To improve our understanding of PCI expertise, a research programme to investigate the role of imitation and cognition in PCI knowledge acquisition is warranted. Employing modern techniques of cognitive neurosciences including behavioural testing and neuroimaging, the development of procedural expertise together with the corresponding neurological substrates could be studied during a period of training, based on either approach. The acquired PCI competence may then be assessed by measuring, e.g. speed, accuracy, and reliability of analysis of standardized interventional coronary angiograms and/or by evaluating eye tracking tracings.

In addition neural substrates may be investigated using functional and structural magnetic resonance imaging (analogous to its use in other learning programmes, e.g. in Taubert et al. J Neurosci 2010, Wenzel et al. PLOS One, 2014). Preliminary observations suggest complementary roles of the implicit, imitation-based, and explicit, cognition-based approaches in the development of PCI expertise.

The implementation of the cognitive approach in PCI holds promise to raise 'the state of practice' in PCI to further improve outcomes, despite the increasingly challenging patient populations.

The challenge of delivering reliable science and guidelines: opportunities for all to participate

Graham D. Cole* and Darrel P. Francis

International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College London, 59-61 North Wharf Road, London W2 1LA, UK

Erroneous research results have potential for harm far beyond the patients recruited into the study, because the leverage of leadership permits them to alter behaviour of thousands of clinicians for many years. Medical research does not yet have as well-developed and rapid-responding safety systems as other spheres where lives are at stake. As doctors we would all leap to assist in a medical emergency for a patient not formally under our care. Yet, when problems occur in clinical research, we do not yet have good mechanisms to all work together quickly and decisively to improve patient safety. In this article we describe how each of us, regardless of our role, can (and perhaps should) take active steps to support delivery of reliable science for patient care.

Keywords
Scientific misconduct • Clinical governance • Patient safety

The complex challenge of undoing erroneous research

Rarely has the ESC faced as difficult a challenge as handling the recommendations regarding perioperative beta-blockade initiation in patients undergoing non-cardiac surgery. On 17 November 2011, a family of perioperative care trials were announced to have suffered from ‘data fabrication’ and ‘academic misconduct’. The host university investigated and reported that the study authors ‘made a number of incorrect or contradictory statements’ which were ‘not very credible’. The press release described the data as ‘fictitious’ and ‘unreliable’. In 2012, the university issued a second report specifying that one of the studies used by the guidelines for its beta-blockade recommendation was ‘negligent’ and ‘scientifically incorrect’. Nevertheless, one week later on 23 November 2011, a press release announced ESC Guidelines are based on the contributions of many European experts and on available evidence based medicine, including many studies from different nations. They are, therefore, the result of a group discussion and not of an individual position. We are saddened by Prof Poldermans’ situation and, although we are confident that our Guidelines are supported by reliable data, we are carefully looking into the Guidelines for Pre-operative Cardiac Risk Assessment.

Neither the details of the expert contributions to the beta-blocker recommendation, nor the identity of the ‘confident’ persons, nor the reason for their confidence have been made public yet.

* Corresponding author. Tel: +44 2075941093, Fax: +44 2080825109, Email: g.cole@imperial.ac.uk

Un fortunately, Guidelines recommending initiation of beta-blockade for non-cardiac surgery, for patients with ischaemic heart disease (IHD) or positive ischaemia tests or high-risk surgery, had incorporated these studies which the host University, Erasmus, had now declared ‘negligent’ and ‘scientifically incorrect’. Nevertheless, one week later on 23 November 2011, a press release announced ESC Guidelines are based on the contributions of many European experts and on available evidence based medicine, including many studies from different nations. They are, therefore, the result of a group discussion and not of an individual position. We are saddened by Prof Poldermans’ situation and, although we are confident that our Guidelines are supported by reliable data, we are carefully looking into the Guidelines for Pre-operative Cardiac Risk Assessment.
Re-analysis removing non-credible trials

The press release did not mention that, without the non-credible DECREASE trials, meta-analysis of the remaining trials suggests initiating beta-blockade specifically for surgery increases perioperative mortality. With ~40 million operations per year in Europe, it is important we ensure that any guidance we distribute to clinicians is secure.

