NCIN
national cancer intelligence network
Using information to improve quality & choice

Extracting cancer specific data

Sean McPhail, Steve Davies, NCIN

NCRI
National Cancer Research Institute

NCIN
national cancer intelligence network
Using information to improve quality & choice

Life with small numbers

Sean McPhail

Senior Analyst, NCIN

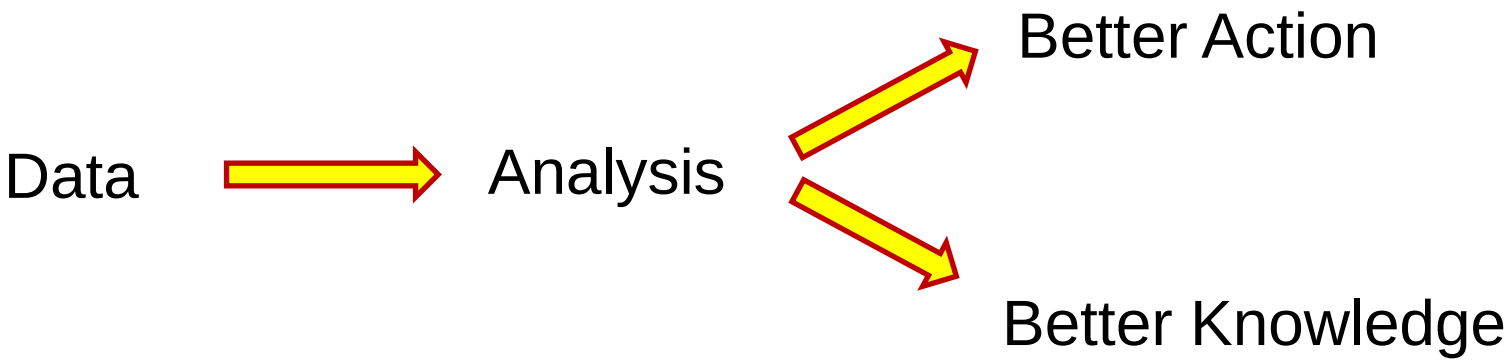
NCRI
National Cancer Research Institute

Big Picture

NCIN
national cancer
intelligence network
Using information to improve quality & choice

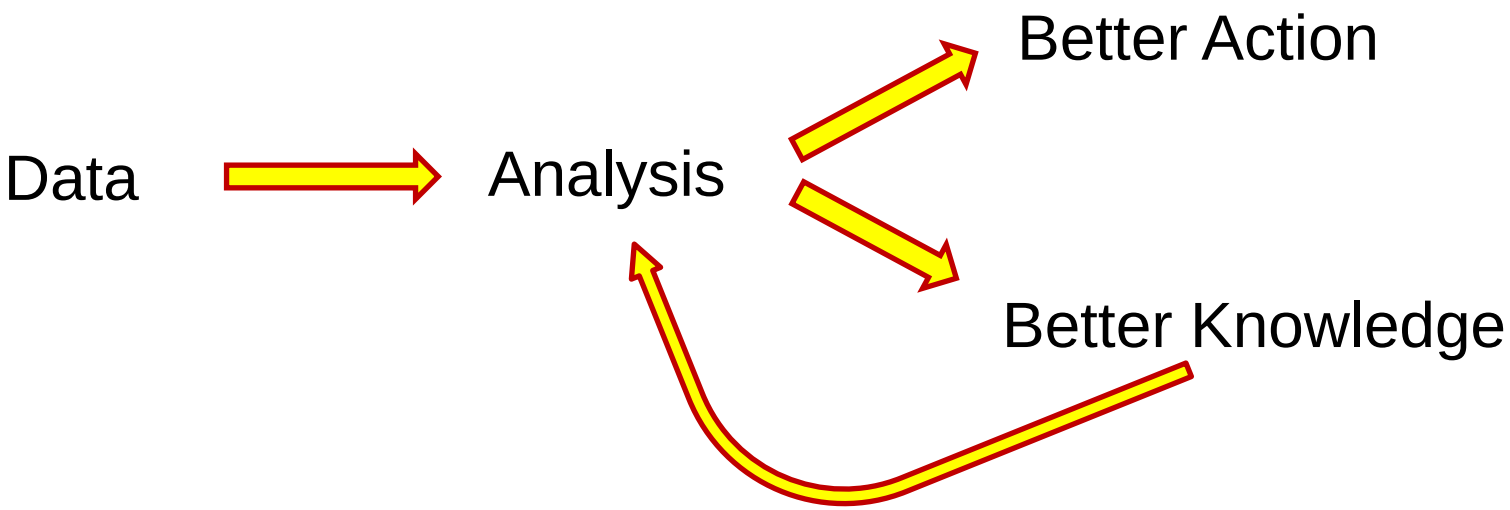
Big Picture

NCIN national cancer intelligence network
Using information to improve quality & choice



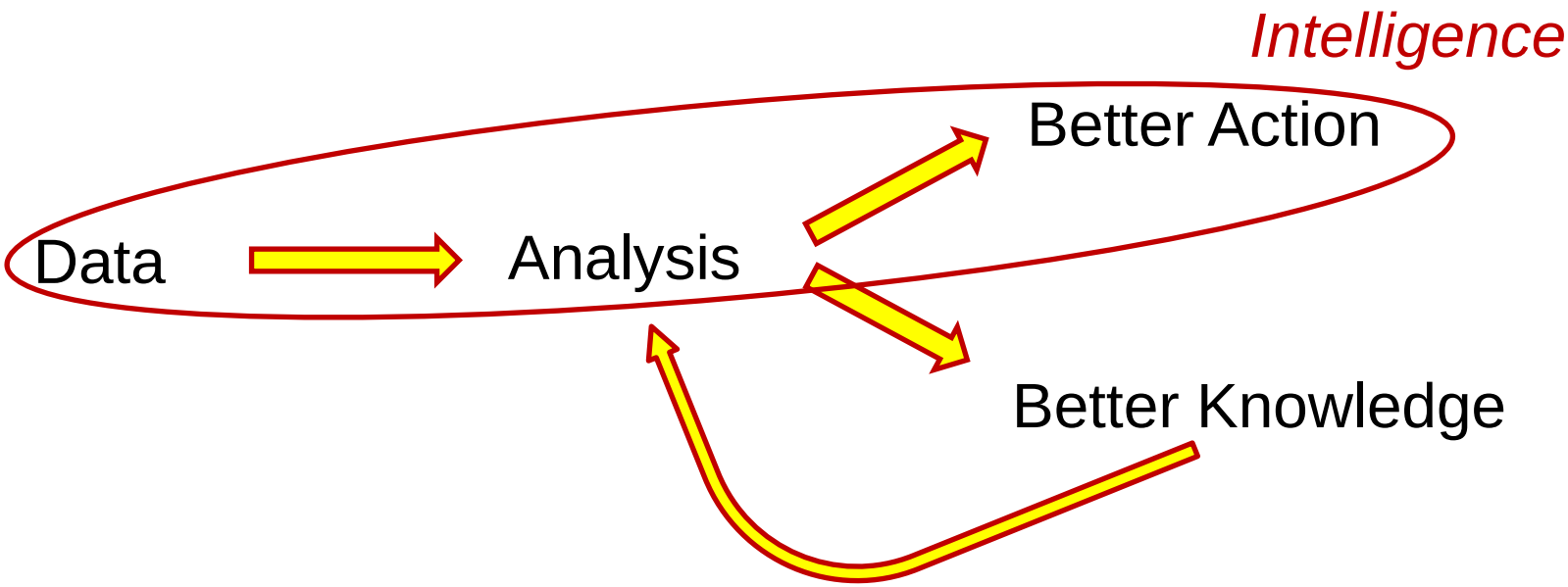
Big Picture

NCIN national cancer intelligence network
Using information to improve quality & choice



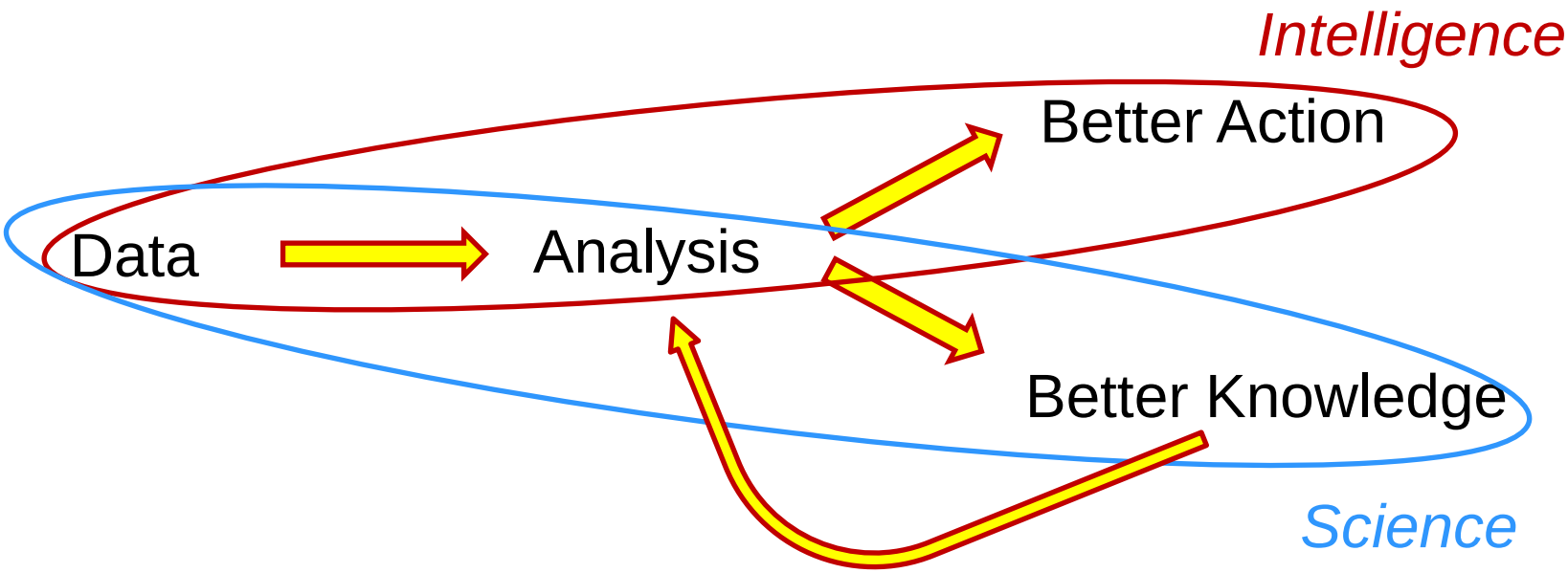
Big Picture

NCIN national cancer intelligence network
Using information to improve quality & choice



Big Picture

NCIN national cancer intelligence network
Using information to improve quality & choice

