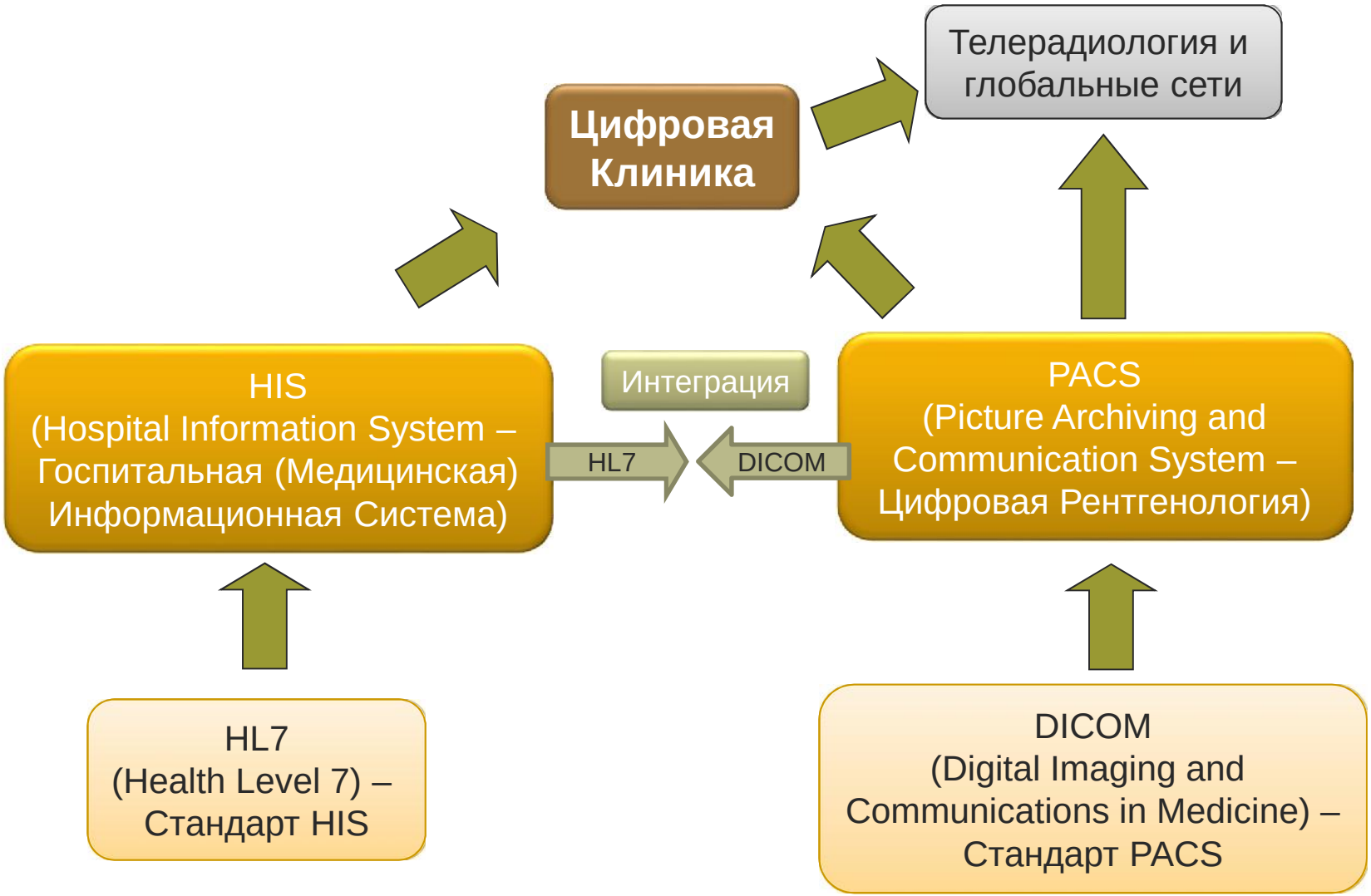
Медицинская Информатика ?

Олег Пьяных
opiany@gmail.com

18.05.2015

Цифровая медицина



Олег Пьяных, opiany@gmail.com

Стандарт HL7

ADT (Admission, Discharge, Transfer) message

Pipe

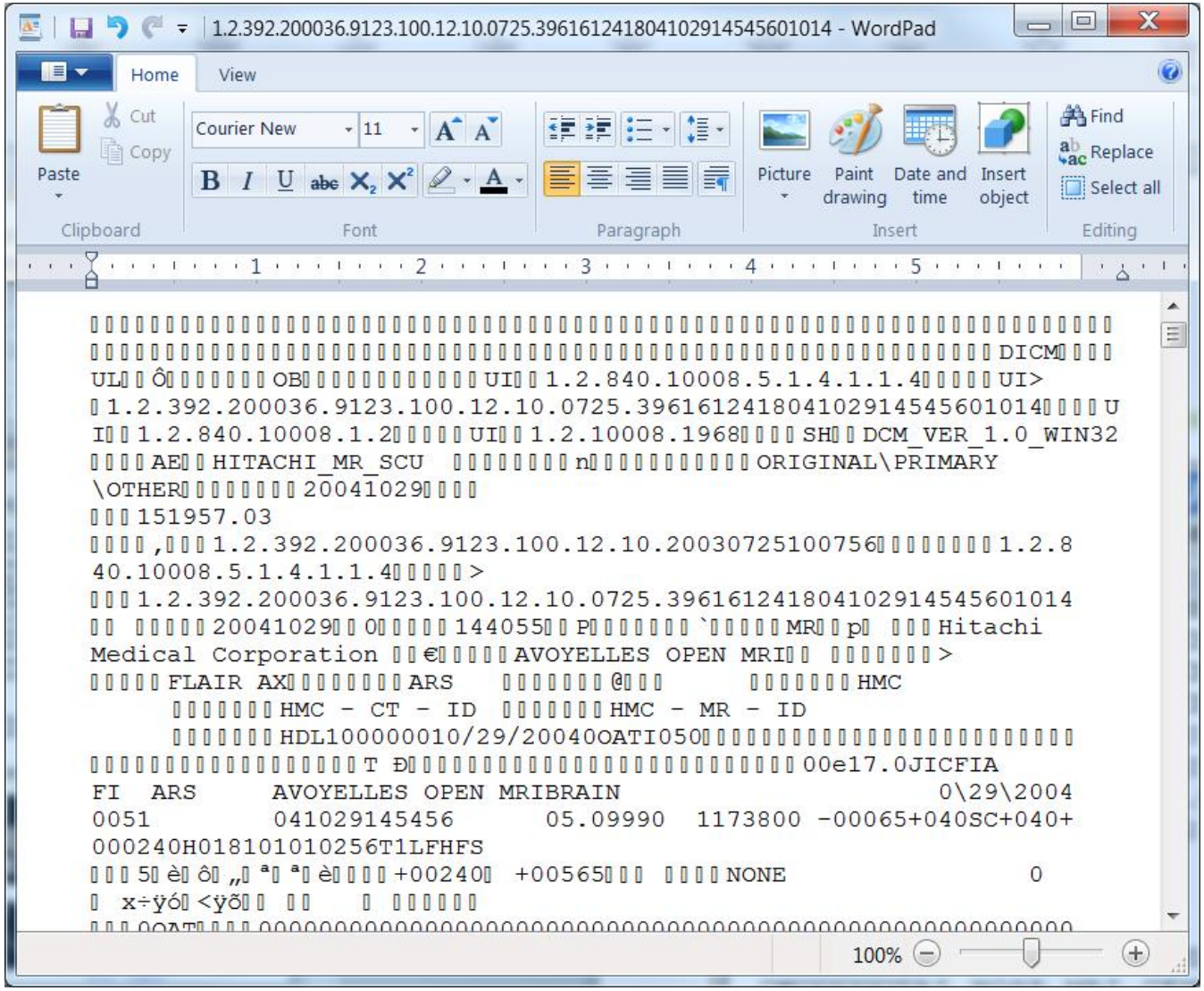
- MSH|^~\&|ADT|N|ADT|MEDSC|200601081527||ADT^A08|RE|P|3.2|||||ASCII|
- EVN|A08|200601080823|||||PID|1||3175875|1127278|SAMPLE^JOE^^^^||19901334|M|||5400 Lake Villa Dr^^Metairie^LA^70001-1230||(405)555-2920|||SINGLE||||||||||||N|
- MSH|^~\&|ADT|N|ADT|MEDSC|200601081527||ADT^A08|RE|P|3.2|||||ASCII|EVN|A08|200601080832|||||PID|1||3057088|1051999|INCOGNITO^MONICA^ANN^^^||19780117|F|||PO Box 1324^^Jefferson^LA^83625-3184||(555)423-1423|||OTHER|||512-11-1425||||||||||N|

Найдите ошибку

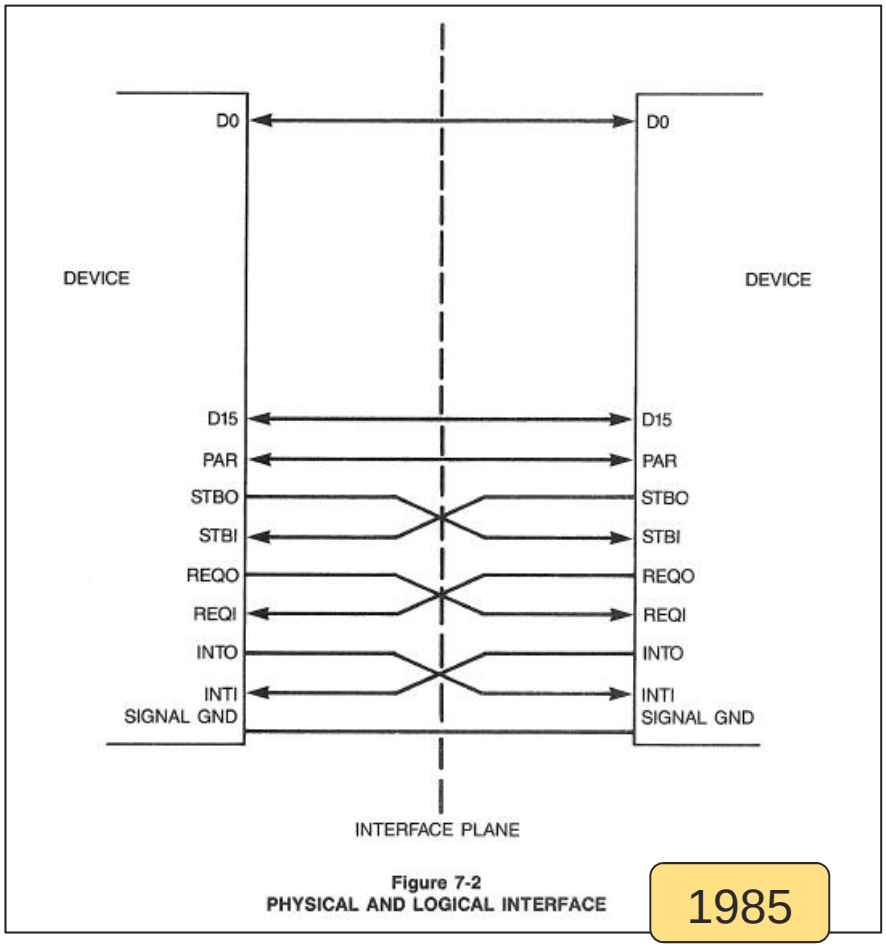
See http://www.hosinc.com/products/interfaces/interface_documentation.htm