However, any meta-analysis of perioperative beta-blockade that limits itself to credible sources faces serious limitations. Only one trial, POISE, had adequate sample size to detect a clinically important difference in mortality, and therefore dominated the result. We therefore have little idea whether other agents in the class, other doses, different patient types, or earlier initiation with uptitration, could cause total mortality to decrease rather than increase.

The Guideline does specify that "high dose beta-blockers without titration are not recommended." This concept of dose of beta-blockade may ultimately prove crucial.

Unfortunately, the Guideline presents a potentially misleading comment about POISE, adding up metoprolol doses on two successive days as though they were part of a single daily dose, while in contrast not making the same miscalculation for DECREASE, i.e. not claiming to have administered 20 mg of bisoprolol. When the lead Guideline author first made this allegation in public in February 2009 he was promptly corrected in person by POISE’s lead author. Despite this publicly witnessed refutation, 7 months later his 2009 guideline rebroadcast the error.

Karl Swedberg had, years before, protested that the pioneering NEJM trial reporting mortality reduction with perioperative beta-blockade had inexplicably excluded deaths within hospital, which were twice as frequent in the beta-blocker arm. He was ignored. The DECREASE family then provided further evidence of benefit elegantly but (at least in part) fictitiously.

Timing, dose, and patient group

Timing of initiation is recommended to be weeks before surgery with careful uptitration in the hope that this will permit mortality benefit. Just two substantial trials reported doing this: DECREASE and DECREASE IV. They seemingly completed a unique (yet undescribed) multi-week uptitration which no other endpoint trial had even attempted. Unfortunately when DECREASE IV was investigated ~2 years after publication, there was no evidence that any patient had received any drug. Moreover, all evidence of the monitoring visits, measurements, and uptitration had been mislaid. We may never know how it happened.

Dosing is also controversial. POISE initiated with 100 mg long-acting once-daily metoprolol, the manufacturer-recommended starting dose for IHD and aimed for a target maintenance of 200 mg od. Even expert anaesthetists often do not realize that the relative bioavailability of this preparation is only 70–75% so these doses correspond to three times daily dosing with 25 and 50 mg, respectively. It is unknown whether even lower dosing results in a reversed effect on mortality.

Risk stratum may also be important. Unfortunately, in the only adequately powered credible study, the higher-risk patients showed a trend to suffer not less but more harm.

Observational data

Peer review of this article required discussion of two non-randomized reports to provide balance. On the spectrum of bias-resistance for guiding clinical decisions, accurately calculated results of reliably conducted RCTs do indeed lie at one extreme. Whether fully representing this applicability spectrum is essential, or indeed wise, is controversial.

One non-randomized study’s raw data shows a ~40% higher rate of 30-day mortality in beta-blocker recipients than in non-recipients as well as higher rates of cardiac arrest, Q-wave myocardial infarction, stroke (each P < 0.001). After processing to create cohorts more balanced for registry-documented characteristics, the relationship reversed with mortality and cardiac complications all significantly lower in the beta-blocker recipients and stroke neutral. When observational data conflict with randomized trials, there should be careful inspection for discrepancies.

In the second major non-randomized study for patients in whom perioperative indication might be contemplated, namely those with IHD but no heart failure or prior myocardial infarction, the odds ratio for all-cause mortality on beta-blockade was 1.3 (P < 0.05). For those with the long-term beta-blocker indication of heart failure, the odds ratio was 0.82 (P < 0.05). For patients with MI, the picture was more neutral. Combining all these groups together, the association was neutral 0.95 (0.85–1.06, P = ns).

