Recruitment Modelling Predicting the COVID-19 Pandemic Impact of Clinical Trial Recruitment

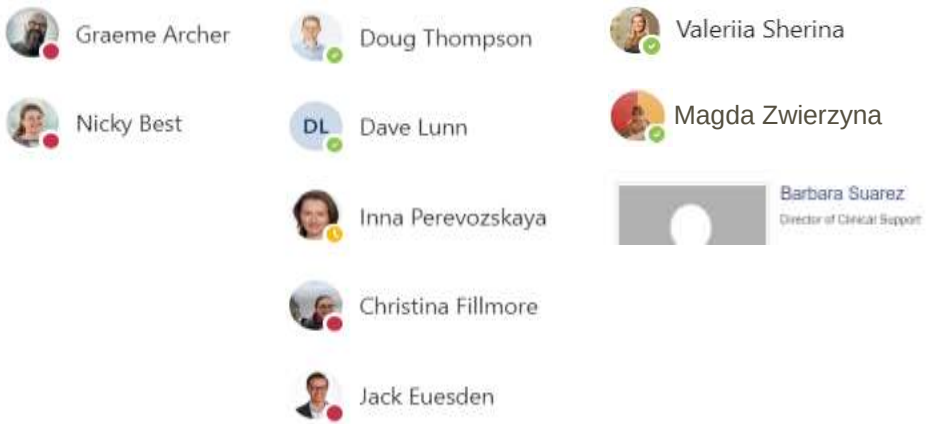
Inna Perevozskaya &
Magda Zwierzyna
10 September 2020

Joint work: The Biostats COVID-19 modelling taskforce

Formed 25 March 2020 from various groups

- Team led by **Nicky Best**, head of the Advanced Biostatistics (ABDA) team
- Barbara Suarez is Project Manager
- Sponsored by **Graeme Archer**, NCTS & Biostats LT
- Members from ABDA:
  - Dave Lunn
  - Inna Perevozskaya
- Members from Research Statistics
  - Valeriia Sherina
  - Jack Euesden
- Members from Statistical Data Science
  - Doug Thompson
  - Christina Fillmore
  - Magda Zwierzyna
- Many thanks to Jamie Lorimer from ClinOps for help and insight

A spontaneously-formed team drawn from three different groups, working as a single, ruthlessly-focussed team



Outline

- Modeling recruitment in clinical trials (Inna)
  - Deterministic vs. probabilistic
  - Anisimov model and StudyOptimizer (SO)
  - QDM in recruitment
- Modelling of COVID pandemic
  - Mechanistic vs statistical models
  - Short term vs long term forecast
- Putting the two together: how COVID affected recruitment planning
  - Recruitment plan from SO + recruitment data + state of pandemic + social isolation measures=>future predictions

- How to make a practical tool out of all this complex modelling: power of data science (Magda)

Predicting Recruitment in Clinical Trials

Why is it important?

- Most of us working in clinical statistics rarely deal with recruitment predictions
  - Its owned by Clinical Operations colleagues
- We (stats) are more focused on study design: overall number of patients to be recruited to deliver desired operating characteristics of that design
- If the design is adaptive (AD) , we worry a little bit more about recruitment:
  - Speed of recruitment vs. endpoint availability time directly affects efficiency of AD
  - If recruitment is too fast, it can wipe out the benefits of AD
- But even in AD case, most statisticians just “get the number” from ClinOps colleagues without thinking too much about where it came from
- Turns out, the latter is quite a complex process involving uncertainty, and we better make sure the number we get is right !
  - It feeds into everything, from study design to eNPV calculation (affects time to filing and cost)
  - Decisions to fund a clinical trials involve balance between quality of clinical information attained and cost
  - Cost, in turn , depends on speed of recruitment and time to complete it (achieve target sample size)

Recruitment modelling: a closer look

Overview*

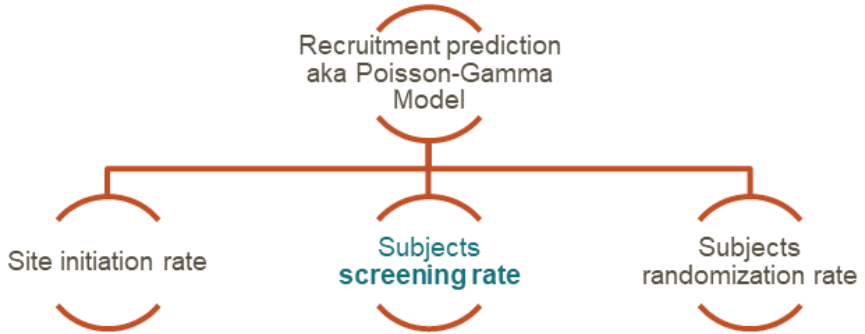
- Deterministic Models (used in the past)
  - Assume fixed or piece-wise constant recruitment rate over time
  - May or may not account for recruitment rate difference between centers
  - Advantage: simple to understand and easy to implement
  - Disadvantage: may be inaccurate, as trial complexity grows
- Stochastic Models (increasingly used today)
  - Treat recruitment rate as a random variable, properly incorporating uncertainty arising from
    1. Recruitment rate varying across multiple centers & countries
    2. Center initiation times varying
    3. Recruitment rates varying over time

*Full review available @ E. Gkioni et al. / Journal of Clinical Epidemiology 115 (2019) 141-149.

Poisson–Gamma Stochastic Model

Ref: Anisimov & Fedorov, 2005-2007; Anisimov, 2009-2017

- Commonly used across Pharma companies
  - Originally developed @GSK by V. Anisimov and validated using actual data
  - Implemented in IQVIA StudyOptimizer Software which forms the basis for our recruitment predictions
- Methodology:
  1. Each center is initiated at random (uniform, gamma, beta)
  2. Screening arrival time follows Poi(λ) process with rate $\lambda \sim Gamma(\alpha, \beta)$
  3. Screened patients “transition” into randomized or screen-fail status (binomial distribution)



Output from Poisson-Gamma Model

3 stochastic processes put together yield predicted recruitment curve over time

Curves like this are useful at initial study *planning stage* to answer questions like:

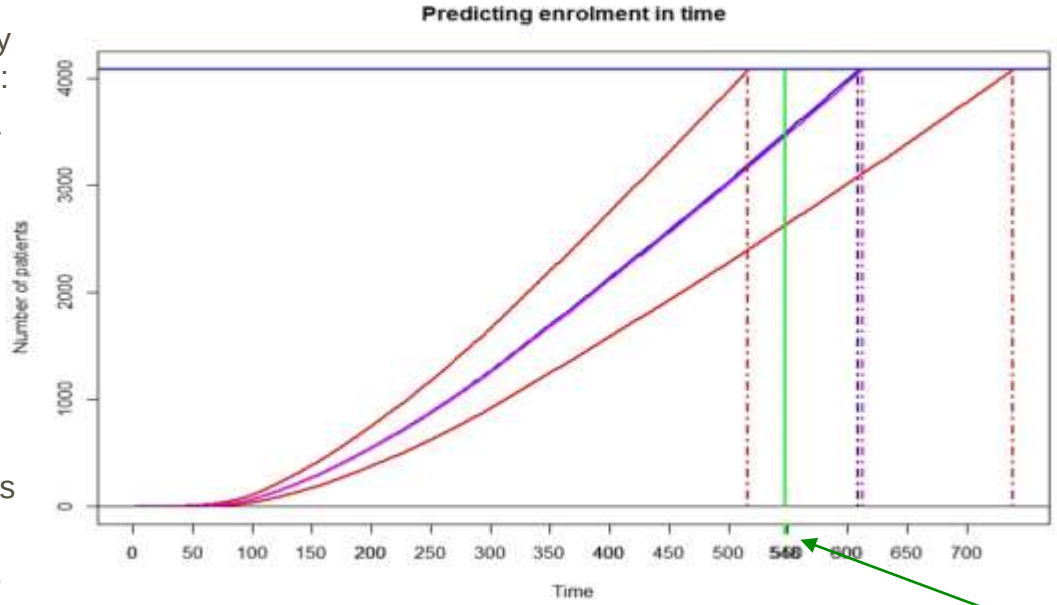
– Is my enrollment likely to meet a pre-defined target ?

– If not, and I add more sites, how much faster can it be?

What we <u>really</u> want to know is how to set the *realistic target* –the one we can achieve with pre-set probability (e.g. 80%)

– StudyOptimizer produces similar plots with pre-set 95%CB's

– But it's not very useful to answer the completion time question above-boundaries too wide to be meaningful



Plot from ADSWG KOL Lecture by V. Anisimov

Monitoring recruitment progress with Poisson-Gamma Model

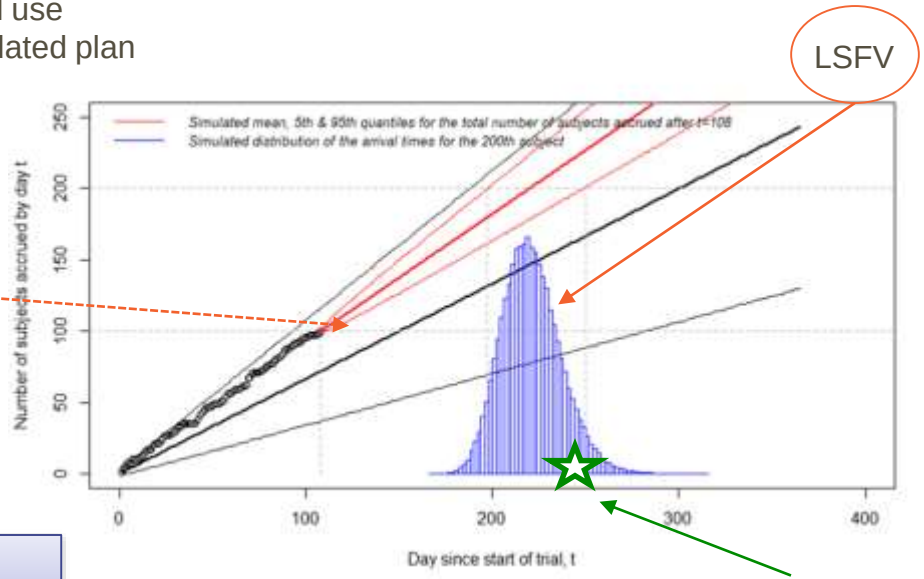
In order to answer “target” question, need to track the distribution of “recruitment complete” milestone or Last-Subject-First-Visit (LSFV)

Hypothetical example of Bayesian Poisson-Gamma model use
N=200 target recruitment, **black** is original plan, red is updated plan

Original plan generated based on **prior distributions** of model parameters (black lines)

Based on **enrollment data** of first 100 patients, the distribution of model parameters is updated (estimation)

And new enrollment projections generated (prediction) based on **posterior distributions** of parameters (red lines)



This allows us to answer the question:
What’s the likelihood of enrolling the study by target date X?

How does all this modelling fit into Clinical Operations?

