Welcome to this INTERACTIVE workshop

Why?

How?

How (Not to do it)?

Consent?

Anil Mehta, Registry implementer:
(who was told in 1993, it was impossible)

CRITICS/Criticism/Scepticism

Sibelius

statues

Goethe

evidence

George Bernard Shaw

progress

INTERACTION

PLEASE PLEASE FEEL FREE TO DISAGREE: IT'S OK WITH (Anil) Me

Three Why (s) are all numerical, sorry, if like me, numbers are difficult

Lord Kelvin:

if you put numbers onto something, that something is changed for good

Leonardo Da Vinci:

Make measurable that which is not (currently) measurable.

Transendental approach (e, pi are such numbers):

Give a person a task that is just out of reach (Tantalus), they will never stop till they die

How(s) ?

This how is integrated with how-not -0

Experience

is just another word for the number of mistakes you make

(provided you learn something)

Otherwise you make the same mistakes again and again..

Challenges and Consent

Changing Consent landscape is the common challenge

www.ecfs.eu/registry/consent is how we cope with change

Challenges in the utility of data you face are all different

Chinese Book of War…and Napoleon's campaigns…and Alexander the Great

AT THEIR HEART: All share logistics as their core claims to (transient) success

BUT MY APPROACH IS DIFFERENT:

Chinese source book of philosophy (1000 pages); Plato's Republic , Descartes/Locke

All these great generals/philosophers say to me:

1. You need to perform a *systems analysis SA*
2. *SA means you need to tell me what you want to achieve (NEED)*
3. I need to analyse that NEED and rephrase it into numbers
4. I need to rearrange those numbers into *log scale of utility for you*
5. I need to overcome 'the rate limiting steps' (Napolean/Alex. Grt)
6. You need to understand that as you increase the number of items,

QUALITY FALLS EXPONENTIALLY AS NUMBER OF DATA ITEMS GOES UP (RLS-data entry)

YOU CAN DO A GREAT DEAL WITH 10 (DEMOGRAPHIC) data items – <u>to start with</u>

AND VERY LITTLE MORE with 100 items…inverse log scale of failed utility (trans No e)

WHAT HAPPENS BETWEEN 10 FIELDS AND 100 FIELDS

HOW CAN WE COPE WITH COMPLEXITY

How can you start affordably

A rate limiting step is poor training of the poorly paid data entry staff

STAFF THINK: Why should I collect data which I never need or understand?

Who checks anyway...who cares...missing data..etc ?

That which is unchecked, decays – who said that?

This gives rise to the principle:

If the data is useful, it will be auto-checked (by those using it daily)

Usefulness is a time limited quantity – feedback the data in real time
to the person entering the data provided they can use it for a purpose

Back to square 1- what do you want from the data??

The power of simple demographics

If you know the
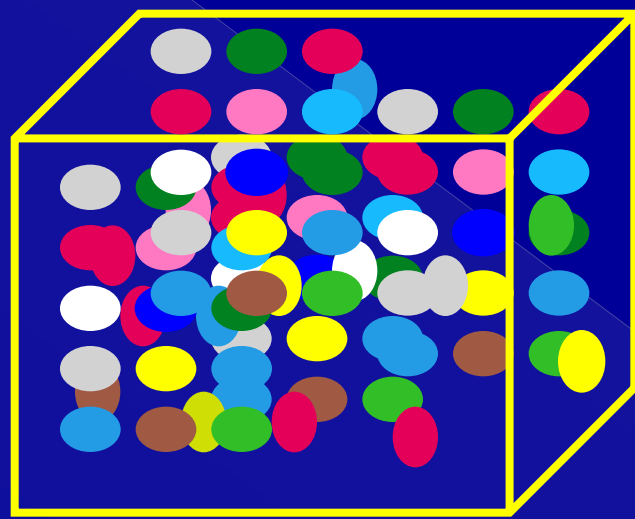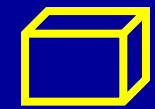
age

alive status

age when dead..