Neither observational study addressed beta-blocker initiation for surgery. Neither reported why the doctors made their decisions, nor should we assume that undocumented features were equal between groups. Balancing both documented and undocumented features is the raison d’être of RCTs.

Challenges to reaching physicians and conveying the alarm

Physicians monitoring ESC press releases will have seen the notice of 5 August 2013 indicating that the ACC, AHA, and ESC writing committees were updating the guidelines. It indicated that ‘the initiation of beta blockers in patients who will undergo non-cardiac surgery should not be considered routine, but should be considered carefully by each patient’s treating physician on a case-by-case basis’. In August 2013, it became accessible from the ESC guideline webpage by clicking a link worded ‘August 2013’ under the heading ‘Regarding the situation of Professor Don Poldermans . . .’ with a subheading ‘Read the ESC statement’.

In February 2014, the text of the notice became prominently displayed on the guideline webpage itself. At the same time the EHJ inserted this statement as a new first page within the PDF of the guideline document. Nevertheless, clinicians did not always consider that the meaning of this advice was clear.

Learning and growing

Aviation professionals have pioneered systems to identify, examine, and improve practice from professional failures. Medicine is now entering this path in clinical practice but must do the same for research.
Focus invites fraud

Researchers focusing on a single novel idea, investing deeply and facing a tight career funnel demanding positive results, could easily be overwhelmed if the long-cherished hypothesis emerges to be incorrect.

Starved of funding, they might in desperation persuade themselves that a little selectivity with data is acceptable to pay salaries of loyal staff.

As a society we need to agree that it is better science to reliably report an effect as small, than to unreliably report it as large. The current convention unfortunately favours the opposite tendency.

Researchers are safer with a broad portfolio than betting on one idea alone. This may be difficult for single-concept companies.

Co-authors: a proposed ‘inverse secrecy’ duty

In interpreting our meta-analysis, we hit upon the possibility that perhaps many of the patients in the multi-centre DECREASE-IV had been recruited from outside the centre with doubtful data. This might have partly rehabilitated DECREASE IV. However, we could not determine which multiple centres had performed it, from the article or the journal. We contacted the non-Erasmus authors asking whether their centre had been involved in recruitment. The only answer received was: “I would like to know the authority from which you ask this question?” We must not criticize them for this, since there is currently no requirement for authors to answer such questions. However, perhaps our community should agree that, for future publications, any author should be willing to answer simple questions: a duty of ‘inverse secrecy’.

This would make scientific research more credible. The approximately 3000 patients reported to have participated in the DECREASE randomized trials, knowingly or otherwise, must have had tens of thousands of interactions with numerous staff—yet staff involved in this extensive activity seem to have fallen silent. Even the first author of two publications, who was asked to speak to the investigation, in person, by telephone or by video-link, repeatedly refused.

Employer duty to prevent moral hazard

The financial collapse of 2008 brutally taught us the term ‘moral hazard’. If gains are received personally but losses jettisoned, financiers may gamble progressively greater amounts, until losses can no longer be concealed. Co-authorship, analogously, brings status and metrics but no harm when the study is found misconducted. A university can accept hundreds of shared publications simultaneously to have enough involvement for co-authorship yet not enough to carry responsibility for integrity.

Co-authors are far better placed than readers, editors, or even institutions, to identify misconduct. If we made them all share the consequences when research is misconducted, they would try harder to prevent it: ‘crowdsourcing’ scientific integrity.

Clinicians transitioning to research: the traps

‘Fabrication’, creation of false data, is widely recognized as misconduct, but other types are more insidious.

A second form, ‘falsification’, can occur from common clinical and teaching habits. For example, when advising patients we often focus on the test results consistent with our believed diagnosis, in order not to confuse them. When teaching, we deliberately choose clear examples. However, in research, suppressing non-fitting measurements or selecting patients to support a hypothesis is falsification. We must recognize these standard clinical and educational behaviours as poisonous to science and curtail them during research. This is particularly problematic in larger studies because the effect of small innocent biases become more (not less) statistically significant.