Quantitative Decision Making (QDM) at GSK

- Data-driven approach to decision–making has been a broad initiative within GSK
  - Endorsed by leadership in a “top-down” approach
  - Started in clinical trial design space with requirement to follow fundamental principles:

- **define your success criteria in advance (e.g. Go/No-Go decisions)**
- **document your assumptions, acknowledge uncertainty (e.g. priors, historical data etc.)**

→ **Quantify probability of success/failure**

- The “best“ design is selected based on the above properties (e.g. operating characteristics)
  - The work has won Royal Statistical Society Award for Pharmaceutical Excellence in 2019 and is now embedded in a trial design space
- We are now expanding the concepts into clinical operations space-***that’s what stochastic recruitment modelling work is about***

Ref: Crisp, A, Miller, S, Thompson, D, Best, N. Practical experiences of adopting assurance as a quantitative framework to support decision making in drug development. *Pharmaceutical Statistics*. 2018; 17: 317– 328. https://doi.org/10.1002/pst.1856

Why do we need QDM in recruitment space

- Across industry, there is increased utilization of data-driven approaches incorporating advanced analytics in recruitment planning space, but …. people don't always know how to translate model output into a decision!
  - Even though stochastic recruitment models provide uncertainty estimates (via 95% CB's), the tendency is just to ignore it and still use means/medians (as in deterministic models)
  - 95%CB's can be too wide to make a meaningful decision
  - When someone does attempt to use it, it can be subjective and approaches vary from team to team
- **Without formulation of a decision–making framework based on probabilistic statements, full utilization of power of stochastic models is likely to remain low**
  - **We re-defined "probability of success" as Pr (to complete recruitment by date X)**
  - The work started in early 2020 and then COVID happened…..

The pandemic and urgent need to get a good grip on recruitment prediction in a rapidly changing environment served as a "catalyst" for improving decision-making within Clinical Operations

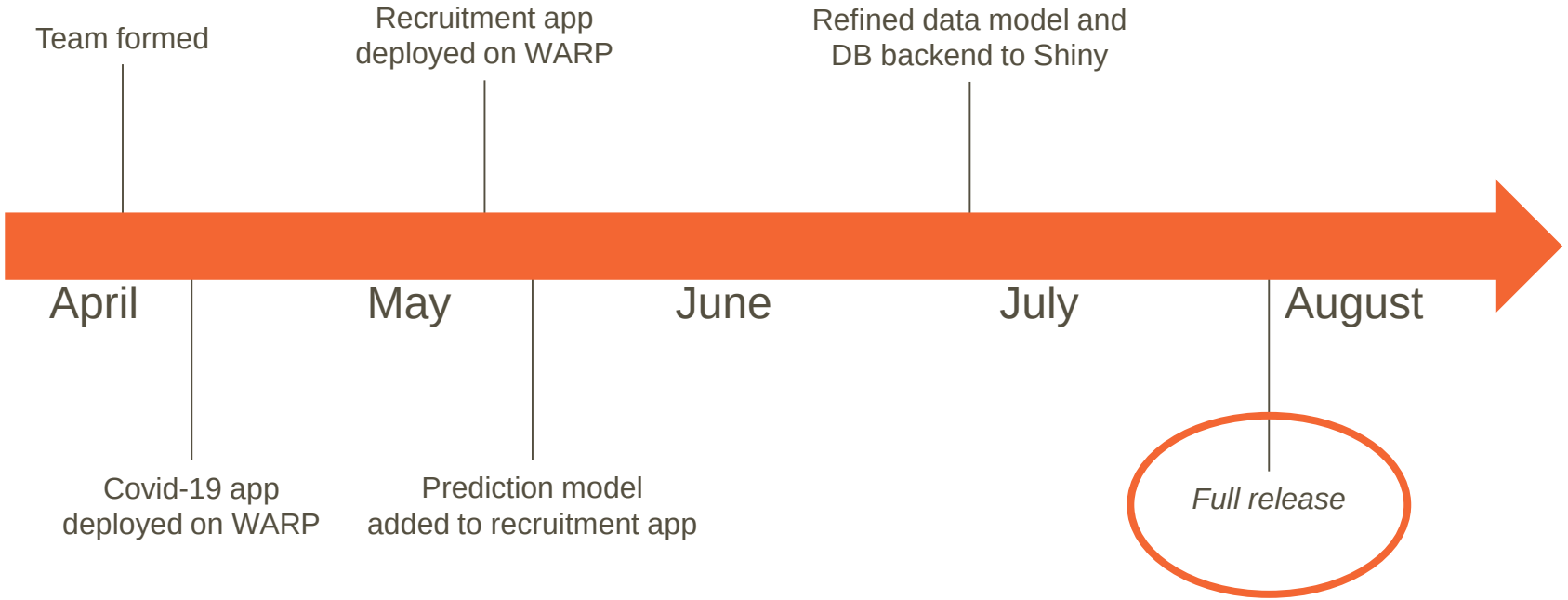
COVID predictions and its impact on recruitment

GSK COVID-19 Biostat Taskforce Objectives

1. Predict COVID-19 patterns as a function of time
   - Scraped from external sources
   - Augmented with our own implementation of the Flaxman (Imperial) model, to permit temporal extrapolations and to examine NPI impact on future death rates
   - We've also produced a user-guide to the various commonly applied forecast models
2. Predict the impact of COVID-19 on recruitment across geographies and time
   - "Recruitment" measured by the standard metrics: # sites initiated, # patients screened, # randomised
   - Inputs to this model are Non-Pharmaceutical Interventions (NPIs), which measure the stringency of social exclusion measures, and predicted COVID-19 cases (so the outputs from the first prediction model are inputs to the second)
   - We will provide predictions under a simplified summary of NPIs ("all removed" "all remain" "switching" etc)

– Produce visualisation/summary tools that teams can use

Project Lifecycle



Models for epidemic estimation and forecasting

1 slide summary!

- **Mechanistic / dynamic / compartmental models**
  - Based on differential equations representing population or transmission dynamics
  - Useful to forecast prevalence levels in the short and long-term (assuming dynamics remain constant)
  - Can assess the effects control interventions by first estimating/making assumptions/putting priors on how key parameters change due to implementation of an intervention
- **Agent-based / microsimulation models**
  - Combines microsimulation models of individuals/groups in a population with mechanistic model governing disease transmission
  - Ideal for short and long term forecasting and predicting impact of multiple interventions
- **Phenomenological / Statistical (growth curve / time series) models**
  - Empirical approach without specific basis on physical laws or mechanisms that give rise to the data
  - Useful for short term predictions, but long term predictions rapidly become very uncertain
- **ML / AI approaches**
  - Use big data from multiple diverse sources (e.g. social media, internet search, travel/phone networks etc.) and ML algorithms to select features predictive of observed location and spread of cases
  - Google Flu Trends is an (unsuccessful) example!

GSK app for country-specific short- and long-term projections of COVID-19 deaths

Tool to visualise COVID-19 projections by country (and US state) from multiple models

Model	Methods	NPI assumptions for projections*	Outcomes predicted	Projection window
IHME (Chris Murray model)**	Combination of mechanistic (SEIR) and statistical curve fitting	Oscillating: current easing of NPI controls continues until deaths reach 8 per million, then NPIs re-imposed for 6 weeks	Deaths, Infections, Hospital admissions	4 months ahead
COVID projections (Youyang Gu model)	Mechanistic (SEIR) simulation model calibrated to fit observed deaths using ML methods	Gradual easing: stringent NPIs until state/country "reopens", after which moderate NPIs are assumed. Reopening assumed to be mid-May (Europe), country or state specific elsewhere (e.g. date US stay-at-home order lifted)	Deaths, Infections	4 months ahead
Imperial College (Flaxman model)	Semi-mechanistic disease transmission model **(GSK bespoke implementation)**	1. Maximum NPIs: NPIs continue at, or are increased to, the maximum stringency level seen in that country since control measured introduced, and remain at this level 2. Minimal NPIs: NPIs continue at, or are decreased to, the minimum stringency level seen in that country since control measured were first eased, and remain at this level 3. Oscillating: NPIs increased to the country-specific maximum if daily new deaths exceed 1 per million, and relaxed back to the country-specific minimum-level when deaths fall below 1 per million	Deaths	To Jan 2022

*None of the models currently account for seasonality, reinfection, virus mutations, effective treatments or vaccines. Wearing face-masks and test-and-trace measures are implicitly accounted for in estimates of current disease transmission parameters but are not explicitly defined in future projection scenarios

**IHME model, while very useful, is no longer used due to licensing restrictions

Overview of NPIs

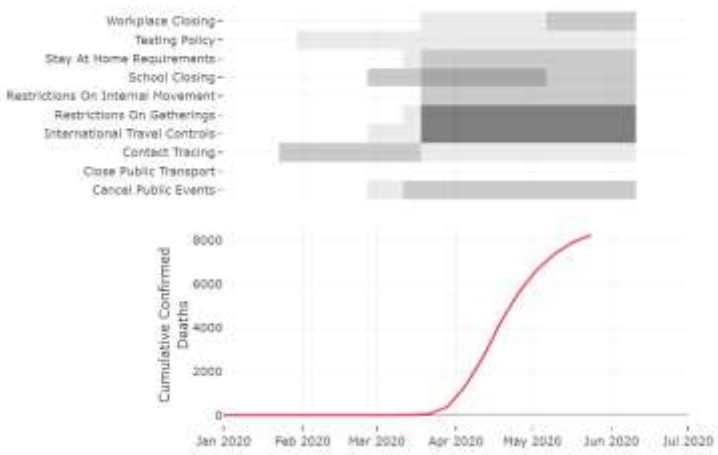
Non-Pharmaceutical Interventions

Measured on all Countries and each US State

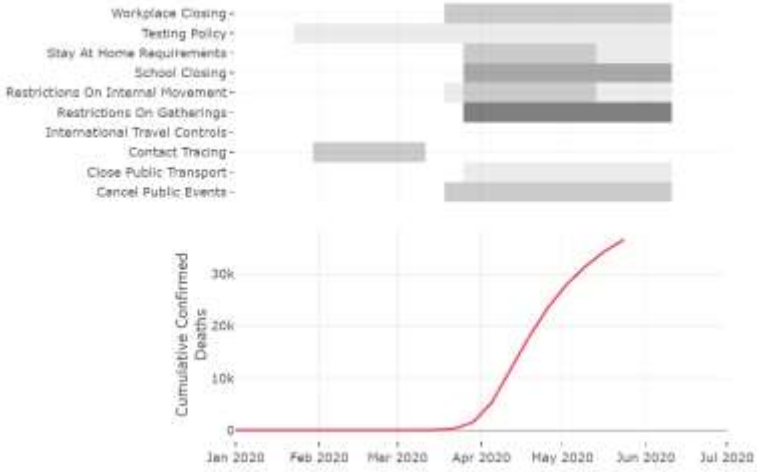
- School closing
- Workplace closing
- Cancel public events
- Restrictions on gatherings
- Close public transport
- Stay at home requirements
- Restrictions on internal movement
- International travel controls

Overview of Government Intervention Strategies

Germany



United Kingdom



The government strategies drive the forecasts of daily deaths/ cases – which then drive government strategies in the future