over time of say 5 years
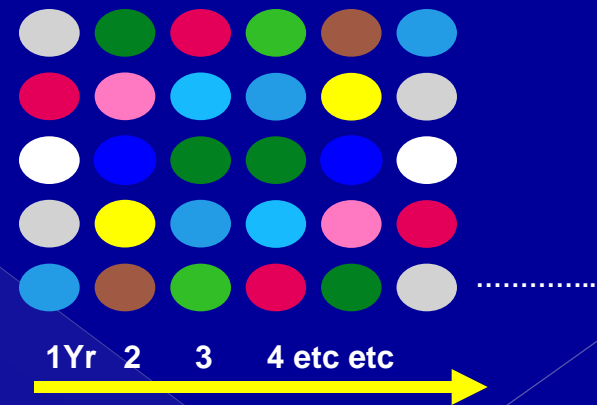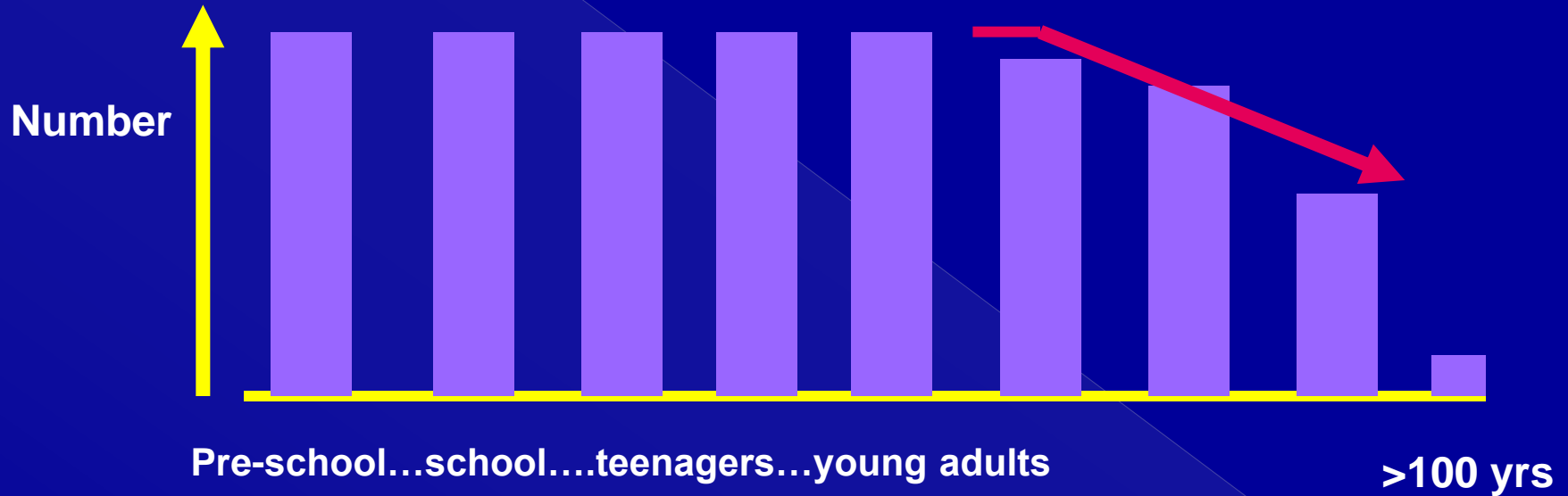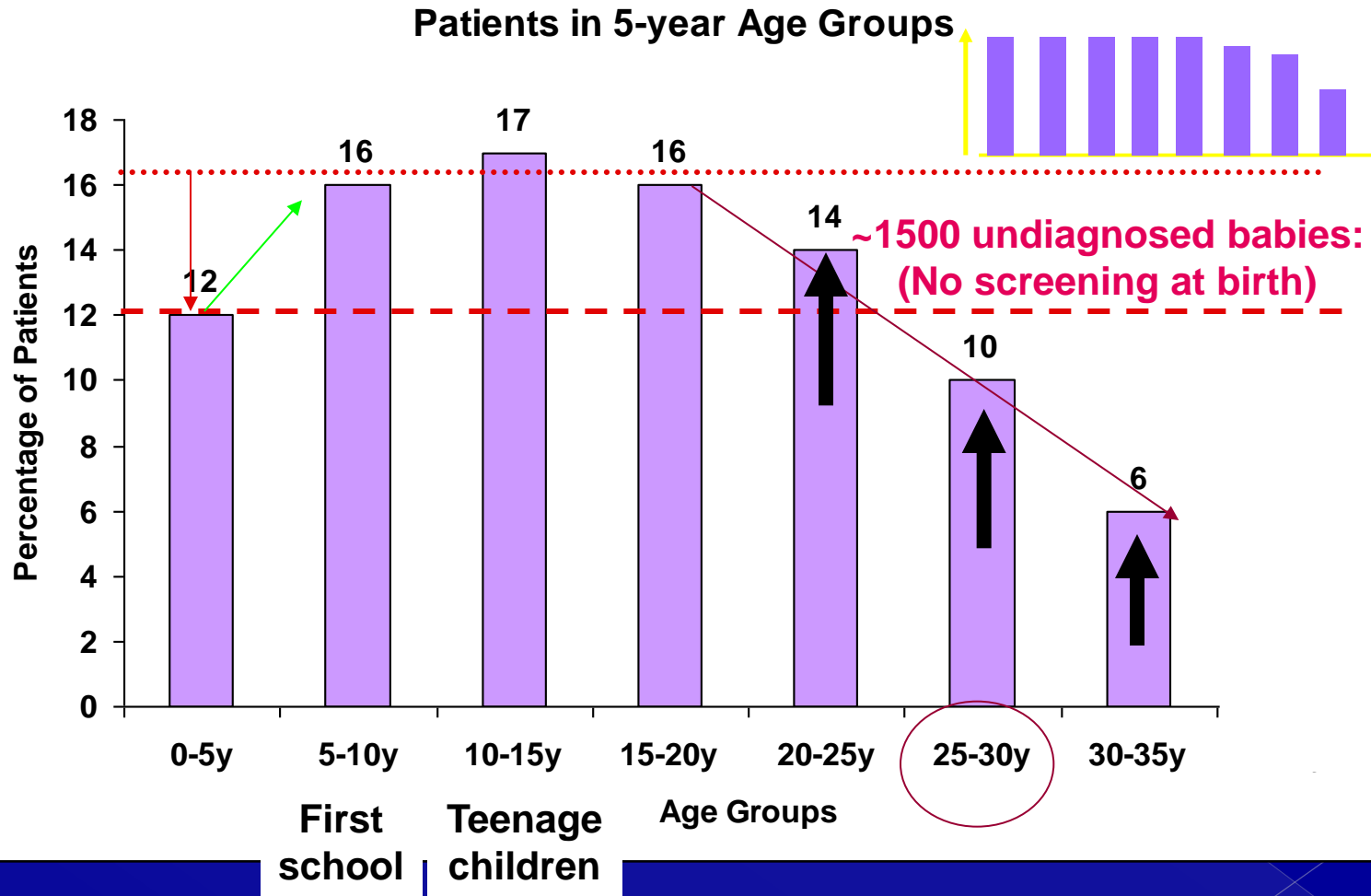
Imagine a census in the town where you were born

