Challenges in Medical Statistics

Rumana Omar
Professor of Medical Statistics
Department of Statistical Science
University College London
UK

Challenges:

To describe challenges I have divided Medical Statistics into categories:

- Skills
- Data
- Regulatory environment
- Responsibility
- Safeguarding medical research
- Methodological innovations
- Placement of a medical statistician

Medical Statistics plays an integral role in modern medicine

It is "the science of obtaining, analysing and interpreting data in order to understand and improve human health".

Contribution of statisticians in Medical Statistics

- Novel statistical methods are developed to meet complex data challenges in Biomedicine.

- Statistical Contribution in clinical investigations (e.g. if a treatment will work, risk prediction of a health outcome, identify risk factors for a disease, monitoring the spread of disease)

  Input of Medical Statistics in health research:
  - Designing studies,
  - Helping to decide what data to collect,
  - Developing/selecting methods for analysis,
  - Analysing data,
  - Helping to interpret the results of the analyses,
  - Collaborating in writing articles to describe the results of medical research.

Challenges:

Skills required to work in Medical Statistics

Medical Statisticians require mathematical ability combined with skills to understand disease process, conceptualise, and interpret. Not just be a mathematician or an analyst. There is more to it.

Developing new statistical methods cannot be done in a vacuum – need to have a connection with real issues that arise in real biomedical application areas - need to understand what the problems are.

Awareness of Medical terminology:
The better statisticians understand the medical processes and terminology, the better they communicate with clinicians/health researchers.

Multi-disciplinary team
Ability to work collaboratively within a diverse team

Know it all:
Different a range of study designs, e.g. CRM. Stepped Wedge, Multistage multi arm, therapist clustering

Different methods of analysis

Non-statistical issues

Ability to meet Deadlines
Research grant applications

Responding to reviewer's comments

Study deadlines

It can be very challenging when the deadlines are so tight that there is insufficient time for a particular problem to be thought through properly, or when the statistician is not included in the relevant discussions.

Biomedical Data:

Data may not be nicely shaped (bell-shaped) distributions. mixed distributions, bivariate skewed distributions.

Data transformation does not solve the problem of the shape of a distribution and may affect the default, well-established meaning of the quantity,

e.g. quality of life outcome, cost outcome

Health economic evaluations are now commonly included in pragmatic clinical trials that inform policy decisions

Despite the usual skewness in the distribution of costs, it is the arithmetic mean that is the most informative measure

Measures other than the arithmetic mean do not provide information about the cost of treating all patients, which is needed as the basis for healthcare policy decisions

Thompson and Barber, BMJ.

Presence of outliers, e.g. biochemistry results from studies on cancer patients in palliative care.

Alanine_transaminase range: 7 1694
C reactive protein range: 4 531

Removing outliers is not appropriate, they often bring important information about the process.

Important for example when developing a risk model that predicts survival in palliative care patients.
.

Missing observations

e.g. Missing covariates in observational studies. Follow-up not complete in trials

In ten top tips trial patients were randomised to either 10 Habit formation tips or 'usual care'. The primary outcome was weight loss at 3 months. 30% dropout

Missingness is not at random:
e.g.
People who did not lose weight dropped out??

Other examples:
Physical activity score of dementia patients, patients lost due to death

Mental and physical wellbeing score of cancer patients, patients lost to death

Sample size is small: recruitment problems in trials, recruitment problem of people delivering the intervention (e.g. surgeons, therapists, rare events, missing data)

Regulatory environment

Confirmatory analysis.

To be objective & to produce reproducible research analysis is planned a'priori & written down in a *Statistical Analysis Plan*.

When unexpected data issues arise: e.g. We have to switch to non-parametric methods (due to violated assumptions) but this is not mentioned in SAP. Any change requires amendments to SAP & Protocol with explanation & justification.

Write an all encompassing SAP!!

Statisticians analyse blind to the allocation of randomised groups for trials

SOPs for trials. Medical statisticians have to work to Standard Operating Procedures for trials. SOPs regulate almost everything we do.

- calculation of sample size,
- writing statistical programs,
- validating programs,
- storing data,
- writing report,
- organising and managing files

Trials of medicinal products: Every possible aspect of the data analysis is going to be regulated. e.