Oleg Pianykh opiany@gmail.com

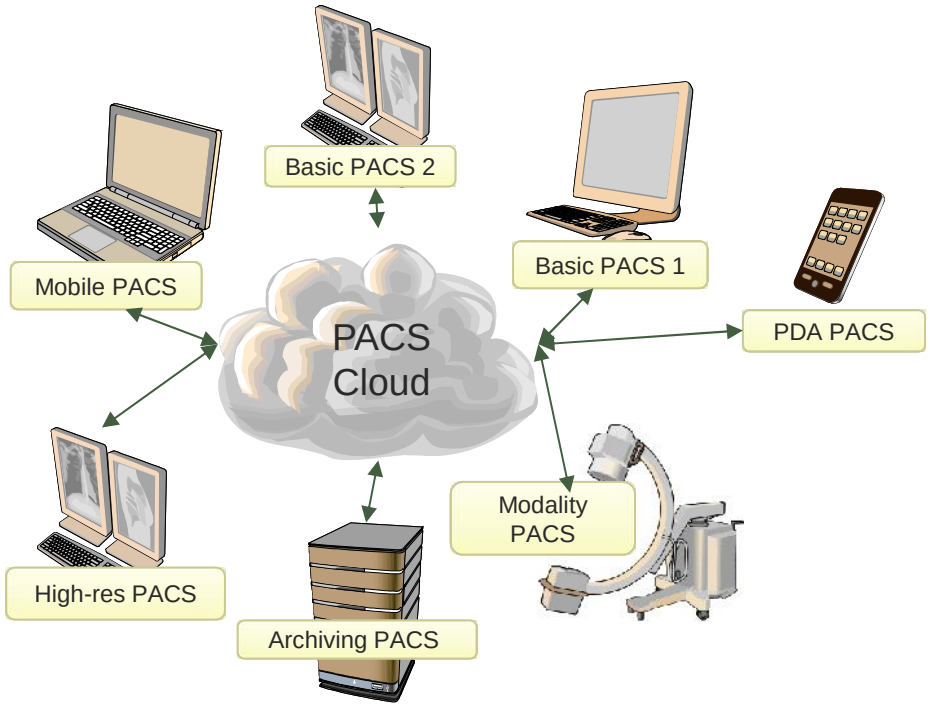
Стандарт DICOM



Эволюция цифровой медицины

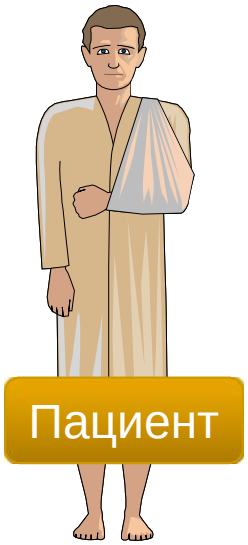


1985

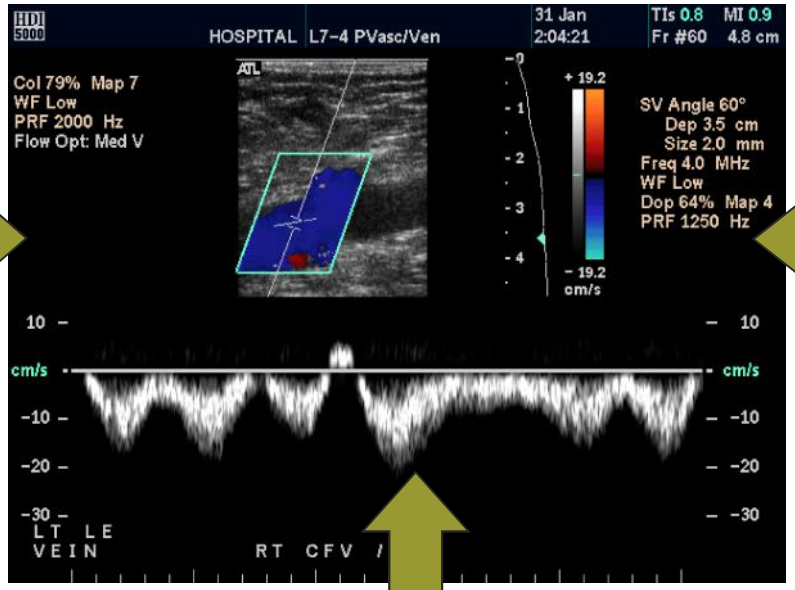


Теперь

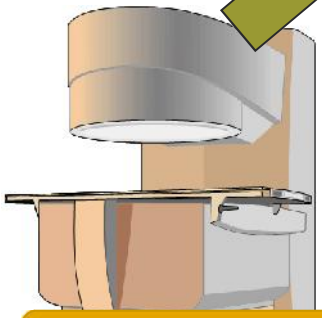
Сбор информации в медицине



Пациент

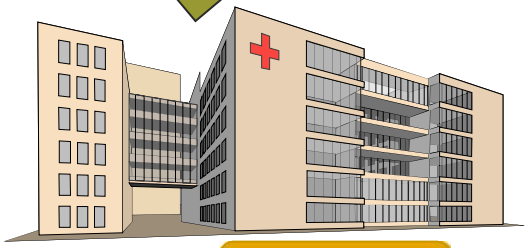


Врач



Модальность

Параметры исследования



Клиника

Олег Пьяных opiany@gmail.com

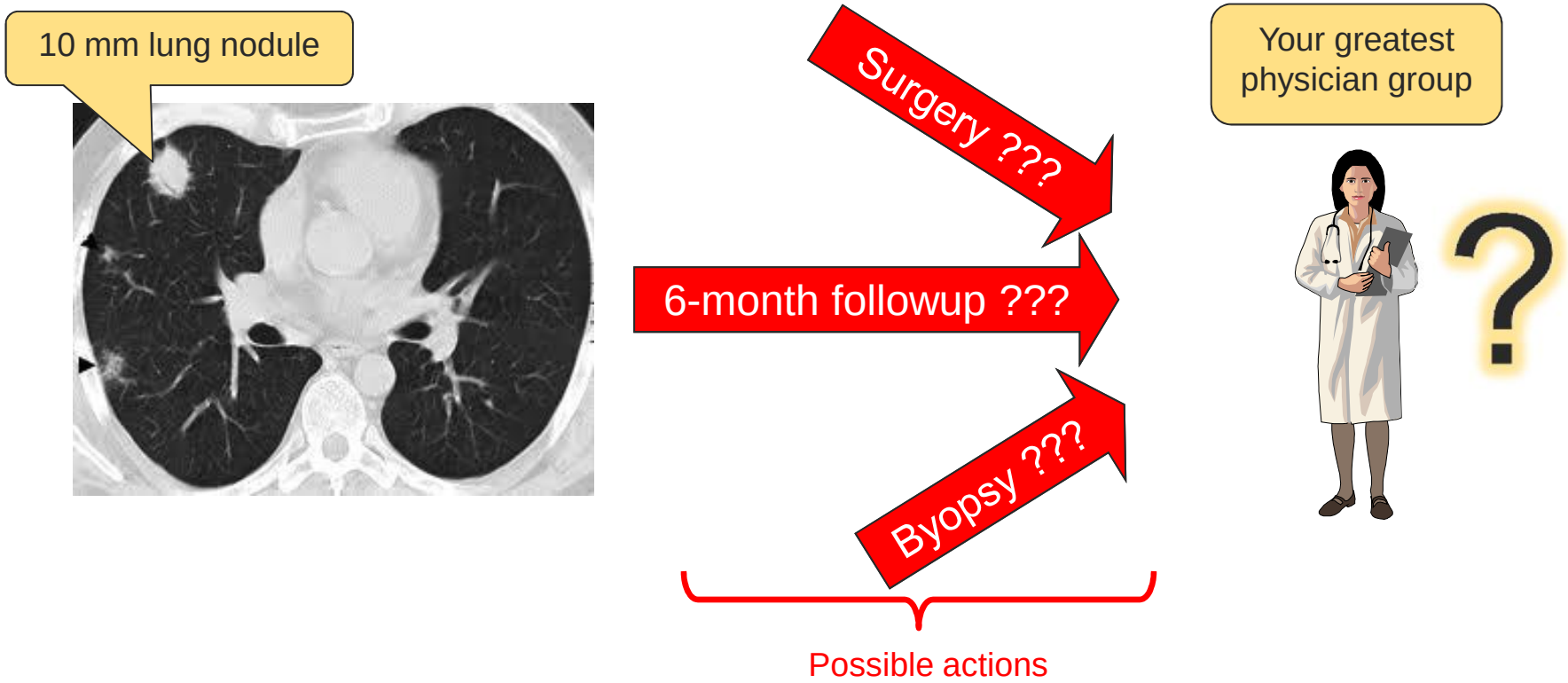
Информация собрана. *Что дальше?*

- Мы уже 20 лет как накапливаем информацию о всевозможных параметрах и процессах в здравоохранении. Не пора ли начать её использовать?
  - Выявление новой логики и закономерностей
  - Показ отклонений от нормы
  - Управление процессами и принятие решений
  - Оптимизация и стандартизация медицины
- Все это требует эффективных математических решений

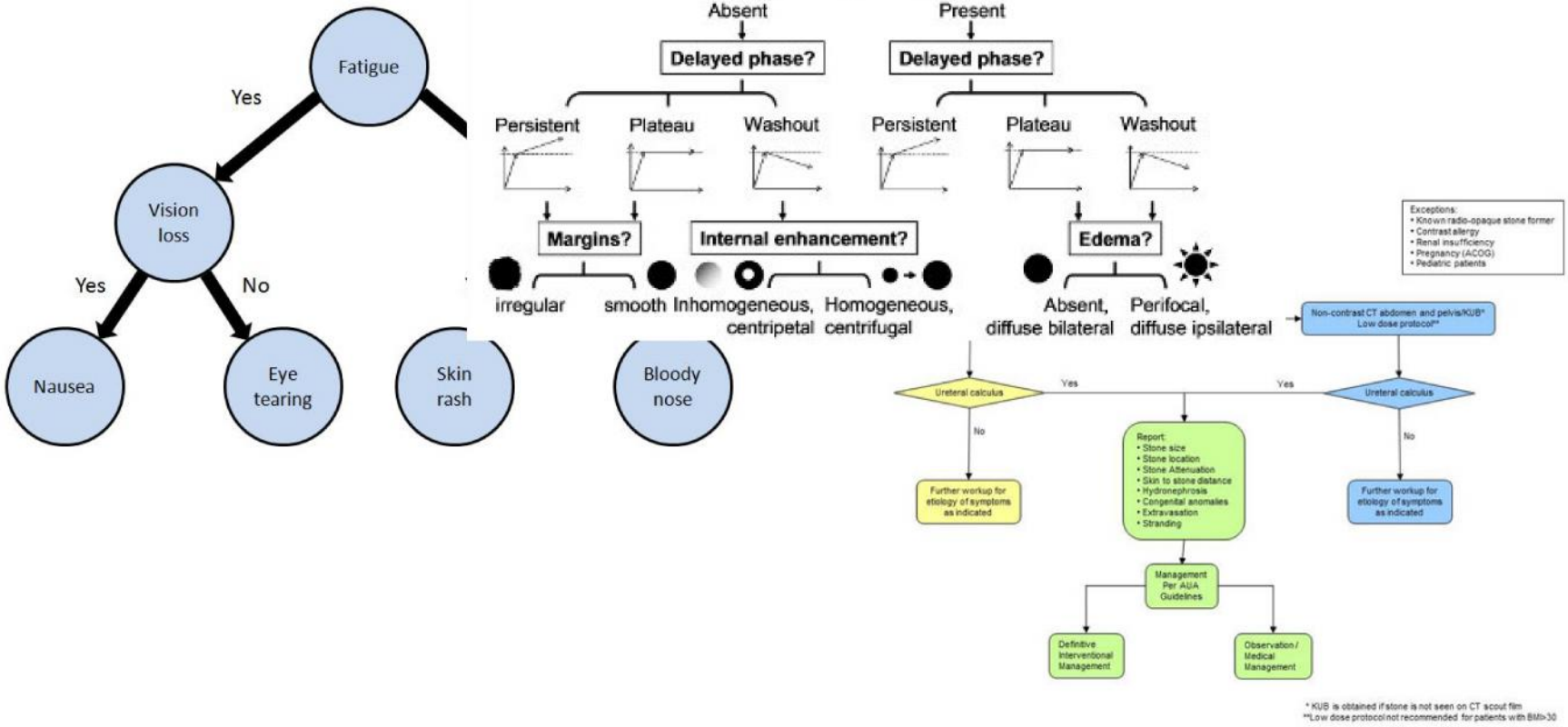
Олег Пьяных, opiany@gmail.com

Очевидный пример: принятие решений в медицине

- The same clinical observations should lead to the same standard actions, but physicians' opinions may differ:

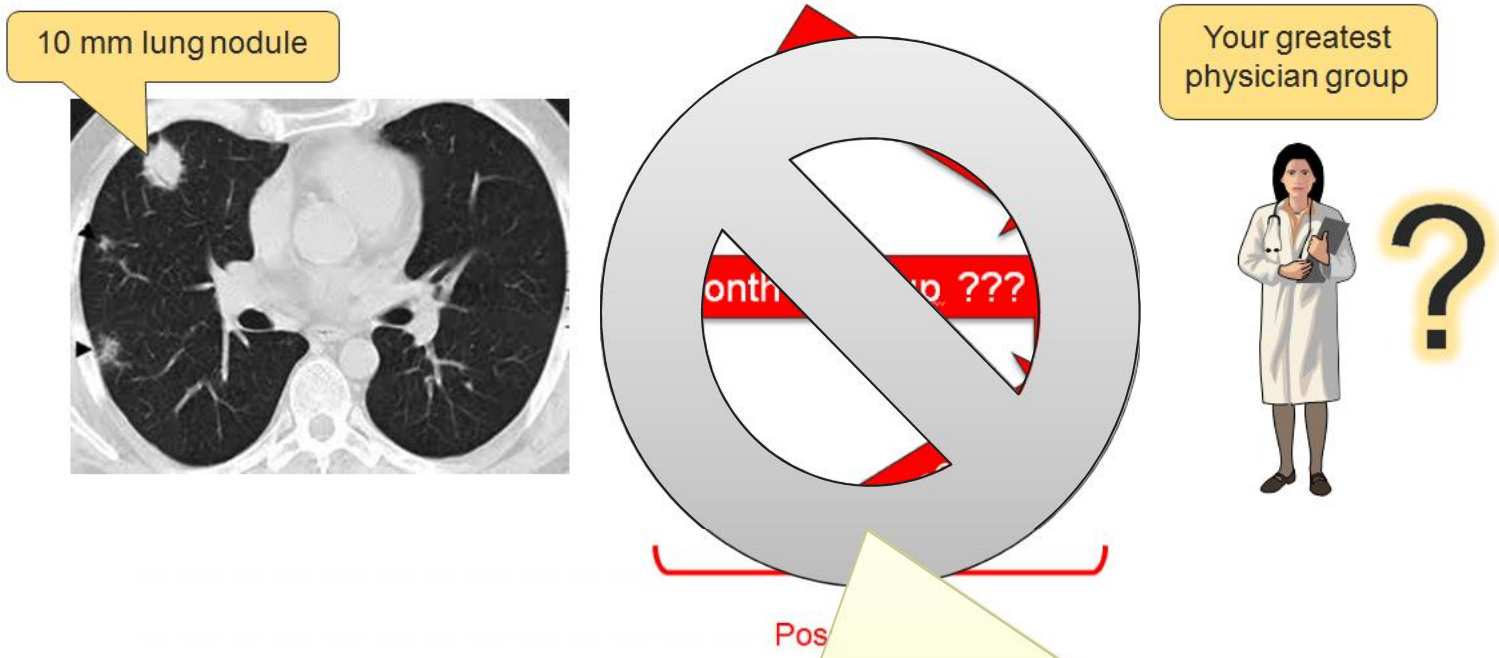


От «больших данных» к алгоритмам решений



Олег Пьяных, opiany@gmail.com

Результат: алгоритмы принятия решений в медицине

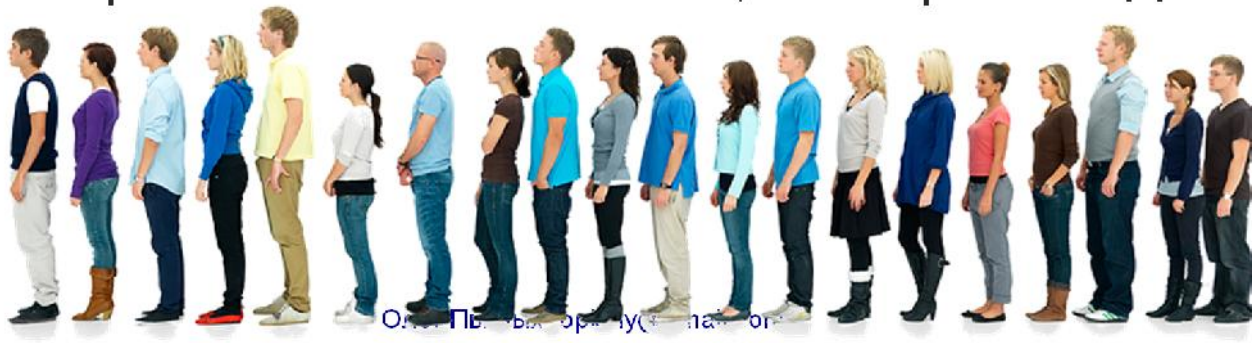


Decision support software suggests the best option based on the HIS/PACS data (such as tumor size).

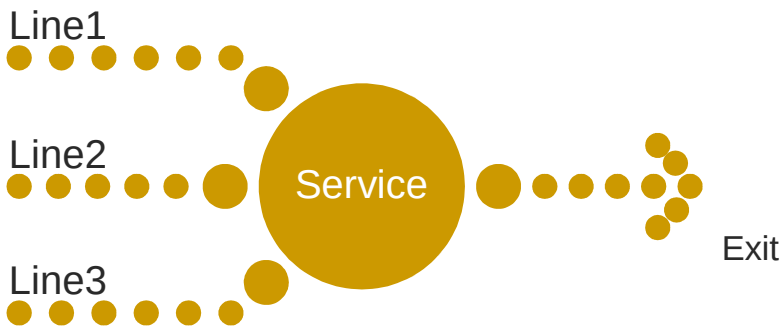
Банальный пример: Обслуживание пациентов в клинике

Зачем это нужно?