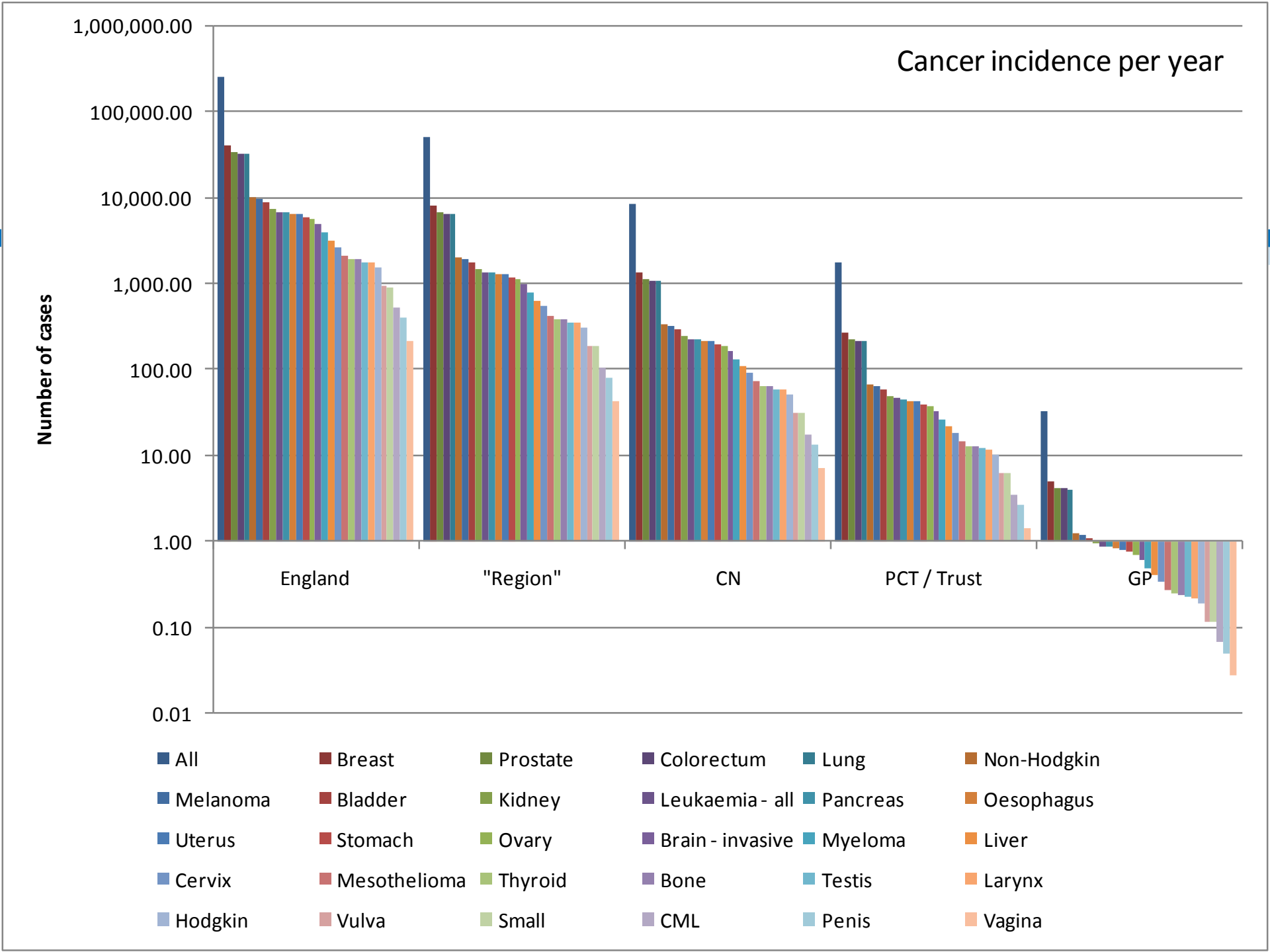


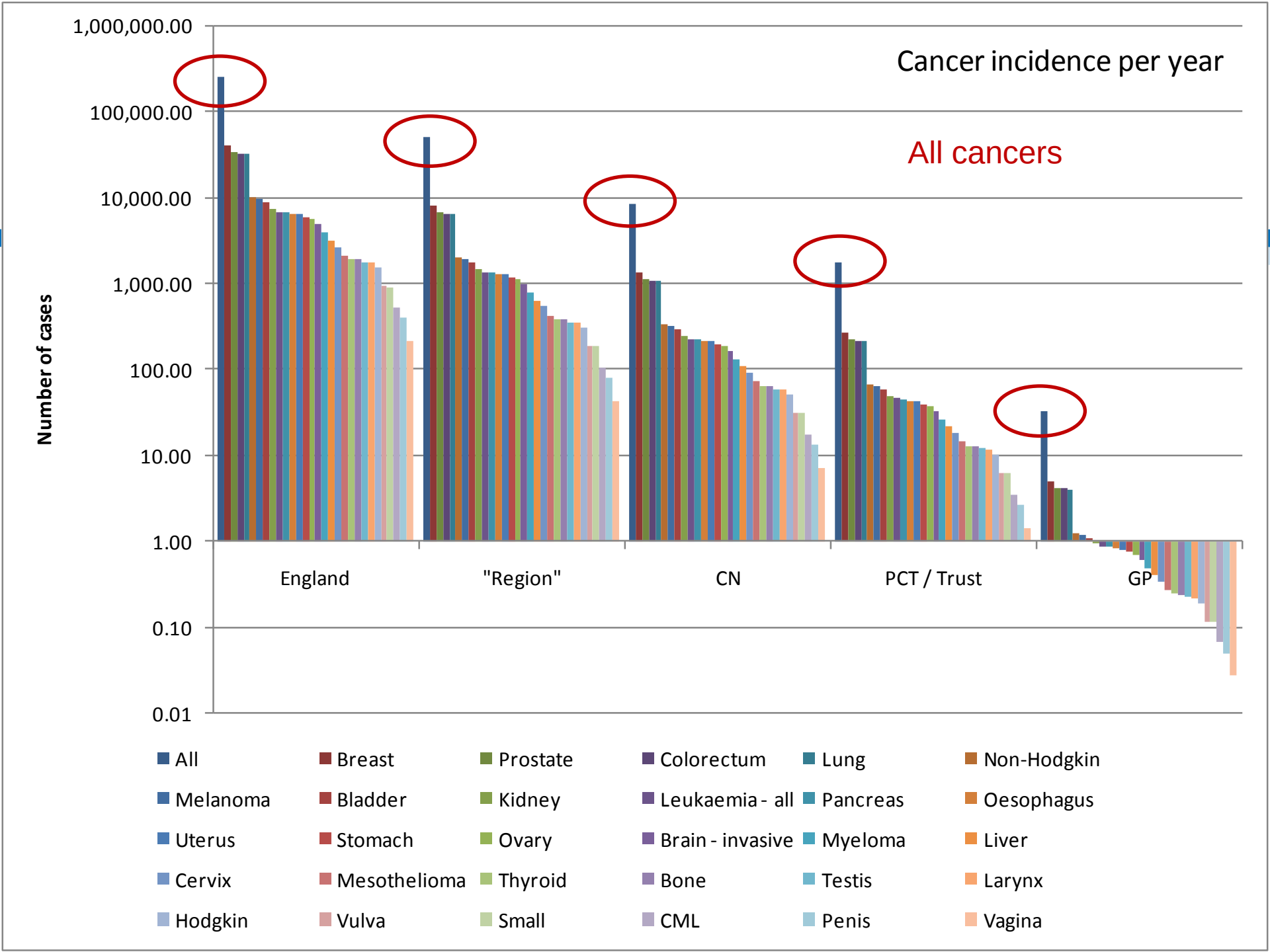
Introduction

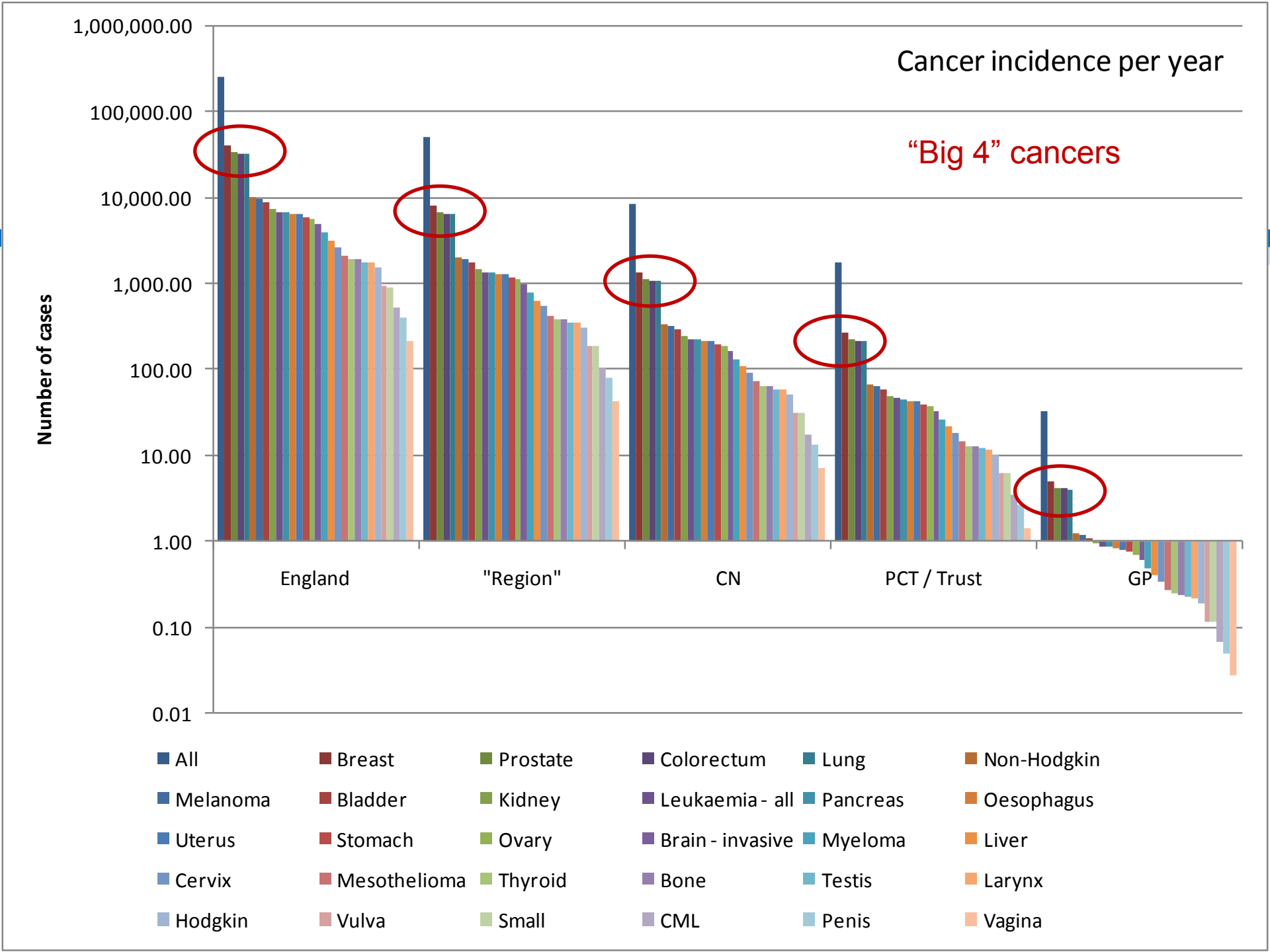
NCIN national cancer intelligence network
Using information to improve quality & choice

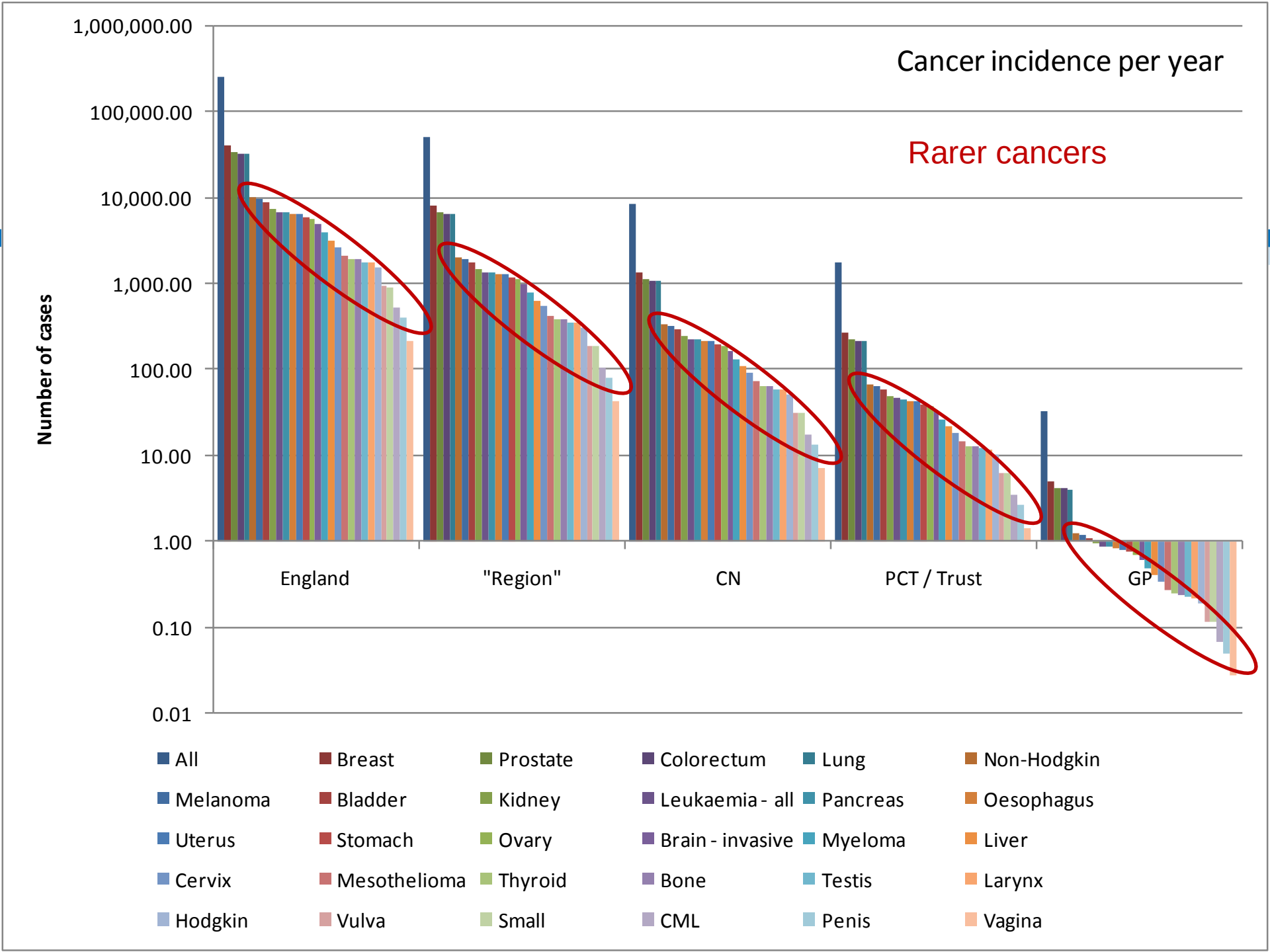
What impact on ‘useful analysis’ does the small number of cancers in ‘rarer’ types have?