Babies Children   Adults   Elderly

So what happens if you take a census of CF-affected citizens?

Number

Pre-school…school….teenagers…young adults

>100 yrs

Each Bar shows a 5 year age band for CF patients

Patients in 5-year Age Groups

~1500 undiagnosed babies:
(No screening at birth)

Bar values by age group:
- 0-5y: 12
- 5-10y: 16
- 10-15y: 17
- 15-20y: 16
- 20-25y: 14
- 25-30y: 10
- 30-35y: 6

Y-axis: **Percentage of Patients** (0 to 18)

X-axis: **Age Groups** — 0-5y, 5-10y, 10-15y, 15-20y, 20-25y, 25-30y, 30-35y

First school

Teenage children
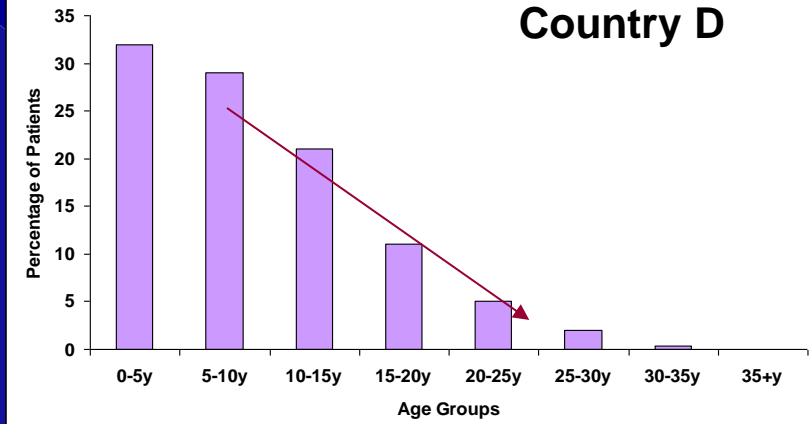
Mehta et al Journal of Cystic fibrosis special issue

The issue is open access at:
http://cysticfibrosisjournal.com/issues/contents?issue_key=S1569-1993(10)X0007-8
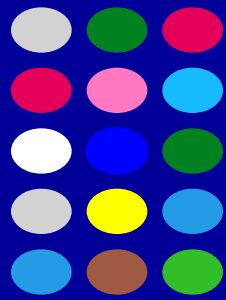
Genetics provide a good idea of this:

Par example: 250 cases per annum in UK

CF Deaths

2008
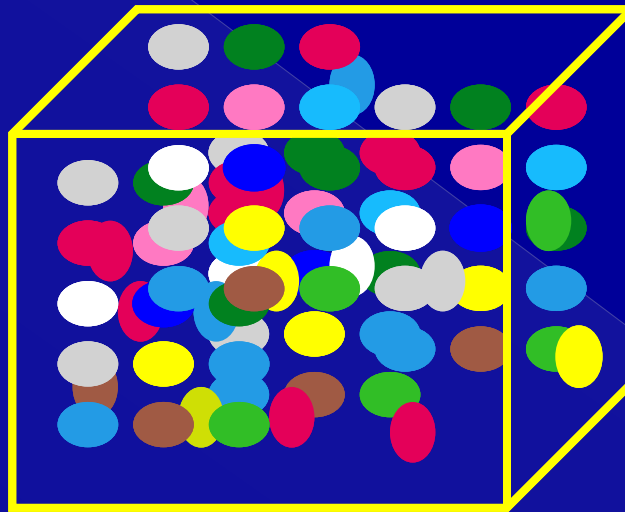
Pool of survivors
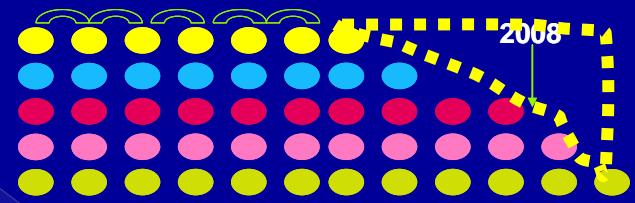
Concept of the Median Age at Death