- Балансировка нагрузки
- Устранение проблем обслуживания, оптимизация работы
- Удовлетворенность пациентов и персонала. Качество обслуживания
- Отказ от бестолковых проектов («а давайте примем в два раза больше!»)
- Элементарное понимание того, что происходит

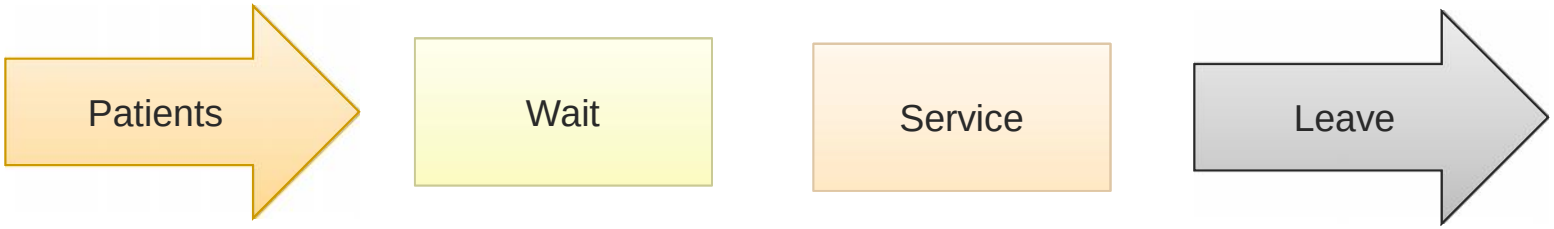


Классика жанра: Queuing theory (Теория массового обслуживания)



- Начало: A.K. Erlang, “The theory of probabilities and telephone conversations,” 1909
- Расчет параметров различных моделей очередей
- Универсальность: применима к любым очередям, будь то расчет задач на процессоре или очередь в кассу магазина

QT в медицине



Demand
(λ patients per minute)

Queue

Capacity
(μ patients per minute,
ρ=λ/μ – utilization,
ρ<1 – stable system)

The formula states that the mean queue length L is given by

$$L = \rho + \frac{\rho^2 + \lambda^2 \operatorname{Var}(S)}{2(1-\rho)}$$

where

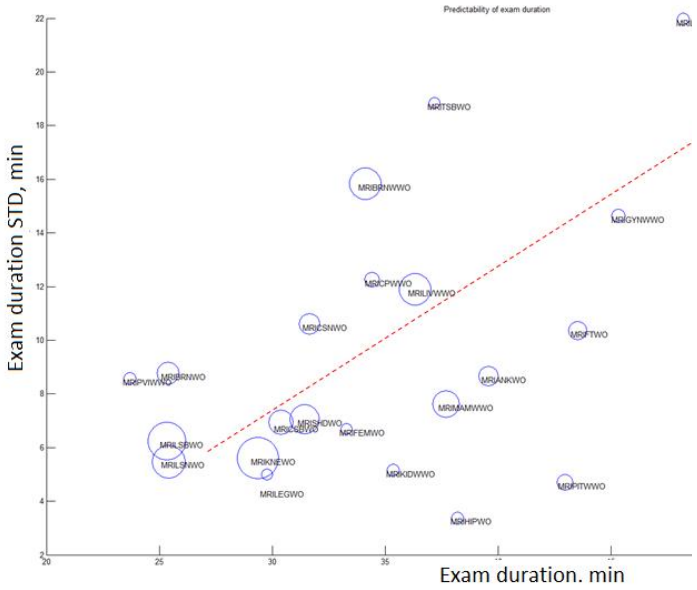
- λ is the arrival rate of the Poisson process
- $1/\mu$ is the mean of the service time distribution S
- $\rho = \lambda/\mu$ is the utilization
- Var(S) is the variance of the service time distribution S.

See “Queuing for Healthcare”, R. Kannapiran Palvannan & Kiok Liang Teow

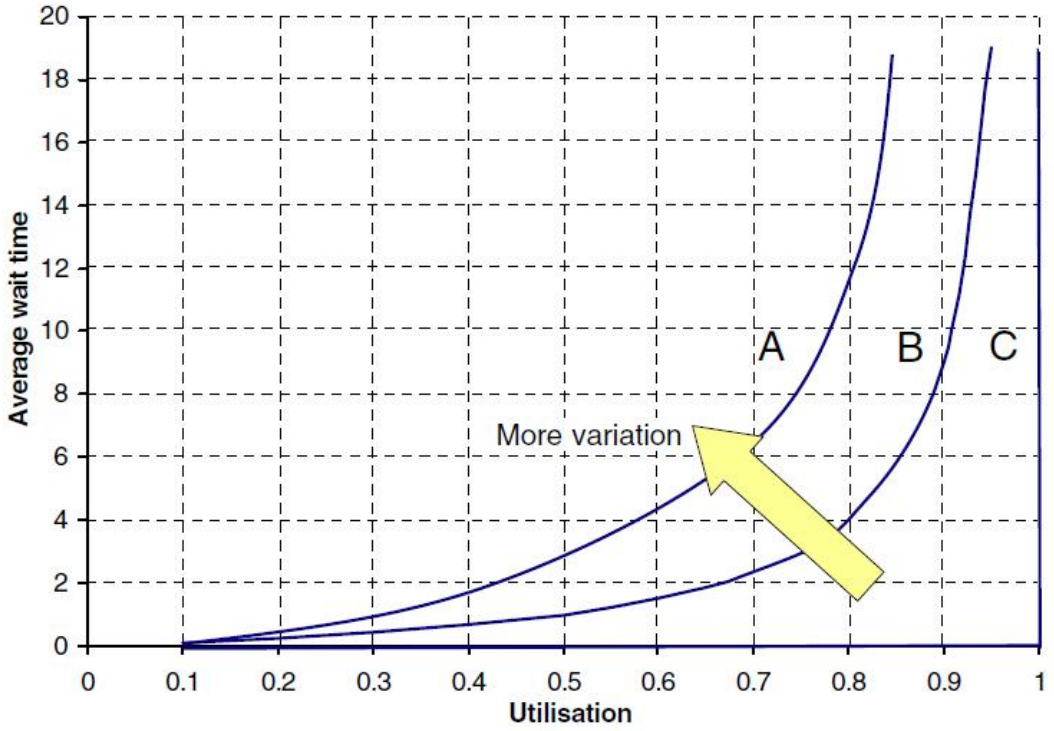
Oleg Pianykh opianykh@mgh.harvard.edu

Уроки QT

- L1: Queues are formed even when service rate *exceeds* demand. Variation in demand and service rate affects patient wait time significantly

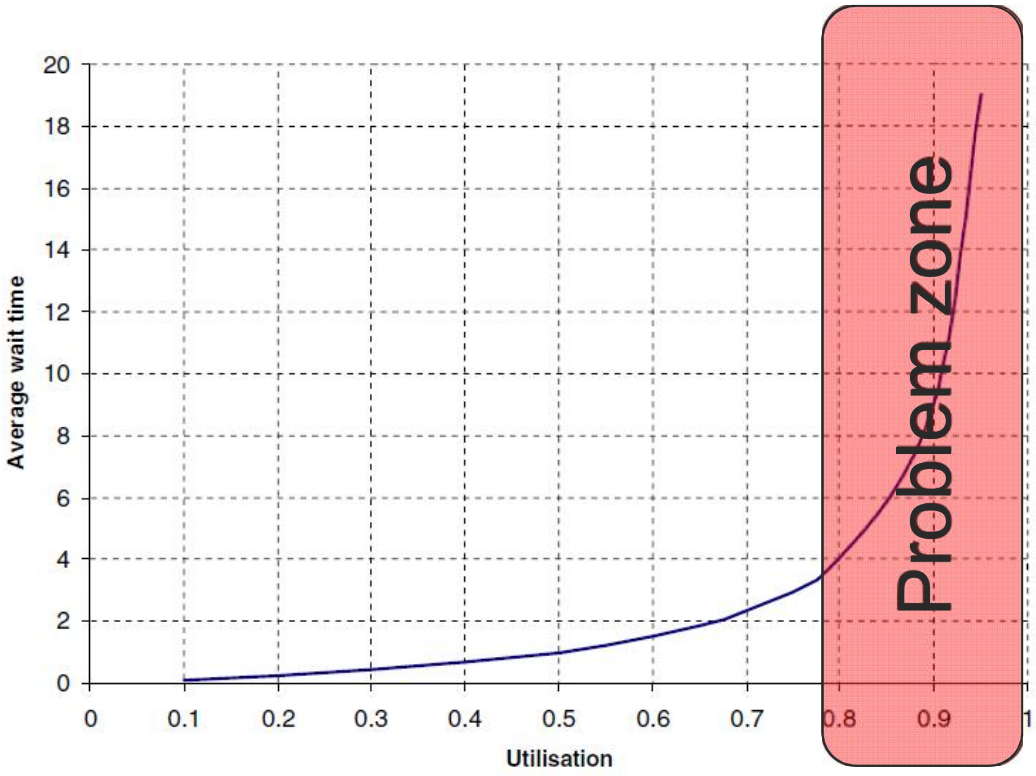


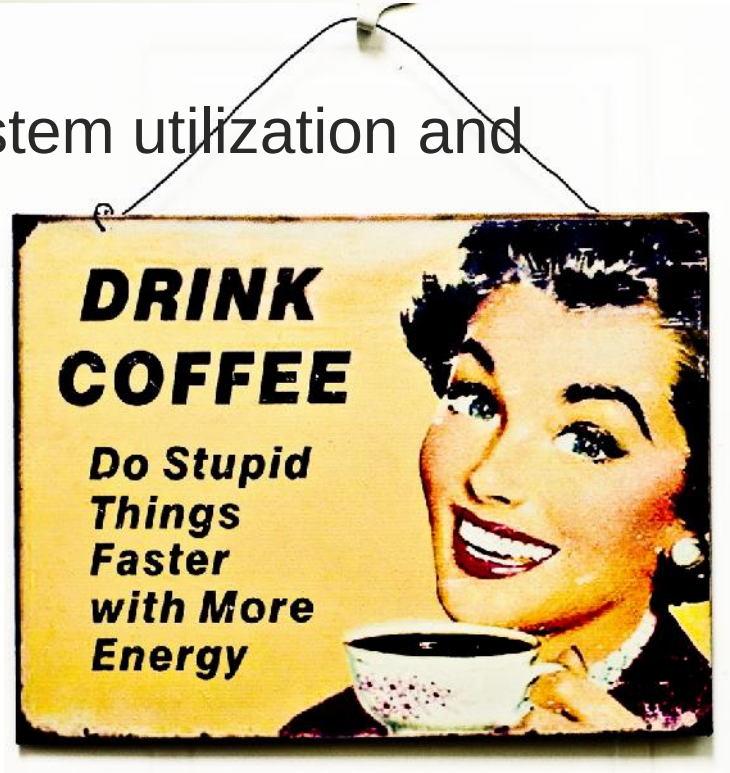
$$L = \rho + \frac{\rho^2 + \lambda^2 \mathrm{Var}(S)}{2(1-\rho)}$$



Уроки QT

- L2: Short wait time requires low system utilization and high cost structure.



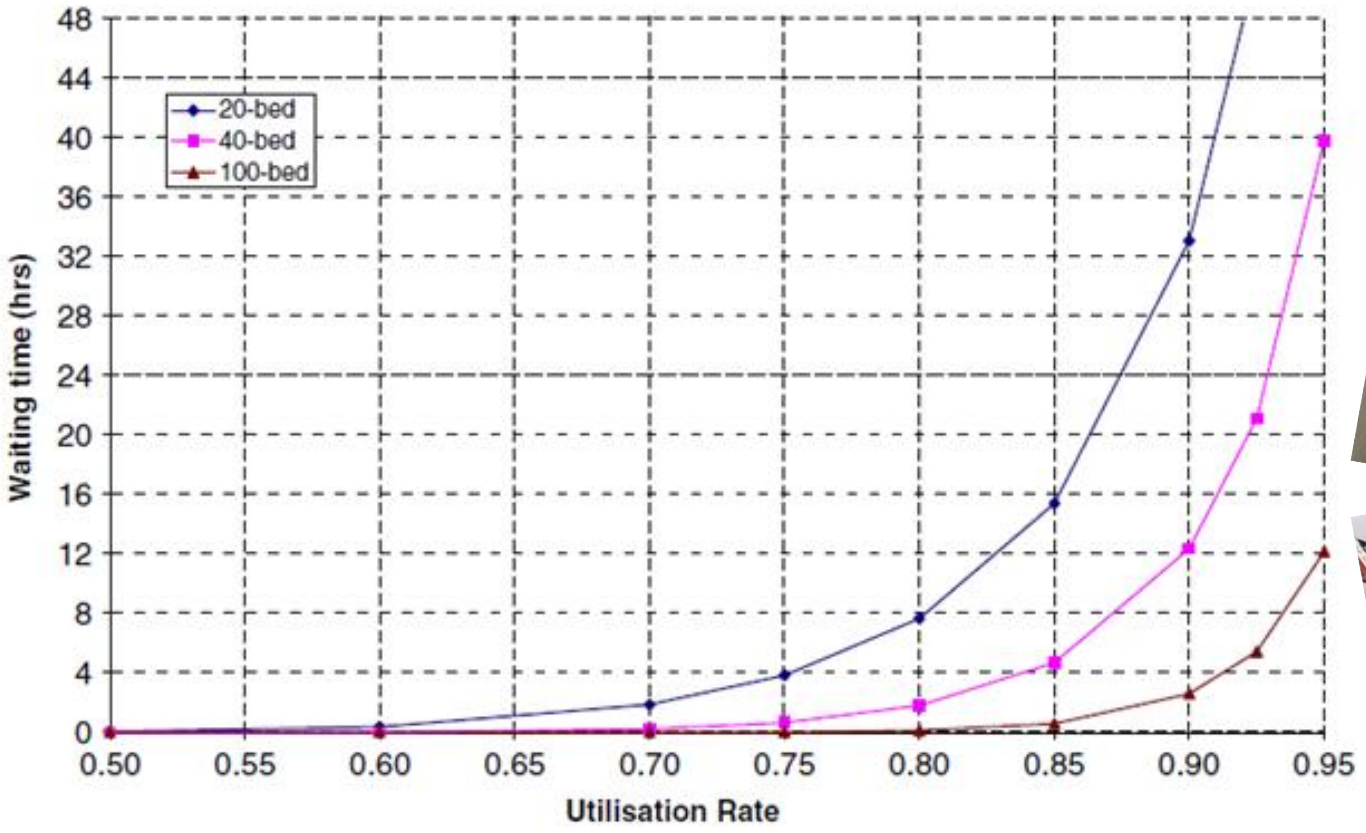


$$L = \rho + \frac{\rho^2 + \lambda^2 \operatorname{Var}(S)}{2(1-\rho)}$$

Oleg Pianykh opianykh@mgh.harvard.edu

Уроки QT

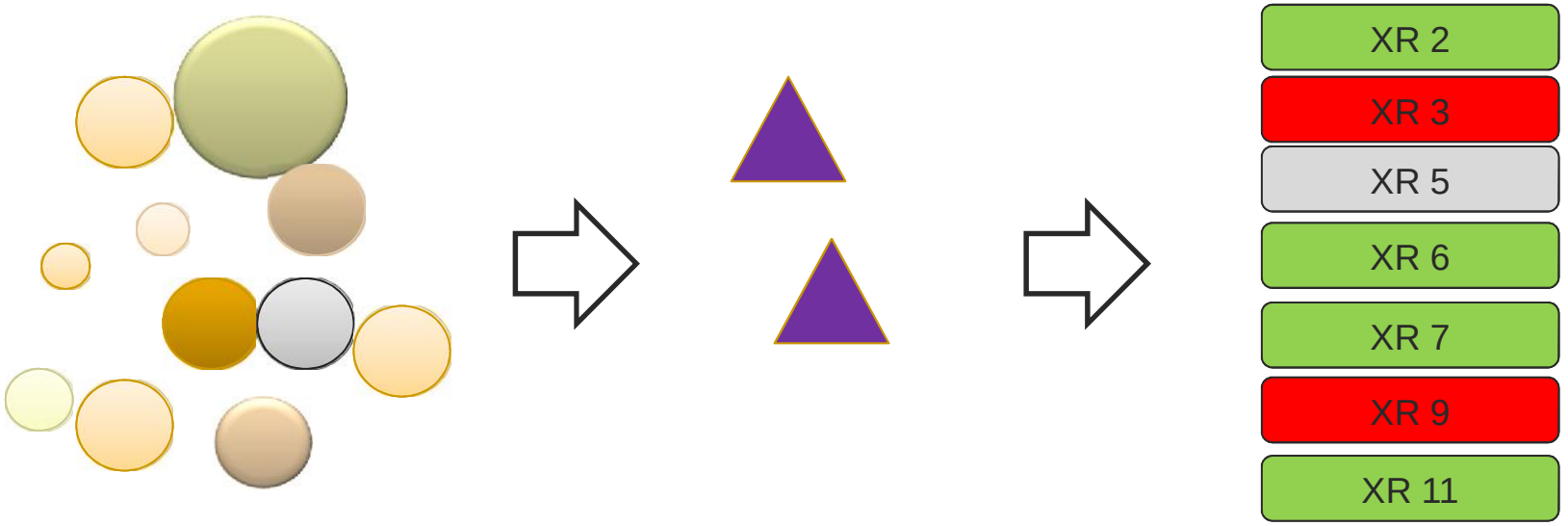
- L3: Resource availability and allocation strategies matter:



Oleg Ivanykh oivanykh@mgh.harvard.edu

Реальность: всё гораздо сложнее

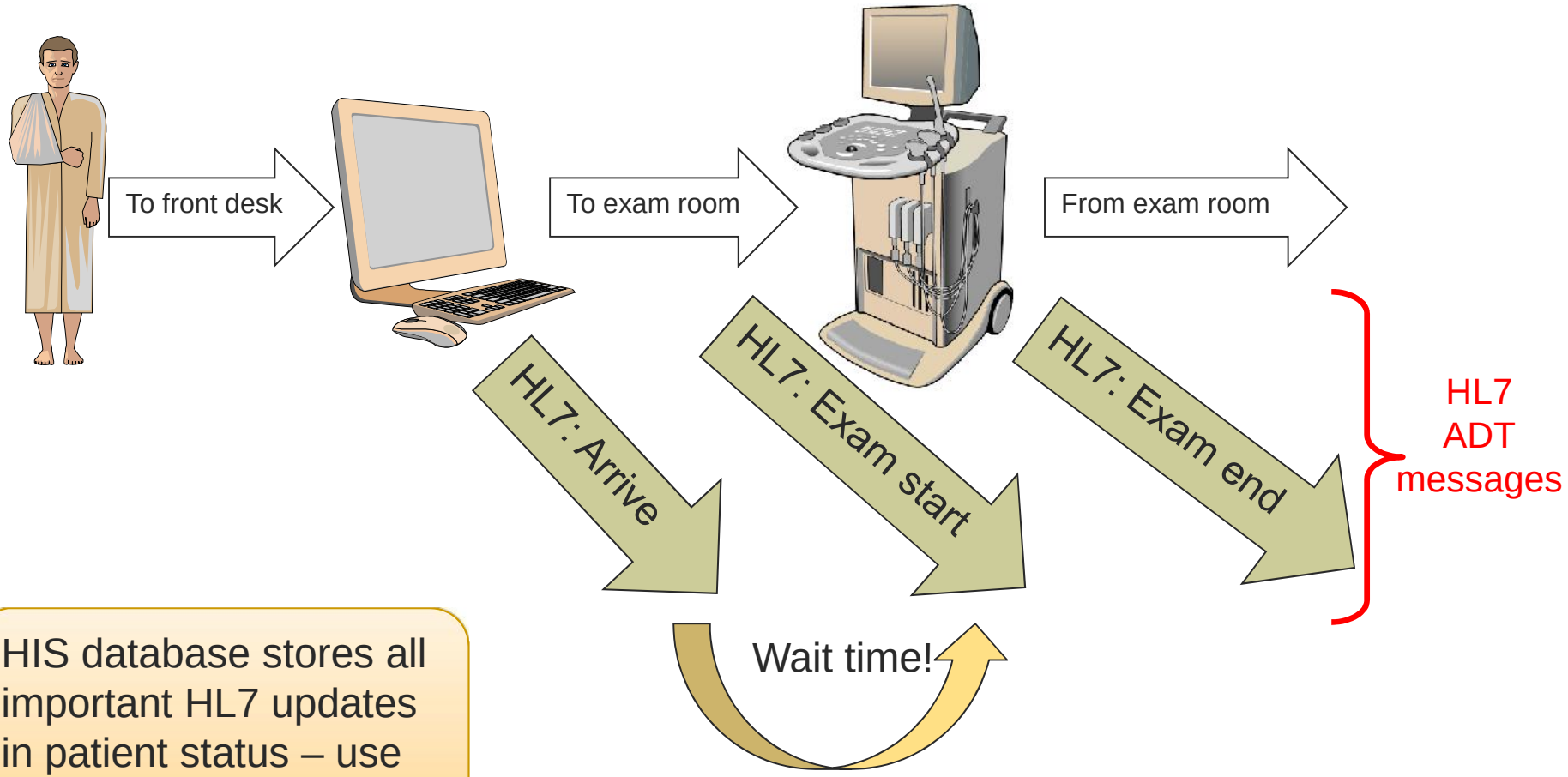
- Простейшая задача:



- Можем ли мы контролировать время ожидания пациента?

Oleg Pianykh opianykh@partners.org

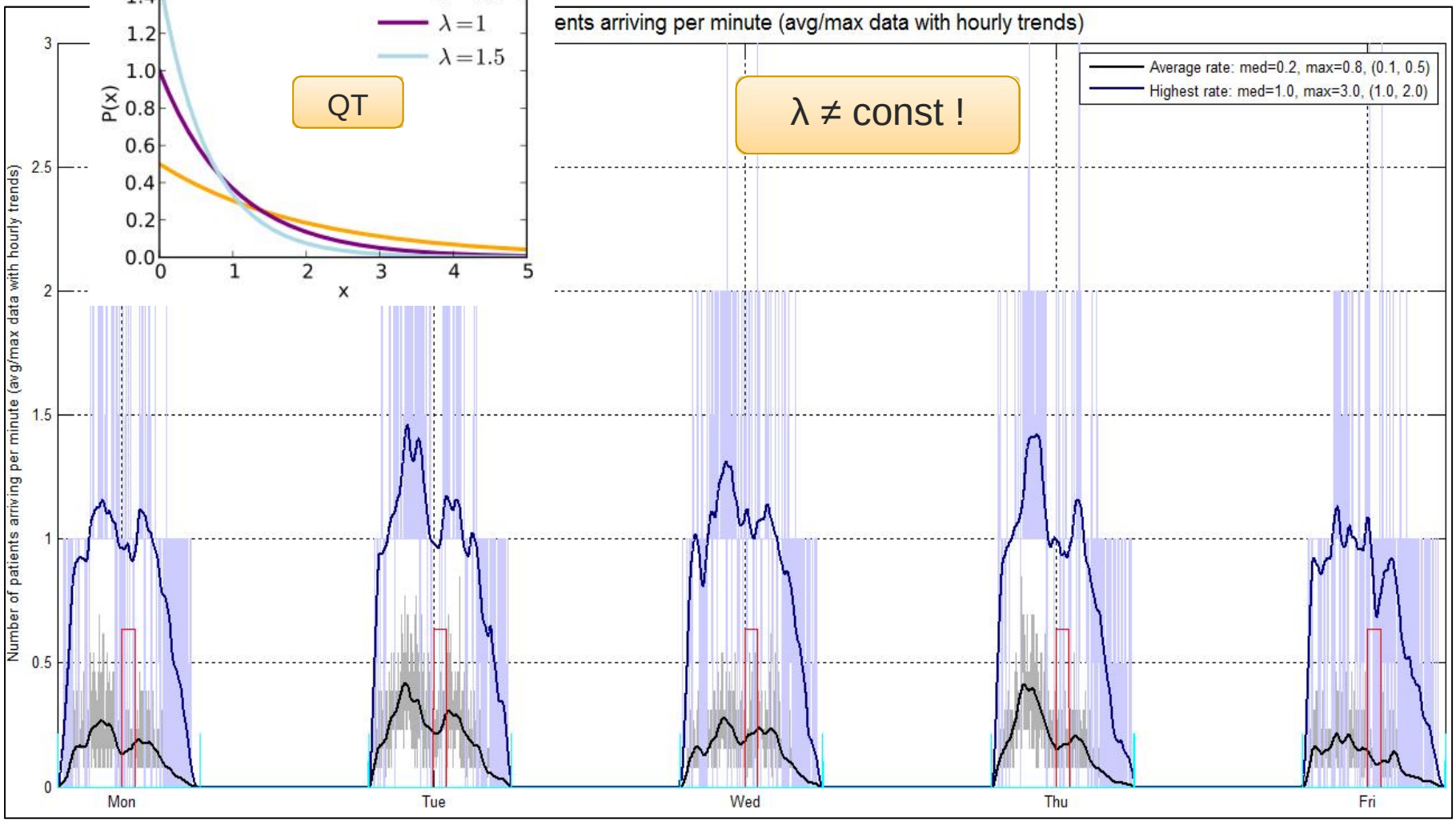
Источник «больших данных»: ГИС



HIS database stores all important HL7 updates in patient status – use them!