Government Intervention Strategies for each Country

Forecasting COVID19

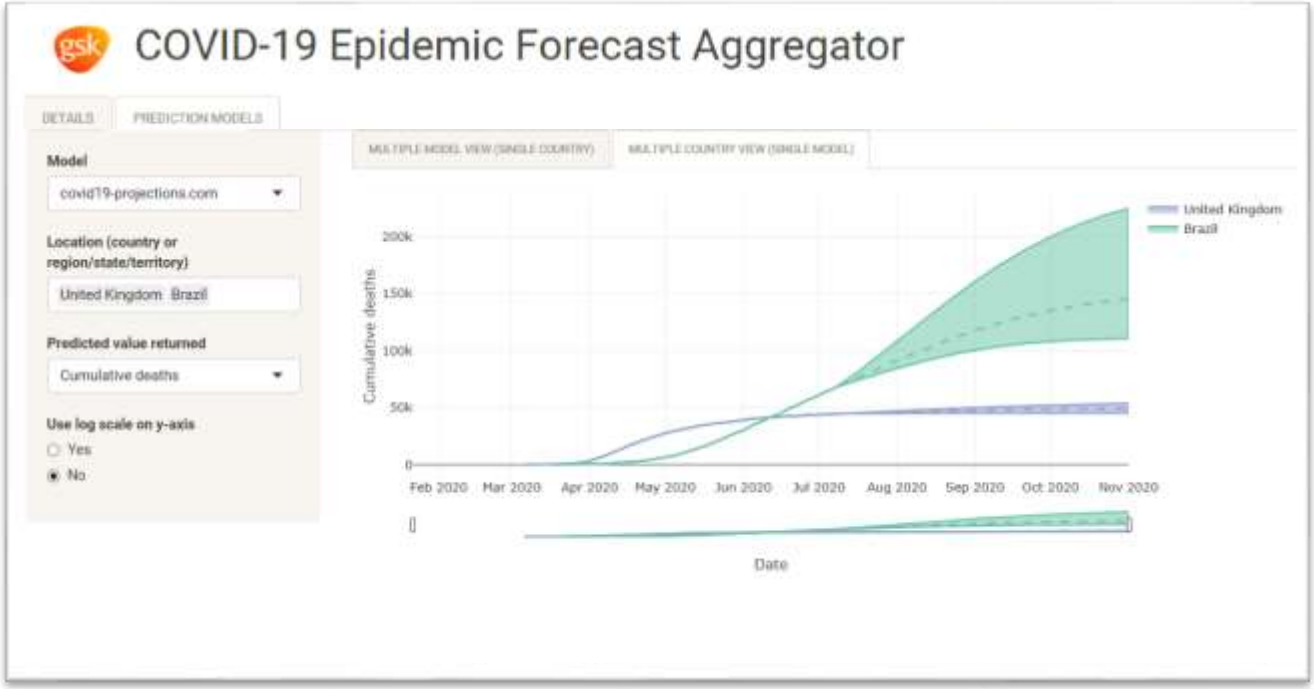
1. Most relaxed government interventions observed once the most stringent were relaxed – unlocking

2. Most stringent government interventions ever observed – relocking

3. Reactive NPIs – unlock until daily deaths pass some threshold, then re-lock until deaths are below threshold again

Epidemic Forecasts

Key features of the epidemic visualisation app

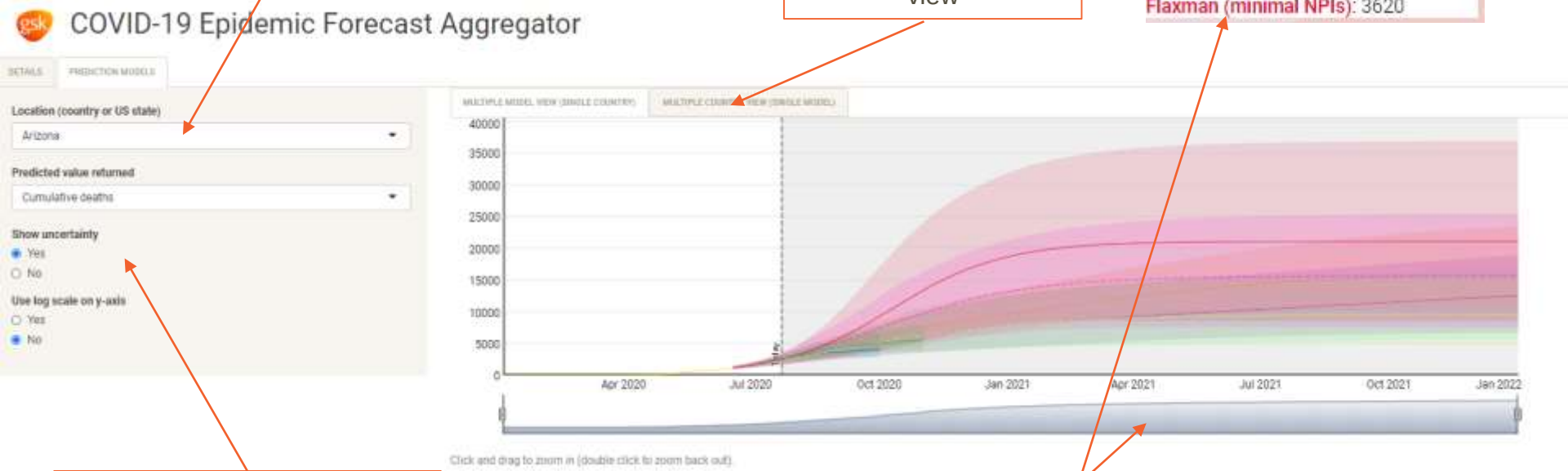


- Two viewing options provided: (i) multiple model, or (ii) multiple country
- Select a location and a value to plot (e.g., cumulative deaths)
- Plotting options can be tweaked to aid interpretation, switching off uncertainty bands or plotting on the log scale
- Interactive plotting features are also provided showing data under the plot and a slider along the bottom
- Multiple country view offers the same level of interactivity

Select Country/US state etc

Select a single or multiple geography view

Jan, 13, 2021:
Flaxman (reactive NPI 5 per mill): 3620
Flaxman (reactive NPI 2 per mill): 2658
Flaxman (reactive NPI 1 per mill): 2079
Flaxman (cyclic NPIs): 2589
Flaxman (max NPIs): 1878
Flaxman (minimal NPIs): 3620

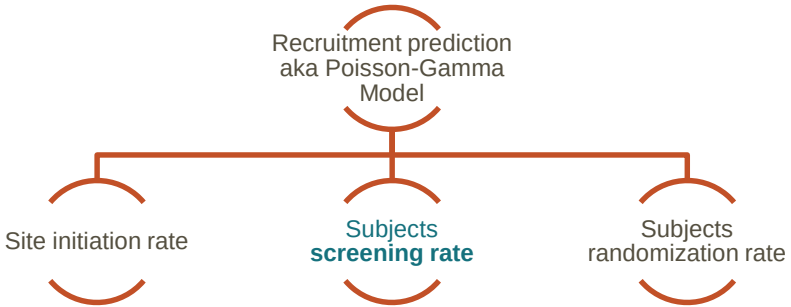


Cumulative or absolute deaths

Moving timeline, estimates provided

Recruitment modelling with COVID: Data

- Recruitment data extracted from StudyOptimizer software
  - Used for study planning and re-planning
- Actual vs planned recruitment from 22 studies across 49 countries
- Recruitment comprises 3 main processes:
  - Site initiation
  - Screening
  - Randomization



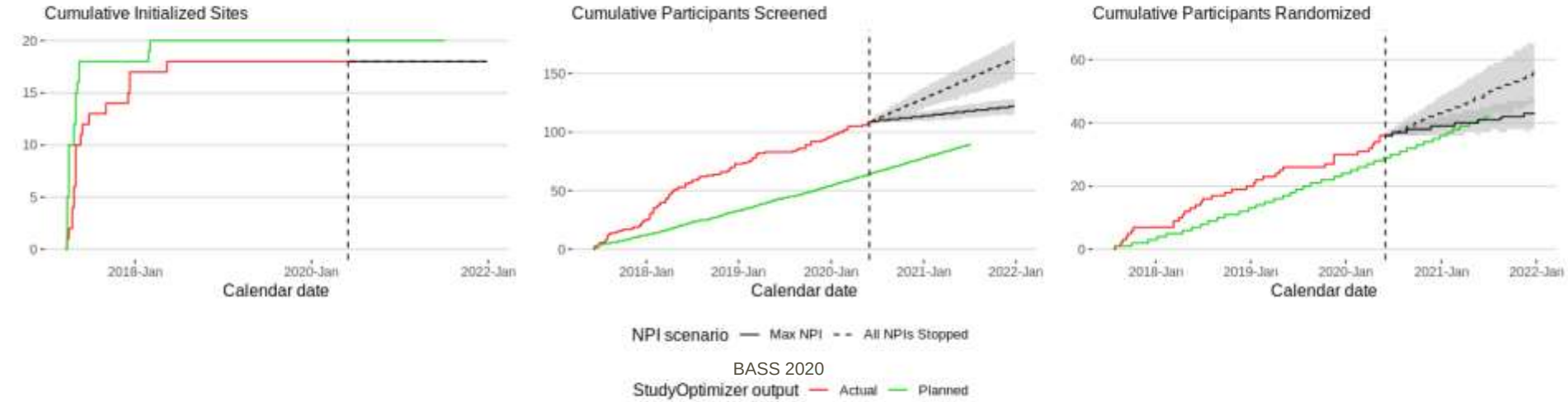
Recruitment modelling with COVID: Data

Key focus: understanding the “gap” between actual and planned recruitment curves

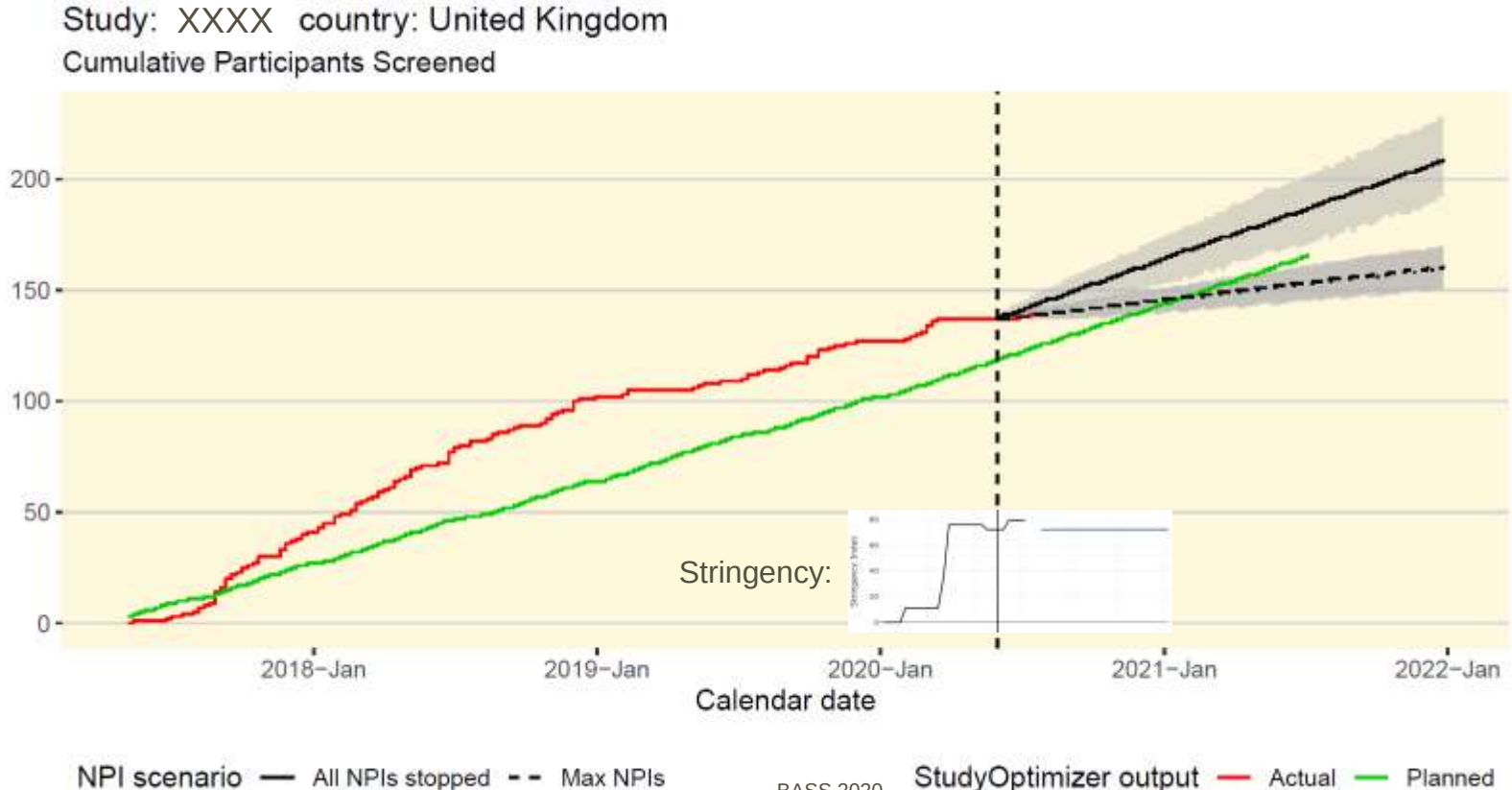
Q: How much of planned/actual gap can be explained by COVID impact ?

