Name: Professor Sheila M. Bird OBE FRSE
Job title: Programme Leader (biostatistics at the interface of public health and other jurisdictions), MRC Biostatistics Unit, Cambridge CB2 0SR (and visiting professor at the Department of Mathematics and Statistics, Strathclyde University in Glasgow).
Is this input submitted as an organisational or individual response? Individual
Are you happy for your response to be published by the Academy? Yes

---

Background and conflicts of interest:

I have served as a Medicines Commissioner, on the research and ethics committee for GP Research Database (now CPRD), and as the first statistician on the Appraisal Committee for the National Institute for Health and Care Excellence (NICE). NICE appraisals during my tenure included initial evaluations of Tamiflu, hip replacements, drugs for multiple sclerosis and for alzheimers disease – among others. I have designed record-linkage studies to elucidate, and meta-analyse, prisoners' risk of drugs-related death soon after release from prison; and to discover if those referred by the criminal justice system into treatment for opioid dependency derive equivalent benefit – in terms of mortality reduction – as other clients who receive opioid substitution therapy. Record-linkage studies have enormous potential to shed light on risks that occur with a frequency of less than 1 in 1000 (so-called rare), especially those which result in hospitalization or mortality.

I hold shares in GSK.

I was a member of UK's Scientific Pandemic Influenza Advisory Committee prior to and during the H1N1 pandemic. I served on the Royal Statistical Society's Working Party on "Statistical issues in First-in-Man Studies" (1) and chaired the RSS Working Party on Performance Monitoring in the Public Services (2). Finally, I lead for the Royal Statistical Society on the need for legislation in England and Wales to end the late registration of fact-of-death which undermines the timeliness of record-linkage studies in England and Wales where fact-of-death for one in five of all deaths aged 5-44 years is not registered for at least six months (3). Finally, I co-supervise the doctoral thesis of a pharmacist at Makerere University who seeks to improve Uganda's pharmaco-vigilance.

1. The overarching aim of the workstream is to better understand how society uses evidence to judge the risks and benefits of medicinal products. In your view, what are the key factors underpinning this process that the Academy should consider?

Context matters when balancing the risks and benefits of medicinal products as illustrated by i) Tamiflu for seasonal versus pandemic influenza, ii) statins for prevention versus treatment, and iii) myriad drugs for advanced versus early disease. Simply stated, the reason is that risks (serious adverse events) which one is prepared to tolerate for benefit in extremis are viewed differently in the context of prevention or by those who are early in their disease-course.

Risks which arise in the longer-term cannot be identified by randomized controlled trials in which patients have been exposed to the trial medication for a shorter period. Similarly, serious risks that occur with a frequency of 1 in 10,000 – or less often – are very unlikely to be discerned in randomized controlled trials unless they manifest in a very distinctive manner. For example, the

announcement by UK of then-new variant Creutzfeldt-Jakob Disease in young people required 10 cases (to be exceptional in terms of frequency at a young age) together with distinctive neuropathology (which was confirmed internationally before UK's announcement).

In terms of the public's understanding of risks and benefits, for example of licensed medicines, it would be useful to make explicit the statistical power (or lack of it) on the basis of which licensing decisions have been made. For example, a statement such as:

"When ABC was licensed, 5,000 patients world-wide had been randomized to receive drug ABC for at least 1-year and were followed-up for at least a further six months after they ceased to use drug ABC. This evidence base is sufficient to confer 99% assurance that a serious adverse event which occurs at a rate of 1 in 1000 patients would have materialized at least once in 5000 patients.

No re-assurance can be given in respect of serious adverse events which might occur at a rate of 1 in 10,000 patients as we'd need to have randomized at least 30,000 patients and thereby delayed other patients' access to an efficacious medication. Nor can we guarantee that there are no serious adverse events in association with use of drug ABC for substantially longer than 1-year."

2. When evaluating the risks and benefits of medicinal products, what are the strengths of evidence that originates from different sources?

How informative pre-clinical studies are about risks to humans depends on whether or not the pharmaceutical concerned is a monoclonal antibody, as the Academy of Medical Sciences' own working party warned well ahead of TGN1412. The RSS's Working Party report in the aftermath of TGN1412 (1) highlighted the importance of expert opinion in bridging from pre-clinical to first-in-man studies: as, indeed, cytokine release storm was an anticipated-risk by immunologists as set out in the TGN1412 protocol and as confirmed by Te Genero's immunologist.