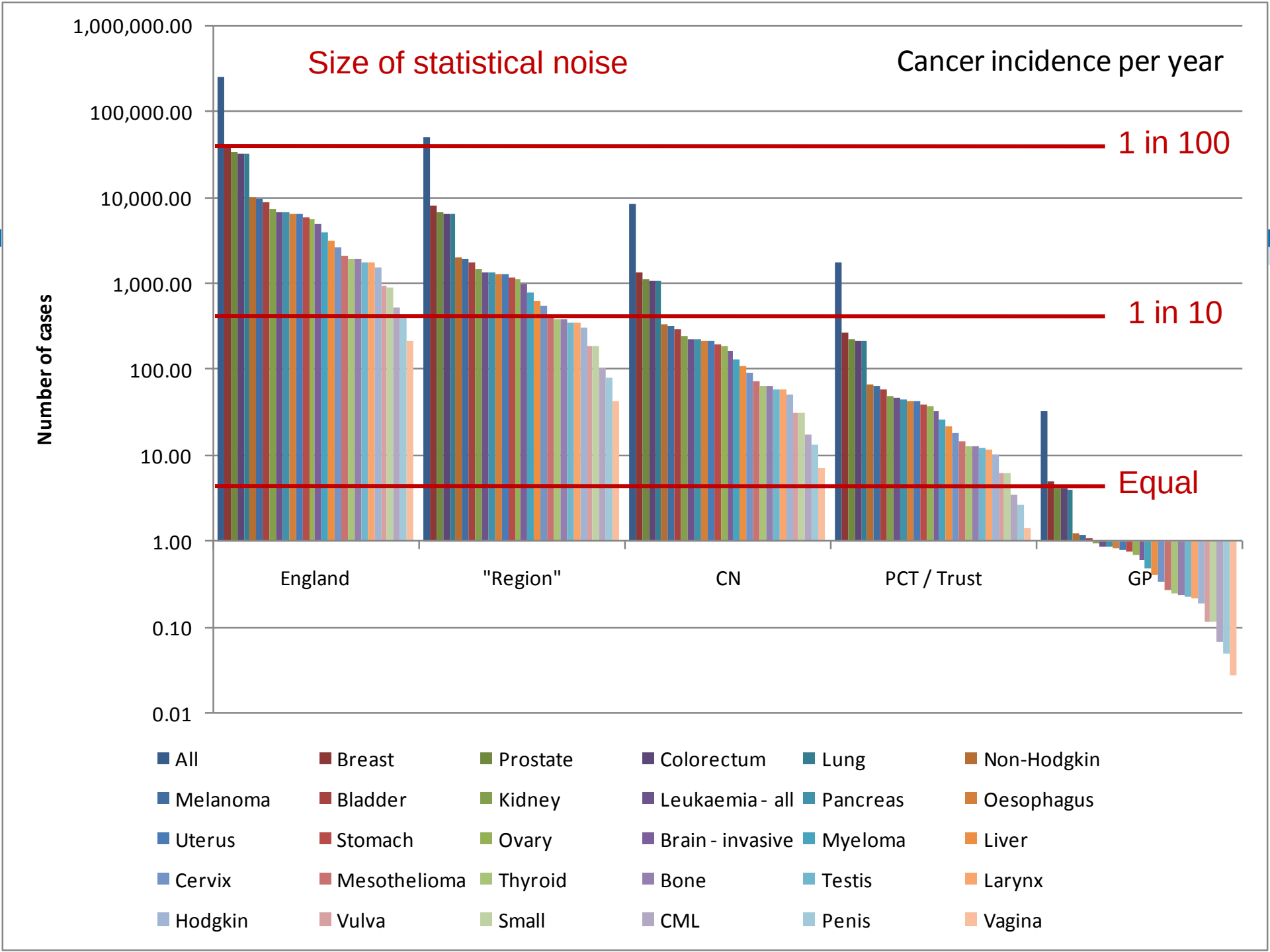
- Contextualising small numbers
- Signal to noise
- Aggregation
- Other Approaches
- Benefits of small numbers

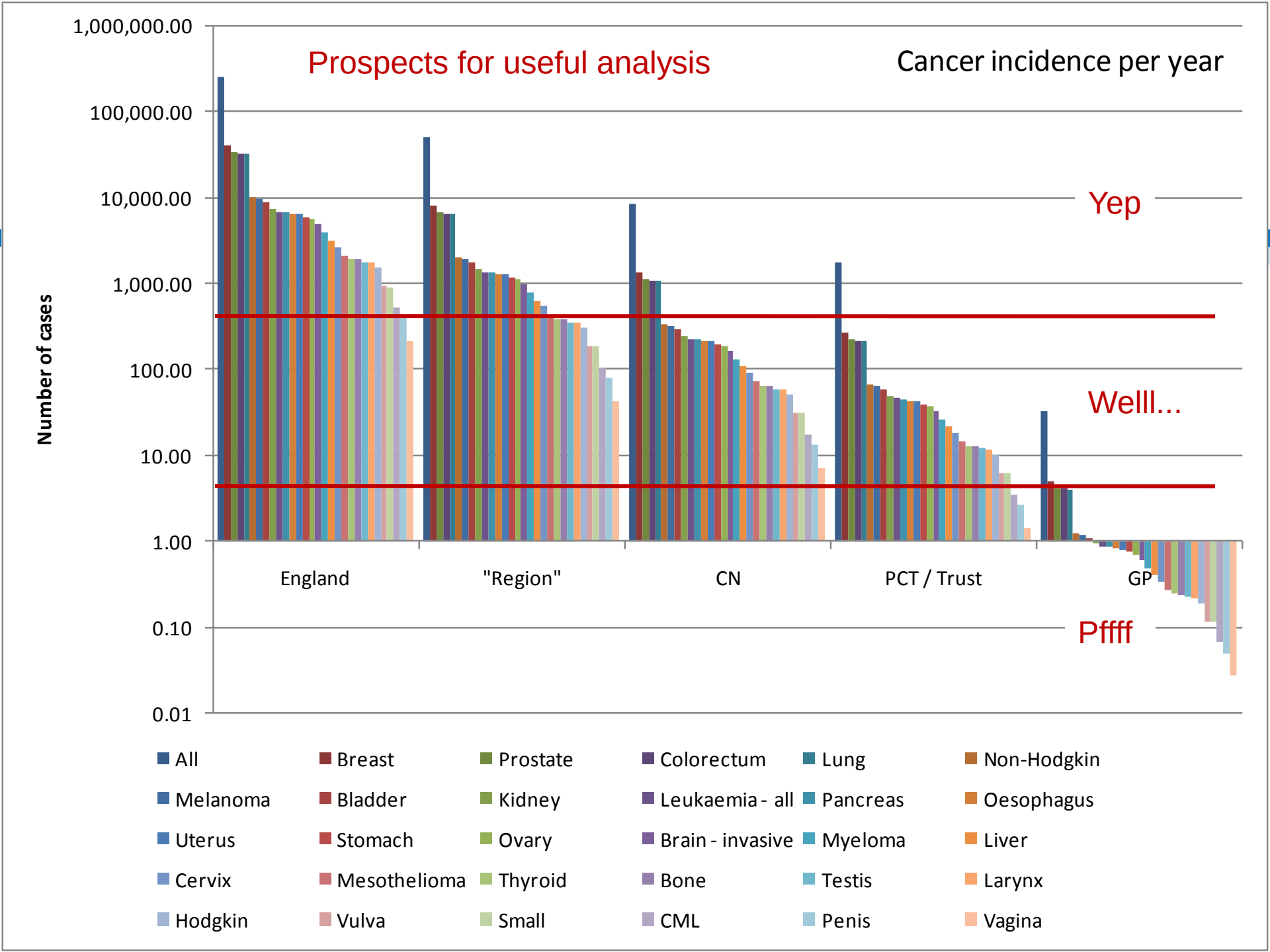

Cancer incidence per year
Number of cases
1,000,000.00
100,000.00
10,000.00
1,000.00
100.00
10.00
1.00
0.10
0.01
England
"Region"
CN
PCT / Trust
GP
All
Breast
Prostate
Colorectum
Lung
Non-Hodgkin
Melanoma
Bladder
Kidney
Leukaemia - all
Pancreas
Oesophagus
Uterus
Stomach
Ovary
Brain - invasive
Myeloma
Liver
Cervix
Mesothelioma
Thyroid
Bone
Testis
Larynx
Hodgkin
Vulva
Small
CML
Penis
Vagina


Cancer incidence per year
All cancers
Number of cases
1,000,000.00
100,000.00
10,000.00
1,000.00
100.00
10.00
1.00
0.10
0.01
England
"Region"
CN
PCT / Trust
GP
All
Breast
Prostate
Colorectum
Lung
Non-Hodgkin
Melanoma
Bladder
Kidney
Leukaemia - all
Pancreas
Oesophagus
Uterus
Stomach
Ovary
Brain - invasive
Myeloma
Liver
Cervix
Mesothelioma
Thyroid
Bone
Testis
Larynx
Hodgkin
Vulva
Small
CML
Penis
Vagina


Cancer incidence per year
"Big 4" cancers
Number of cases
1,000,000.00
100,000.00
10,000.00
1,000.00
100.00
10.00
1.00
0.10
0.01
England
"Region"
CN
PCT / Trust
GP
All
Breast
Prostate
Colorectum
Lung
Non-Hodgkin
Melanoma
Bladder
Kidney
Leukaemia - all
Pancreas
Oesophagus
Uterus
Stomach
Ovary
Brain - invasive
Myeloma
Liver
Cervix
Mesothelioma
Thyroid
Bone
Testis
Larynx
Hodgkin
Vulva
Small
CML
Penis
Vagina