- Each process measured in terms of weekly numbers, for both planned and actual recruitment
- Wish to compare actual to planned recruitment but main focus on understanding the impact of COVID-19

Study XXX; country: Czechia

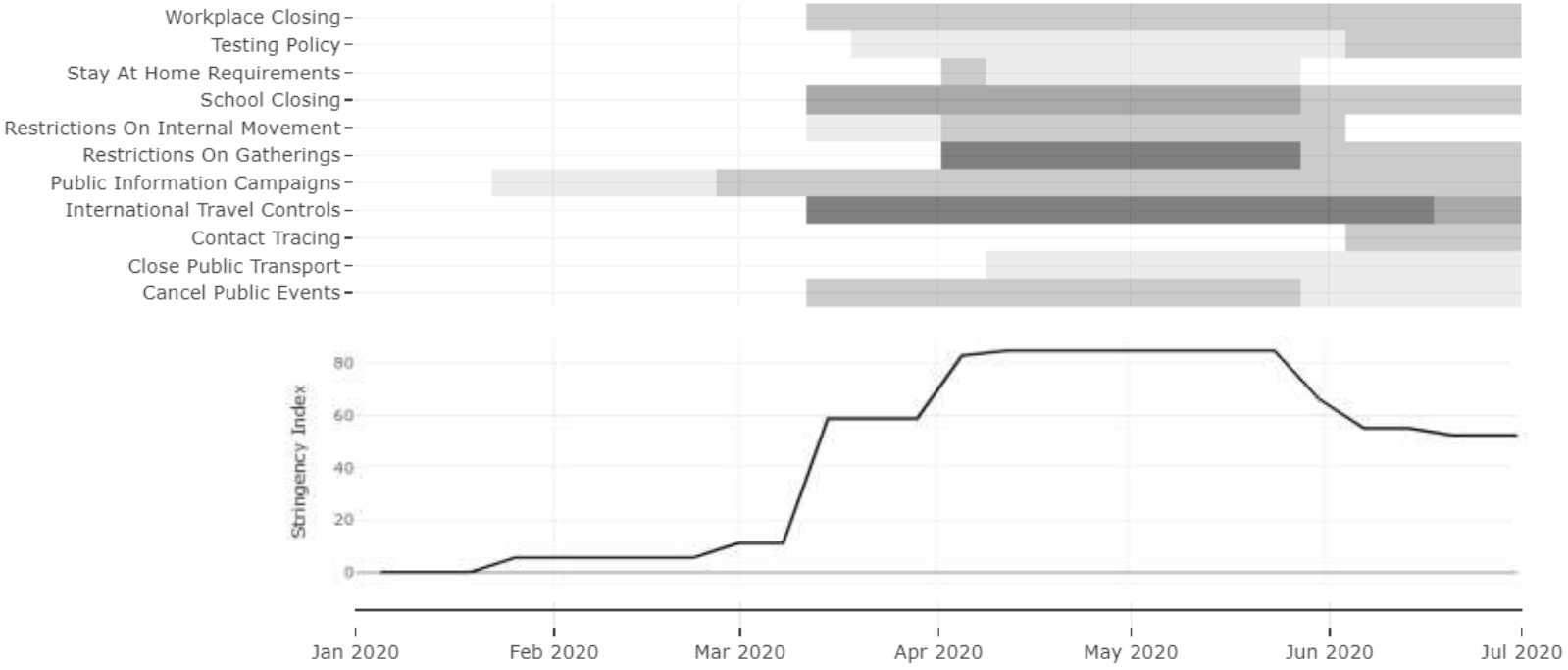


COVID-19 impact



External data: COVID19 statistics

NPI Stringency index (Poland as example)



Model overview

- Each component of recruitment process (initiation/screening/randomization) is characterised by a rate parameter, e.g. number of sites initiated per week
- For each process we say:
  - Actual rate = Planned rate x Multiplier
- So, for example, a screening-rate multiplier of 0.5 means that actual screening is progressing at half the planned rate (assuming the same number of sites)
- We allow the multiplier to depend on various COVID-19-related metrics, e.g.
  - Number of cases/deaths
  - Stringency of government-imposed restrictions

  ← Come form COVID prediction model
- Multipliers are of direct interest (modelled) but main focus is on recruitment milestones (e.g. LSFV); that's what is used in decision -making

Model is complicated…..(see next slide)

Here is snapshot at **screening component** to illustrate modelling of multipliers

- $n_{scj} \sim Poisson(\mu_{scj} e_{scj})$
- Here μ_{scj} is the screening rate *per-site*, *per-week*
- And e_{scj} is the total "exposure to screening":
- $e_{scj} \approx \sum_{t=1}^{j} \nu_{sct}$ (the cumulative number of sites initiated)
- We relate the actual screening rate to the planned rate via
  - $\mu_{scj} = B_{cj} \times \mu^{P}_{sc}$
  - $\log B_{cj} = \alpha_{\lambda_c} + X_{cj}\beta_{\lambda}$

COVID-related

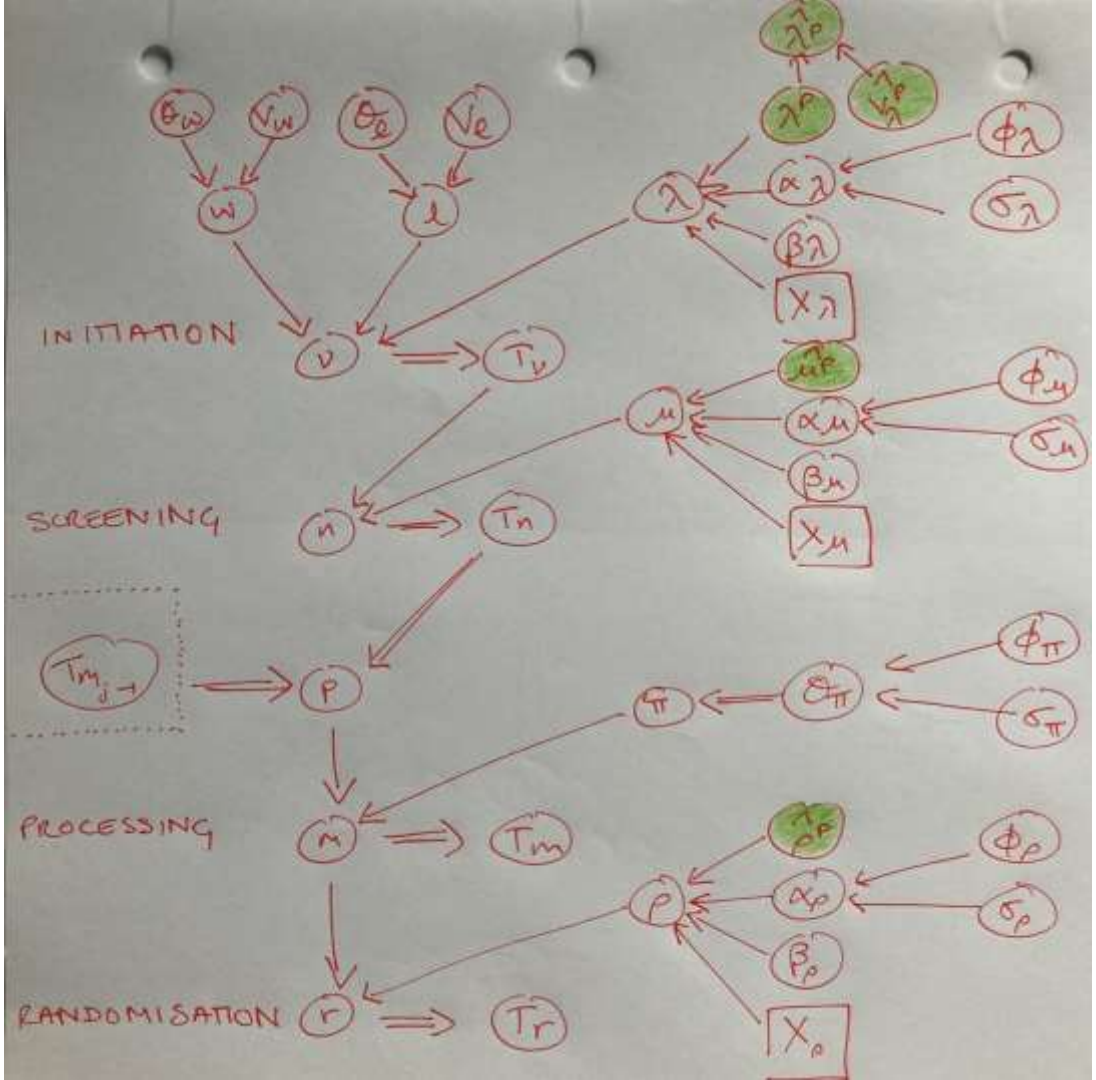
- Here μ^{P}_{sc} is estimated empirically from the plan data (but we could have used the above)

Notation:

- s index studies
- c index countries
- j index time-periods (weeks)
- $\nu_{\cdot}$ denotes the (weekly) number of sites initiated [rate = $\lambda_{\cdot}$]
- $n\_{\cdot}$ denotes the (weekly) number of individuals arriving for screening [rate = $\mu\_{\cdot}$]
- P superscript and color green denotes "planned" as opposed to "actual" (red)

Actual in red, planned in green, with P index

DAG:

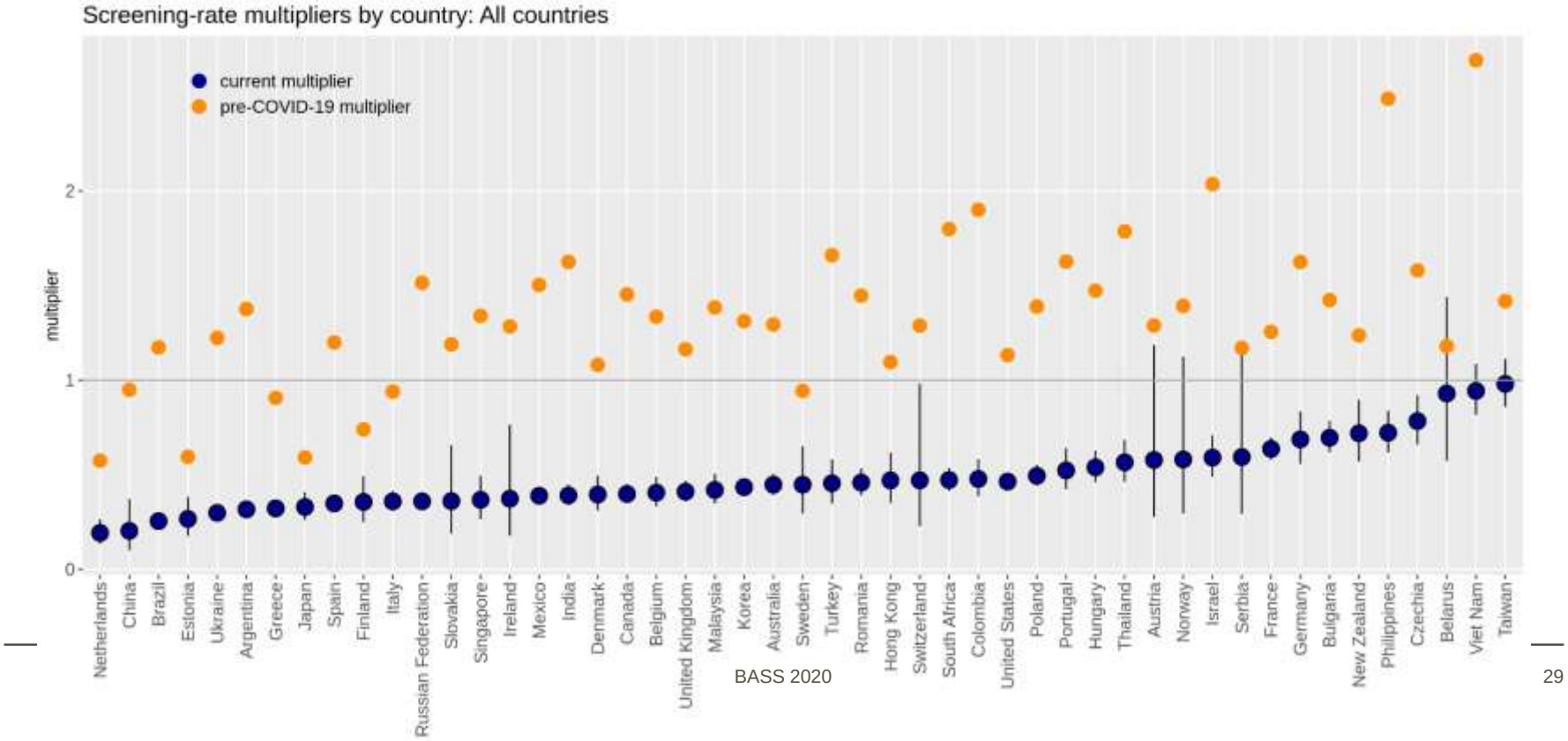


Predictions

- Multipliers vary over time with COVID-19 metrics
- Can predict how recruitment will progress under various pandemic scenarios, e.g.
  - No restrictions going forwards
  - Maximal restrictions
  - Restrictions cycle over the next year or two
- Primary focus on estimating dates at which recruitment targets achieved, e.g.
  - Full recruitment, 75%, 50%, etc
- Best estimate + confidence interval, say, available
- Or percentile-based estimates, e.g. we can be 80% confident that we will have achieved 50% recruitment by Aug 2021

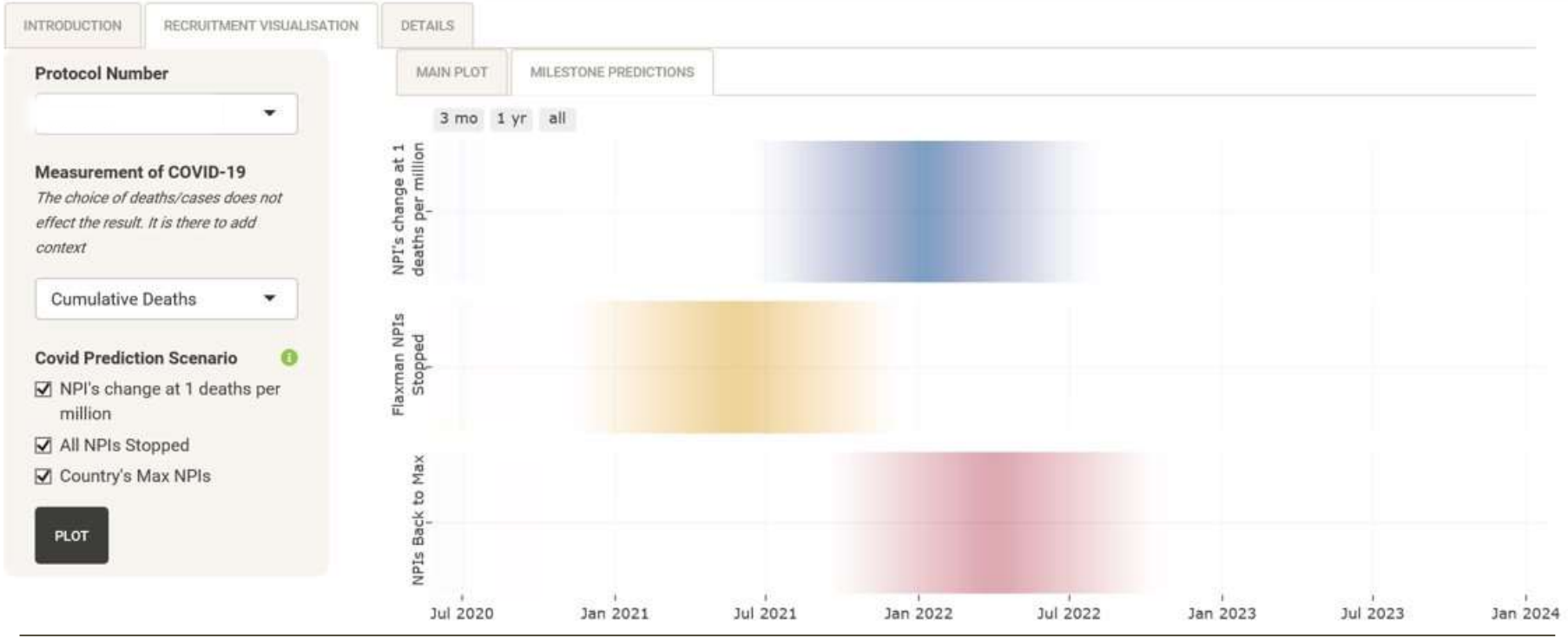
Predicted screening-rate multipliers by country

All countries…



Milestone estimates in Recruitment App

View probability distributions (density strips) for the LSFV visit milestone



Milestone estimates: decision making

LSFV percentiles

Probability of Reaching LSFV (%):

75

These could be used to set target dates

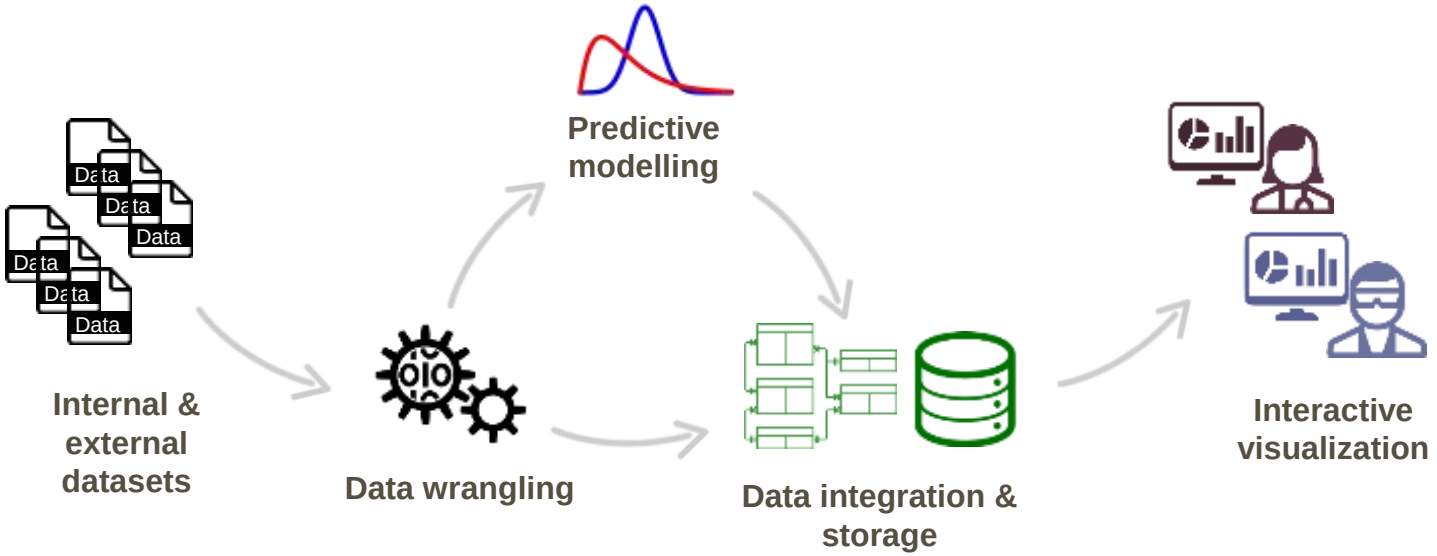
Target Probability	Scenario	Date Closest to Target Probabilty	Probability on Date
5%	All NPIs Stopped	2021-05-02	3.4%
	Country's Max NPIs	2022-02-27	4%
	NPI's change at 1 deaths per million	2021-11-28	4.8%
50%	All NPIs Stopped	2021-05-23	47.8%
	Country's Max NPIs	2022-03-27	47.9%
	NPI's change at 1 deaths per million	2022-01-02	43.3%
75%	All NPIs Stopped	2021-05-30	70.2%
	Country's Max NPIs	2022-04-10	78.4%
	NPI's change at 1 deaths per million	2022-01-16	71.7%
95%	All NPIs Stopped	2021-06-13	94.8%
	Country's Max NPIs	2022-04-24	95.9%
	NPI's change at 1 deaths per million	2022-02-06	96.1%

How to make a practical tool out of all this complex modelling:

View from a Data Scientist perspective (Magda)

Data science workflow

Turning data into insights



Data science workflow

Constant communication and feedback enabled by technology

Data wrangling and integration

Preparing input for the models and the interactive applications

Datasets we used:

- **COVID-19 statistics**: deaths and cases per country over time
- **NPIs** per country over time
- **NPI stringency index** per country over time
- **Recruitment data** per study/country over time
  - Randomized participants
  - Screened participants
  - Initiated sites
- **Pre-COVID19 recruitment plans** over time

Derived values and model predictions integrated with the data:

- Future predicted COVID-19 deaths and cases
- Future predicted NPIs and NPI stringency indexes
- Recruitment predictions for individual scenarios
- Milestone probabilities for individual scenarios

All these datasets have been created for different purposes and they do not follow the same data standards.