4 groups, of equal size
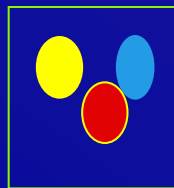
Frame a new question:

Of those (few) who die, At what age are half of them dead?

Divide the age at which patients die into 4 groups

50% have died (UK=23)

Europe (Non-UK)

Age at Death

4 Countries
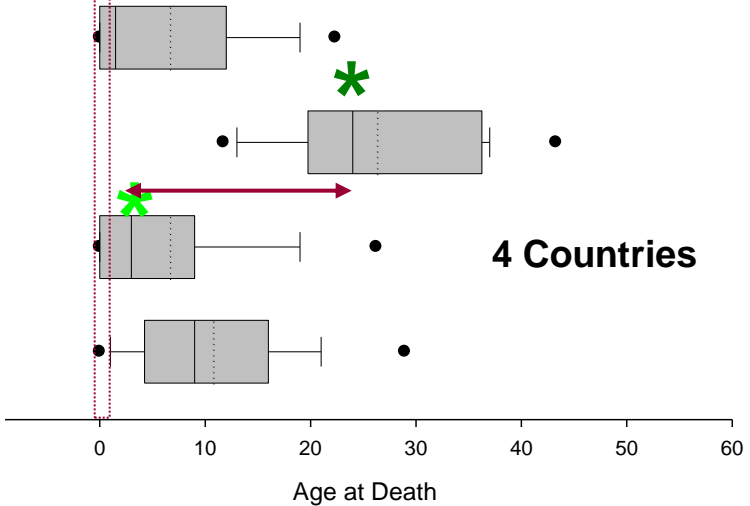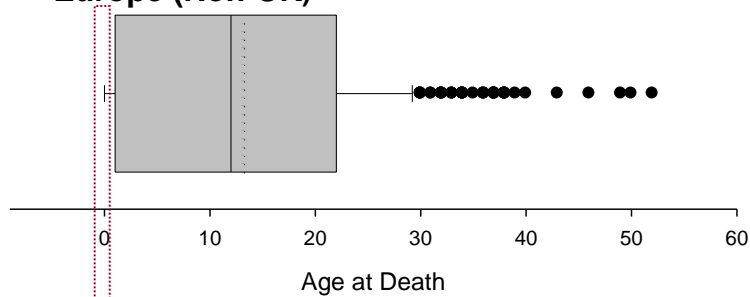
Age at Death

1 in 4 deaths are in the first year
1 in 2 deaths are below 10 years of age

50%

2nd

1st quarter
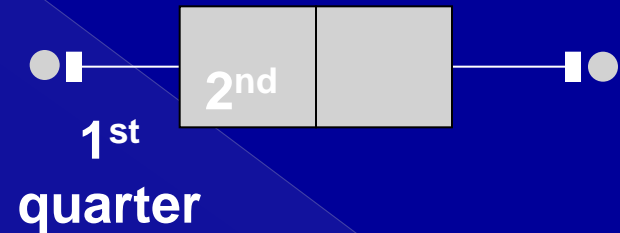
Age at Death

FIRST RULE OF DEMOGRAPHICS

DEATH IS A POWERFUL WEAPON

IN POLITICS…DEATH (AND TAXES)