- Analytical software (including the process of updates and configuration.
- We have to set up a library of packages/codes/scripts, it should be tested and versions - frozen.
- Backup policy.
- The process of versioning documents and tracking changes.

Attend inspection meeting with regulators

e.g. MHRA in the UK

.
Responsibility.
Decisions on health care of patient and the public will be taken upon the results of the statistical analysis.

Sleepless nights!!

Safeguarding medical research

In 1994 statistician Professor Doug Altman published an editorial in the *BMJ* arguing that much medical research was of poor quality and misleading.

In his editorial entitled, "The Scandal of Poor Medical Research," Altman wrote that much research was "seriously flawed."

Altman's conclusion was: "We need less research, better research, and research done for the right reasons. Abandoning using the number of publications as a measure of ability would be a start."

Richard Smith (CBE, British Medical Doctor, ex BMJ Editor) in January, 2014 writes:

Sadly, the *BMJ* could publish this editorial almost unchanged again this week.

Small changes might be that ethics committees are now better equipped to detect scientific weakness and more journals employ statisticians.

These quality assurance methods don't, seem to be working as much of what is published continues to be misleading and of low quality.

In 2014 *The Lancet* published an important collection of articles on waste in medical research.

Arbitrary choice of analyses and an overemphasis on random extremes might affect the reported findings.

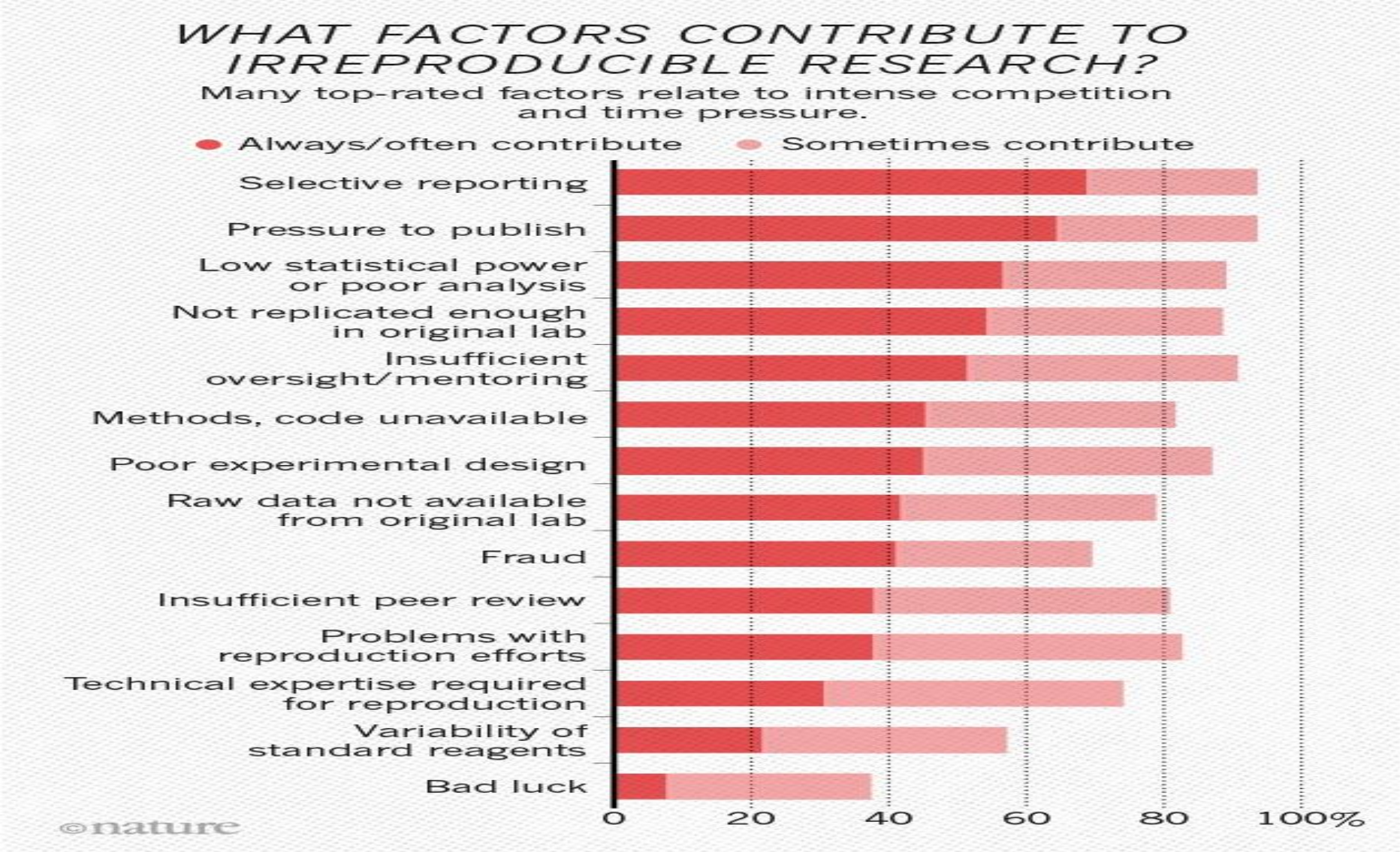
HOUSES OF PARLIAMENT – *INTEGRITY IN RESEARCH*