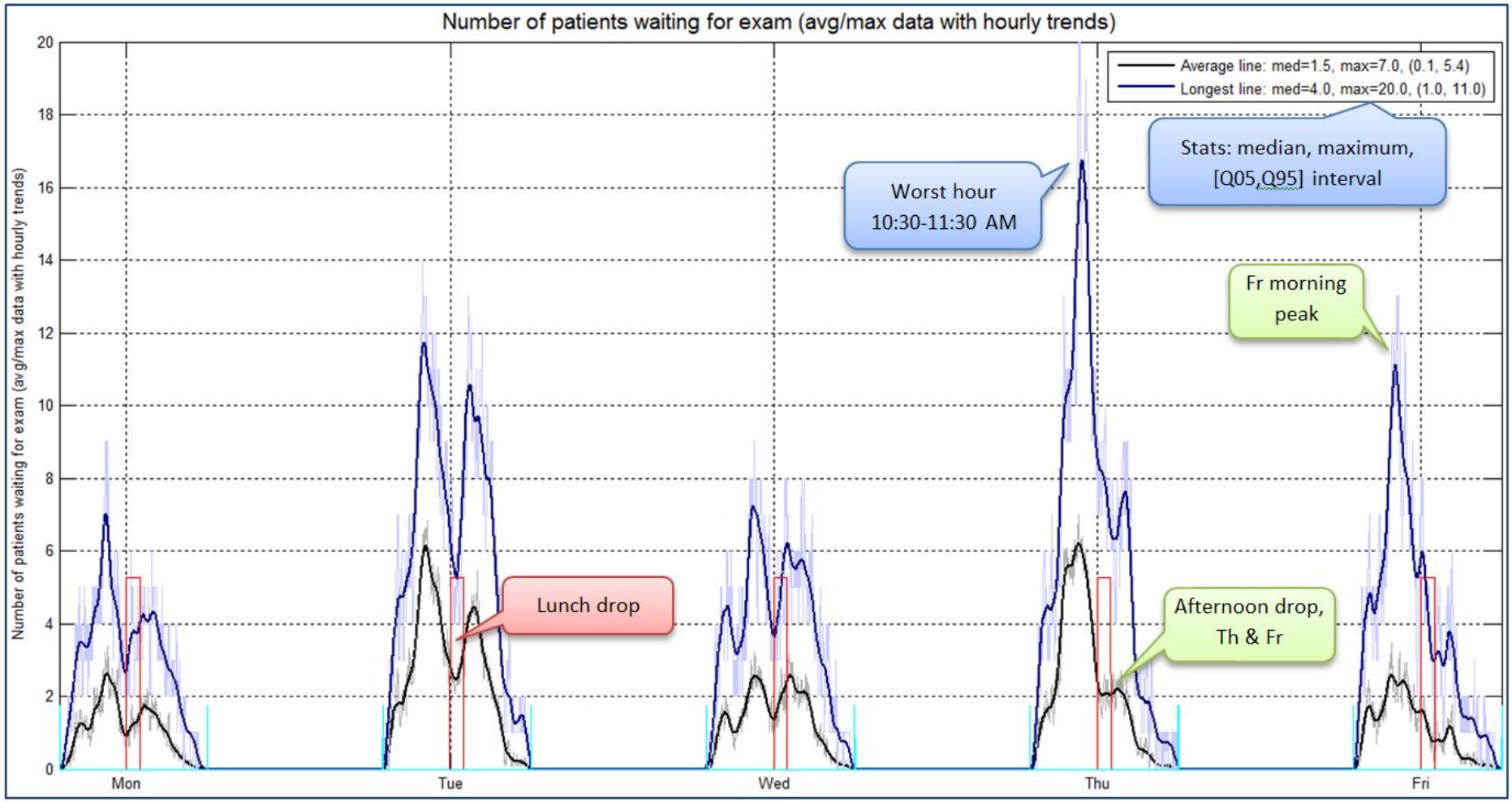
Oleg Pianykh opiany@gmail.com

Реальность: Patient arrival rate λ



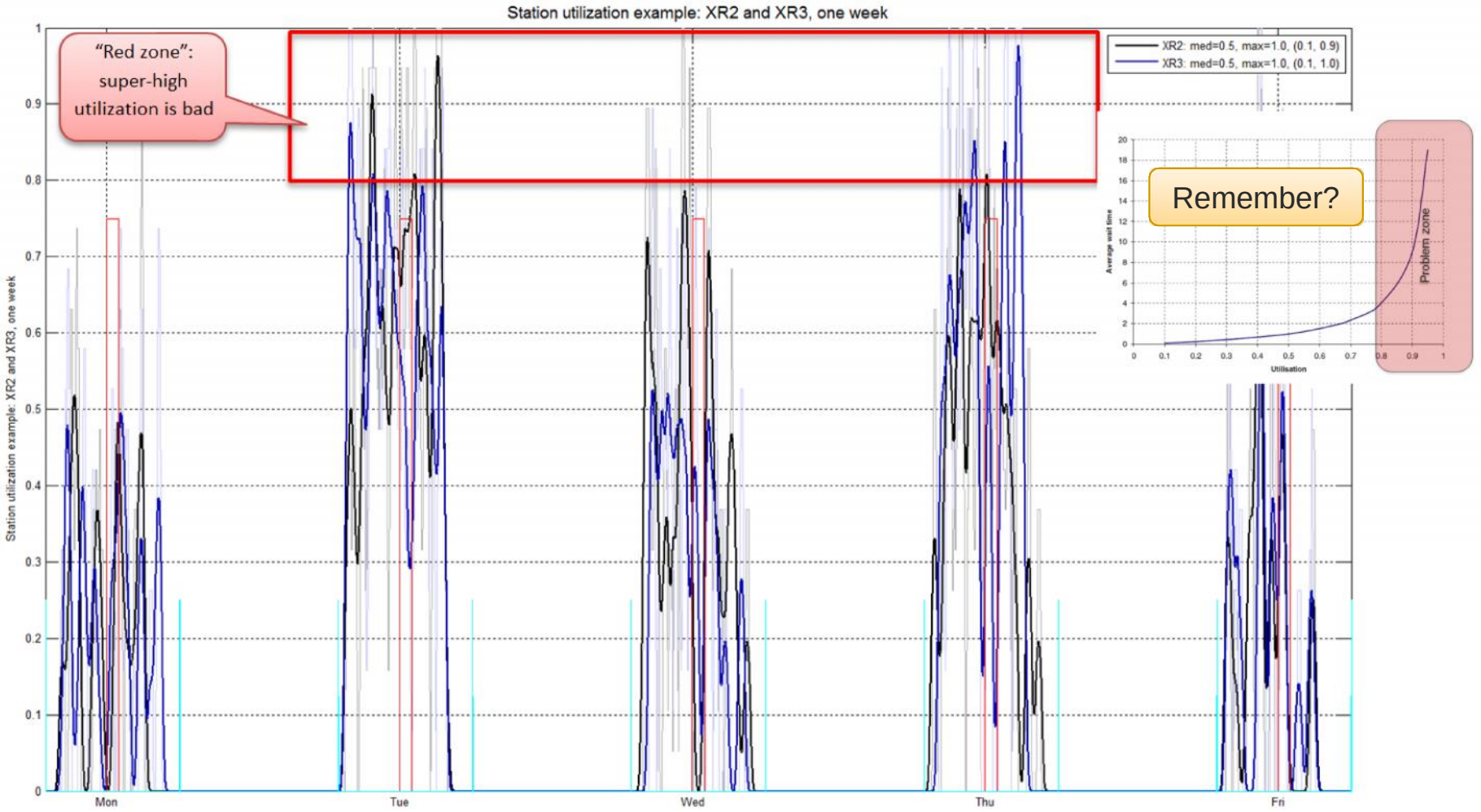
Oleg Pianykh opianykh@mgh.harvard.edu

Реальность: Patient waiting time W



Oleg Pianykh opianykh@mgh.harvard.edu

Реальность: Station utilization



Oleg Pianykh opianykh@mgh.harvard.edu

Итак, на самом деле

- “Big Data” значительно отклоняется от классических распределений и моделей. Например, в нашем случае корреляция QT с наблюдаемым временем оказалась и вовсе отрицательной
- Применение моделей сильно ограничено тем, что мы знаем на данный момент, и достоверностью этого знания.
- Значительное присутствие непредсказуемых и случайных факторов (особенно в медицине, где многое меняется «на лету»)

Неутешительный вывод

- Классические модели обслуживания малопримени́мы в сложных реальностях клиники

- Приходится *начинать* с простых, но надежных эмпирических моделей – например, линейной регрессии. Они строятся на основе накопленных «больших данных», и ложатся в основу практических алгоритмов обслуживания.

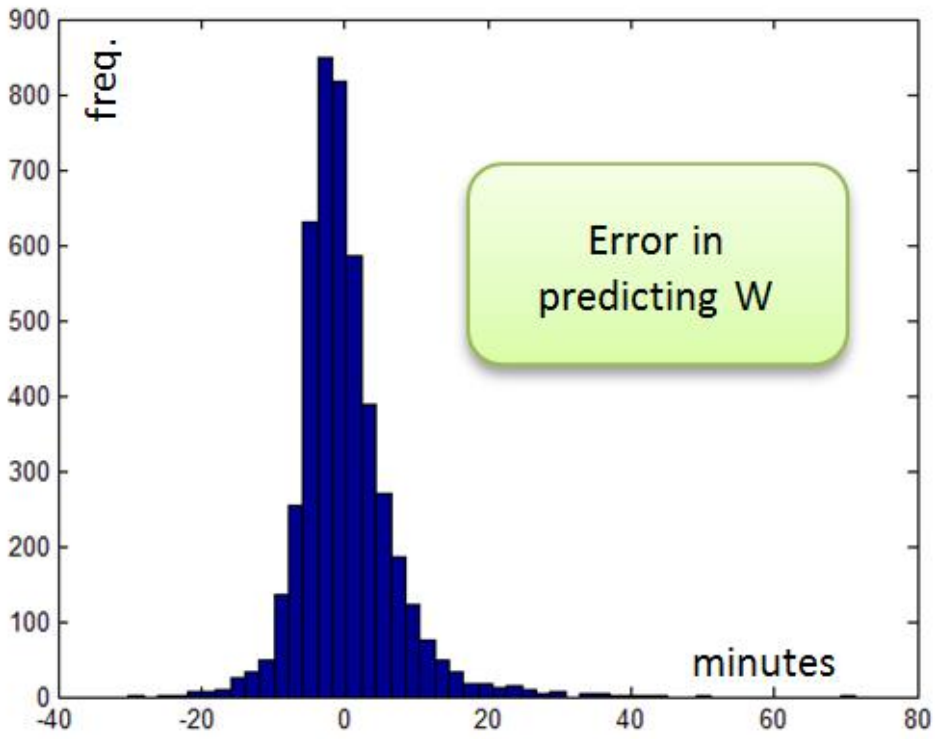
Как же предсказать время ожидания пациента **W?**

- QT formula **W=L/λ** – based on idealized model

- Empirical formula, derived with best-predictor regression analysis from 3 months of real clinic data:

$W = 4.5 + L + 0.5L_5 + 0.25L_{10}$,

where L_5 is the size of the waiting line 5 minutes ago, and L_{10} – 10 minutes ago.



Oleg Pianykh opianykh@mgh.harvard.edu

Проблемы с простыми моделями

- Статичны: не подстраиваются под изменяющуюся реальность
- Ограничены по точности
- Работают лишь в простых процессах

Можем ли мы улучшить нашу модель?

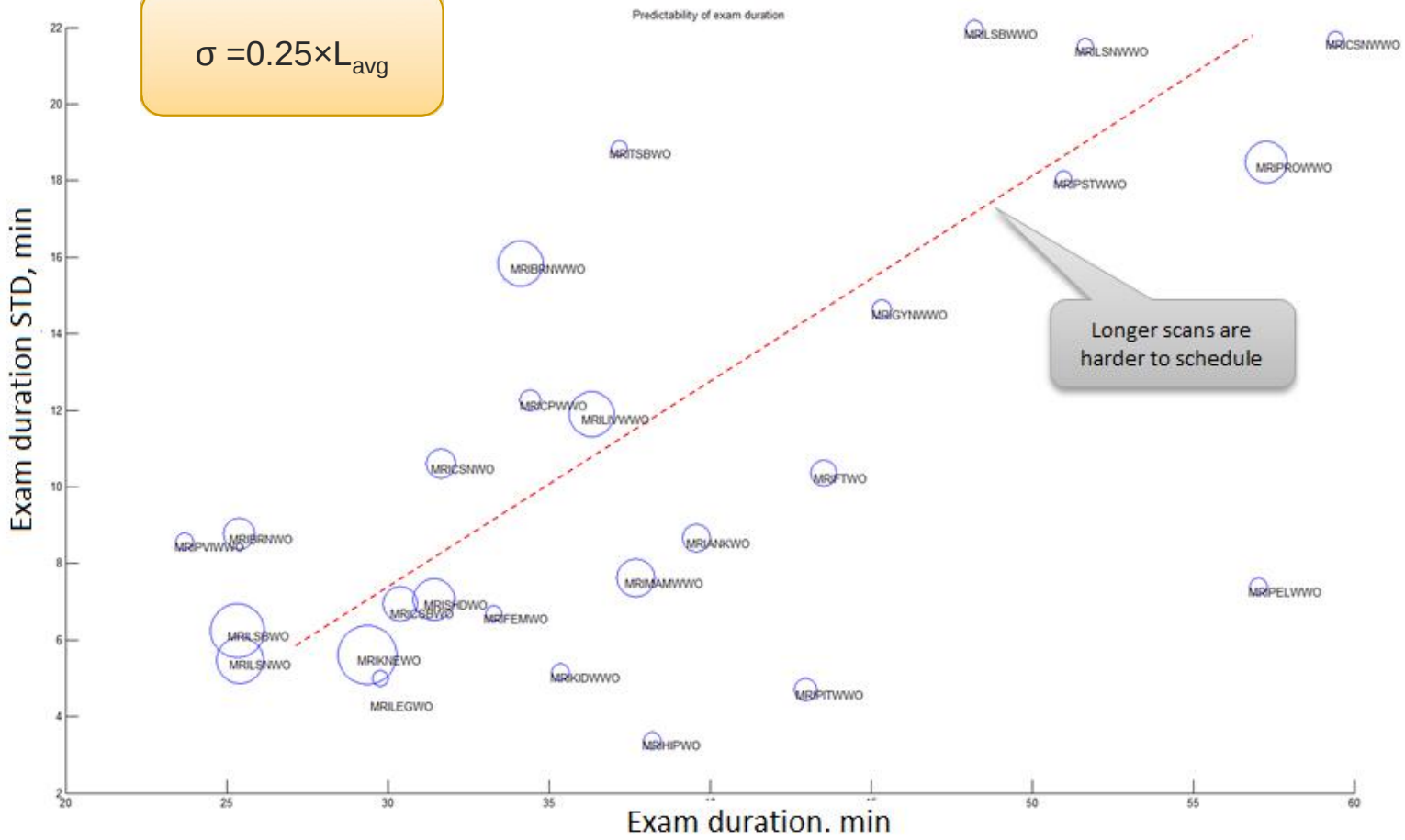
- Регрессия – ограничена
- Нелинейная регрессия – не получилось!
- Искусственный интеллект (нейронные сети) – не получилось!
- Почему?

Жизнь гораздо сложнее….

- Никаких иллюзий: «большие данные» не гарантируют «большой смысл»! В них отсутствует структура, корректность, достоверность. *«Большие данные» - это просто всё то, что у нас есть, и с чем надо как-то работать.*
- Большие данные могут содержать большие горы мусора:
  - Случайные, непредсказуемые события
  - Неверные данные (ручной ввод!!!)
  - Данные не в реальном времени (естественная задержка в получении данных, или данные post factum)
  - Бесполезные или неполные данные

Накопление случайных событий

$\sigma = 0.25 \times L_{avg}$

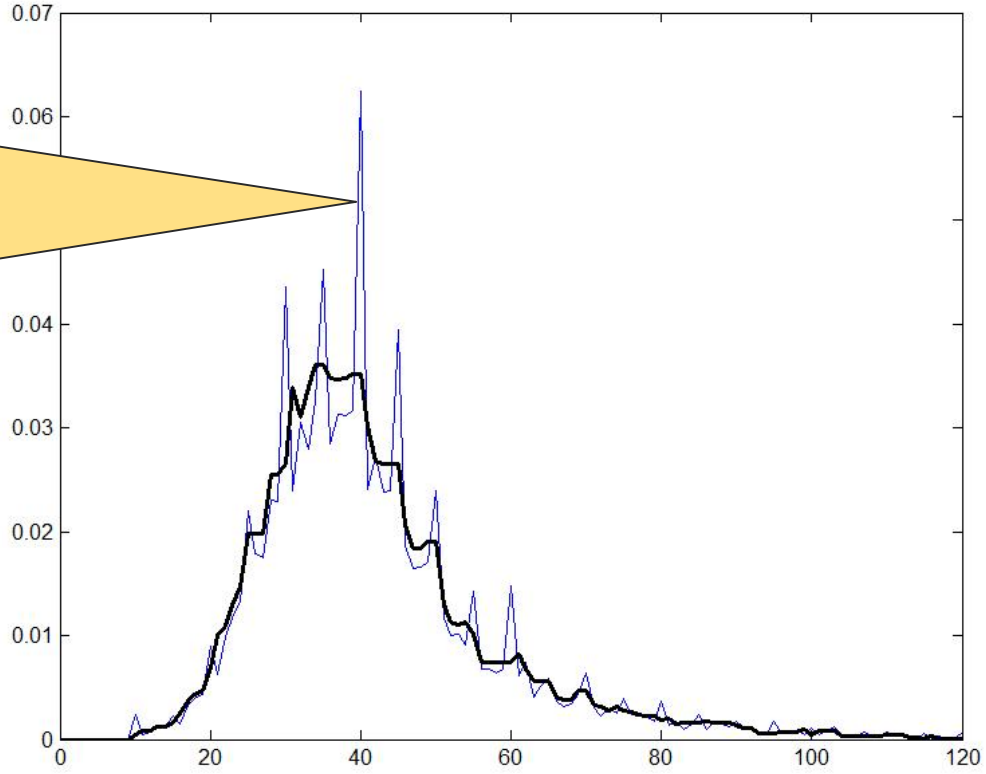


Oleg Pianykh opianykh@mgh.harvard.edu

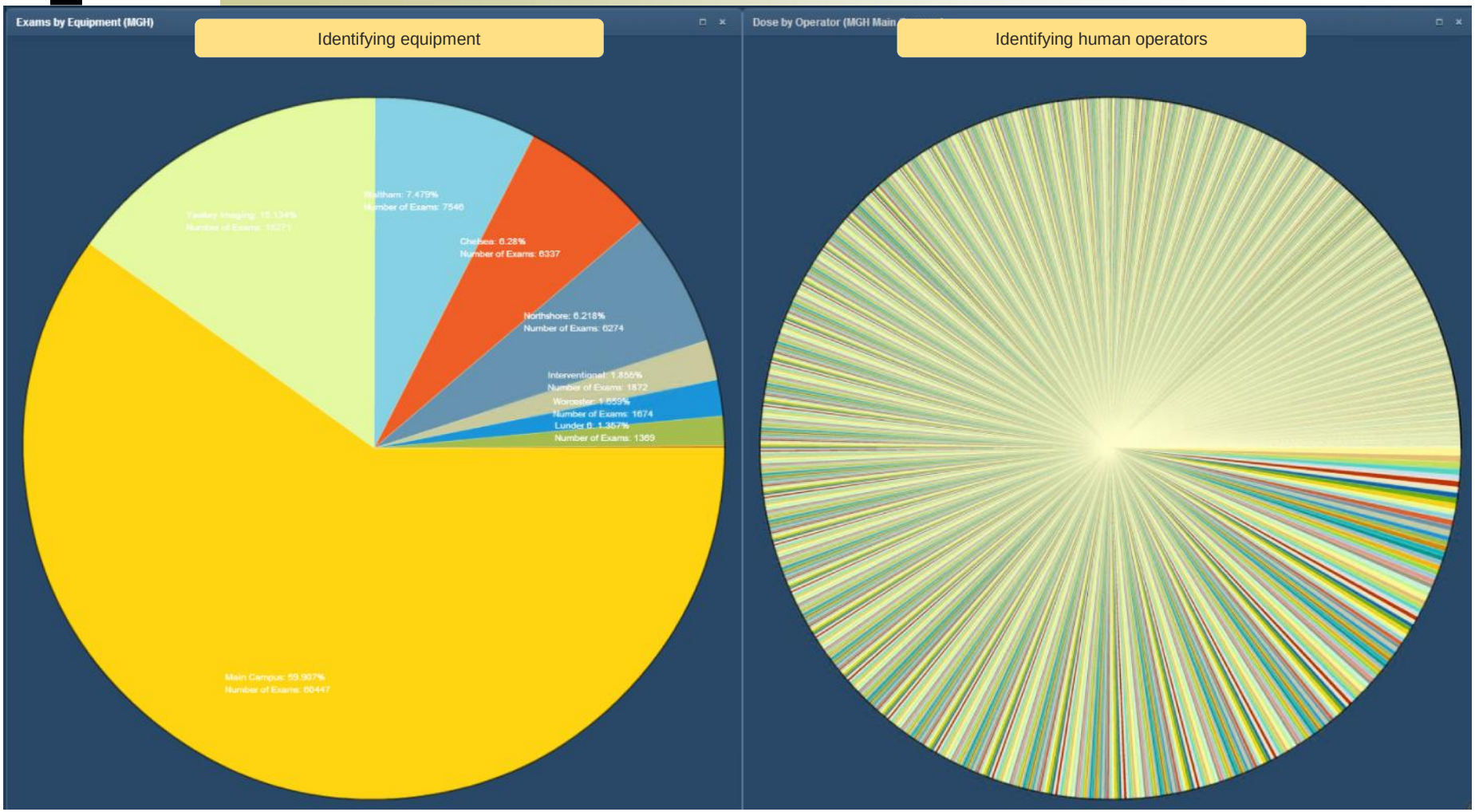
Неверные данные

- Manual entry supplies most of HIS data!
- Result: incorrect or strangely-biased numerical results