95% of the common CF form –is diagnosed in early childhood, if health care is good
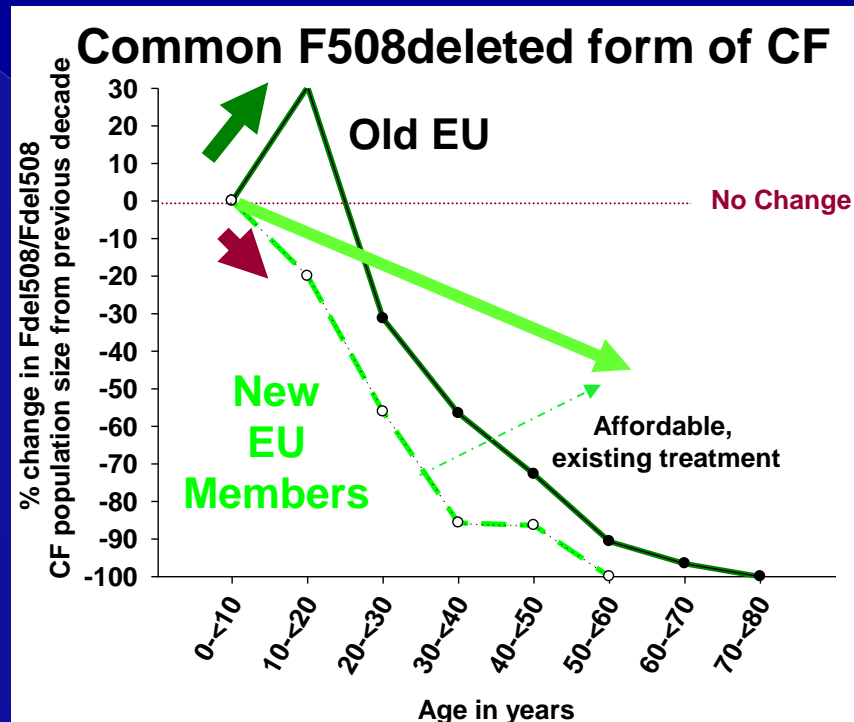
Per decade of life, how much does the population fall in %?

McCormick et al Lancet 2010

In summary at the beginning, KISS:
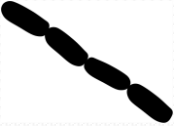
1. Elocutory intent……Person A due to disease X Needs a registry

2. Illocutory impact 1 …Person B says I can get a career from this NEED (Me)

3. Illocutory impact 2 – Company/Charity C says I can get impact on Gov

4. **Beware of different impacts to different people with different needs**

   – confusing aims, ideals, language…**contract lawyer is needed**

IT ALL COMES BACK TO WHAT YOU WANT, Quantitatively

Lawyers: The tricky problem of data ownership

- Who owns the collected data?

- Who owns the analysed /processed data? (IP issue)

- Who has permission to do what with each type?

- Consent is key but is inadequate, insufficient, ill-understood
  - Ownership has to be agreed
    - Are you paying for storage space?
    - Are you paying for analysis?
    - Are you paying for processing the data?
    - What do you want to be free of charge?

Registry: Evolution of ideas

1. The spreadsheet registry (Dundee 1994)

2. Multiuser access (has taken until 2014)

3. Structure and security (highest possible)

4. Communicating with health systems (Not easy)

5. Virtual centre of expertise (the future)

European Cystic Fibrosis Society (ECFS): Patient Registry NEEDS; www.ecfs.eu

- **To measure, survey and compare** aspects of cystic fibrosis and its treatment in the participating countries
  - Annual cross-sectional data gathering GET PHASE = consent
- **Providing data** for epidemiological research
  - ANALYTICS i.e. Act on the data gathered in some way
  - This means storage, access, control, the lawyers (OpenApp)
- **Identifying special patient groups**
  - suitable for multi-centre trials.
  - This means work-partners with different NEEDS

SURVEY a range of rare diseases seen in a clinic in 1 year

Cardiac Abnormality	Number (N)	As a % of Total Consultations
DNA	78	17.3
VSD/ASD/AVSD	64	14.2
Valvular	55	12.9
Innocent Murmur/Normal heart	48	10.6
Complex Cardiac	45	10.0
Palpitations/SVT	34	7.5
Family History of Cardiac Disease	32	7.1
Named Genetic Syndrome	23	5.1
Tetralogy of Fallot	20	4.4
Hypertrophic Cardiomyopathy	10	2.2
Infective	9	2.0
Heart Block	8	1.8
Coarcation of the Aorta	8	1.8
Chest Pain	6	1.3
Syncopal Episodes	6	1.3
Renal/Hypertension	6	1.3
PDA	5	1.1
Transposition of Greater Vessels	4	0.9
Other	23	5.1