Elicitation to prior opinion matters – on benefits and risks: When it comes to conflict of interest, the best should not be made the enemy of the good: as the best insight may come from those most closely involved with a drug's development and their industry-affiliation should not exclude them as consultees when it comes to the elicitation of prior opinion about either benefits or risks. When designing RCTs or other studies, analysts should be mindful that innovators (whether in the pharmaceutical industry or not) are likely to be more optimistic about their intervention than are external scientists. Thus, when – in the 1980s - I elicited the prior opinion of non-Cambridge paediatrician peers about the likely impact of Artificial Lung Expounding Compound on the mortality of 25-29 weeks old babies, their view was that a reduction to from 36% to 28% was a priori plausible whereas the Cambridge team's prior belief centred on 24% (4). The BREATHE trial (5) was designed to discern a reduction of a quarter from 36% to 27%, swayed by peer-opinion as uptake requires that peers be persuaded: BREATHE observed a reduction to 21%!

In the early 1990s, we stopped the MRC-funded RCT of neutron versus photon therapy for pelvic cancer because of higher mortality in those randomized to neutrons, of which the minority prior-opinion had forewarned (6).

Randomized controlled efficacy-trials matter: Typically, more than efficacy RCT is conducted prior to a licence-application but the evidence portfolio may relate to a range of doses, which has

gradually narrowed as considerations of benefit versus risk have homed-in on a licensable-dose. The “homing-in” process is being refined, and made more explicit, by smarter Bayesian designs for early RCTs.

Meta-analysis of RCTs is important for evidence-synthesis on efficacy but - in the end – is only as good as its constituent RCTs are. Constituent RCTs may differ not only in dose but in duration and in target-population. Analysts can generally do a more thorough job of evince-synthesis if they have access to individual patient data or, as in the case of NICE’s re-appraisal of Alzheimers’ drugs, triallists have agreed to a meta-analysis protocol whereby they each provide pre-agreed summary-statistics from their trials which enable, for example, dose-response studies or investigation of how long efficacy persists after the drug us stopped or in accordance with the randomized duration of prescribing which may differ within and/or between trials. In summary, meta-analysts also make a range of decisions, which is why the actual decisions and the justifications for them are set out in a meta-analysis protocol before the actual data are brought together.

Network meta-analysis takes evidence-synthesis on efficacy an important step further because it facilitates indirect comparison of the efficacy of A versus B when, for example, there has been a suite of RCTs which compare A versus C1 or A versus C2 or B versus C1 or B versus C2 where the same set of UK triallists has undertaken the C1 comparisons but (say) a range of different triallists internationally has contributed to the C2 comparisons. It is immediately clear that different sets of assumptions may be appropriate when meta-analysing the C1 comparisons versus the C2 comparisons and that the extent of difference mediates how, if at all, the C1 and C2 comparisons are to be pooled (if at all) in reaching a summary of the available evidence concerning the relative efficacy of A versus B.

3. When evaluating the risks and benefits of medicinal products, what are the limitations of evidence that originates from different sources?

Different RCT conduct, different centres, different doses, different target populations, different control treatments, different durations of treatment and/or follow-up, and differently-specified primary and secondary outcomes - let alone their associated statistical power – make for limitations of evidence on efficacy from different sources. However, the value of evidence from different sources may be radically different for risks versus benefits.

For example, a specific serious risk pertaining to patients randomized to drug ABC may be the same whether the drug was deployed for disease X, Y or Z, which means that meta-analysis across randomized patients with disease X, Y or Z could be instructive. Alternatively, the specific serious risk that pertains to drug ABC may be thought to pertain to all drugs in the class Axx. Accordingly, meta-analysis across patients who were randomized in RCTs of any drug (including ABC) belonging to class Axx could be important.

If the above specific serious risk pertaining to patients randomized to intervention ABC does not materialize until 5-10 years after randomization (such as: early-failure rates for replacement hip) but RCTs for efficacy followed patients for 2-years only then unless patients were invited to give explicit consent for later follow-up by record-linkage (to discover hospital-episodes in the longer-term and survival-status) or a hip-replacement registry was set up (as in UK following NICE’s recommendations), serious adverse events risk being undiscovered even in the longer-term (7).

Evidence-synthesis on risks therefore matters - but typically requires, besides RCTs, data from post-marketing surveillance, disease-registries or observation (such as CPRD). This type of meta-analytical evidence-synthesis is not routinely conducted – even by drug-class – by licensing authorities at the time of granting a product-licence. The reason is that information is not routinely or proactively pooled across submissions for different drugs in the same drug class (Axx in my example).