Data wrangling and integration

Preparing input for the models and for the interactive apps

Data standardization:

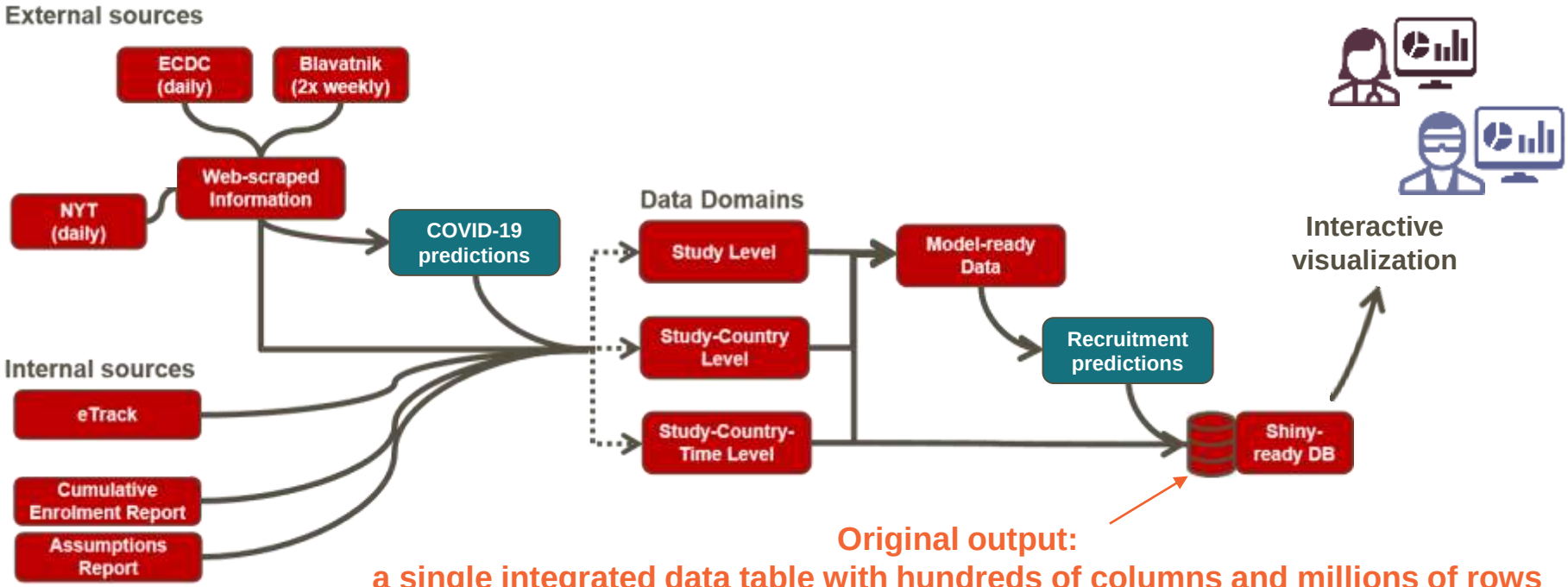
- Country vs state-level data
  - US state-level data require aggregation
  - Georgia (the state) vs Georgia (the country)
- Different country names in individual datasets
  - Russia -> The Russian Federation
- Different approaches to missing values
  - 0s, NAs, missing rows, inferred values
- Different approaches to date formatting
- Weekly vs daily statistics
- Different update cadence
  - weekly vs biweekly updates
- Long datasets vs wide datasets
- ….

Data transformation:

- Deriving new variables or summaries
- Filtering
- Calculating rates, cumulative values, etc.
- Removing duplicates
- Deriving missing values
- …

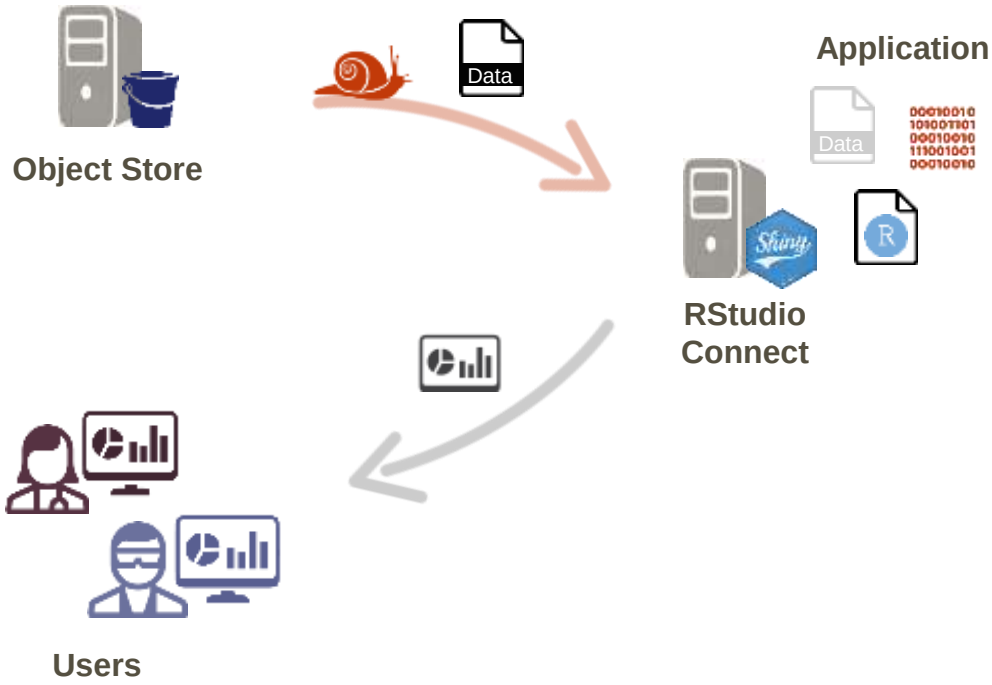
Data integration overview

Putting it all together



Original output:
a single integrated data table with hundreds of columns and millions of rows

Drawbacks of working with a single large dataset



Workflow:

- Big dataset file copied from the object store
- All data loaded into the memory
- Additional smaller datasets generated for individual plots

Issues:

- **Dependency** on the object store being alive and well
- **Time** to **copy** the data and **load** them into the memory

Ultimate problem: 1) dataset is **too big** and 2) data are **not used efficiently**

Improving data storage by data normalization

As explained on a dummy dataset

Student id	Student Name	Student Gender	Subject	Professor id	Professor Name	Professor Phone Number	Professor Room Number
100	Percy Vere	Male	Maths	501	Prof. Anita Letterback	501-340-376	31
100	Percy Vere	Male	CS	612	Prof. Laura Biding	509-362-575	23
123	Constance Noring	Female	Maths	546	Prof. Chris Anthemum	748-934-645	54
123	Constance Noring	Female	Engineering	546	Prof. Chris Anthemum	748-934-645	54
123	Constance Noring	Female	CS	612	Prof. Laura Biding	509-362-575	23
256	Hugo First	Male	-	-	-	-	-

Problems: data redundancy & Null values

Improving data storage by data normalization

The same dummy dataset – normalized (decomposed into tables)

Student

Student id	Student Name	Student Gender
100	Percy Vere	Male
123	Constance Noring	Female
256	Hugo First	Male

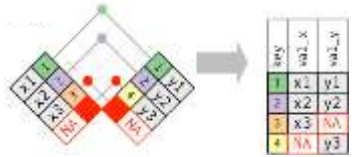
Lecture

Student id	Subject	Professor id
100	Maths	501
100	CS	612
123	Maths	546
123	Engineering	546
123	CS	612

Professor

Professor id	Professor Name	Professor Phone Number	Professor Room Number
501	Prof. Anita Letterback	501-340-376	31
612	Prof. Laura Biding	509-362-575	23
546	Prof. Chris Anthemum	748-934-645	54

Joining tables



Improving data storage by data normalization

Dummy dataset before and after data normalization

6 rows X 8 cols = **48**

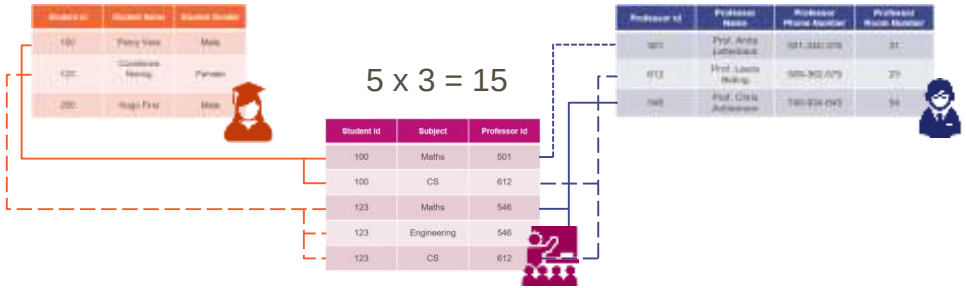
Student id	Student Name	Student Gender	Subject	Professor id	Professor Name	Professor Phone Number	Professor Room Number
100	Percy Vere	Male	Maths	501	Prof. Anita Letterbeck	561-340-570	31
100	Percy Vere	Male	CS	612	Prof. Laura Biding	509-362-575	23
123	Constance Nesing	Female	Maths	546	Prof. Chris Anderson	749-934-643	34
123	Constance Nesing	Female	Engineering	546	Prof. Chris Anderson	749-934-643	34
123	Constance Nesing	Female	CS	612	Prof. Laura Biding	509-362-575	23
256	Hugo Pond	Male	-	-	-	-	-

9 + 15 + 12 = **36**

3 x 3 = 9

3 x 4 = 12

5 x 3 = 15



Advantages of data normalization

- Less redundant data and fewer null values
- Reduced storage size
- Increased data integrity
- Intuitive data schema

Byte-wise representation:

Professor name vs Professor id

“Prof. Laura Biding” = **18 bytes**

612 = **2 bytes* / 4 bytes ****

* SQLite ** R

Improving data storage by data normalization

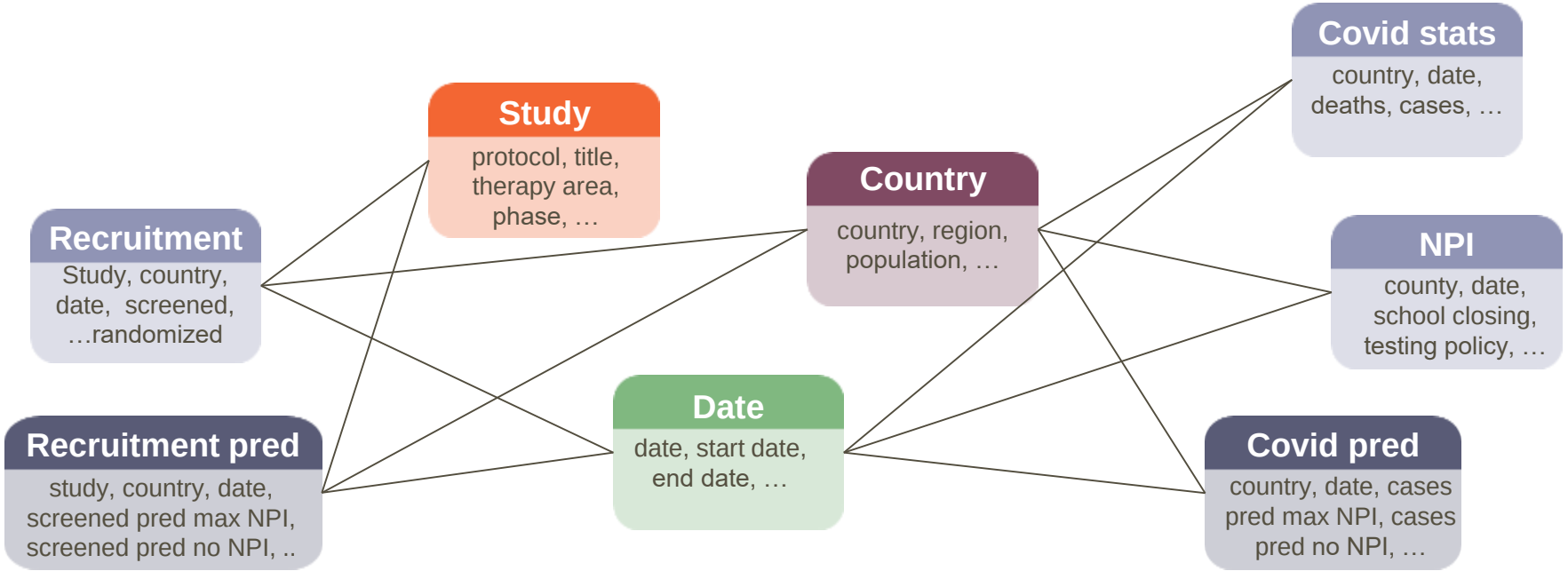
Dummy version of the recruitment dataset (a very tiny subset of it)