POST NOTE NO 544 JAN 2017

Pressure to publish increases fabrication because

- Pressure to reduce timescales so adopt less rigorous methods
- Perception that publishing in high impact journals requires positive results leads to falsification or fabrication of data
- Pressure to exaggerate the application of research

New House of Commons Science and Technology Committee - enquiry Jan 2017



Nature's survey of 1,576 researchers ..More than 60% of respondents said that each of two factors — pressure to publish and selective reporting.

Experience of the Biostatistics Group in UCL and UCLH studies: a few examples

1. Data errors neglected by Chief Investigator
Objective: To derive a risk prediction model for a health outcome in children, funded by ARC.
Data kept on changing. The CI did not inform us when they found there were errors in the gender variable which is one of the risk factors.

2.Inappropriate study design
Crossover trial to assess the therapeutic benefits of an intervention in in an Intensive Care Unit.

The trial protocol states: one of the likely reasons for drop out was “clinical deterioration needing re-intubation during 8 hours” suggesting a rapidly changing health condition.

Another statement in the protocol:
“informed consent should be given within 6 hours due to changing condition of patients”.

BSG expresses concerns as a Crossover trial design is appropriate for chronic and stable conditions

Their response

“I appreciate your concerns and explanations regarding the possible problems of a cross over design but my superiors wish to continue as planned”.

After several discussions following changes were made in the protocol:

“Although the patients are acutely ill, the clinical condition of the patients is thought unlikely to change over the eight hour period. Patients are expected to return to their baseline clinical condition immediately after …….is removed. “

No clinical justification given for these changes made in the protocol

.

