Hierarchies of Evidence

Michael Rawlins
Honorary Professor, London School of Hygiene and Tropical Medicine
President, Royal Society of Medicine, London
Experimentation-Deduction

Robert Hooke
1635-1703

Robert Boyle
1627-1691
Deductive Falsification

Sir Karl Popper
1902-1994
Observation-Induction

Francis Bacon
1561-1626

Rene Descartes
1596-1650

Thomas Hobbes
1588-1679
The Problem of Induction

“Probability is founded on the presumption of resemblance, betwixt those objects of which we have experience, and those of which we have not; and therefore ‘tis impossible this assumption can arise from probability”.

David Hume
1711-1778
Juvenile Huntington’s disease
Age of onset of Huntington’s disease

[Graph showing the relationship between CAG repeat length and age of onset]

- CAG repeat length
- Age of onset
“Philosophers of science are as useful to science as are ornithologists to birds”
## Hierarchies of Evidence

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1++</td>
<td>High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias.</td>
</tr>
<tr>
<td>1+</td>
<td>Well-conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias.</td>
</tr>
<tr>
<td>1-</td>
<td>Meta-analyses, systematic reviews of randomized controlled trials or RCTs with a high risk of bias.</td>
</tr>
<tr>
<td>2++</td>
<td>High quality systematic reviews of case-control or cohort studies with a low risk of confounding, bias or chance and a high probability of causality.</td>
</tr>
<tr>
<td>2+</td>
<td>Well-conducted case-control or cohort studies with a low risk of confounding, bias or chance and a significant chance that the relationship is not causal.</td>
</tr>
<tr>
<td>2-</td>
<td>Case-control or cohort studies with a high risk of confounding, bias or chance and a significant risk that the relationship is not causal.</td>
</tr>
<tr>
<td>3</td>
<td>Non-analytical studies (for example case records, case series)</td>
</tr>
<tr>
<td>4</td>
<td>Expert opinion, formal consensus.</td>
</tr>
</tbody>
</table>

Harbour R, Miller J (2001)
Observational evidence for determining drug safety
Is no substitute for evidence from randomised controlled trials

“Only properly randomised trials can provide truly reliable evidence on adverse events, just as these are the only convincing data on drug efficacy”.

“Observational studies may provide some limited reassurance that a drug is safe, or they may provide an early indication of a problem, but by design they cannot provide reliable evidence on questions of drug safety”.

Freemantle N, Irs A. BMJ 2008;336:627
### Safety of thiazolidinediones

<table>
<thead>
<tr>
<th>Event</th>
<th>Rosiglitazone (OR and 95% CIs)</th>
<th>Pioglitazone (HR and 95% CIs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>1.18 (0.89 to 1.55)</td>
<td>0.92 (0.76 to 1.11)</td>
</tr>
<tr>
<td>- all causes</td>
<td>1.64 (0.98 to 2.74)</td>
<td>-</td>
</tr>
<tr>
<td>- cardiovascular causes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>1.43 (1.03 to 1.98)</td>
<td>0.81 (0.64 to 1.02)</td>
</tr>
<tr>
<td>Stroke</td>
<td>-</td>
<td>0.80 (0.80 to 1.04)</td>
</tr>
<tr>
<td>Composite (death/MI/stroke)</td>
<td>-</td>
<td>0.82 (0.72 to 0.94)</td>
</tr>
</tbody>
</table>

After Nissen and Wolski 2007; Lincoff et al 2007

**Thiazolidinedione Intervention with Vitamin D Evaluation (TIDE)**
The TIDE study

Primary objectives

1. To test the cardiovascular effects of long-term treatment with rosiglitazone or pioglitazone.

2. To compare the effects of long-term vitamin D supplementation on death and cancer.
The TIDE study

Design

Type 2 diabetics
n = 16,000

Placebo
n = 5,333

Pioglitazone
n = 5,333

Rosiglitazone
n = 5,333

Composite endpoint
Myocardial infarction, stroke, death

Type 2 diabetics
n = 16,000

Placebo
n = 8,000

Vitamin D
n = 8,000

All-cause death or cancer
Such trials only justified if:
• they answer critically important public health questions;
• the potential risks are acceptable and minimised;
• there is explicit informed consent.
Randomised Controlled Trials