GET THE DATA PHASE IS DONE BUT WE CANNOT COMPARE

Legend:
- DNA
- VSD/ASD/AVSD
- Other
- Valvular
- Innocent Murmur/Normal Heart
- Complex Cardiac
- Palpitations/SVT
- Family History
- Named Genetic Syndromes
- Tetralogy of Fallot
- Hypertyrophic Cardiomyopathy
- Infective
- Heart Block
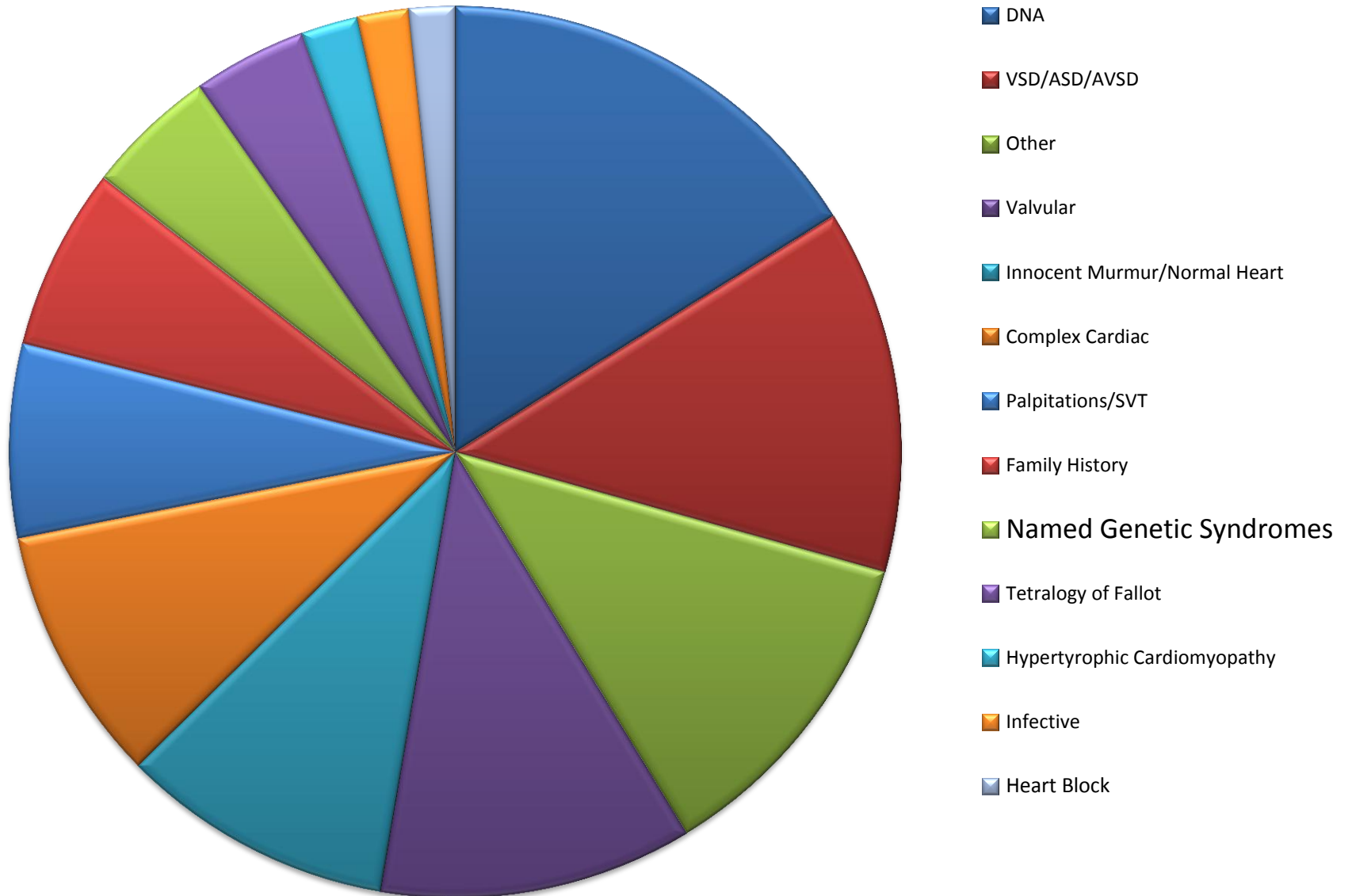
ACT ON THE DATA: SURVEY
Percentage of Patients Discharged with Certain Diagnoses
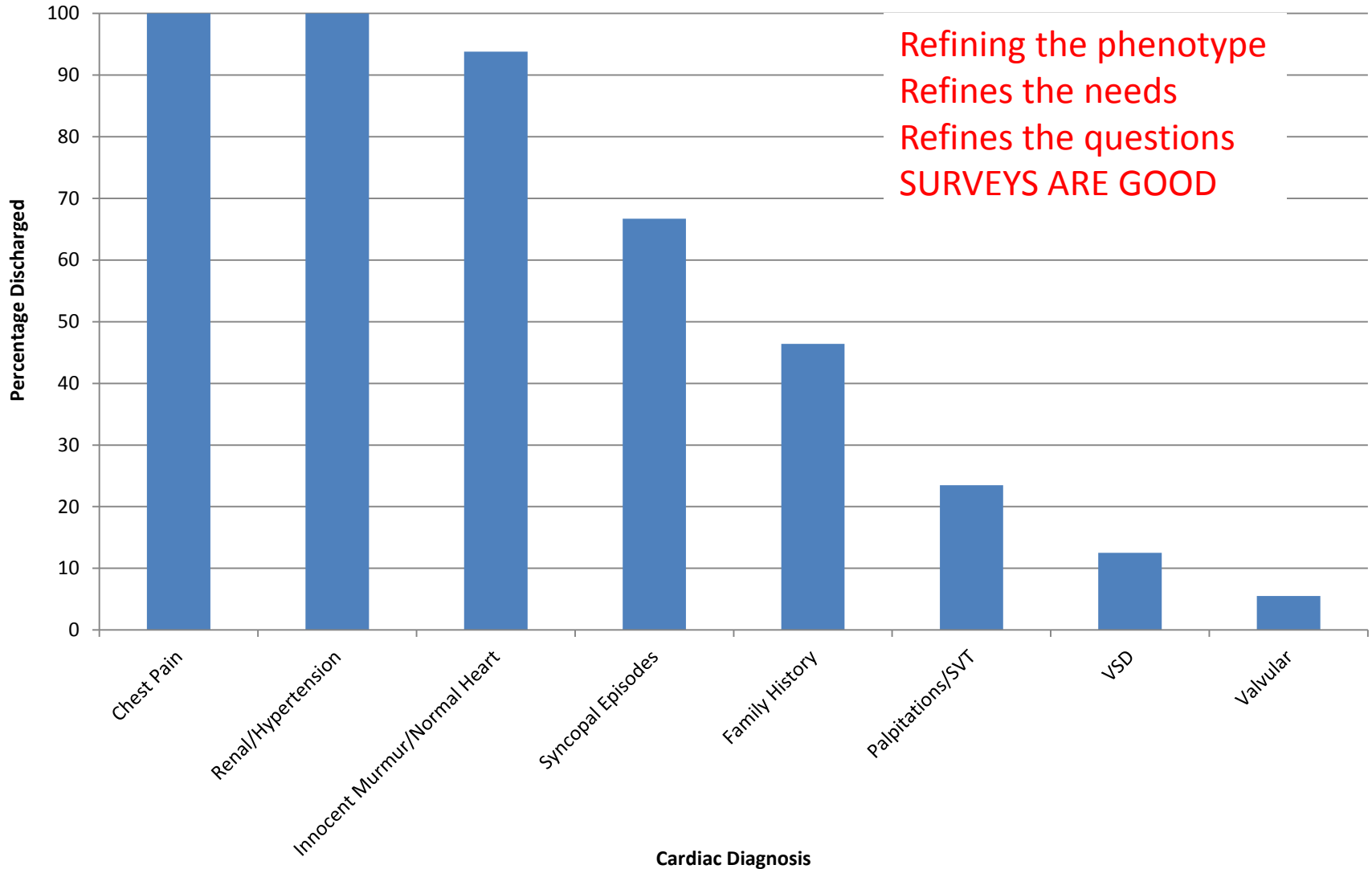
What question am I asking and how will the data answer it?

