-gamma model and homogeneity testing criteria, arXiv: *2111.17393*, 26 Nov 2024, 23 pages
19. Anisimov V., Forecasting the process of screened/randomized patients and optimal early stopping time in multicentre clinical trials, 11th International Conference on Advances in Statistics, 25-27 April 2025, p.18, Bologna, Italy.
20. Anisimov V., Oliver L., Modeling/forecasting patient recruitment in multicentre clinical trials using time-dependent models, in book: Data Analysis and Related Applications 5, Models, Methods and Techniques, Y. Dimotikalis, C. H. Skiadas Edt., vol. 13, Wiley, 22 Aug 2025, pp.1-22.
21. Anisimov V., Gortzak L, Joint forecasting of screening/randomization processes in multicentre clinical trials and predicting optimal enrolment stopping time, chapter in: “Data Analysis and Related Applications 5”, Wiley, 2026, 14pp.