Δ P<0.05
Randomised Controlled Trials

**Strengths**

1. Minimises bias

2. Minimises confounding

3. Minimises random error
Alfie
Randomised controlled trials

**Weaknesses**

1. Statistical issues
2. Generalisability
3. Resource implications
The null hypothesis is tested by estimating the probability of obtaining a result as extreme or even more extreme, as the one observed, were the null hypothesis to be true.
Statistical Issues

The null hypothesis

1. Definition of "extreme"
   - Arbitrary
   - Inconsistent

2. Ignores previous studies
   - Drug development
   - Previous trials

3. Clumsy
   - Equivalence
   - Non-inferiority
   - Futility (!)
Statisticl Issues

Multiplicity

Multiple testing:
- Interim analyses
- Subgroup analyses
- Safety analyses
Generalisability

Main problems:

1. Relatively small patient numbers
2. Homogeneous patient population
3. Limited period of time
4. Under-representation
   - Young
   - Elderly
   - Ethnic minorities
   - Co-morbidity
"Between measurements based on randomised controlled trials and benefit in the community there is a gulf which has been much under-estimated".
"Any belief that the controlled trial is the only way would mean not that the pendulum had swung too far but that it had come right off the hook".
Solutions?

1. Pragmatic trials.

2. More active comparator trials.

3. Greater use of Bayesian approaches.

4. Increasing dependency on observational studies (pharmacoepidemiology)
Thomas Bayes

(1701-1761)
Bayesian Statistics

What's the Problem?

1. Statistical prejudice
2. The concept of subjective probability
3. Establishing priors
4. Computationally difficult
5. Drug regulatory authority resistance
6. Some statisticians can’t do it
Observational approaches

1. Historical controlled trials
2. Case-control studies
3. Concurrent cohort studies
4. Before-and-after designs
5. Databases/registries
6. Case reports
Observational Studies

Strengths

• Assessment of benefits

• Assessment of harms

• Generalisability
Observational Studies

Weaknesses

- Selection bias
- Confounding by indication
Historical Controlled Trials

Comparison(s) between:

• A group of patients treated with a (usually) new intervention

• A “historical” cohort (implicit or explicit)
## Historical Controlled Trials

### Evidence of Benefit

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyroxine (1891)</td>
<td>Myxoedema</td>
</tr>
<tr>
<td>Streptomycin (1948)</td>
<td>Tuberculous meningitis</td>
</tr>
<tr>
<td>Defibrillation (1948)</td>
<td>Ventricular fibrillation</td>
</tr>
<tr>
<td>Ganglion blockers (1959)</td>
<td>Malignant hypertension</td>
</tr>
<tr>
<td>Oestrogen + progestogen (1960)</td>
<td>Oral contraception</td>
</tr>
<tr>
<td>N-acetylcysteine (1979)</td>
<td>Paracetamol poisoning</td>
</tr>
<tr>
<td>Ganciclovir (1986)</td>
<td>CMV retinitis</td>
</tr>
<tr>
<td>Imiglucerase (1990)</td>
<td>Gaucher’s disease</td>
</tr>
<tr>
<td>Laser therapy (2000)</td>
<td>Port wine stains</td>
</tr>
<tr>
<td>Imatinib (2002)</td>
<td>Chronic myeloid leukaemia</td>
</tr>
<tr>
<td>Imatinib (2005)</td>
<td>Gastrointestinal stromal tumours</td>
</tr>
</tbody>
</table>
Cytomegalovirus retinitis

Ganciclovir
Port wine stains
Historical Controlled Trials

Criteria for Acceptance

1. Biological plausibility
2. No reasonable comparator
3. Predictable natural history
4. Adverse effects not expected to compromise benefits
5. Substantial effect size (signal-to-noise ratio)
# Case-control studies