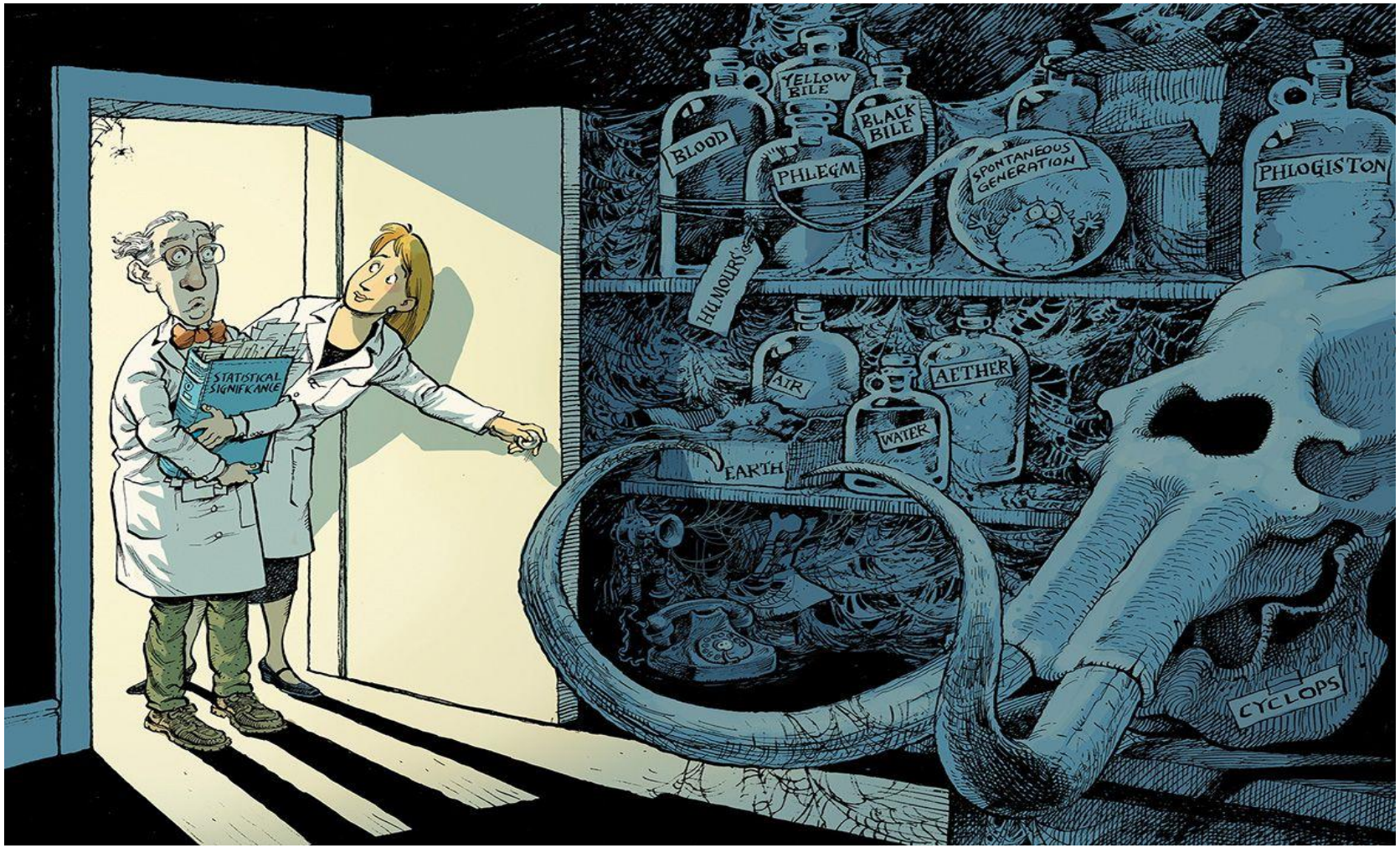
-
-

3. Over use of P-values in a study

	White	Asian	Black	Other	Total	P-value
N	2321 (81.2%)	273 (9.5%)	169 (5.9%)	96 (3.4%)	2859	
Male	65.2% (1513)	78% (213)	66.3% (112)	70.3% (68)	66.7% (1906)	0.011
Age	48.5 (35.4 - 59.8)	47.8 (36.7 - 58.3)	47.5 (39.7 - 58.2)	49.2 (33.6 – 59.7)	48.4 (35.7 - 59.3)	1.000
BMI (kg/m2)	26.8 (23.9 - 30.3)	26.2 (24.3 - 29.4)	28 (25.1 - 31.3)	26.4 (22.6 – 29.8)	26.9 (24 - 30.2)	0.067
Missing	4.6% (107)	4.4% (12)	4.7% (8)	4.2% (4)	4.6% (2728)	
Systolic BP (mmHg)	129 ± 22	129 ± 21	137 ± 23	127 ± 23	130 ± 22 a	0.026
Missing	17.1% (396)	16.5% (45)	12.4% (21)	18.8% (18)	16.8% (480)	
Diastolic BP (mmHg)	77 ± 11	77 ± 11	80 ± 11	76 ± 13	77 ± 11	1.000
Missing	24.5% (569)	26.7% (73)	17.2% (29)	27.1% (26)	24.4% (697)	
Hypertension	25.7% (597)	35.9% (98)	56.2% (95)	32.3% (31)	28.7% (821)	<0.001
Missing	3.8% (88)	3.3% (9)	0.59 (1)	2.1% (2)	3.5% (100)	
Atrial Fibrillation	15.6% (362)	7.3% (20)	9.5% (16)	9.4% (9)	14.2% (407)	<0.001
Missing	10.7% (248)	2.9% (8)	3.6% (6)	8.3% (8)	9.4% (270)	
Diabetes Mellitus	4.4% (102)	13.2% (36)	8.9% (15)	3.1% (3)	5.5% (156)	<0.001
Missing	17.6% (409)	7.7% (21)	1.8% (3)	12.5% (12)	15.6% (445)	
Stroke	2.9% (68)	4.0% (11)	6.5% (11)	4.2% (4)	3.3% (94)	<0.001
Missing	11.3% (262)	2.9% (8)	3.6% (6)	8.3% (8)	9.9% (284)	
Coronary disease	4.1% (95)	11.4% (31)	4.7% (8)	3.1% (3)	4.8% (137)	<0.001
Missing	10.9% (252)	3.7% (10)	3.0 (5)	7.3% (7)	9.6% (274)	
Cardiac arrest	1.3% (31)	1.1% (3)	1.2% (2)	1.0% (1)	1.3% (37)	1.000