Study	Country	Population	Region	Date	Cases	Deaths	Initiated actuals	Screened actuals	NPI: School closing	Screened Predicted
XXX	Argentina	44494502	South America	2017-04-09	NA	NA	1	0	NA	NA
XXX	Argentina	44494502	South America	2017-06-04	NA	NA	12	5	NA	NA
XXX	Argentina	44494502	South America	2020-03-08	9	1	16	47	0	NA
XXX	Argentina	44494502	South America	2020-04-12	2137	89	18	273	3	NA
XXX	Argentina	44494502	South America	2020-10-06	NA	NA	NA	NA	NA	593
XXX	Argentina	44494502	South America	2020-11-05	NA	NA	NA	NA	NA	824

Problems: data redundancy & Null values

Improving data storage by data normalization

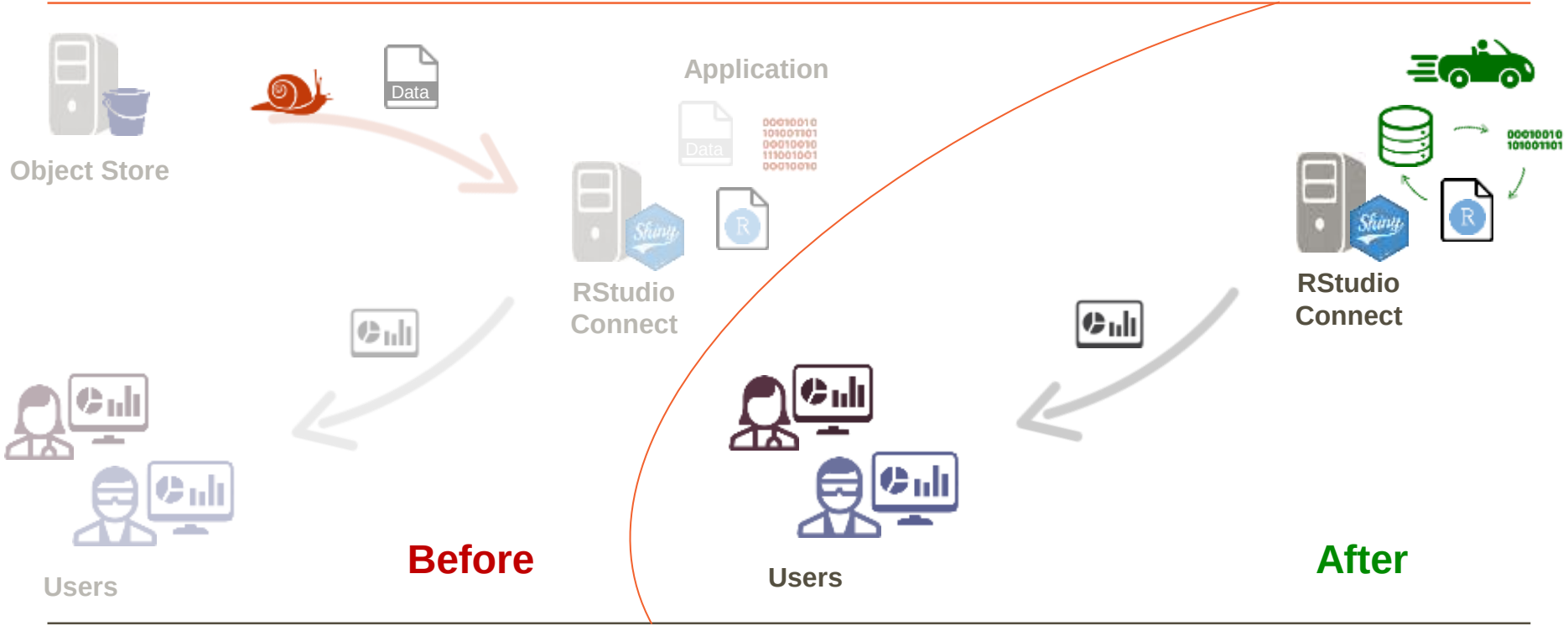
Normalized data model for Covid-19 data (simplified)



Speeding up dataset preparation with a SQLite database

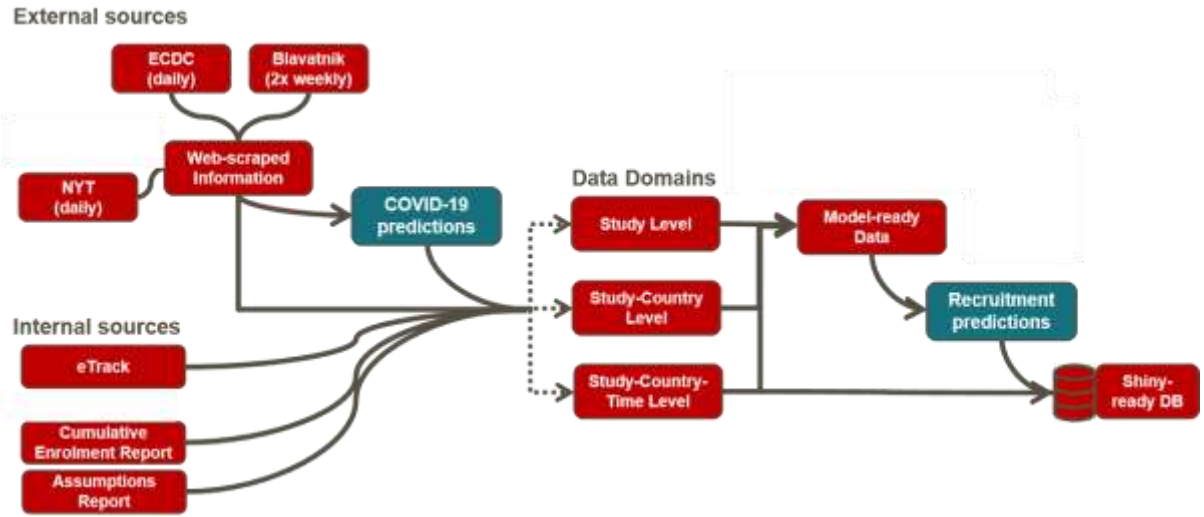
… because only a small subset of the data is needed to generate a single plot

Database-powered interactive visualization app

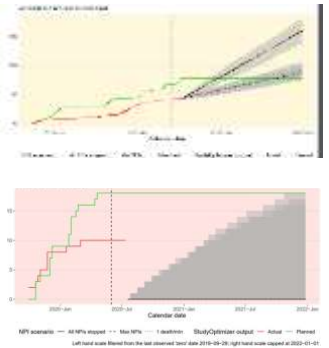


Problem:

We often find data issues at the very end of the process



QA plots

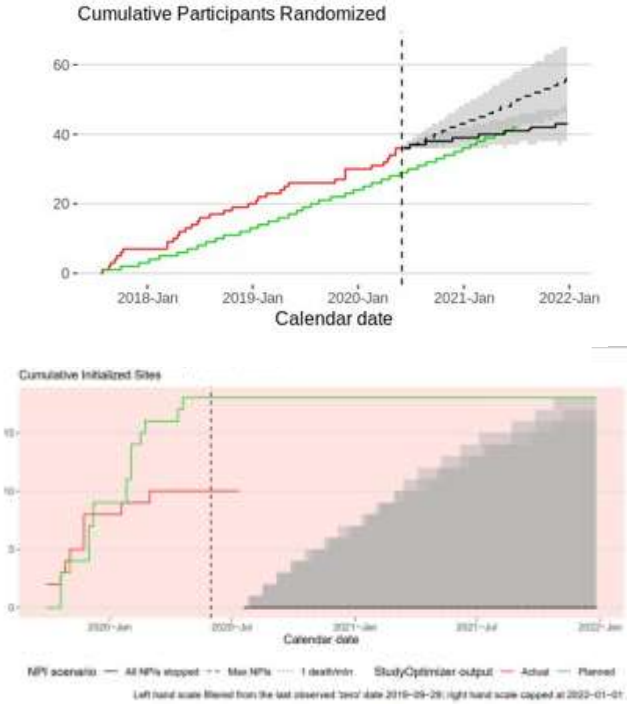


DATA PROCESSING STEPS

WHERE WE SEE MOST ERRORS

Problem:

We often find data issues at the very end of the process



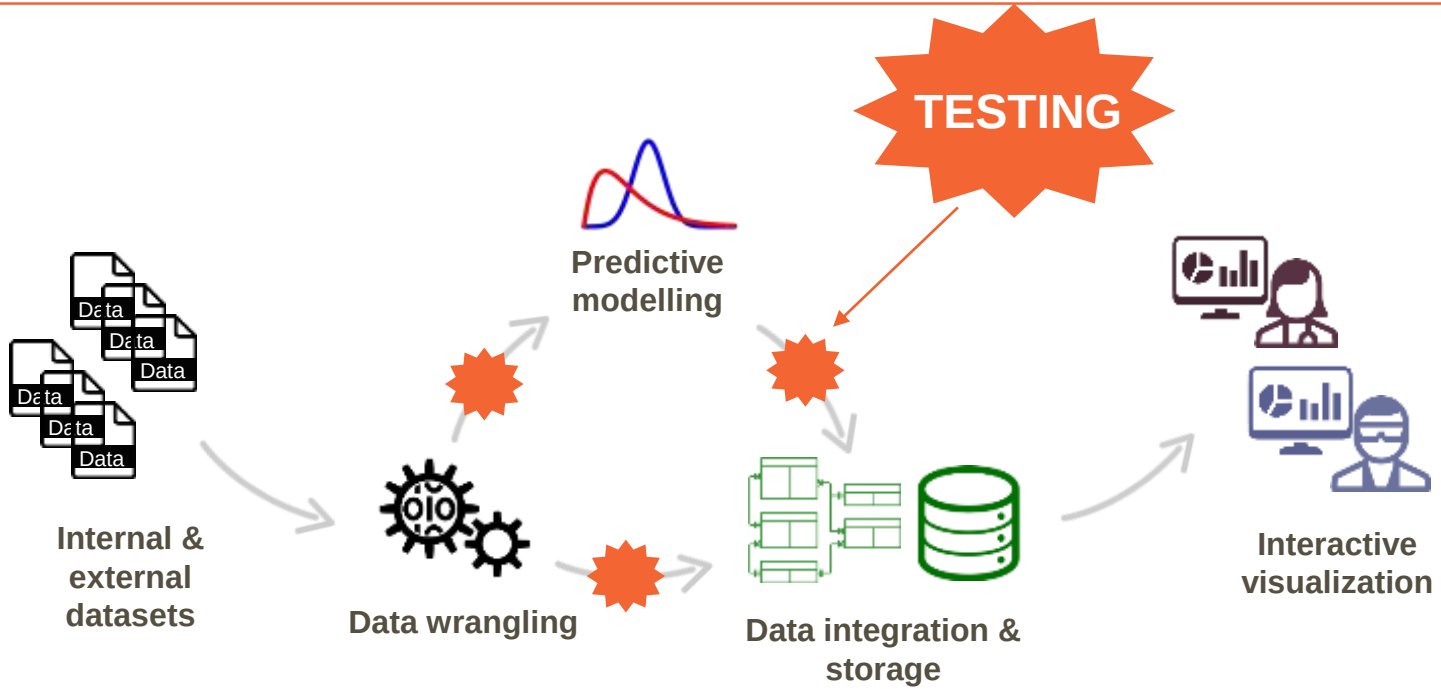
Possible data errors

- Lack of predictions for a given study-country combination
- Predictions all equal to 0 (despite higher last observed value)
- Non-monotonic predictions (**cumulative** initialized sites)
- …

All these errors can be detected automatically similar to how queries are run on clinical data.