## Harms

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Harm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral contraceptives</td>
<td>Venous thromboembolism</td>
</tr>
<tr>
<td>Diethylstilboestrol in pregnancy</td>
<td>Genital tract cancer (in offspring)</td>
</tr>
<tr>
<td>Non-steroidal anti-inflammatory drugs</td>
<td>Upper gastro-intestinal bleeding</td>
</tr>
<tr>
<td>Aspirin in children</td>
<td>Reye’s syndrome</td>
</tr>
<tr>
<td>Hormone replacement therapy</td>
<td>Venous thromboembolism</td>
</tr>
<tr>
<td>Hormone replacement therapy</td>
<td>Breast cancer</td>
</tr>
<tr>
<td>Anticonvulsants</td>
<td>Stevens-Johnson syndrome</td>
</tr>
<tr>
<td>Olanzepine</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>Fluoroquinolones</td>
<td>Ruptured Achilles tendon</td>
</tr>
<tr>
<td>Biphosphonates</td>
<td>Atypical femoral fractures</td>
</tr>
</tbody>
</table>
## Some pharmacogenetic associations

<table>
<thead>
<tr>
<th>Genetic marker</th>
<th>Population prevalence</th>
<th>Drug</th>
<th>Adverse reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-B*3101</td>
<td>2-5%</td>
<td>Carbamazepine</td>
<td>Hypersensitivity</td>
</tr>
<tr>
<td>HLA-B*5701</td>
<td>5-9%</td>
<td>Abacavir</td>
<td>Hypersensitivity</td>
</tr>
<tr>
<td>HLA-B*5701</td>
<td>5-9%</td>
<td>Flucloxacillin</td>
<td>Hepatitis</td>
</tr>
<tr>
<td>HLA-DRB1*1501</td>
<td>21.4%</td>
<td>Co-amoxiclav</td>
<td>Cholestasis</td>
</tr>
<tr>
<td>HLA-B*1502</td>
<td>5-12% (Asians)</td>
<td>Carbamazepine</td>
<td>Stevens-Johnson syndrome</td>
</tr>
<tr>
<td>HLA-B*5801</td>
<td>0.2%</td>
<td>Allopurinol</td>
<td>Stevens-Johnson synrome Toxic epidermal necrolysis</td>
</tr>
<tr>
<td>m.1555A&gt;G</td>
<td>0.2%</td>
<td>Aminoglycosides</td>
<td>Permanent deafness</td>
</tr>
<tr>
<td>Factor V Leyden mutation</td>
<td>≈5% (Caucasians)</td>
<td>Oral contraceptives</td>
<td>Venous thromboembolism</td>
</tr>
</tbody>
</table>
“I cannot be certain that the sun will rise every morning for the next month.

But I am sufficiently confident to have purchased a monthly season ticket to get me to work each day”.
## Hierarchies of evidence

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Systematic review of randomised controlled trials with homogeneity</td>
</tr>
<tr>
<td>1a-</td>
<td>Systematic review of randomised controlled trials worrisome heterogeneity</td>
</tr>
<tr>
<td>1b</td>
<td>Individual randomised controlled trial with narrow confidence interval</td>
</tr>
<tr>
<td>1c</td>
<td>All or none effects</td>
</tr>
<tr>
<td>2a</td>
<td>Systematic review of cohort studies with homogeneity</td>
</tr>
<tr>
<td>2a-</td>
<td>Systematic review of cohort studies with worrisome heterogeneity</td>
</tr>
<tr>
<td>2b</td>
<td>Individual cohort study including randomised controlled trials with &lt; 80% follow-up</td>
</tr>
<tr>
<td>2c</td>
<td>Outcomes research or ecological studies</td>
</tr>
<tr>
<td>3a</td>
<td>Systematic review of case-control studies with homogeneity</td>
</tr>
<tr>
<td>3a-</td>
<td>Systematic review of casecontrol studies with worrisome heterogeneity</td>
</tr>
<tr>
<td>3b</td>
<td>Individual case-control studies</td>
</tr>
<tr>
<td>4</td>
<td>Case series and poor quality cohort or case control studies</td>
</tr>
<tr>
<td>5</td>
<td>Expert opinion without explicit critical appraisal; or based on physiology or “first principles”</td>
</tr>
</tbody>
</table>
William Blake
(1757-1827)

“God forbid that truth should be confined to mathematical demonstration”