Missing	0.2% (5)	0	0	0	0.2% (5)	
Age at Diagnosis	44 (30 - 56)	45 (34 - 55)	46 (36 - 56)	44 (23 - 57)	44 (31 - 56)	1.000
Missing	2.2% (51)	0.7% (2)	0	1.0% (1)	1.9% (54)	
Presentation route:						0.148
Family Screening	16.2% (376)	11.4% (31)	5.9% (10)	15.6% (15)	15.1% (432)	
Symptoms	53.3% (1236)	61.9% (169)	66.9% (113)	47.9% (46)	55.7% (1564)	
Incidental	28.7% (666)	25.6% (70)	26% (44)	35.4% (34)	28.5% (814)	
Missing	1.9% (43)	1.1% (3)	1.2% (2)	1.0% (1)	1.7% (49)	
Family history of HCM	39.2% (909)	28.6% (78)	20.1% (34)	35.4% (34)	36.9% (1055)	<0.001
Missing	0.4% (9)	0.4% (1)	0	0	0.4% (10)	
Family history of SCD	22.7% (526)	23.8% (65)	18.3% (31)	13.5% (13)	22.2% (635)	1.000
Missing	0.4% (10)	0.4% (1)	0	1.0% (1)	0.4% (12)	
Asymptomatic	20.6% (478)	13.2% (36)	12.4% (21)	21.9% (21)	19.5% (556)	0.122
Chest pain	47.1% (1093)	61.5% (168)	56.8% (96)	52.1% (50)	49.2% (1407)	0.009
Missing	0.2% (5)	0	0	0	0.2% (5)	
Shortness of breath	55.9% (1297)	59% (161)	62.7% (106)	62.5% (60)	56.8% (1624)	1.00
Missing	0.1% (3)	0	0	0	0.1% (3)	
Palpitation	40.5% (939)	28.9% (79)	42% (71)	31.3% (30)	39.1% (1119)	0.259
Missing	0.2% (4)	0	0	0	0.1% (4)	
Syncope	17% (394)	15% (41)	18.3% (31)	15.6% (15)	16.8% (481)	1.000
Beta Blocker	41.3% (959)	48.4% (132)	49.7% (84)	47.9% (46)	42.7% (1221)	0.74
ACE/ARB	13.4% (311)	19.1% (52)	26.6% (45)	15.6% (15)	14.8% (423)	<0.001
CCB	12.5% (291)	11% (30)	10.1% (17)	9.4% (9)	12.1% (347)	1.000
Thiazide	5.4% (126)	5.9% (16)	20.7% (35)	8.3% (8)	6.5% (185)	<0.001