Thus, consideration might be given by licensing authorities for establishing a sort of meta-analytical protocol for the pooling of evidence about rare serious adverse events pertaining to drugs in the same class. However, doing so would be non-trivial and would need piloting and international collaboration – not only by regulatory authorities but by industry.

More immediately feasible would be to offer patients who have collaborated in RCTs the safeguard of longer-term follow-up by record-linkage for follow-up in terms of hospital episodes and survival status.

As personalized medicine comes into its own, personalization in terms of risks as well of benefits may further complicate the scene – so that risks may be common across phenotypes but benefits differential or the reverse.

4. Please provide details of any further examples or case studies that it would be useful for the project to consider.

Context matters – thus, the side-effects of Tamiflu mattered more because H1N1 was, generally, a much milder thing than H5N1, the wild thing that had been prepared for.

Statistical power and the RCT evidence-base matter – thus, Pandemrix (GSK vaccine against H1N1, see http://www.theguardian.com/world/2015/jul/01/swine-flu-jab-narcolepsy-linked-autoimmune-response) had its vaccine-delivery-mechanism pre-tested in some 5000 patients but the actual-vaccine in only a few hundred before being administered to some 6 millions in UK, including teenagers. The risk to young people of developing narcolepsy was identified by surveillance in Scandanavia and confirmed in UK by an approved, special study under the auspices of the British Paediatric Association's Surveillance (8).

Lay administration of take-home naloxone or naloxone-on-release to reduce the risk of fatality from opioid overdose was subject to science-led before/after evaluation of a national public health policy in Scotland (9) where one-seventh only of the take-home naloxone was administered to the person for whom it was prescribed and one-third of the naloxone-on-release, as the MRC-funded prison-based pilot N-ALIVE Trial's interim data also showed. The consistency of Scottish and N-ALIVE Trial's reported-back data meant that instead of individually-randomizing over 50,000 eligible prisoners, the main N-ALIVE Trial would have had to randomize over 150,000 and could not rule out substantial contamination at administration between eligible ex-prisoners who had been prescribed naloxone-on-release and randomized controls. Public health interventions for prevention cannot always be subject to individualized randomization but – even in emergencies – consideration might be given to randomized step-wedge designs.

The arguments against RCTs in how to manage Ebola were misguided but oft-repeated.

5. Please highlight any broadly applicable principles that should govern the presentation, interpretation and weighting of evidence about medicinal products.

Broad principles should apply well beyond medicinal products: for example, to devices and to social or criminal justice interventions and ministerial policies.

Parliament, press and public should understand that denominators matter: so that 10% has a different inferential value if the estimate is based on 1 out of **10** versus 999 out of **10,000**.

Parliament, press and public should signify that study-design matters by always reporting it: in particular, the principle of randomization matters as safeguard against unmeasured biases.

Parliament, press and public need transparency about the primary beneficial-effect that is considered a priori plausible (and why), about known potential adverse-events (and whether measured), and about what rate of unknown serious-risks would be confidently discerned: in particular, there should be transparency at the outset about the plausible benefit that a new medicinal product brings (and the likely cost per quality-adjusted life-year); about the serious risk-rate that the research-programme can reliably elucidate; and – for government policies – how known potential adverse-events will be monitored, see RSS Working Party on Performance Monitoring in the Public Services (2).

6. Concerns have been raised about how industry funding impacts on the validity, or the perception of validity, of evidence. For example, the ability of academic researchers funded by industry to remain impartial when evaluating evidence has come into question. How should conflicts of interest be addressed? How important is industry funding in generating and analysing evidence? Other than industry sponsorship, what are other potential sources of conflicts of interest?

Declarations about conflicts of interest in respect of published papers will soon be longer than the paper itself as editors cannot set a word-limit on the latter declarations.

For academics, conflicts across research programmes or in terms of influencing policy – whether by industry, government or research-funders – may be as critical as financial conflicts. The current practice of ruling ineligible for chairmanship those who hold shares (for example) but admitting as eligible those who serve on grant-awarding or policy-forming bodies may have got the balance wrong. What matters for the public health is that decision-making bodies have access to the best-informed advice and it is preferable that such advice be given in an open forum/committee, where it can be challenged by fellow-scientists than by informal, undeclared, off-line consultation.

7. Please outline any past, current or planned initiatives to examine how patients, citizens and healthcare professionals (and those who seek to inform them) evaluate scientific evidence about medicinal products.