Cardiac Abnormality	Number (N)	As a % of Total Consultations
DNA	78	17.3
VSD/ASD/AVSD	64	14.2
Valvular	55	12.9
Innocent Murmur/Normal heart	48	10.6
Complex Cardiac	45	10.0
Palpitations/SVT	34	7.5
Family History of Cardiac Disease	32	7.1
Named Genetic Syndrome	23	5.1
Tetralogy of Fallot	20	4.4
Hypertrophic Cardiomyopathy	10	2.2
Infective	9	2.0
Heart Block	8	1.8
Coarcation of the Aorta	8	1.8
Chest Pain	6	1.3
Syncopal Episodes	6	1.3
Renal/Hypertension	6	1.3
PDA	5	1.1
Transposition of Greater Vessels	4	0.9
Other	23	5.1

Next need: Go Compare- Eigen-Value

- The same value across the world ($ and £)
- Agreements for exchange of eigen-values;
  - Between whom?
  - Coding agreement (Orpha-code or Snomed-CT or ?)
  - Definition(s)
- **Quantitation is a problem (numbers again!)**
  - Take the number 1..we know it intuitively
  - $[-e^{i\pi}] = 1$

What is the point of a registry?

- In mathematics, a **matrix** (plural **matrices**) is a rectangular array[1] of numbers, symbols, or expressions, arranged in rows and columns.[2][3]

- The individual items in a matrix are called its *elements* or *entries*.

- Spreadsheets are matrices

- Make a table of numbers by row and by column

- Each number is a value to a scale of some data element

All these people have the same problem of numbers

European CF Society

Child Health in Europe

HSE

Irish Coast Guard
GARDA CÓSTA na hÉIREANN

RCSI

Air Ambulance

Road
Safety
Authority

Irish Thoracic Society

RCPI

VERTEX

National Cancer Control Programme

NOCA National Office of Clinical Audit

NRA
An tÚdarás um Bóithre Náisiúnta
National Roads Authority

CFRI

There are only 3 types of registry data

Numbers attached with precision of measurement

Scale-attached and precise semi-quantitation

Complete rubbish/garbage/detritus/confusion

To summarise, after 20 years of work, by linking with industry/academia

CONSENT TO [RE-]GET

CONSENT TO STORE

CONSENT TO DO

1 Patient Portal

Patient
- Data input
- Surveys
- Lab data
- Other info

2 Physician Portal

Physician   Lab

- Demographic Data
- Clinical Data
- Patient consent
- Care pathway - schedule

3 Id Mgmt
- 3rd party secure
- encryption and hash
- Patient identifiers

Registry Portal

4 Registry Data
- Signs
- Symptoms
- Laboratory values
- Treatments

Biorepository

5

- User management
- Data Access Mgmt

6 Derived Data
Research
Reports

- Survival data
- Outcome predictors
- Treatment effectiveness
- Safety profiles
- Reports
- Analysis tools