	White	Asian	Black	Other	Total
Electrocardiogram					
Heart Rate	64 (57 - 73)	64 (60 - 74)	66.5 (58 - 73)	66 (57 - 74)	64 (57 - 73)
Missing	65.9% (1530)	58.2% (159)	46.7% (79)	63.5% (61)	64% (1829)
PR interval (msec)	168 (150 - 188)	161 (149 - 180)	166 (152 - 184)	164 (153 - 189)	167 (151 - 186)
Missing	58% (1346)	49.8% (136)	42% (71)	55.2% (53)	56.2% (1606)
QRS Duration (msec)	96 (88 - 107)	93.5 (85 - 102)	90 (86 - 100)	96 (89 - 109)	96 (88 - 106)
Missing	55.5% (1289)	48.7% (133)	42% (71)	53.1% (51)	54% (1544)
Bundle branch block					
LBBB	1.9% (45)	2.6% (7)	0.6% (1)	1% (1)	1.9% (54)
RBBB	2% (47)	1.8% (5)	0.6% (1)	3.1% (3)	2% (56)
Missing	56.5% (1311)	50.2% (137)	43.2% (73)	57.3% (55)	55.1% (1576)
Axis					
Right Axis	2.8% (65)	1.5% (4)	0.6% (1)	1% (1)	2.5% (71)
Left Axis	7.2% (166)	9.2% (25)	3.6% (6)	3.1% (3)	7% (200)
Extreme Axis	0.7% (17)	0.7% (2)	0.6% (1)	0	0.7% (20)
Missing	55.5% (1289)	49.1% (134)	42.6% (72)	52.1% (50)	54% (1545)
LVH	22.7% (526)	27.5% (75)	30.8% (52)	26% (25)	23.7% (678)
Missing	63.4% (1472)	57.1% (156)	56.8% (96)	64.6% (62)	62.5% (1786)
ST Depression					
None	32.7% (760)	26.7% (73)	27.2% (46)	31.3% (30)	31.8% (909)
Mild	10.1% (234)	15.8% (43)	16% (27)	8.3% (8)	10.9% (312)
Severe	2.3% (53)	9.2% (25)	15.4% (26)	8.3% (8)	3.9% (112)
Missing	54.9% (1274)	48.4% (132)	41.4% (70)	52.1% (50)	53.4% (1526)
T inversion					
None	12.5% (291)	7.7% (21)	6.5% (11)	14.6% (14)	11.8% (337)
Mild	24.5% (568)	24.5% (67)	20.1% (34)	10.4% (10)	23.8% (679)
Mod	6.6% (152)	11.7% (32)	17.8% (30)	16.7% (16)	8% (230)
Severe	1.6% (36)	7.7% (21)	14.2% (24)	6.3% (6)	3% (87)
Missing	54.9% (1274)	48.4% (132)	41.4% (70)	52.1% (50)	53.4% (1526)

Scientists rise up against statistical significance: Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.



Amrhein & Greenland

Let's be clear about what must stop:

We should never conclude there is 'no difference' or 'no association' just because a *P* value is larger than a threshold such as 0.05 or, equivalently, because a confidence interval includes zero.

Surveys of hundreds of articles have found that statistically non-significant results are interpreted as indicating 'no difference' or 'no effect' in around half.

We're frankly sick of seeing such nonsensical 'proofs of the null' and claims of non-association in presentations, research articles, reviews and instructional materials.

An interval that contains the null value will often also contain non-null values of high practical importance.

Not all values inside are equally compatible with the data, given the assumptions. The point estimate is the most compatible, and values near it are more compatible than those near the limits. This is why we urge authors to discuss the point estimate, even when they have a large *P* value or a wide interval, as well as discussing the limits of that interval..

Example from one of our studies

We said:
“Asian and black patients were more likely to develop heart failure (Asian HR 2.14, 95% CI: 1.27-3.62; black HR 2.01; 95% CI: 0.95-4.28; global p=0.013)"

Reviewers comments
1. However the HR for black patients is in fact not significant as the CI includes 1 so don’t state that Black patients were more likely to develop heart failure.

Was the authors’ (our) statement incorrect?

How do we strike a balance when we have statistically illiterate reviewers making statistical comments with editors not knowing any better?? Clinical investigator wishes to satisfy the editor despite their incompetency.

Amrhein & Greenland
We must learn to embrace uncertainty.