The MRC-funded pilot N-ALIVE Trial (10) randomized over 1600 eligible prisoners in 15 English prisons. The N-ALIVE team invited participants to give consent for self-completion of a recidivist questionnaire if re-incarcerated within 6 months of their N-ALIVE release-date. The self-

questionnaire asked for free-text contributions on whether the trial had presented any difficulties for the participant; and if participants had any suggestions for us about the trial. Prisoners' contributions via this forum were impressive – from concern about children's access to naloxone or insufficient awareness of the trial by local police forces to suggestion that the randomization ratio be more favourable to naloxone than 50:50 or that the wash-out period before re-randomization be shorter than 6-months since last N-ALIVE release-date; and much more. We offered this forum because the N-ALIVE Trial's participants did not have re-contact with their prescriber post-release but the thoughtfulness of participants' replies makes us consider that this sort of open-forum could be useful for engaging patients in research studies more generally in thinking about, and improving, the design of the studies in which they have volunteered to take part.

(In N-ALIVE Trial, we know to which group a recidivist was randomized but - to maximize frankness - not name or RCT-number and participants were given an envelope in which to seal answers).

8. What are the most effective ways of communicating evidence to various stakeholders and engaging with them about such evidence?

Please see work by Cambridge University's Winton Professor for Public Understanding of Risk. Effective ways of communicating evidence that suit one cadre of individuals (highly quantitative, say) does not suit another (also highly quantitative) or a third (not confident quantitatively but heaps of common-sense) or a fourth . . .

How many relevant "cadres" there are may differ by issue, but offering a hierarchy of detail that the more questioning can drop more deeply into and a variety of visual formats to choose from has merit.

References

1. Royal Statistical Society Working Party (chair: Prof Stephen Senn, member: **Bird SM**). *Statistical issues in first-in-man studies. Journal of the Royal Statistical Society, Statistics in Society* 2007; 170: 5 17 – 579.

2. Royal Statistical Society Working Party on Performance Monitoring in the Public Services (chair: **Bird SM**). *Performance indicators: good, bad, and ugly. Journal of the Royal Statistical Society, Series A* 2005; 168: 1 – 27.

3. **Bird SM**. Editorial: Counting the dead properly and promptly. *Journal of the Royal Statistics Society Series A* 2013; 176: 815 – 817.

4. **Gore SM**, Morley CJ. ALEC saves lives of very premature babies. *MRC News* 1987; 36: 6-7.

5. TEN CENTRE STUDY GROUP (**Gore SM** statistician-member)
   Ten-centre trial of artificial surfactant (artificial lung expanding compound) in very premature babies. *British Medical Journal* 1987; 294: 991-996.

6. Errington RD, Ashby D, **Gore SM**, Abrams KR, Myint S, Bonnett DE, Blake SW, Saxton T.High energy neutron treatment for pelvic cancers: study stopped because of increased mortality. *British Medical Journal* 1991; 302: 1045-1051.

7. Smith AJ, Dieppe P, Howard PW, Blom AW, on behalf of the National Joint Registry for England and Wales. Failure rates of metal-on-metal hip resurfacings: analysis of data from the National Joint Registry for England and Wales. *Lancet* 2012; 380: 1759-1766.

8. Miller E, Andrews N, Stellitano L, Stowe J, Winstone AM, Shneerson J, Verity C. Risk of narcolepsy in children and young people receiving AS03 adjuvanted pandemic A/H1N1 2009 influenza vaccine: retrospective analysis. *BMJ 2013; 346 doi: http://dx.doi.org/10.1136/bmj.f794 (Published 26 February 2013) Cite this as: BMJ 2013;346:f794*

9. **Bird SM**, Parmar MKB, Strang J. Take-home naloxone to prevent fatalities from opiate-overdose: protocol for Scotland's public health policy evaluation, and a new measure to assess impact. *Drugs: Education, Prevention and Policy* 2015; 22: 66-76. Published online on November 18, 2014. (doi:10.3109/09687637.2014.981509) http://informahealthcare.com/eprint/y2XmCCCbJFkUyJMyp4aG/full.

10. Strang J, **Bird SM**, Parmar MKB. Take-home emergency naloxone to prevent heroin overdose deaths after prison release: rationale and practicalities for the N-ALIVE randomized trial. *Journal of Urban Health* 2013; 90: 983 – 996. Published online: 01 May 2013; doi:10.1007/s11524-013-9803-1.