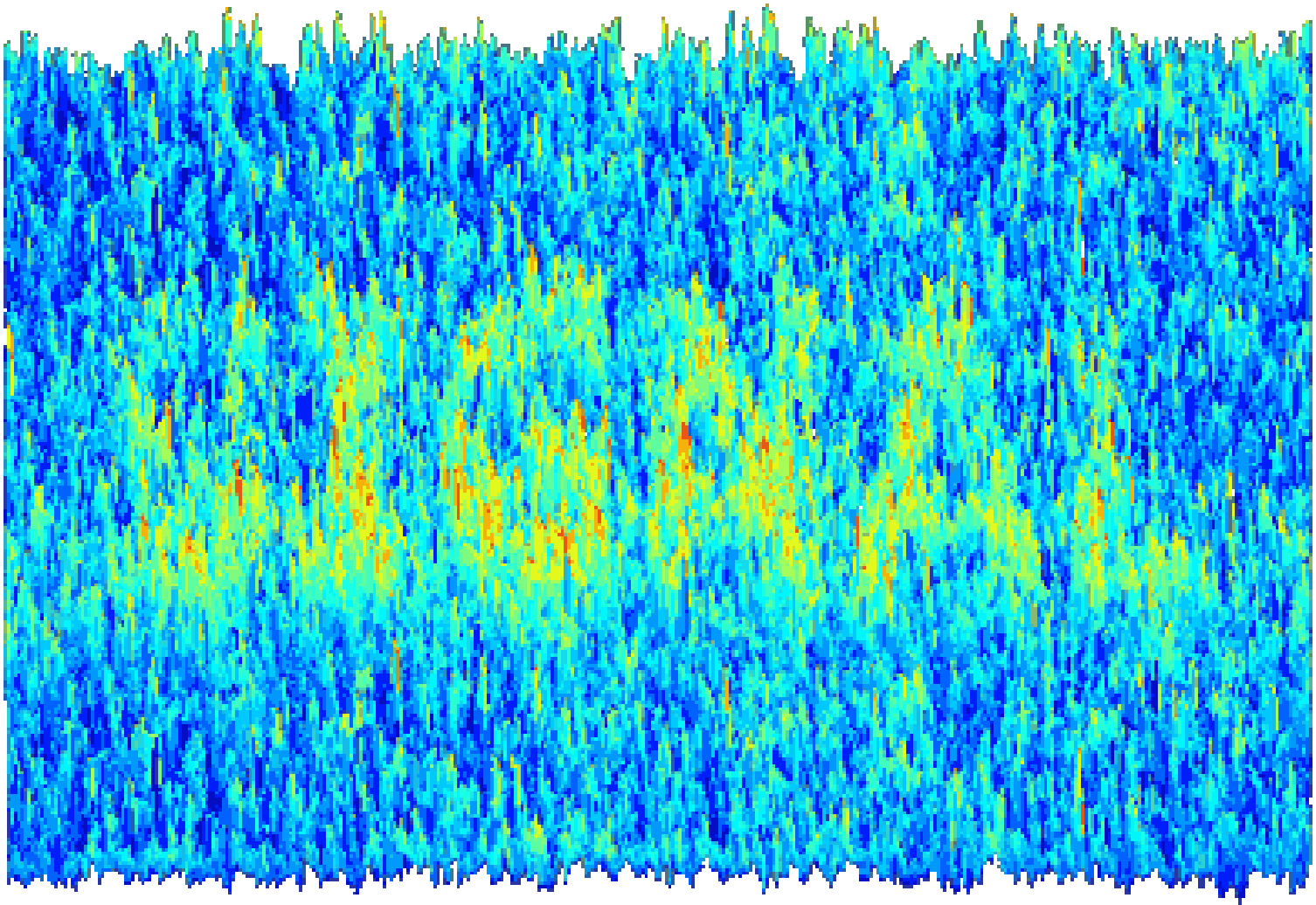

Cancer incidence per year
Rarer cancers
Number of cases
1,000,000.00
100,000.00
10,000.00
1,000.00
100.00
10.00
1.00
0.10
0.01
England
"Region"
CN
PCT / Trust
GP
All
Breast
Prostate
Colorectum
Lung
Non-Hodgkin
Melanoma
Bladder
Kidney
Leukaemia - all
Pancreas
Oesophagus
Uterus
Stomach
Ovary
Brain - invasive
Myeloma
Liver
Cervix
Mesothelioma
Thyroid
Bone
Testis
Larynx
Hodgkin
Vulva
Small
CML
Penis
Vagina


Size of statistical noise
Cancer incidence per year
1 in 100
1 in 10
Equal
Number of cases
1,000,000.00
100,000.00
10,000.00
1,000.00
100.00
10.00
1.00
0.10
0.01
England
"Region"
CN
PCT / Trust
GP
All
Breast
Prostate
Colorectum
Lung
Non-Hodgkin
Melanoma
Bladder
Kidney
Leukaemia - all
Pancreas
Oesophagus
Uterus
Stomach
Ovary
Brain - invasive
Myeloma
Liver
Cervix
Mesothelioma
Thyroid
Bone
Testis
Larynx
Hodgkin
Vulva
Small
CML
Penis
Vagina


Prospects for useful analysis
Cancer incidence per year
Number of cases
1,000,000.00
100,000.00
10,000.00
1,000.00
100.00
10.00
1.00
0.10
0.01
Yep
Welll...
Pffff
England
"Region"
CN
PCT / Trust
GP
All
Breast
Prostate
Colorectum
Lung
Non-Hodgkin
Melanoma
Bladder
Kidney
Leukaemia - all
Pancreas
Oesophagus
Uterus
Stomach
Ovary
Brain - invasive
Myeloma
Liver
Cervix
Mesothelioma
Thyroid
Bone
Testis
Larynx
Hodgkin
Vulva
Small
CML
Penis
Vagina

Signal to noise

NCIN
national cancer intelligence network
Using information to improve quality & choice



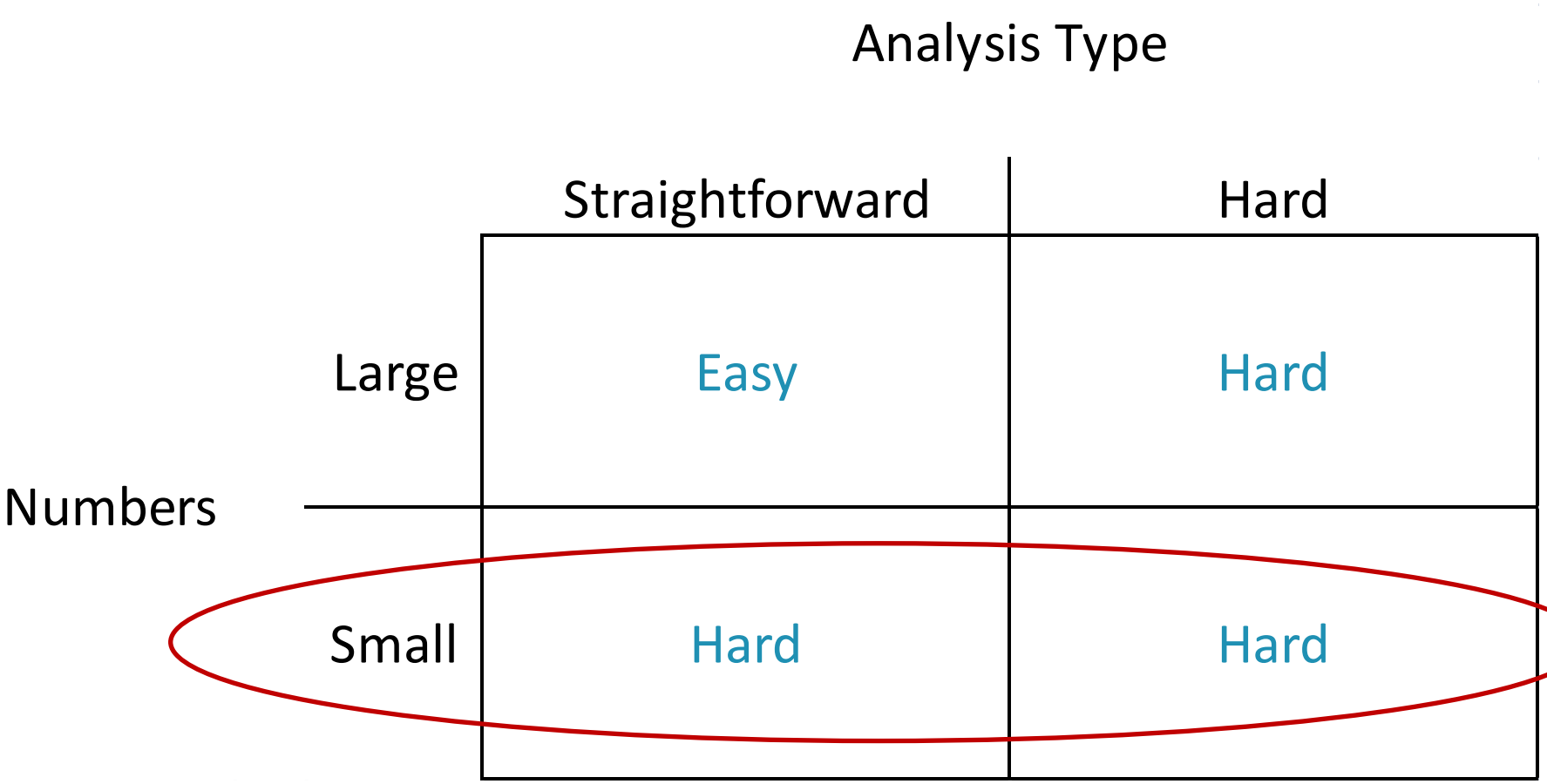
Prospects for useful analysis

NCIN national cancer intelligence network
Using information to improve quality & choice

Numbers \ Analysis Type	Straightforward	Hard
Large	Easy	Hard
Small	Hard	Hard

Prospects for useful analysis

NCIN national cancer intelligence network
Using information to improve quality & choice



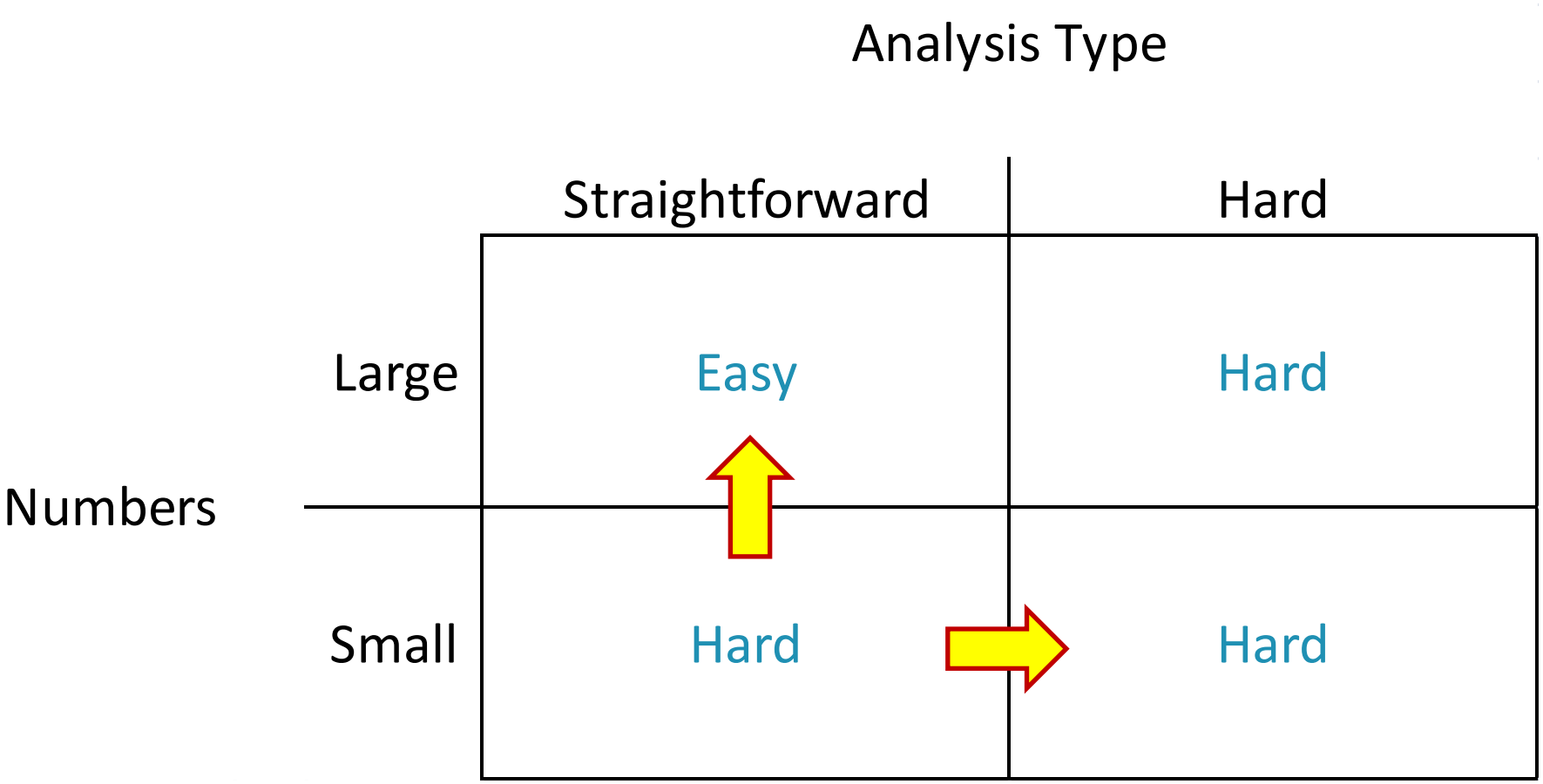
Signal to noise

NCIN national cancer intelligence network
Using information to improve quality & choice

Can’t reduce noise in cancer incidence, so...