Researcher
Registry or
Industry

NEXT we must think about where data comes from (in groups) and Data storage and protection of data identity – Identity Management

1 Patient Portal

Patient
- Data input
- Surveys
- Lab data
- Other info

2 Physician Portal

Physician   Lab
- Demographic Data
- Clinical Data
- Patient consent
- Care pathway - schedule

3 Id Mgmt
- 3rd party secure
- encryption and hash
- Patient identifiers

Registry Portal

4 Registry Data
- Signs
- Symptoms
- Laboratory values
- Treatments

Biorepository

5
- User management
- Data Access Mgmt

6 Derived Data
Research Reports
- Survival data
- Outcome predictors
- Treatment effectiveness
- Safety profiles
- Reports
- Analysis tools

Researcher
Registry or Industry

GET THE INITIAL DATA PHASE (STARTING UP)

Fixed cost basis of working…pay for storage
Rack space on a server such as Amazon

Sales Pitch from Companies

- No license copy costs. No license management
- You can walk away at no cost from the
    - Proliferation of new technologies
    - BENEFITS TO YOU:
    - Open standards, open data, open access, open source
    - You can negotiate your own needs

Anil Mehta working with OpenApp – here is an approach for all of you

- Different Users have Different Needs
- Easy to extend applications:
  - Different country datasets
  - Flexible views: Children / Adult
  - CF Newborn Screening
  - Clinical trials
- Flexible security model
- Business rules for summarizing / consolidating data
- Integrated real time reporting

- Same security level as online banking
- All changes including deletes are user & date stamped
- Log allowing audit trails
- Multiple consent models – paper or electronic signature
- Forthcoming pan European Data Protection Law:
  - Explicit Consent will require online approval

To summarise, after 20 years of work, by linking with industry/academia

CONSENT TO [RE-]GET

CONSENT TO STORE

CONSENT TO DO

1 Patient Portal
Patient
- Data input
- Surveys
- Lab data
- Other info

2 Physician Portal
Physician   Lab
- Demographic Data
- Clinical Data
- Patient consent
- Care pathway - schedule

3 Id Mgmt
- 3rd party secure
- encryption and hash
- Patient identifiers

Registry Portal

4 Registry Data
- Signs
- Symptoms
- Laboratory values
- Treatments

Biorepository

5

- User management
- Data Access Mgmt

6 Derived Data
Research
Reports

- Survival data
- Outcome predictors
- Treatment effectiveness
- Safety profiles
- Reports
- Analysis tools

Researcher
Registry or
Industry

Use a secure online web browser solution as recommended by EPIRARE

Do not develop on Excel or Access

Standardize and use a Common Data Element set based on for example on the NIH model registry ORDR data set

as well as coding standards such as Orphanet, ICD-10 SNOMED CT

Design intra country interoperability from the start.

Pooling and sharing data with other patient organizations throughout Europe is much more important than integrating with hospital EHR systems or national rare disease networks.

4)   **Legacy systems:** Migrate existing data and use this as a Proof Of Concept to your RD centres and funders.

5)   **A web based portal** can easily be customized to add the required disease specific data elements. Agree the broad specifications for project deliverables for phase 1, phase 2 and phase 3. Your vendor can give you a fixed quote for phase 1, an estimate for phase2 based on programmer days and lump all other features into phase 3.

6)   **Have a clear registry purpose** so you only collect data that you really need. Otherwise you will end up with a lot of missing data.

7)   Consult with an epidemiologist to compare intra country RD standards of care and medical equipment used so that data can be combined without bias. For example different for diagnostic criteria will adversely affect disease populations in different countries.

8)   Don't forget to budget for data migration. There is extremely valuable data within old standalone systems, spreadsheets and databases. Use the epidemiologist to define the rules for migrating this historic data. For example to define outlier data and what to do about missing and combining data.

9)   Recognize from the outset that some centres will be reluctant to share data. Data protection will be used to cover up a concern over loss of control.

10)   **The registry should offer patient secure online login** to allow patients to review lab results and update data to include PROM quality of life surveys and to act as a hub for mHealth data integration. Clinicians can spend more time reviewing patient entered data and less time asking basic clerical questions.

11)   **A patient portal** can also be used to create a dynamic consent model which could address concerns about the forthcoming Personal Data Protection Legislation where explicit consent is required for secondary uses.

12)   The biggest single improvement in data quality can be achieved by removing paper from data collection. **Disease centres should be provided with low cost network enabled tablets running the browser software.** These can: (1) reduce bad on missing data entry (2) provide data validation (3) bypass restrictive hospital data protection and ICT policies and (4) provide patients with real time variance reporting against the natural history of the disease.