There is a third problem, Russian-doll publication (undeclared data set overlap) which can occur innocently when experts understandably report growing cohorts. Less understandable is that publishers accept duplicate publication with trial names altered and standably report growing cohorts. Less understandable is that data set overlap, which can occur innocently when experts under-report an effect as small, than to unreliably report it as large. The current convention unfortunately favours the opposite tendency.

Researchers are safer with a broad portfolio than betting on one idea alone. This may be difficult for single-concept companies.

How to negotiate the bumpy roads?

The noted bumpiness of the road to translation of pioneering findings has many causes. Ioannides has calculated that most published pioneering findings are incorrect. Some clinicians may not understand that larger sample sizes do not eliminate bias. We believe it is an error that many guideline systems accept an observational study, if large, as the same level of evidence (B) as a randomized trial. It is easy to be impressed by strong associations between baseline predictors and outcomes or response to intervention, without seeking out their true origin.

Clinicians entering research should invest time and thought to design projects that do not propel them onto the slippery slope to unreliability.

An environment open to discussion and critical examination of perceived authority figures is vital. Doctors stumbling upon incorrect clinical research or guidelines should consider themselves responsible for taking action. For example, if a report states that 70 patients are divided into two groups of 80 and 35, curiosity should be triggered. If a guideline currently in force relies on unreliable data, perhaps one should alert one’s fellow cardiologists who through obeying it might be causing harm?

The complexity of the medical research translation process is, for once, not a challenge but an opportunity. For research failure to evoke enduring harm with global reach, multiple potential guardians must assist, actively, or passively. We now explore these frequently missed opportunities.
Duty to appreciate seriousness of unreliable diagnostic or therapeutic claims

Prison terms have been given to groups of people who promoted ineffective non-medical diagnostic equipment that may indirectly have cost lives. Enthusiastic officials had presented rigged unblinded demonstrations, and rejected negative data as reflecting inadequate training. While political authorities accepted the claims uncritically and offered support and credibility to a baseless concept, it was left to Internet-enabled journalists to expose the truth.

In contrast, we physicians enjoy unparalleled freedom to praise diagnostic or therapeutic technology with impunity. For example, much of our field has been engaged in claims on how to identify patients who benefit from cardiac resynchronization therapy, sometimes presenting mathematically implausible observational data and rejecting negative results of better-conducted studies as reflecting inadequate training. Sadly, well-conducted randomized controlled trial data show that implanting on this criterion raises mortality by +81% (95% CI: +11 to +193%). We will never be called to account.

One parallel with non-medical misguidance, however, is that when medical authorities exhibit inertia, the task may fall to Internet-enabled journalists.

Readers’ duty to stay alert

Even when merely reading papers, we should not accept claims that lie so far outside the range of plausibility that they are likely to be incorrect. For example, our community lacked curiosity when a 2002 report of a randomized trial of atorvastatin vs. usual care gave its start date vaguely as ‘4 years ago’. No-one questioned why ‘two independent committees, one for each patient group, recorded and validated primary and secondary endpoints’, or why ‘each committee was blinded to the results of the other group’. The narrowness of the standard deviation of the body mass index, 0.6 kg/m², was not considered bizarre.

This left the door open to ‘substudies’ reporting, for example, astonishing improvement in renal function from statin therapy, a concept requiring a major meta-analysis to discredit.

The next substudy claimed that liver function tests can be improved by atorvastatin. The text describes ‘intention-to-treat’ analysis, contradicted by its first figure. Many stated median drug doses appear mathematically impossible. The liver function test results are contradictory between figures and text. Half the patients with abnormal baseline liver function achieved final LDLs between 2.41 and 2.44 mmol/L, a tightness that is incredible.