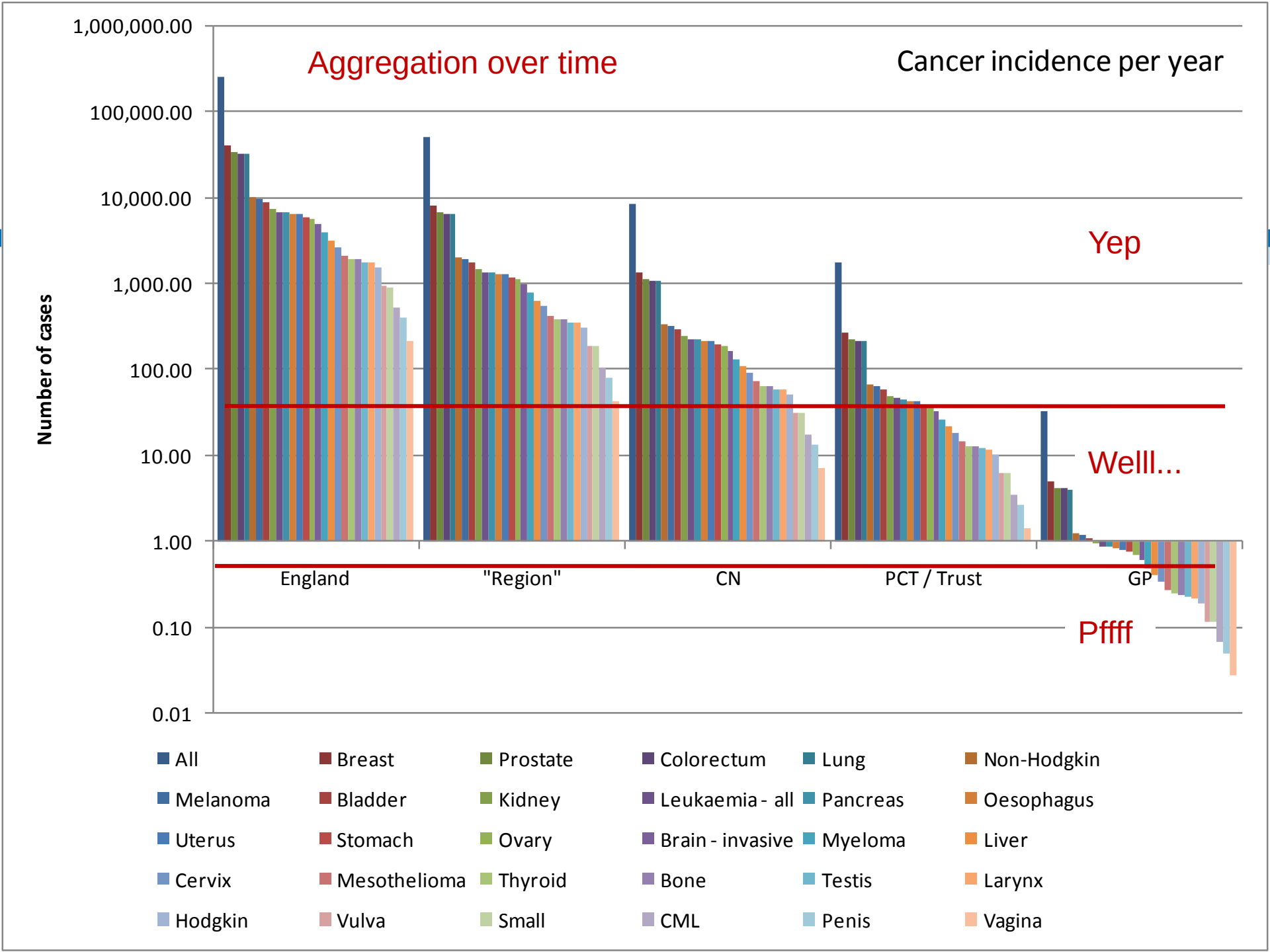
- Increase signal (aggregate)
- Go looking for a different signal