One practical way to do so is to rename confidence intervals as ‘compatibility intervals’ and interpret them in a way that avoids overconfidence.

Specifically, we recommend that authors describe the practical implications of all values inside the interval, especially the observed effect (or point estimate) and the limits. In doing so, they should remember that all the values between the interval’s limits are reasonably compatible with the data, given the statistical assumptions used to compute the interval. Therefore, singling out one particular value (such as the null value) in the interval as ‘shown’ makes no sense.

The objection we hear most against retiring statistical significance is that it is needed to make yes-or-no decisions.

e.g. Our statement:
"We do not present any threshold for the P-values. All the results are presented in a transparent manner in the tables for the reader to make their own judgement."

Reviewers' comments:
Although the scale of the study is a strength and this topic is of great interest, the presentation of certain results could be misconstrued by readers accustomed to expecting only statistically significant results to be highlighted.

Amrhein & Greenland
But for the choices often required in regulatory, policy and business environments, decisions based on the costs, benefits and likelihoods of all potential consequences always beat those made based solely on statistical significance.

- What will retiring statistical significance look like?

- We hope that methods sections and data tabulation will be more detailed and nuanced. Authors will emphasize their estimates and the uncertainty in them — for example, by explicitly discussing the lower and upper limits of their intervals.

- They will not rely on significance tests. When *P* values are reported, they will be given with sensible precision (for example, $P = 0.021$ or $P = 0.13$) — without adornments such as stars or letters to denote statistical significance and not as binary inequalities ($P < 0.05$ or $P > 0.05$).

- Decisions to interpret or to publish results will not be based on statistical thresholds.

- People will spend less time with statistical software, and more time thinking.

Methodological innovations in Medical Statistics
As new and more complex biomedical problems emerge, biostatistics faces a challenge in terms of both the novel application of existing methods and the development of new superior method.

Medical Statistics, Data science, Artificial Intelligence, Big data

New thinking: AI, machine learning techniques are the answer of everything

Predicting scheduled hospital attendance with artificial intelligence
npj Digital Medicine, **volume 2**, Article number: 26 (2019)

Hospital develops AI to identify patients likely to skip appointments

London's UCLH creates tool predicting 90% of no-shows – potentially saving NHS millions
The Guardian

The wide multiplicity of potential causes, and the poor performance of systems based on simple, linear, low-dimensional models, suggests complex predictive models of attendance are needed. Here, we quantify the effect of using complex, non-linear, high-dimensional models enabled by machine learning.

Big data - solution to everything

Problems with big data

- Big data usually come from routine databases
- Not specifically designed to answer the research question
- Data accuracy
- Bias
- Missing data
- Drawing causal inference

My own methodological research topics:

- Sample size for development of risk prediction models
- Sample size for validation of risk prediction models
- Handling missing values in time varying covariates measured at irregularly spaced time intervals in risk prediction
- Cost effectiveness of the HCM risk model
- Handling multiple primary outcomes in trials
- Methods for detecting outlying hospitals in terms of their performance measured by mortality

Placement of a medical statistician

Several models exist in academic and other research institutions.

a) statistician holds an appointment entirely within a non-medical department, e.g. in a department of statistics, with strong links with the medical research group.

Advantages: helps the statistician to maintain credibility and encourages them to keep up with developments in mainstream statistics. Statisticians remains in close contact with peers.

Disadvantage: time for sufficient commitment to medical research may be difficult.

b) Appointment entirely within a medical setting.
Advantage: total commitment to work on medical problems and it may enhance that continuity of statistical input to the research programme.

Disadvantage: it can limit career prospects. promotion may be limited because of lack of 1st authorship publications, opportunities to act as a PI on grant applications.

Other models: cross- or joint-appointments between inter-disciplinary departments such as medicine and statistics: can be extremely challenging, satisfy both mathematical & medical sciences

or the establishment of a biostatistics department within a medical school.

A mechanism is needed that

recognises the value of collaboration and synthesis across conventional discipline boundaries

ensure that medical statisticians do not fall down the gaps between the walls that departments build around themselves.

Severe skill shortage in the UK

Addressing the challenges: Institutions, Researchers, Journals and funders all have a part to play.