Moreover, despite having originally claimed that all the patients meeting the LDL target were enrolled, the study now remembered that 21 declined. Finally, the baseline body mass index data have completely changed yet readers teased: ‘body mass index did not change’.

Enquirers’ duty to be persistent

Journal correspondents should not assume that questions reaching print will be fully answered, even by world-leading scientists conducting high-quality research. For example, a recent editorial in this journal has recently been reported on the Wall Street Journal website as supporting the exciting C-CURE trial.

However, the mean of the individual patient increments in ejection fraction in its Figure 4 appears distinctly < 5.5 units: how can it simultaneously be 7.0 units as stated in the Abstract and Results and company stockmarket prospectus? Did the authors’ reply solve the puzzle by tabulating the individual patient data?

Institutions: a duty to prioritise future patient welfare

The report of the investigative committee on academic integrity dated 8 November 2011 has done considerable harm to the reputation of the research group involved. This 2012 follow-up investigation has not been able to limit this harm.

We should not trust an institution to swiftly correct science its workers have seriously misrepresented. Harvard University famously cleared a study after examining raw data internally, only for later external investigation to reveal that there had been no study and no raw data. Erasmus University, examining cardiological research misconduct, built a panel with as many lawyers as cardiologists. The panel’s purpose can be inferred by its penultimate conclusion paragraph (quoted above). Yet the panellists’ creditable determination did not translate to retraction for several studies.

Guideline custodians were therefore put in an unenviable position, since they were forced to interpret lack of retraction as endorsement.

Excuses for non-retraction of fictitious studies were extraordinary:

… the clinical and social relevance of the potential withdrawal of this publication is now small, particularly because the Dobutamine Stress Echo has now virtually disappeared from clinical practice as a predictor of perioperative complications.

We must make institutions fear not discovery of misconduct but slowness to retract trial reports that are false.

‘Damage limitation’ should focus on limiting damage to future patients.

Authors: publish all the data, all the time

The investigation into the DECREASE family was hampered by the non-existence of almost all source data for many studies including DECREASE IV. For ECG results, records of the actual medications used, and records of the basis for the clinical outcome findings, the source documentation was almost entirely lacking.

The only DECREASE study whose full data was found turned out to be 97% composed of real patients who were not study participants. Nevertheless, unobtainability of much of the data made the investigation committee unable to retract the entire DECREASE family of publications.

In the modern era, authors should try to publish all the data, all the time. Had the DECREASE studies published anonymised case record forms, they could not have disappeared. Scans, operations,
and endpoint committees would have been recognized as non-existent.

Many journals will agree to show extensive online supplements of individual-patient data.46 Openness also protects all of us from any temptation to edit data to match expectations.28,39

A spreadsheet containing raw data from even tens of thousands of patients can be smaller than a single high-resolution photomicrograph. Similarly, paper documents can be scanned and compressed. Data needed for an inquiry would thereby be safe from being mislaid or accidentally shredded.

Journal responsibility: habeas data for impossible claims

If extraordinary claims require extraordinary evidence,65 impossible claims need even more. Journals informed of internal contradictions should ask the authors to immediately provide the full raw data set so that readers can verify the source of the error.

If authors delay, journals could immediately temporarily retract the article, by executive action of an administrator. Journals should be judged by their determination and efficiency in handling such events. Journal editors should not be trapped by legal restrictions into letting unreliable research stand.29

This would mirror the legal concept of 'habeas corpus', developed in 1215 to curtail abuses of power by leaders making arbitrary decisions in secret.

Decades ago, journal editors could personally arrange replication of incredible claims.61 Modern clinical research journals are far more reluctant to communicate to readers serious problems and even factual impossibilities.29,67

No reader should relive our disappointment of showing a senior journal editor clear evidence of large-scale misconduct and receiving the reply:

We at the journal had gone as far as we can go, sorry.