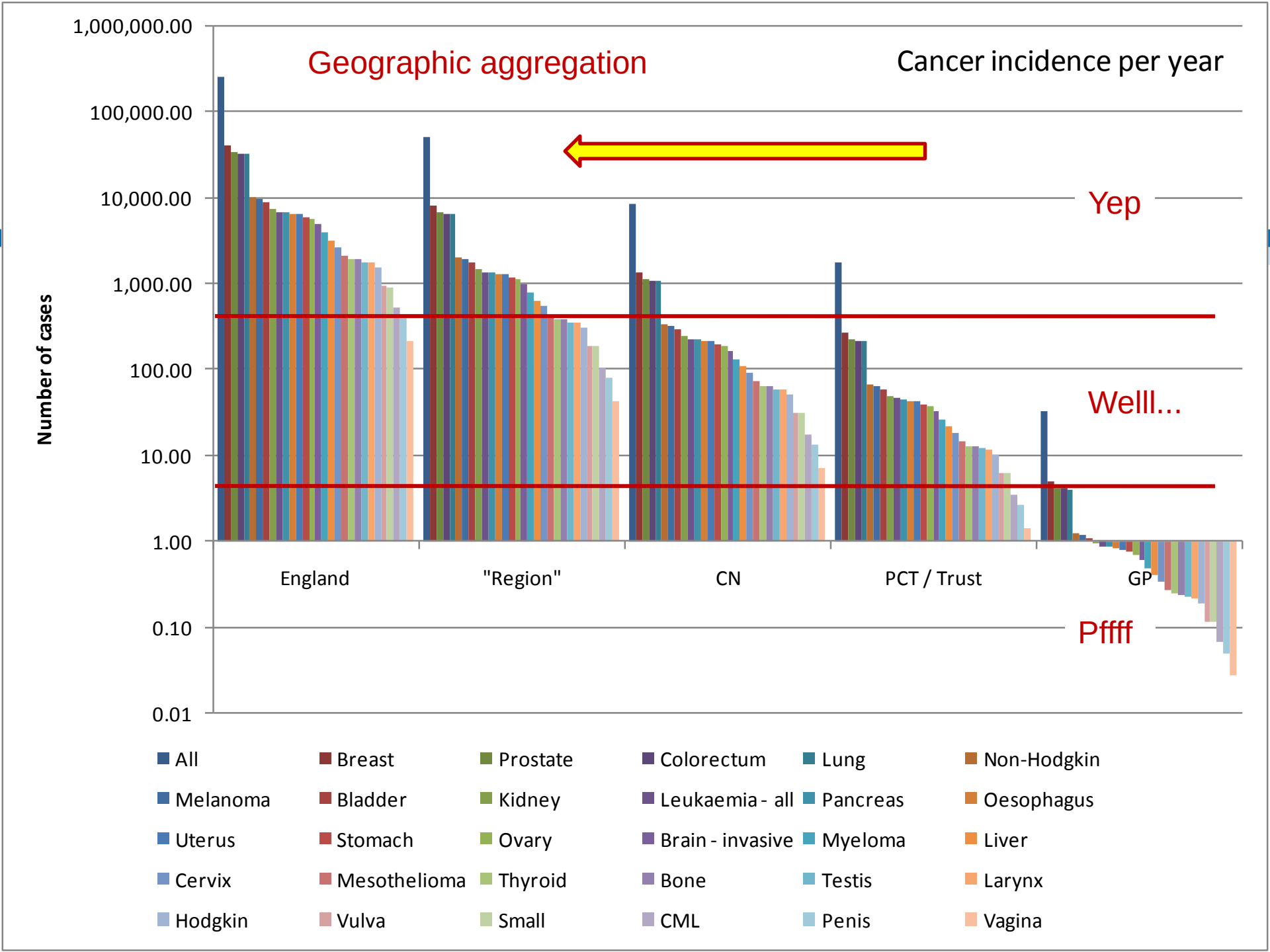
Prospects for useful analysis



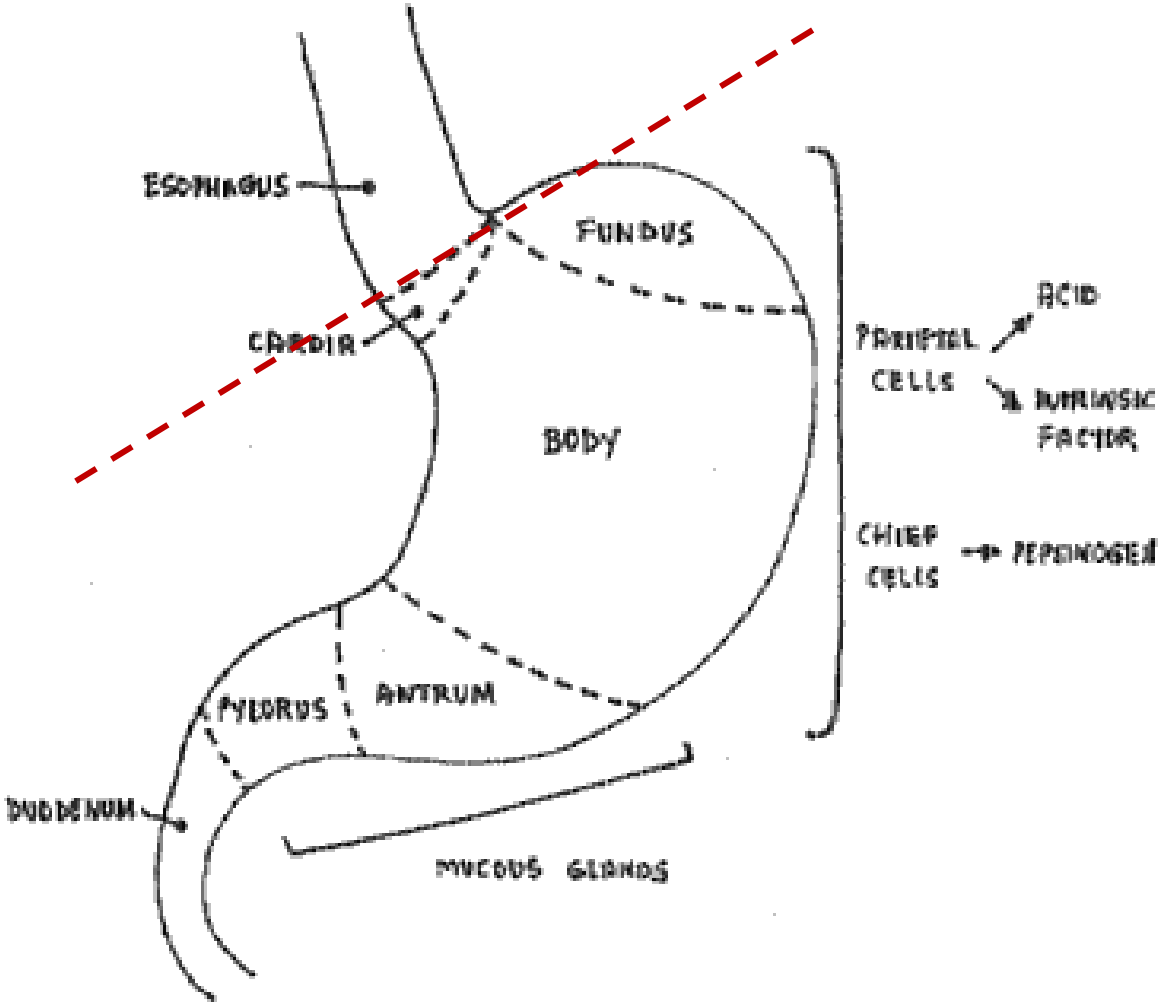
Aggregation

- Over Time
- Geographically
- By cancer type 
- Contextually

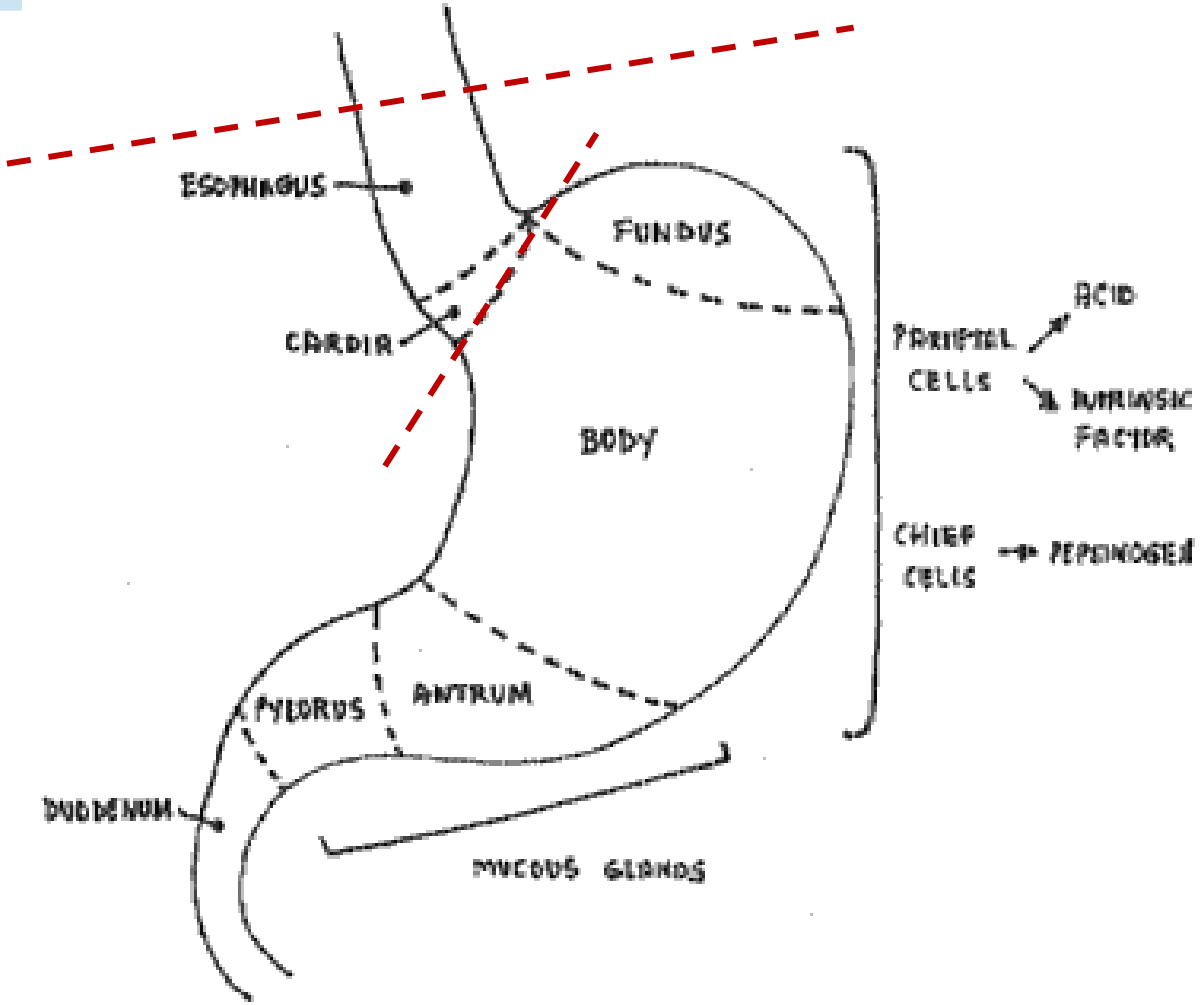

Aggregation over time
Cancer incidence per year
Number of cases
1,000,000.00
100,000.00
10,000.00
1,000.00
100.00
10.00
1.00
0.10
0.01
Yep
Welll...
Pffff
England
"Region"
CN
PCT / Trust
GP
All
Breast
Prostate
Colorectum
Lung
Non-Hodgkin
Melanoma
Bladder
Kidney
Leukaemia - all
Pancreas
Oesophagus
Uterus
Stomach
Ovary
Brain - invasive
Myeloma
Liver
Cervix
Mesothelioma
Thyroid
Bone
Testis
Larynx
Hodgkin
Vulva
Small
CML
Penis
Vagina


Geographic aggregation
Cancer incidence per year
Number of cases
1,000,000.00
100,000.00
10,000.00
1,000.00
100.00
10.00
1.00
0.10
0.01
Yep
Welll...
Pffff
England
"Region"
CN
PCT / Trust
GP
All
Breast
Prostate
Colorectum
Lung
Non-Hodgkin
Melanoma
Bladder
Kidney
Leukaemia - all
Pancreas
Oesophagus
Uterus
Stomach
Ovary
Brain - invasive
Myeloma
Liver
Cervix
Mesothelioma
Thyroid
Bone
Testis
Larynx
Hodgkin
Vulva
Small
CML
Penis
Vagina

Aggregation by cancer type



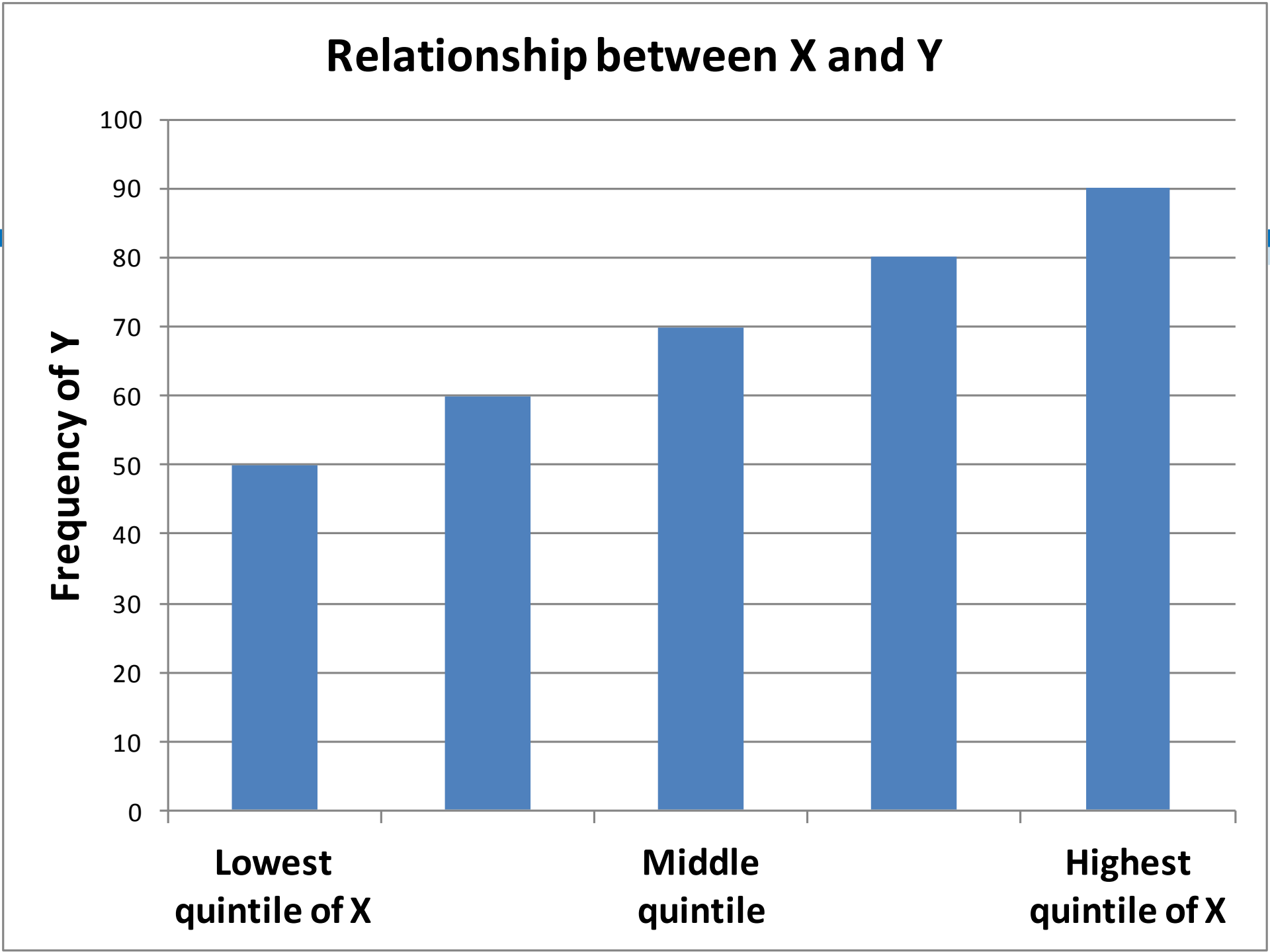
Aggregation by cancer type

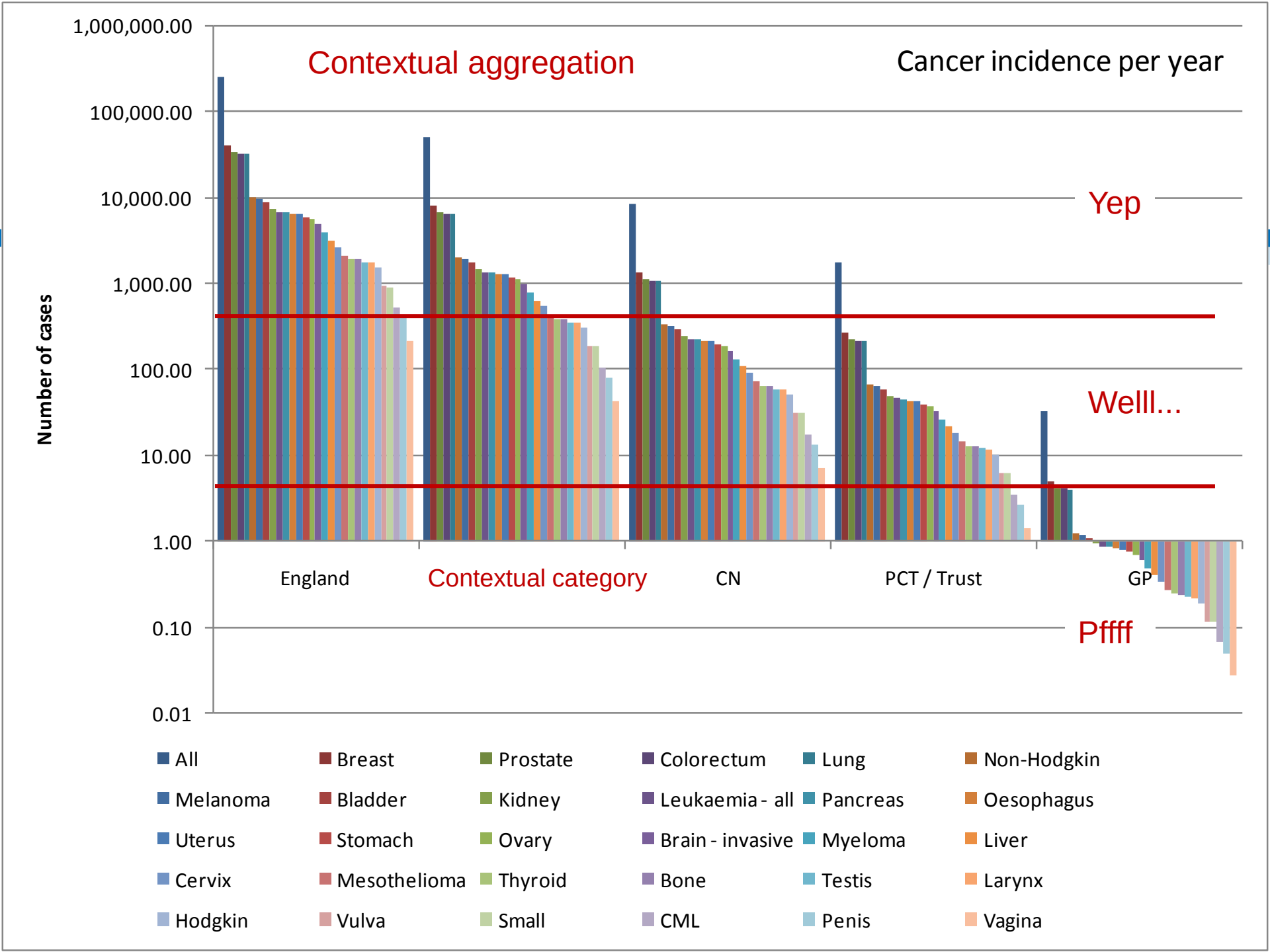


Contextual aggregation

- Use low level data sources
- Pick category of interest
- Aggregate by category of interest