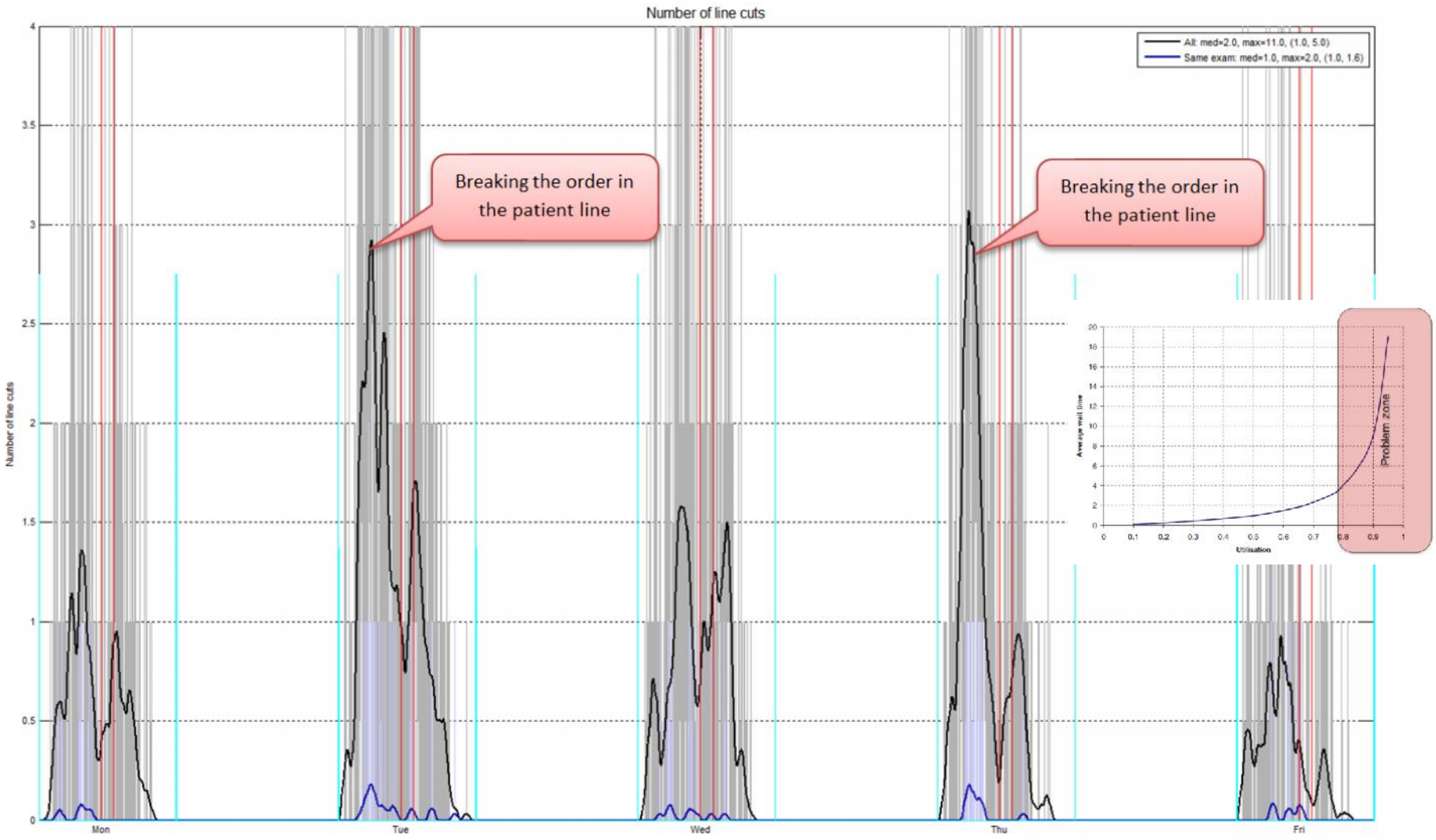
Example: Histogram for MRI exam length distribution (blue) shows unusual spikes for multiple of 5 minutes: data entry with 5-min accuracy. They have to be filtered out (black).



Неверные данные



Хаос в данных (и в работе)



Oleg Pianykh opianykh@mgh.harvard.edu

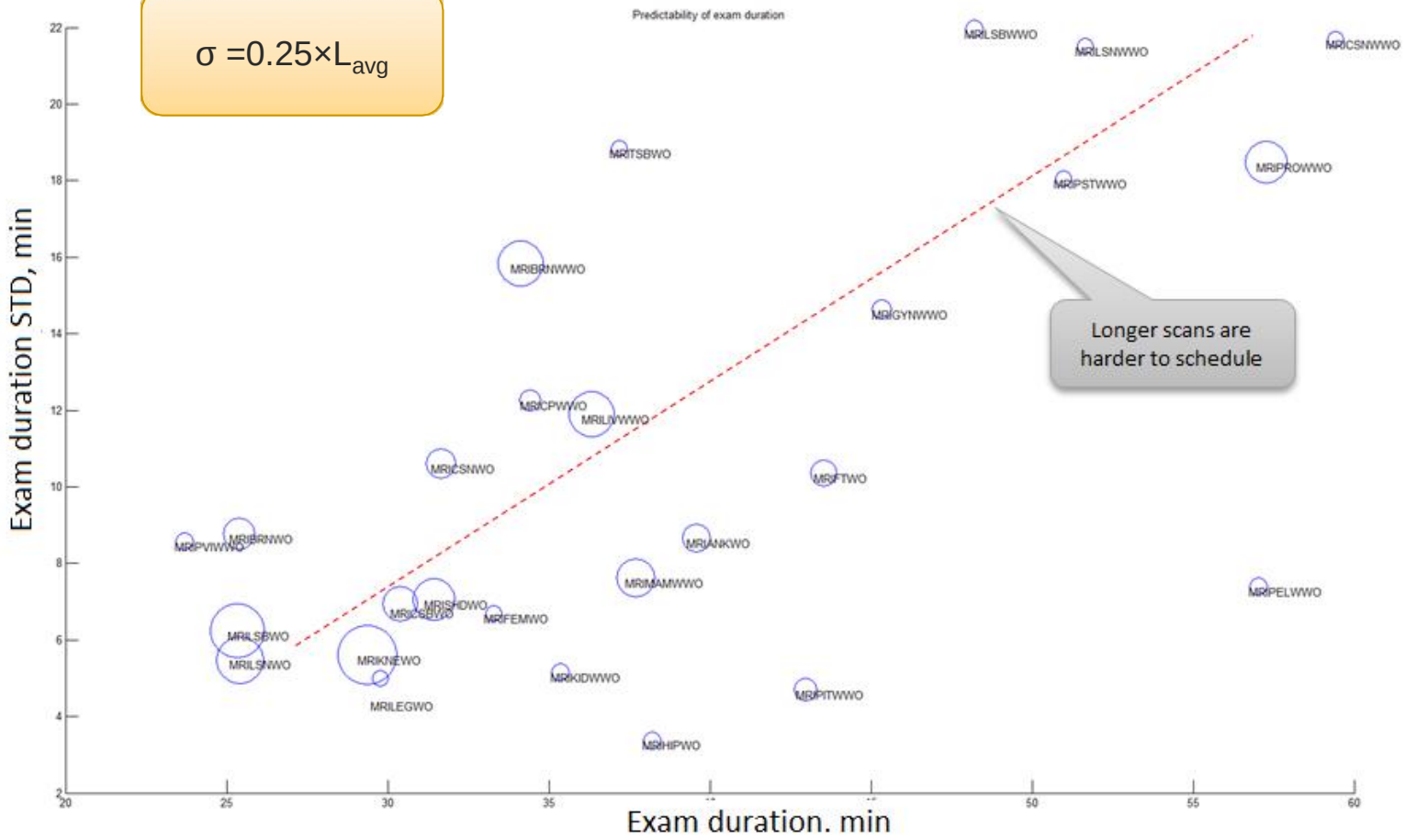
Что делать?

- Случайность, хаос, несвоевременность являются основными врагами отлаженного рабочего процесса
- Логика работы с большими данными должна включать в себя устранение этих аномалий, и превращение любого процесса в наиболее детерминированный
- → Анализируя «большие данные», необходимо:
  - Понимать их
  - Развивать не только логику, но и оптимальные стратегии работы

Пример: понижение случайности

Накопление случайных событий

$\sigma = 0.25 \times L_{avg}$



Oleg Pianykh opianykh@mgh.harvard.edu

Как ускорить обслуживание пациентов ?

- Стандартный подход: уменьшить время обслуживания

- Правильный подход: изменить стратегию обслуживания для понижения случайности

Выбор стратегии обслуживания

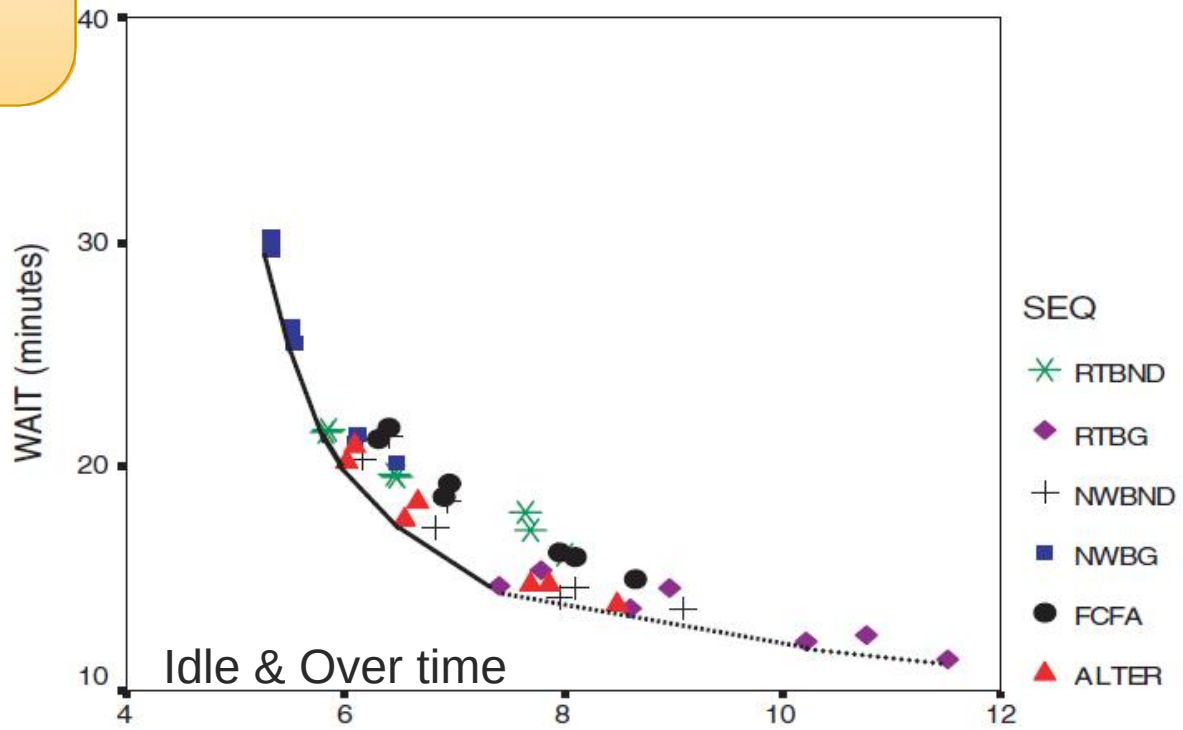
- Основные правила:
  - Никогда не начинать дня с наиболее длинного исследования
  - Чередовать длинные и короткие
  - Назначать исследования пропорционально их реальной длине (средней для данного типа исследования)

Олег Пьяных, opiany@gmail.com

Выбор стратегии обслуживания

Sequencing rules:

FCFS
ALTER (long/short)
Long first (NWBG)
Short first (RTBG)
Long first and last (NWBND)
Short first and last (RTBND)



"Designing appointment scheduling systems for ambulatory care services", Tugba Cayirli · Emre Veral · Harry Rosen

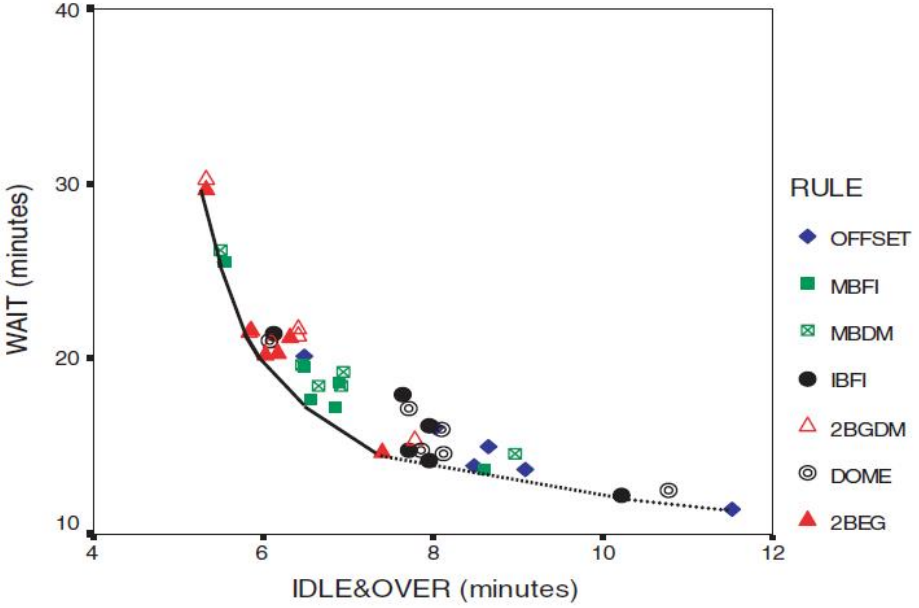
Oleg Pianykh opianykh@mgh.harvard.edu

Выбор стратегии обслуживания

Appointment rules:

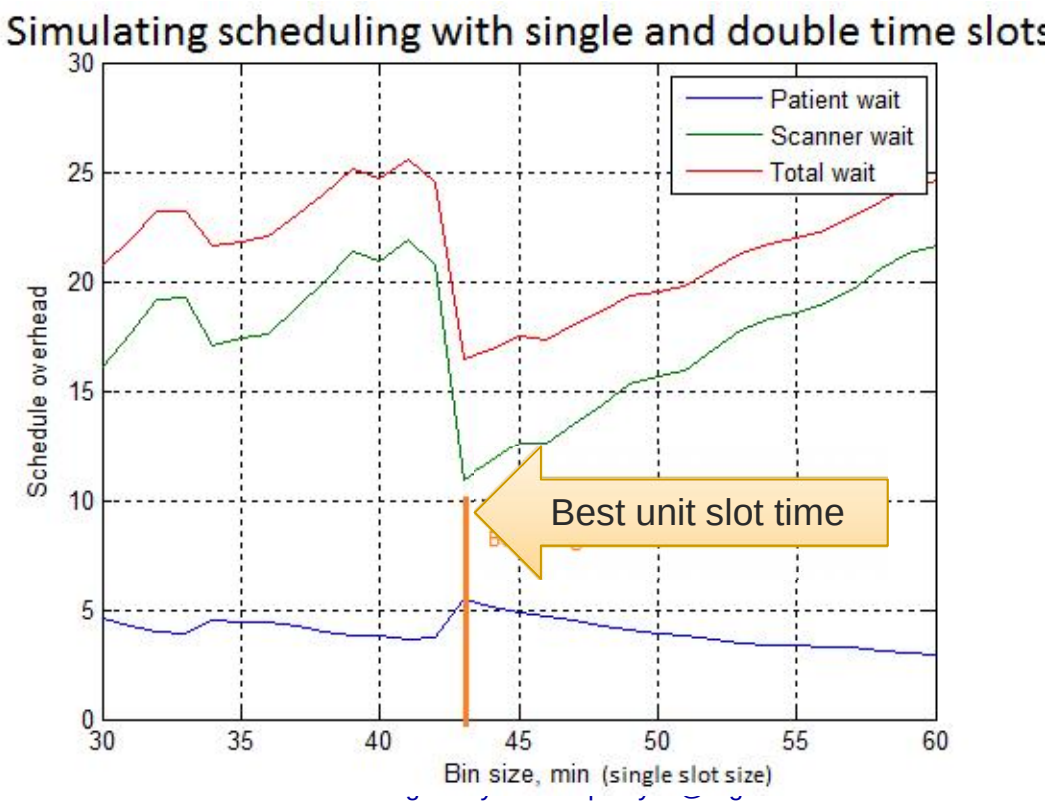
Table 1 Appointment rules

Symbol	*Description*	*Formulations*[1]	
IBFI	Individual-block/fixed-interval rule calls patients individually at intervals equal to the mean service times of patients	$t_1 = 0$ $t_i = t_{i-1} + \mu$	for $i > 1$
	1 1 1 1 1 1 1 1 1 1		
OFFSET	Individual-block/variable-interval rule, where initial $(k_1 - 1)$ patients are scheduled earlier, and the rest are scheduled later compared to IBFI	$t_i = (i-1)\mu - \beta_1(k_1 - i)\sigma$ $t_i = (i-1)\mu + \beta_2(i - k_1)\sigma$	for $i \le k_1$, and for $i > k_1$
	1 1 1 1 1 1 1 1 1 1		
DOME	Individual-block/variable-interval rule, where initial $(k_1 - 1)$ patients are scheduled earlier, patients $(k_1 + 1)$ through $(k_2 - 1)$ are scheduled later, and the rest earlier compared to IBFI	$t_i = (i-1)\mu - \beta_1(k_1 - i)\sigma$ $t_i = (i-1)\mu + \beta_2(i - k_1)\sigma$ $t_i = (i-1)\mu - \beta_3(i - k_2)\sigma$	for $i \le k_1$, for $k_1 < i < k_2$, and for $i \ge k_2$
	1 1 1 1 1 1 1 1 1 1		
2BEG	Individual-block/fixed-interval rule with an initial-block of two patients	$t_1 = t_2 = 0$ $t_i = t_{i-1} + \mu$	for $i > 2$
	2 1 1 1 1 1 1 1 1		
MBFI	Multiple-block/fixed-interval rule calls patients two-at-a-time with intervals set equal to twice the mean service time	$t_i = t_{i+1} = (i-1)\mu$	for $i = 1, 3, 5$
	2 2 2 2 2		
2BGDM	Combination of the 2BEG and the DOME rules	$t_i = t_{i+1} = (i-1)\mu - \beta_1(k_1 - i)\sigma$ $t_{i+1} = (i-1)\mu - \beta_1(k_1 - i)\sigma$ $t_{i+1} = (i-1)\mu + \beta_2(i - k_1)\sigma$ $t_{i+1} = (i-1)\mu - \beta_3(i - k_2)\sigma$	for $i = 1$ for $2 \le i \le k_1$ for $k_1 < i < k$ for $i \ge k_2$
	2 1 1 1 1 1 1 1 1		
MBDM	Combination of the MBFI and the DOME rules	$t_i = t_{i+1} = (i-1)\mu - \beta_1(k_1 - i)\sigma$ $t_i = t_{i+1} = (i-1)\mu + \beta_2(i - k_1)\sigma$ $t_i = t_{i+1} = (i-1)\mu - \beta_3(i - k_2)\sigma$	for $i = 1, 3, .$ for $k_1 < i < k$ for $i \ge k_2$; wh k_2 are odd int
	2 2 2 2 2		

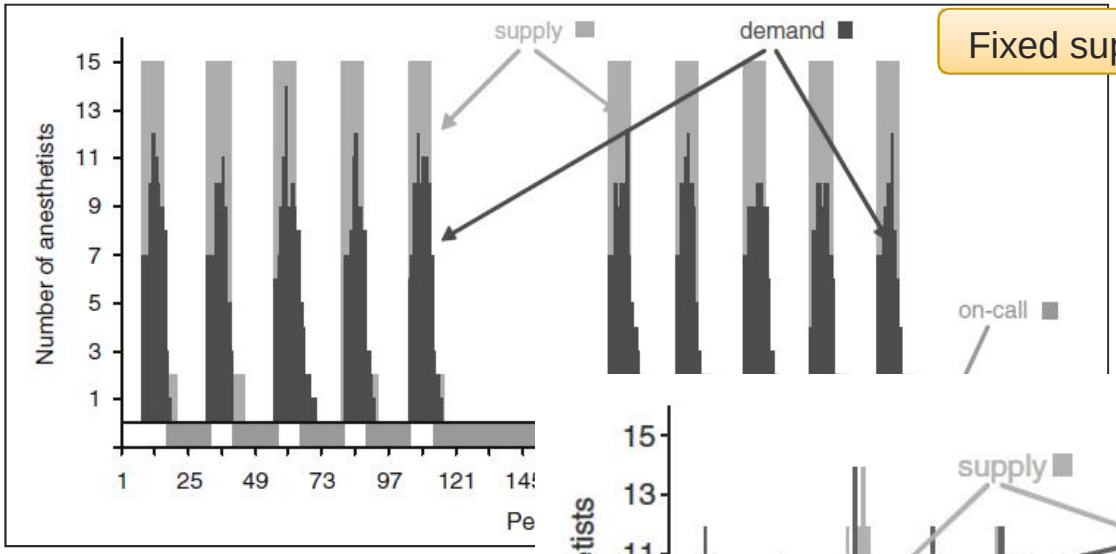


Оптимизация расписания исследований

- Пример: найти оптимальный интервал времени T так что все исследования можно назначать либо на T, либо на 2T времени

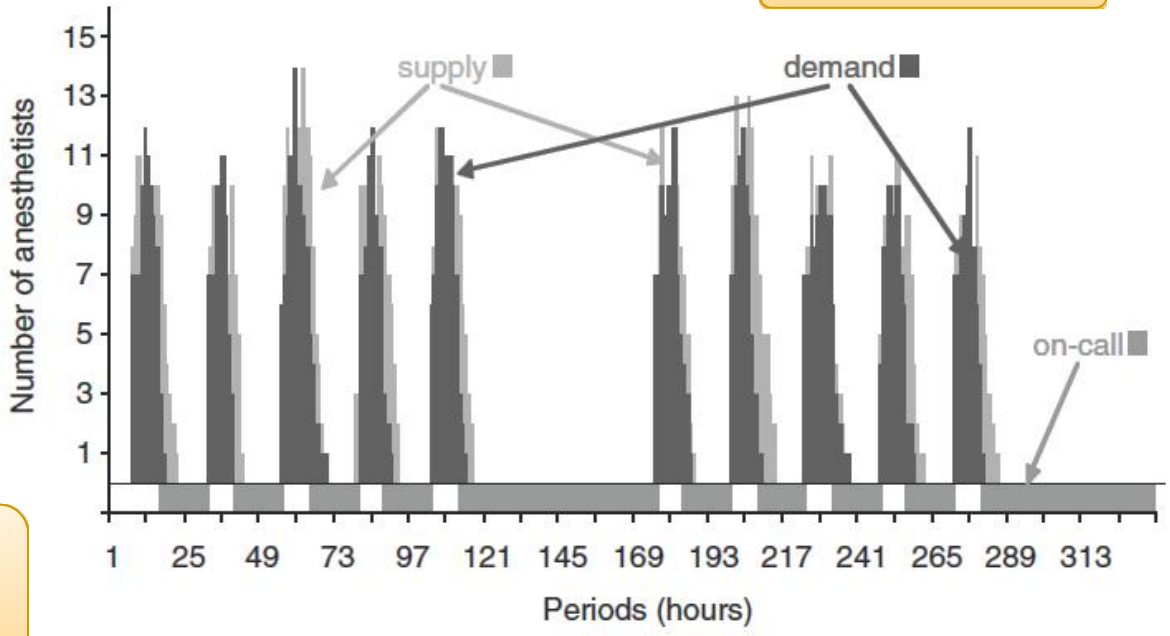


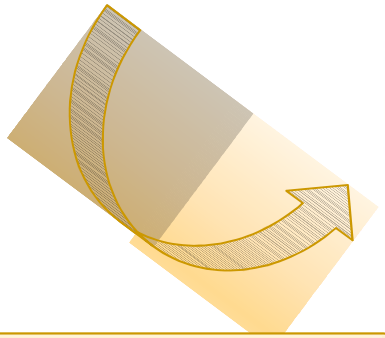
Оптимизация расписания ресурсов



Fixed supply

Optimal supply





Задачи целочисленного линейного/нелинейного программирования

See:
Brunner JO, Bard JF, Kolisch R., "Flexible shift scheduling of physicians", Health Care Manag Sci. 2009 Sep;12(3):285-305.

Oleg Pianykh opiany@gmail.com

Оптимизация расписания ресурсов

Model

Parameters

d_t demand in period t
c_i^{paid} cost per hour of paid out time for physician i
c_i^{over} cost per hour of overtime for physician i
c^{out} cost per hour for outside physicians
O_i^{max} maximal allowed overtime for a physician i in a week
$\overline{T}^{shift}$ maximum shift length
$\underline{T}^{shift}$ minimum shift length
$\underline{T}^{rest}$ minimum rest length after a shift ends
r_i regular working hours per week for physician i according to his individual contract
T^{week} number of periods within a week; $T^{week} = |T|/|W|$
T^{day} number of periods in a day; $T^{day} = T^{week}/7$
T^{oncall} length of an on-call service
N^{oncall} number of physicians required for each on-call service
O^{oncall} maximal number of on-call services for a physician in a week

Function

$f_1(t)$ calculates the number of hours that are charged for any on-call service for each $t \in I$, to regular working time per week

Binary decision variables

x_{it} 1, if physician i works in period t, 0 otherwise
y_{it}^{shift} 1, if physician i begins a shift in period t, 0 otherwise
y_{it}^{rest} 1, if rest period begins for physician i in period t, 0 otherwise
y_{it}^{oncall} 1, if physician i begins an on call service in period $t \in L$, 0 otherwise

General integer decision variables

o_{iw} amount of overtime for physician i in week w
u_{iw} amount of undertime for physician i in week w
h_{iw} amount of paid out time for physician i in week w
x_t^{out} number of outside physicians hours in period t

Model

$$\text{Minimize} \sum_{w\in W}\sum_{i\in I} c_i^{paid}\cdot h_{iw} + \sum_{w\in W}\sum_{i\in I} c_i^{over}\cdot o_{iw} + \sum_{t\in T} c^{out}\cdot x_t^{out} \quad (1)$$

$$y_{it}^{shift} = x_{it}\cdot(1-x_{i,t-1}) \quad \forall i\in I, t\in T \quad (2)$$

$$y_{it}^{rest} = x_{it-1}\cdot(1-x_{it}) \quad \forall i\subset I, t\subset T \quad (3)$$

$$\sum_{\tau=t}^{\min(t+T^{shift}-1,|T|)} x_{i\tau} \geq \min\left(\underline{T}^{shift}, |T|-t+1\right)\cdot y_{it}^{shift} \quad \forall t\in I, t\in I \quad (4)$$

$$\sum_{\tau=t}^{t+\overline{T}^{shift}} (1-x_{i\tau}) \geq y_{it}^{shift} \quad \forall i\in I, t\in\left\{1,\ldots,|T|-\overline{T}^{shift}\right\} \quad (5)$$

$$\sum_{\tau=t}^{\min\left(t+\underline{T}^{rest}-1,|T|\right)} (1-x_{i\tau}) \geq \min\left(\underline{T}^{rest}, |T|-t+1\right)\cdot y_{it}^{rest} \quad \forall i\subset I, t\subset T \quad (6)$$

$$\sum_{i\subset I} x_{it} + x_t^{out} > d_t \quad \forall t\in T \quad (7)$$

$$\sum_{t\in T^w\,|t\in I\,w-1,w\,T^{day}|} x_{it} + \sum f_1\left(i-T^{day}\right)\cdot y_{it-T^{day}}^{oncall} = r_i + o_{iw} - u_{iw} \quad \forall i\in I, w\in W \quad (8)$$

$$h_{iw} \geq o_{iw-1} \quad u_{iw} \quad \forall i\subset I, w\subset W \quad (9)$$

$$\sum_{\tau=\max(i-\underline{T}^{rest},0)}^{t-1}(1-x_t) > \min\{\underline{T}^{rest}, t-1)\cdot y_{it}^{oncall} \quad \forall i\in I, t\in L \quad (10)$$

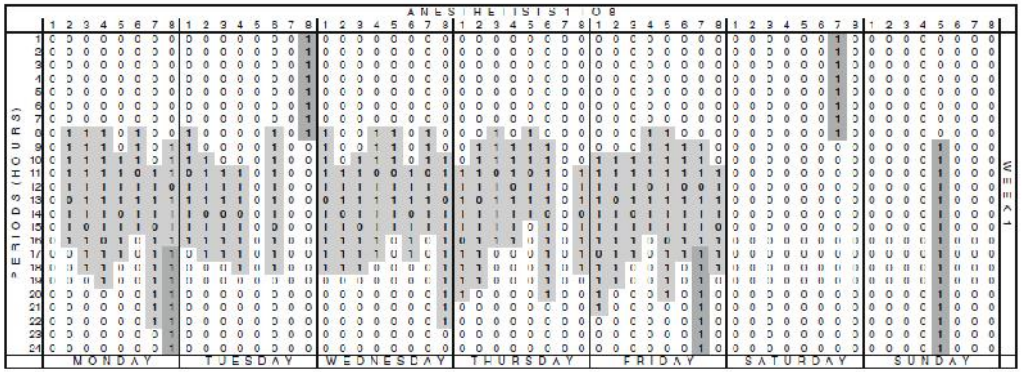
$$\sum_{\tau=i}^{\min(t+\underline{T}^{rest}-T^{oncall}-1,\,T)} (1-x_{i\tau}) \geq \min\left(\underline{T}^{rest}+T^{oncall}, |T|-t+1\right)\cdot y_{it}^{oncall} \quad \forall i\in I, t\in L \quad (11)$$

$$\sum_{i\in I} y_{it}^{oncall} = N^{oncall} \quad \forall t\subset L \quad (12)$$

$$\sum_{t\in L_w} y_{it}^{oncall} \leq O^{oncall} \quad \forall i\in I, w\in W \quad (13)$$

$$y_{it}^{oncall} + y_{i,t+T^{day}}^{oncall} \leq 1 \quad \forall i\in I, t\in\{t\in L.t<|T|-T^{day}\} \quad (14)$$

Optimal roster



See:
Brunner JO, Bard JF, Kolisch R., “Flexible shift scheduling of physicians”, Health Care Manag Sci. 2009 Sep;12(3):285-305.