Guideline committee design: small, fast, independent, and accountable

Guidelines, exerting the leverage of leadership, have enormous potential for benefit or harm. Rapid revision is impeded by the huge number of signatories, unclear lines of emergency authority and, allegedly, general sluggishness of large societies.24 When DECREASE trials were discredited, the 45 authors, reviewers and additional contributors, and 18 additional supervisors and coordinators,1 may have been unsure who should act. Each might point to >50 others.

Meanwhile the press release62 emphasized confidence in the status quo, and ‘sadness’ for the situation of the researcher whose misconduct had now become public knowledge.

Guideline publications elevate signatories’ status and enhance the journal Impact Factors of which editors are proud.68 Their high value to patients could be further enhanced if they had a rapid mechanism to patients.

Correction of the perioperative guideline might have been quicker if the guideline chair and the head of the entire guideline system had not themselves been authors of the trials that became discredited.

Community responsibility: no wilful blindness

If criminals claim ignorance of information that would make them culpable (e.g. 'I did not ask what was in mysterious packages I was handsomely paid to transport'), courts may recognize this as 'wilful blindness' and thereby convict them.69 As doctors we should never give in to the temptation to fail to notice the obvious: whether as co-workers reading trials that we know did not happen in our hospital, co-authors offered credit on misleading manuscripts, readers noticing factual impossibilities, or editors told of serious problems.

It is unpleasant to realize that one’s own actions may have been harmful. The psychological principle of cognitive dissonance70 is the blanking-out of such recognition. Every time we act, speak or write in accordance with an instruction, we slightly adjust our belief system towards making the instruction appear less wrong.

Nor should we fall into the logical fallacy of argument from consequences: 'if this is true, we have been harming people, so it cannot be true'.

Retract and rehabilitate

We should be generous when authors who, having presented misleading reports, later correct them. Everyone makes mistakes: good scientists look for and proactively fix their own errors.67 Researchers correcting or retracting their own misleading paper, and advising others how to avoid the same traps, should be considered rehabilitated so their ‘scientific creativity once again shines’.71 Shame and ostracism should be reserved for authors (or others) in denial,72 obstructing resolution,29 or hiding misdeeds through legal threats or censorship.

Professionals’ duty to prioritize patient safety

Science is the process of replacing untrue beliefs with less untrue ones. The first version of this article used a formula73,74 from a discredited research group to highlight the urgency of reconsidering certain guideline recommendations. The article was taken offline by the journal and then unfortunately re-reported by persons who could not check the original article to see and convey all the cautions we attached to the estimate we arrived at nor understand that this was merely to demonstrate that unreliable research is not necessarily harmless.

Our article also struggled to find another profession where an error could cascade to comparable large scale harm. The example we found, political leadership, threw up examples which were rightly criticised as unpleasant. We apologise to readers who, travelling with us on the bumpy road to reliable science, may have been offended.

Our community’s leaders appeared in newspapers to be criticizing us in isolation15,75 without seeming to simultaneously criticise the original research misconduct.

Four EHJ editors criticized our removed article76 in an editorial77 accompanied by an illuminating ESC press release.78
The Editor-in-Chief, remarking to a newspaper on the difficulty of delivering reliable research results, used broad vision to point out that this was the third in a series of scientific critiques which had travelled a bumpy road.

The 2013 statistical analysis that had predicted that the true blood-pressure reduction by renal denervation would be far less than 30 mmHg, was considered provocative, but the following year, to widespread surprise, it was vindicated.

The second example was the 2010 uncovering of a serious discrepancy in research which, having broken presentation rules, was reprimed by charitably minded leaders on the untrue premise that it had not previously been presented at a cardiology conference. Later 200 more impossible features emerged, which journal editors were unable to relay to readers. The host university has now confirmed misconduct of a magnitude requiring referral to the police.

We appeal to all readers to support journals in their quest to uphold the codex of science: honesty, precision and truth.

Conflict of interest: none declared.

References


44. Leiden University Investigation. (April 2 2012).