- E.g., by deprivation quintile, by size of GP practice, by rurality quintile, anything definable by person or service


Relationship between X and Y
Frequency of Y
100
90
80
70
60
50
40
30
20
10
0
Lowest quintile of X
Middle quintile
Highest quintile of X


Contextual aggregation
Cancer incidence per year
Number of cases
1,000,000.00
100,000.00
10,000.00
1,000.00
100.00
10.00
1.00
0.10
0.01
Yep
Welll...
Pffff
England
Contextual category
CN
PCT / Trust
GP
All
Breast
Prostate
Colorectum
Lung
Non-Hodgkin
Melanoma
Bladder
Kidney
Leukaemia - all
Pancreas
Oesophagus
Uterus
Stomach
Ovary
Brain - invasive
Myeloma
Liver
Cervix
Mesothelioma
Thyroid
Bone
Testis
Larynx
Hodgkin
Vulva
Small
CML
Penis
Vagina

Other approaches: ‘Mild’ data vs ‘Wild’ data

NCIN national cancer intelligence network
Using information to improve quality & choice

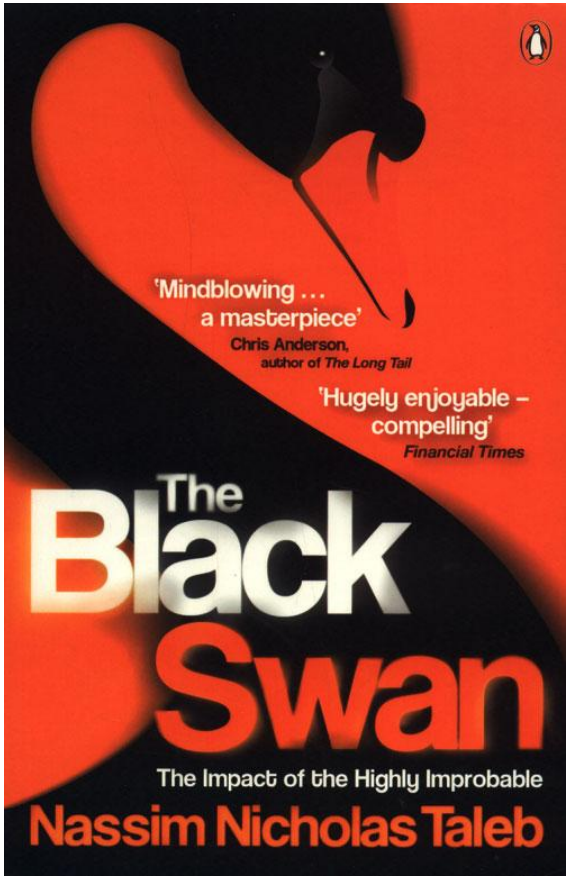
Mild

Independent

Physical basis

History a good guide to future

E.g.: height, weight, cancer incidence



Wild

Non-independent

‘Man-made’ basis

History a poor guide to future

E.g.: wealth, sales, finance, communicable disease

Geek slide

NCIN
national cancer
intelligence network
Using information to improve quality & choice

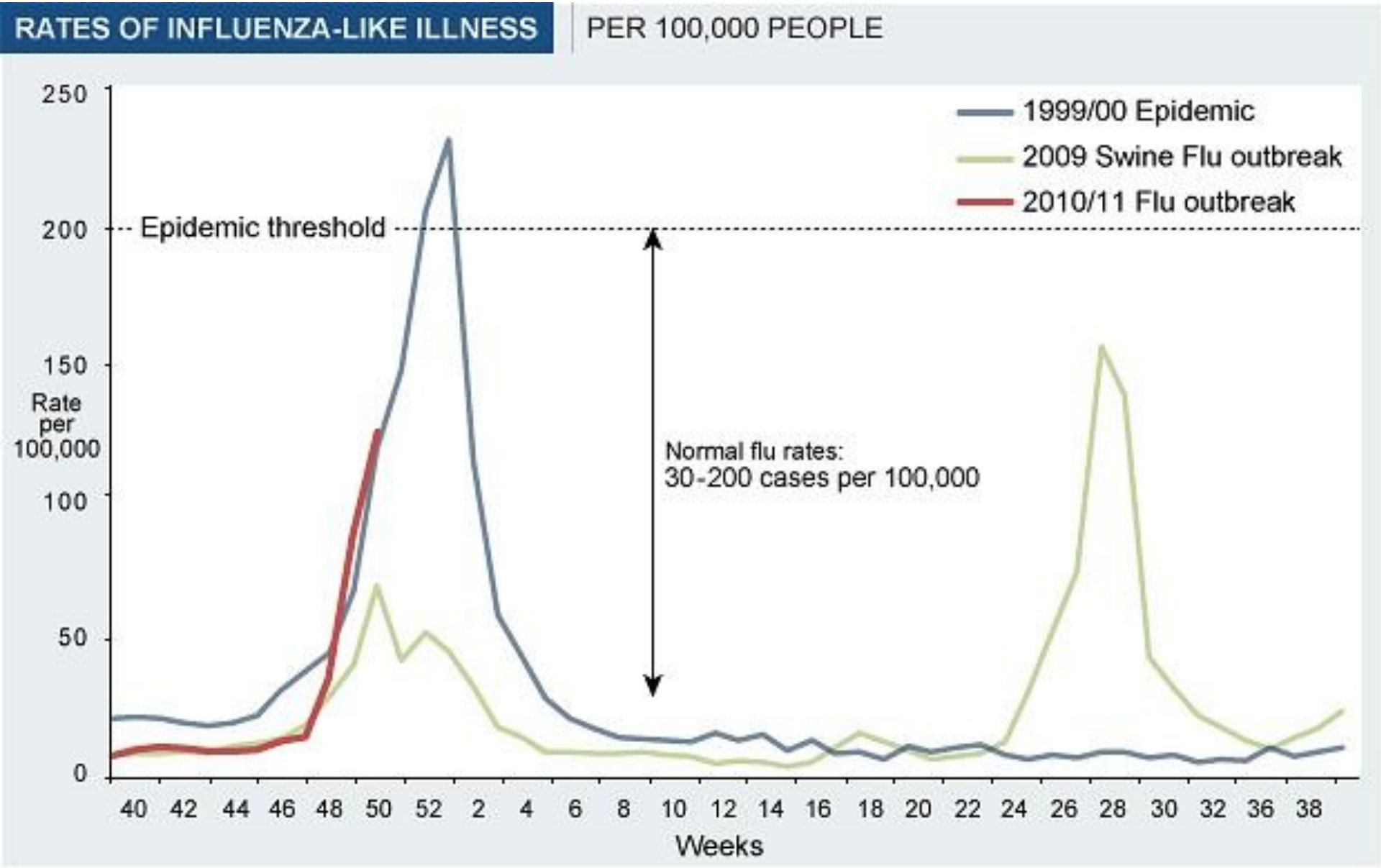
Geek slide

NCIN national cancer intelligence network
Using information to improve quality & choice

- Area of tail of Normal: $P>x \sim \exp(-x^2)/x$
- Area of tail of power law: $P>x \sim x^{-\alpha}$

- For $\alpha<2$ varience undefined
- Even for $\alpha>2$ varience converges very slowly

- So... expect more outliers for wild data


RATES OF INFLUENZA-LIKE ILLNESS
PER 100,000 PEOPLE
1999/00 Epidemic
2009 Swine Flu outbreak
2010/11 Flu outbreak
Epidemic threshold
Normal flu rates:
30-200 cases per 100,000
Rate per 100,000
250
200
150
100
50
0
40 42 44 46 48 50 52 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 36 38
Weeks

What about cancer?

NCIN national cancer intelligence network
Using information to improve quality & choice

Incidence (and mortality) – mild, so...

Increase signal:noise by looking at process measures and 'man-made' or non-independent measures.

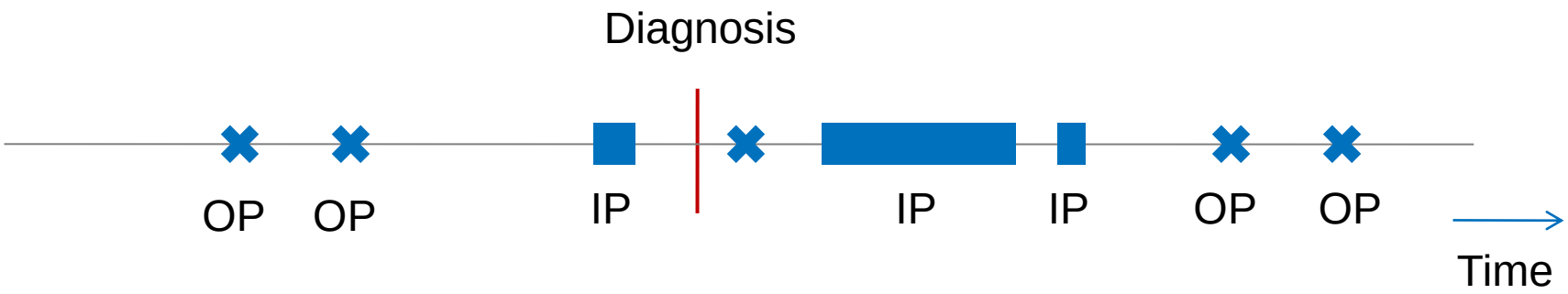
Key question – what are these? Probably cancer specific...

Referral practices? Waiting times? Routes to diagnosis?

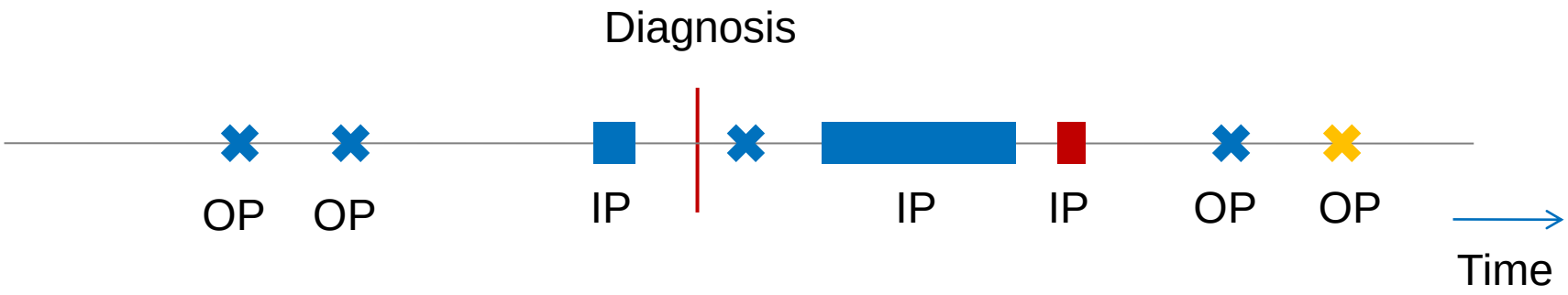
Some other approaches

- Cancer Patient Experience Survey
- Routes to Diagnosis
- Pathway based analysis