Oleg Pianykh opiany@gmail.com

ОК, алгоритм готов. Что дальше?

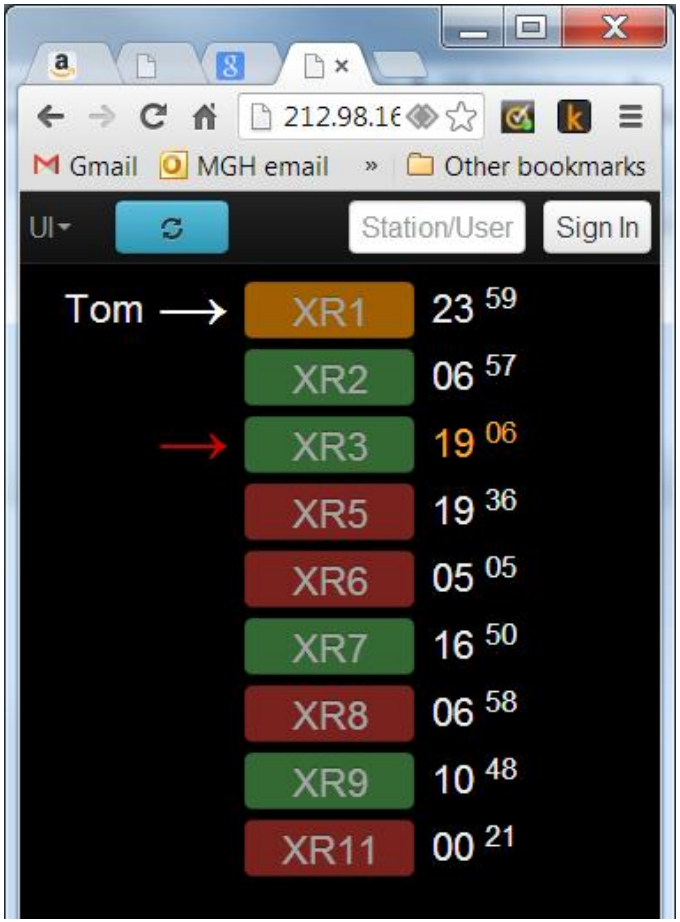
Разработка приложений

- Прекрасно: мы проанализоровали «большие данные» и создали улучшенный алгоритм работы

  НО

- Любой алгоритм необходимо **внедрить**:
  - Доказать преимущества решения на практике
  - Создать надежное, устойчивое решение
  - Создать удобное решение
  - Отработать все возможные случаи, когда решение не срабатывает

Пример: распределение пациентов по рентгенам



- Web application to view and book X-ray stations”
  - Faster patient dispatching
  - Fairness in exam-to-tech assignments
  - Optimal station utilization

→ Reduction of variability

→ Reduction of over-utilization and stress

Oleg Pianykh opianykh@mgh.harvard.edu

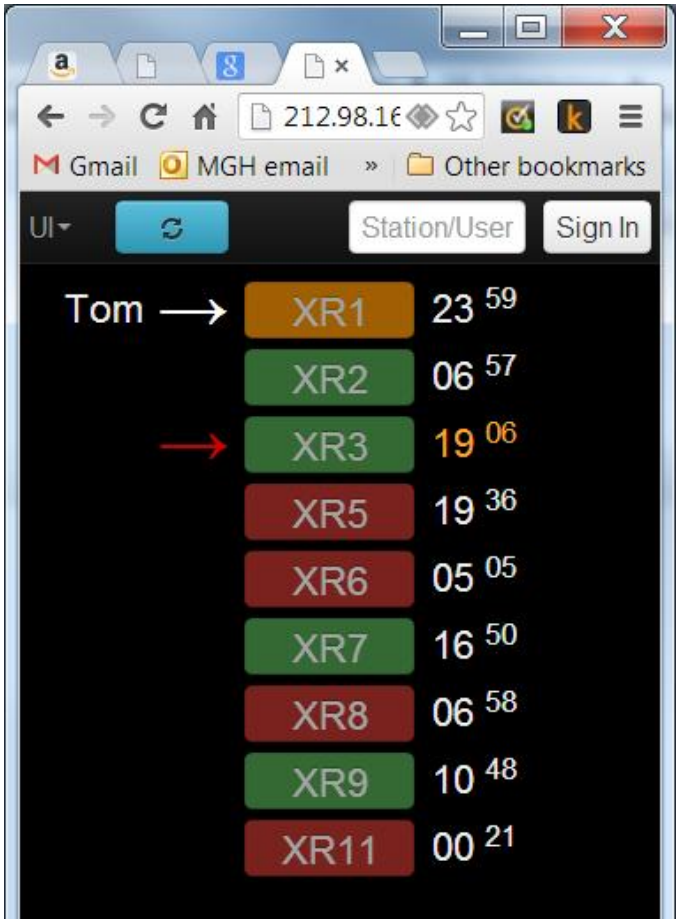
Всё, проблема решена?

- Нет!
  - Интеграция с госпитальными системами (источники данных)
  - Надежность в работе (что делать, если села батарейка, исчез WiFi, ввели не тот пароль, …?)
  - Масштабируемость; переносимость на другие процессы
  - Обучения персонала
  - Работа с пациентами

Человеческий фактор

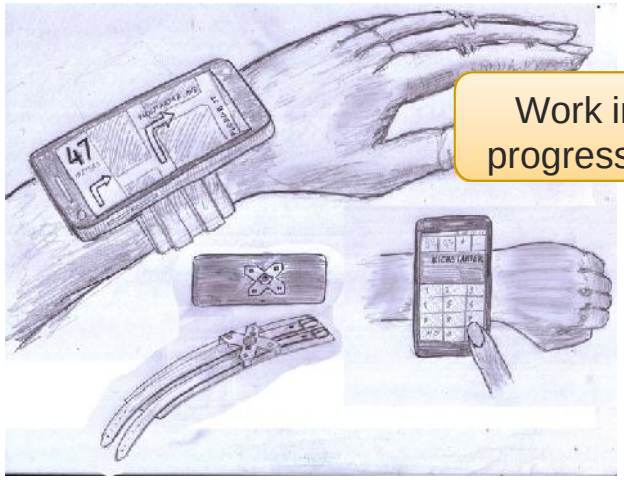
- Любое массовое внедрение алгоритмов неизбежно превращается в «социальную игру»: «Обмани машину»
  - Люди пытаются обмануть алгоритм
  - Люди пытаются обмануть друг друга
  - Люди далеко на просто «интегрируются» с компьютерами и гаджетами!
- Ваше приложение должно включать пользователей-антогонистов

Пример: распределение пациентов по рентгенам



Oleg Pianykh opianykh@mgh.harvard.edu

Пример: «интеграция» с человеком



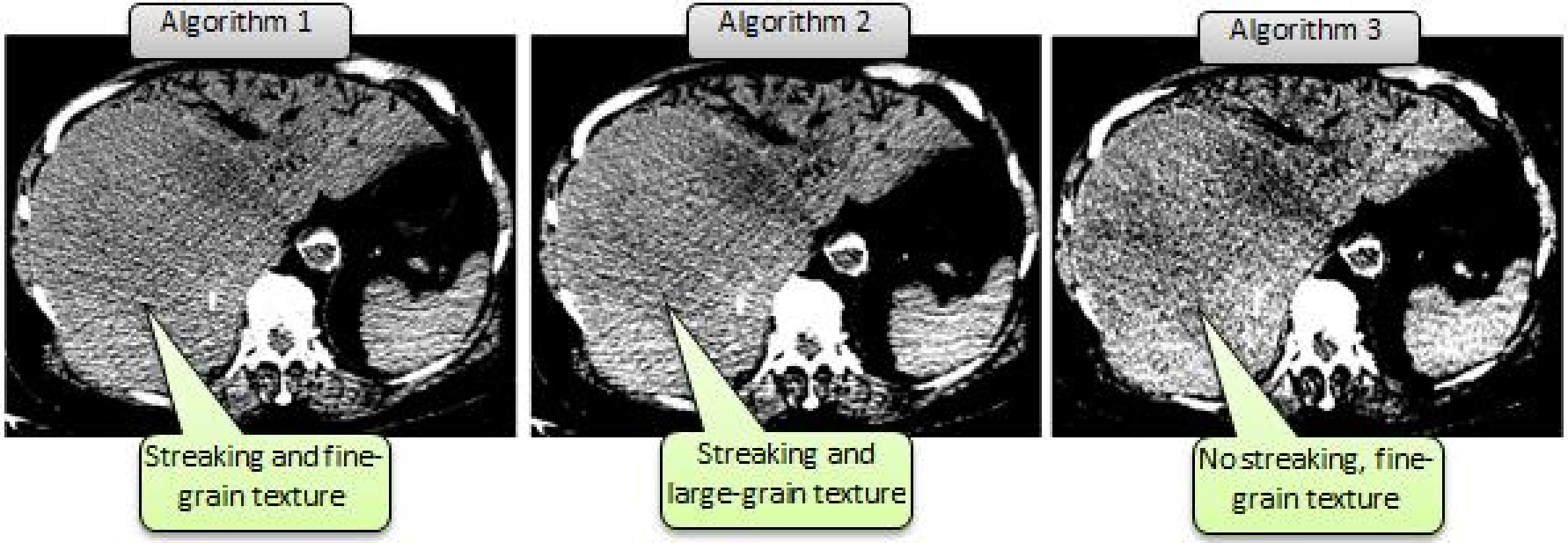
http://www.kickstarter.com/projects/1958024164/connectorr-smartphone-iphone-gps-wearable-mounting?ref=recently_launched

Oleg Pianykh opianykh@mgh.harvard.edu

Заключение

- Да, «большие данные» в медицине абсолютно необходимо анализировать и использовать для оптимизации работы любого процесса.
- Но:
  - Начинать надо с задачи, а не решения
  - Надо вкладываться в понимание накопленной информации еще до того, как создавать какие-либо решения на ее основе
  - В медицине наиболее важно законченное, до конца внедренное решение. Это налагает жесткие требования на качество алгоритма.
  - «Не навреди» → «Сделай значительно лучше»

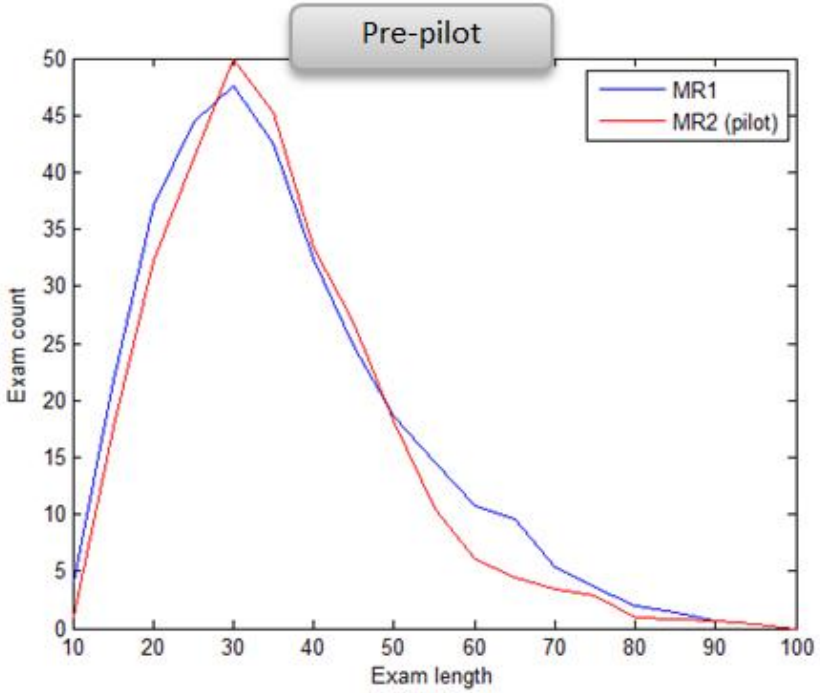
Неправильный анализ

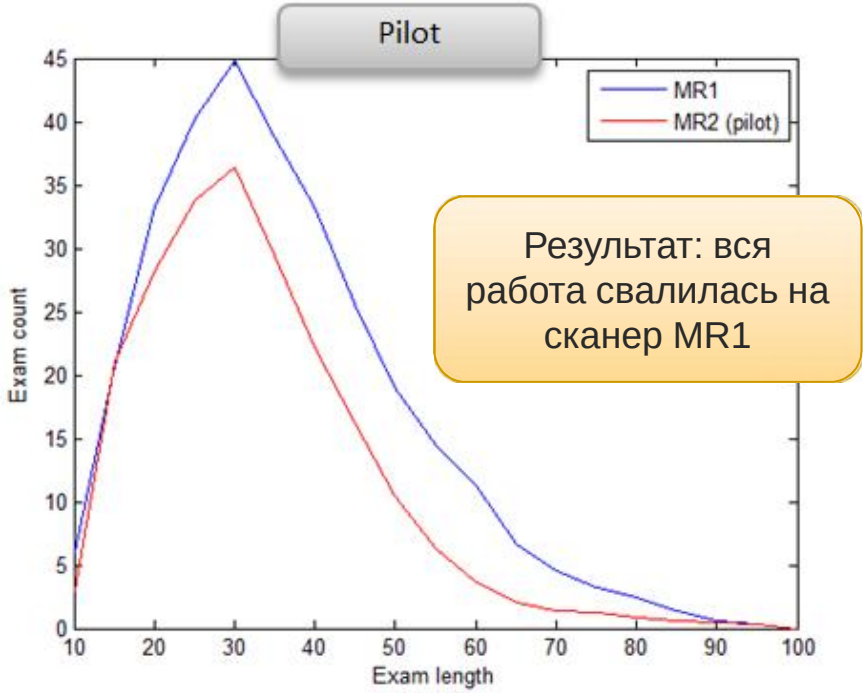


Олег Пьяных, opiany@gmail.com

Пример проблемы со «ускорением» обслуживания

- На сканере MR2 решили обслуживать пациентов на 5 минут быстрее:





Результат: вся работа свалилась на сканер MR1

Oleg Pianykh opianykh@mgh.harvard.edu