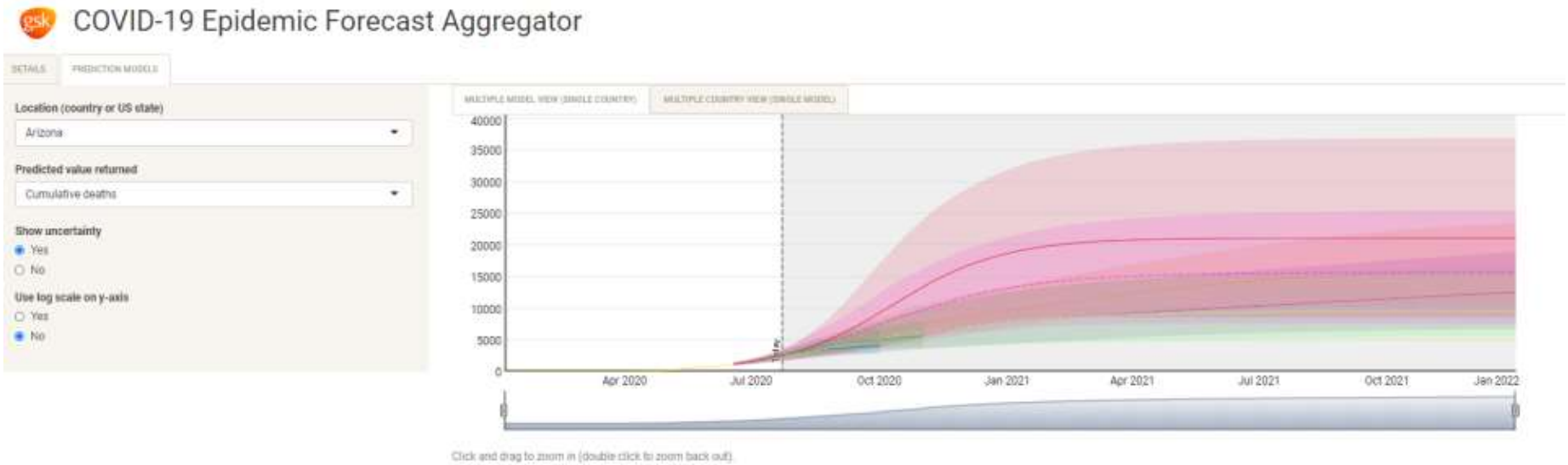
Solution:

Automated data testing to detect data issues as early as possible



Interactive web applications based on a single database

View COVID19 predictions from a variety of models

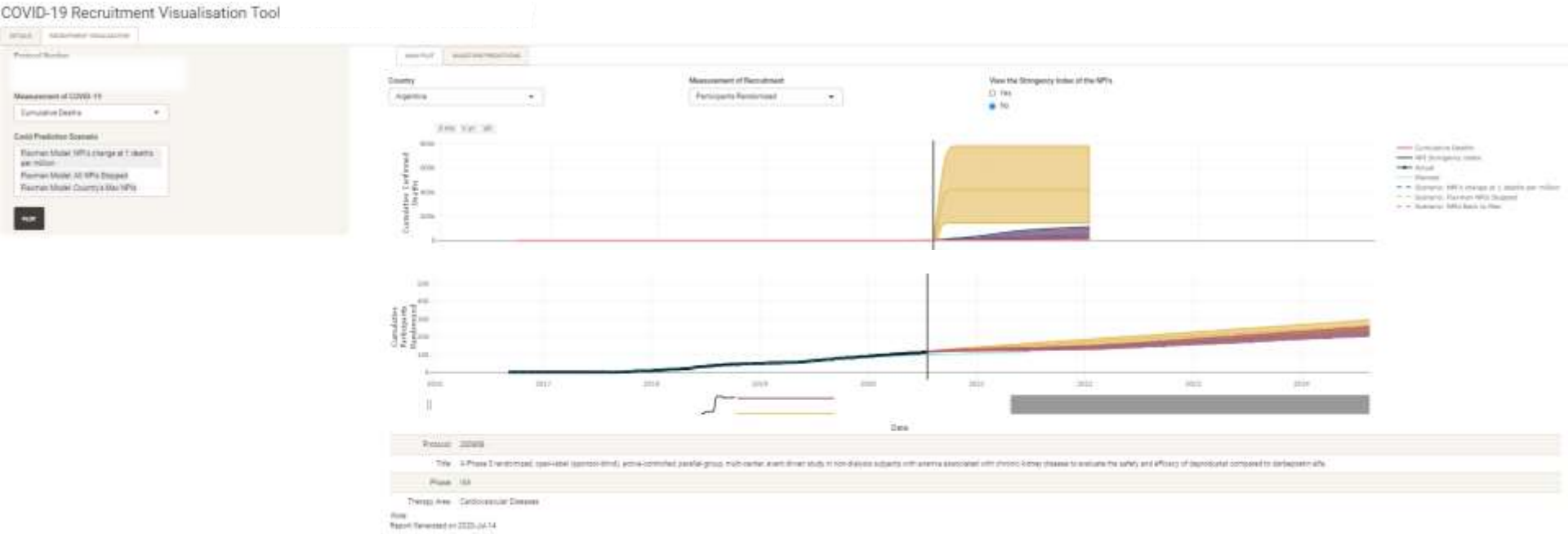


Interactive web applications based on a single database

View Covid19 statistics and predictions on an animated map

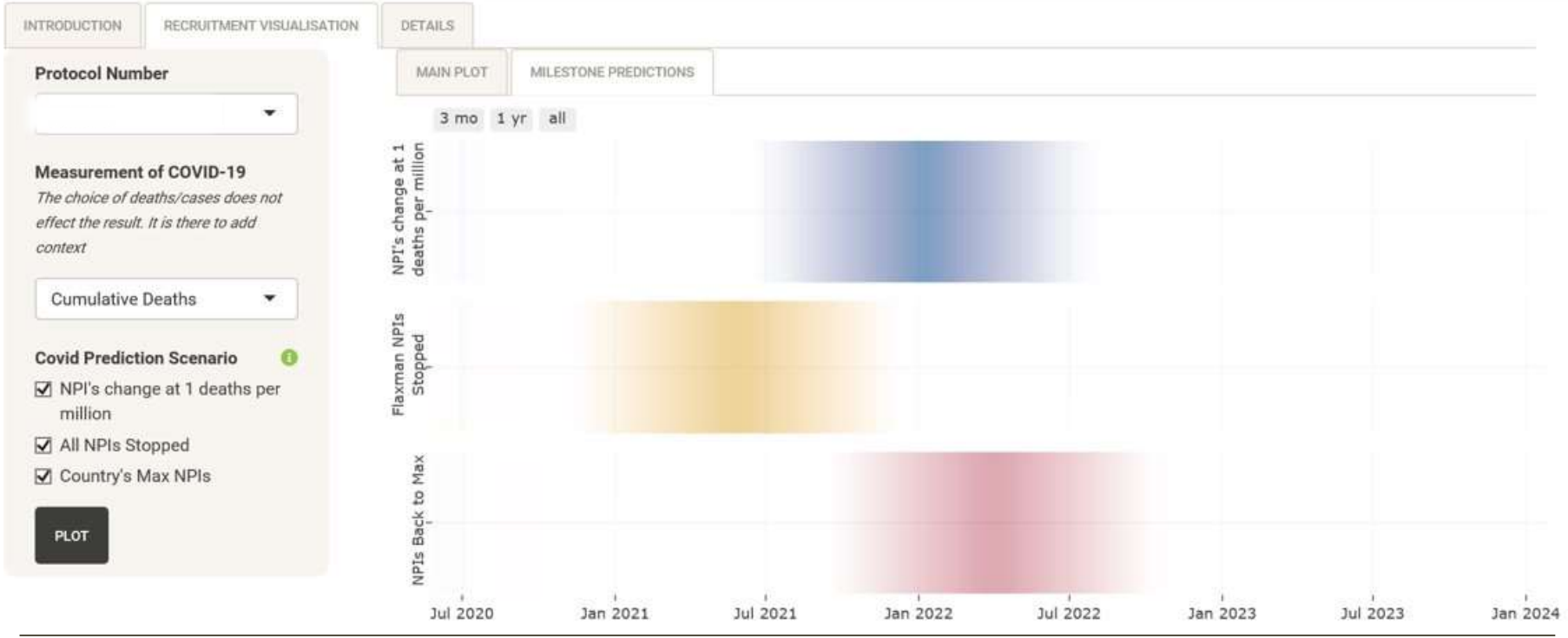
Interactive web applications based on a single database

View detailed recruitment predictions per study



Interactive web applications based on a single database

View probability distributions for the last subject first visit milestone



Data science workflow - enablers

Tools and packages

Data wrangling

Modelling

Data and code storage

Visualization

Data science workflow - enablers

People

- Diverse expertise
- Constant communication
- Knowledge sharing
- Frequent stand ups
- Agile mindset
- Shared codebase
- Pair programming
- Focus, motivation & team work

All of this remotely and across continents!

Summary

- We have presented work that resulted from intense collaboration within a multi-disciplinary team of statisticians, data scientists and clinical operation professionals over the course of 5 month since pandemic started
- It builds on existing well-developed stochastic Poisson-gamma model commonly used for recruitment prediction
  - To properly account for uncertainties in recruitment processes, the model has 3 components: site initiation, patients screening and randomization
- COVID pandemic has affected all 3 sub-processes of recruitment
- We have build a Bayesian hierarchical model that “explains” how the rates of all 3 processes are affected by COVID-related covariates such as NPI stringency index

Summary (cont.)

- In the process, we also built a standalone COVID forecaster App which can be viewed as aggregator of publicly available COVID-19 models and GSK custom implementation of the Imperial College Flaxman model
  - Can be used independently of recruitment app to assist teams with planning (to plan initial footprint of studies based on COVID situation)
- The work is far from being done: this phase was a “pilot” for immediate impact using recruitment data from a subset of GSK studies.
- We are entering implementation phase where study teams “test-drive” it using their own on-the-ground insight into recruitment process and COVID disruptions
  - With more data and more insight from experts, the model is likely to be further refined
- The long term goal is to embed this kind of thinking and tools into routine processes and systems within Clinical Operations

It’s a long journey…..

But an exciting one!!!

Thank you!