Visualising the pathway

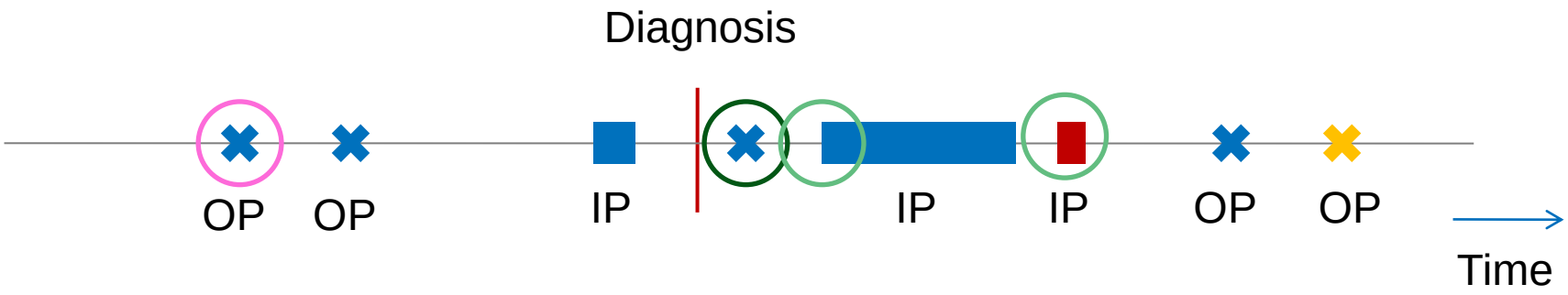


Visualising the pathway



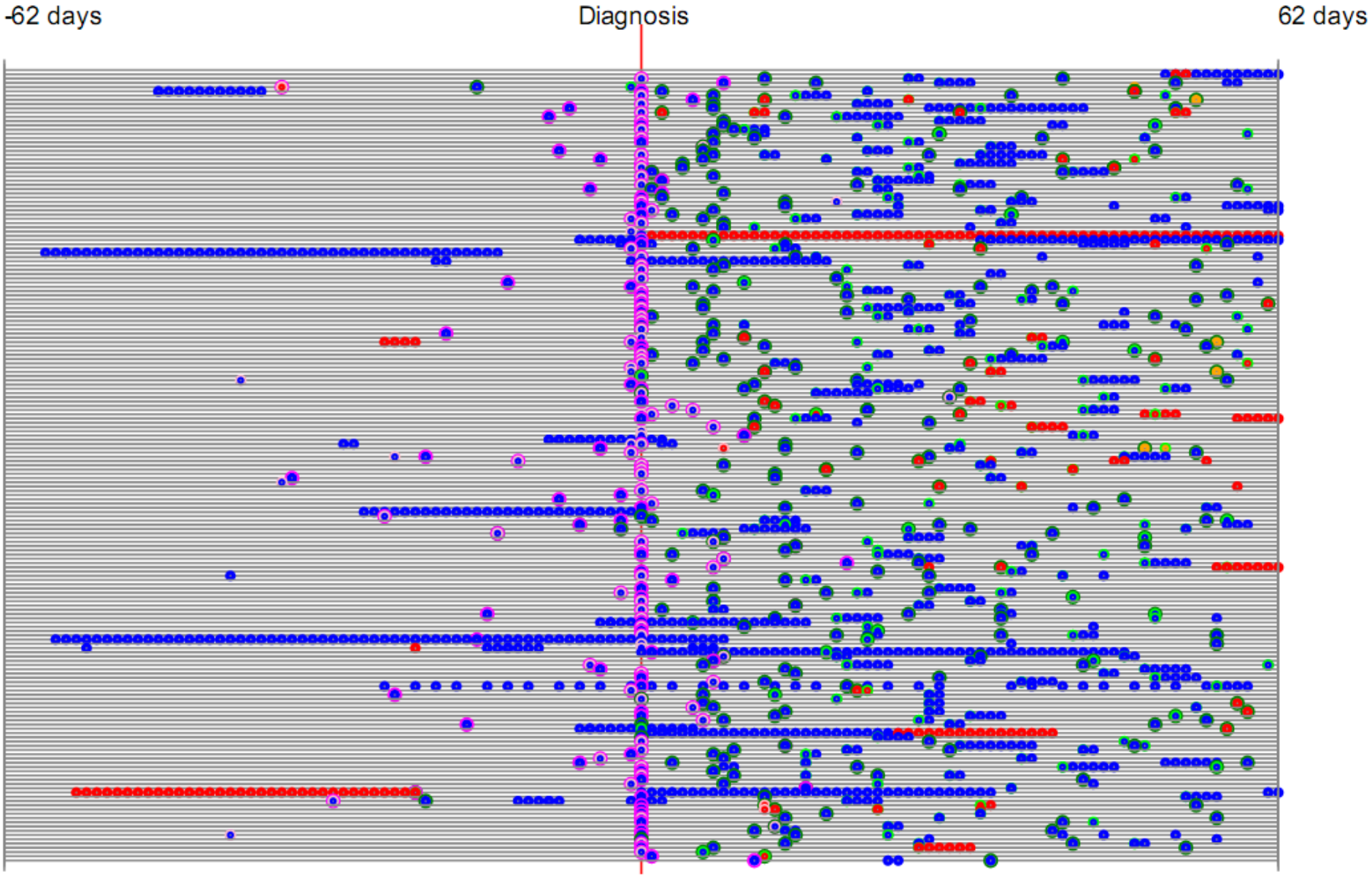
- Closest Trust to Diagnosis Date
- 2nd closest Trust to Diagnosis Date
- 3rd closest Trust to Diagnosis Date

Visualising the pathway



- Closest Trust to Diagnosis Date
- 2nd closest Trust to Diagnosis Date
- 3rd closest Trust to Diagnosis Date
- Place First Seen
- Treatment Period Start
- Treatment Start

Visualising the pathway (breast cancer)



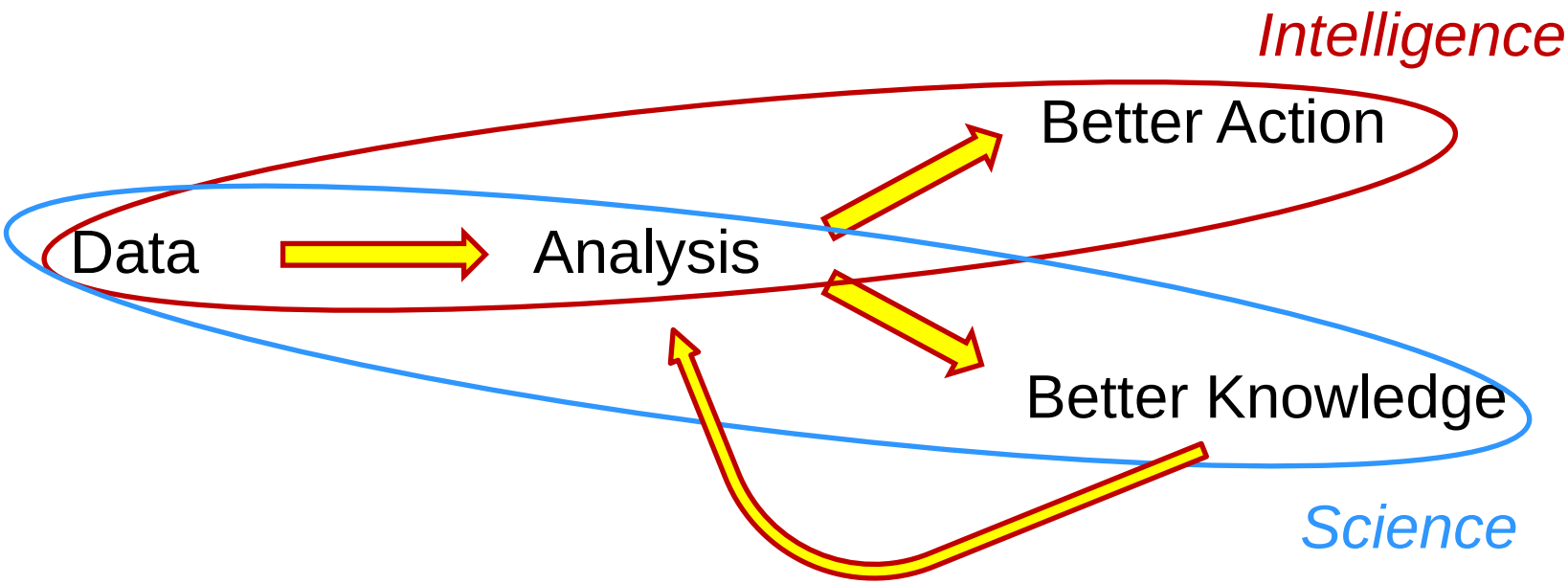
Benefits of small numbers

NCIN national cancer intelligence network
Using information to improve quality & choice

- More complete picture
- (possibility of) better data
- Individual case audit
- Qualitative approaches

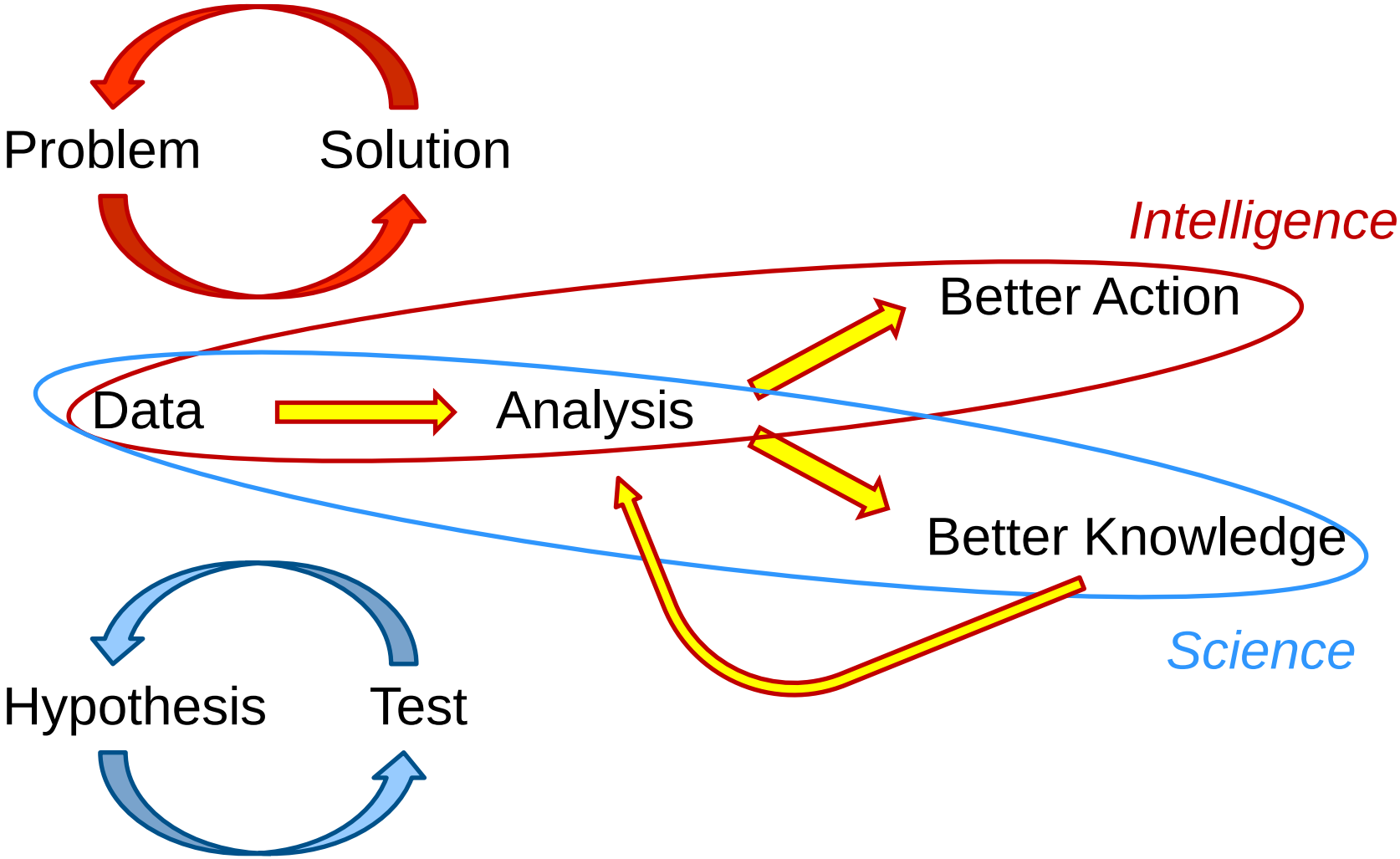
Big Picture - revisited

NCIN national cancer intelligence network
Using information to improve quality & choice



Big Picture - revisited

NCIN national cancer intelligence network
Using information to improve quality & choice



In conclusion

NCIN national cancer intelligence network
Using information to improve quality & choice

- Incidence/mortality has low signal:noise for rarer cancers
- Straightforward analyses are problematic

- Aggregate (especially contextually)
- Look for ‘less mild’ data
- Embrace the small numbers
- What are the questions to be asked?