ORIGINAL RESEARCH Randomised controlled trial

Parachute use to prevent death and major trauma when jumping from aircraft

Robert W Yeh, Linda R Valsdottir, Michael W Yeh, Changyu Shen, Daniel B Kramer, Jordan B Strom, Eric A Secemsky, Joanne L Healy, Robert M Domeier, Brahmajee K Nallamothu, Dhruv S Kazi on behalf of the PARACHUTE Investigators

Objective To determine if using a parachute prevents death or major traumatic injury when jumping from an aircraft.

Design Randomised controlled trial.

Setting Private or commercial aircraft between September 2017 and August 2018.

Participants 92 aircraft passengers were screened for participation. 23 agreed to be enrolled and were randomised.

Intervention Jumping from an aircraft with a parachute versus an empty backpack (unblinded).

Main outcome measures Composite of death or major traumatic injury (Injury Severity Score over 15) upon impact with the ground.

Results Parachute use did not significantly reduce death or major injury (0% for parachute *v* 0% for control; P>0.9). Compared with individuals screened but not enrolled, participants included in the study were on aircraft at significantly lower altitude (mean of 0.6 m for participants *v* mean of 9146 m for non-participants; P<0.001) and lower velocity (mean of 0 km/h *v* mean of 800 km/h; P<0.001).

Conclusions Parachute use did not reduce death or major traumatic injury when jumping from aircraft in the first randomised evaluation of this intervention. However, only participants on small stationary aircraft on the ground were enrolled. When beliefs regarding the effectiveness of an intervention exist, randomised trials might selectively enroll individuals with a lower likelihood of benefit, diminishing the applicability of results to clinical practice.

Full author details are on bmj.com.

Correspondence to: R W Yeh ryeh@bidmc.harvard.edu

Cite this as: *BMJ* 2018;363:k5094

Find the full version with references at http://dx.doi.org/10.1136/bmj.k5094

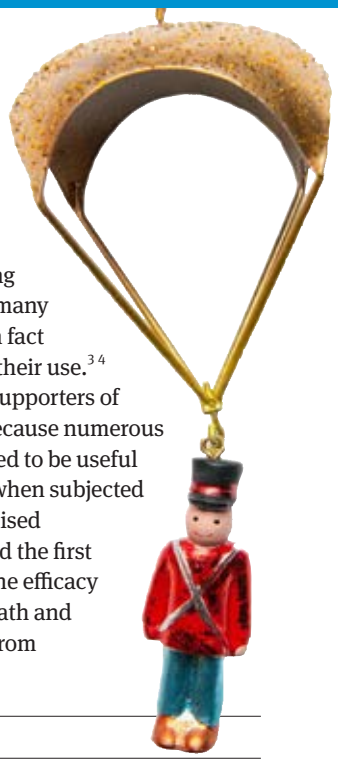
WHAT IS ALREADY KNOWN ON THIS TOPIC

- Parachutes are routinely used to prevent death or major traumatic injury among individuals jumping from aircraft, but their efficacy is based primarily on biological plausibility and expert opinion
- No randomised controlled trials of parachute use have yet been attempted, presumably owing to a lack of equipoise

WHAT THIS STUDY ADDS

- This randomised trial of parachute use found no reduction in death or major injury compared with individuals jumping from aircraft with an empty backpack
- Lack of enrolment of individuals at high risk could have influenced the results of the trial

Introduction



Parachutes are routinely used to prevent death or injury among individuals jumping from aircraft. However, evidence supporting their efficacy is weak,[1 2] and many studies of parachutes have in fact suggested injuries related to their use.[3 4] This may raise concerns for supporters of evidence-based medicine, because numerous medical interventions believed to be useful have failed to show efficacy when subjected to properly executed randomised clinical trials.[6 7] We conducted the first randomised clinical trial of the efficacy of parachutes in reducing death and major injury when jumping from aircraft.

Methods

Individuals were screened for inclusion in the PArticipation in RAndomised trials Compromised by widely Held beliefs aboUt lack of Treatment Equipoise (PARACHUTE) trial. Prospective participants were approached by study investigators on commercial or private aircraft, and were asked whether they would be willing to be randomised to jump from the aircraft at its current altitude and velocity. We randomised patients (1:1) to the intervention (parachute) or control (empty backpack) after obtaining written informed consent (supplementary materials fig 1 on bmj.com). Participants were instructed to jump from the aircraft after being provided with their assigned device.

Researchers recorded the altitude and velocity of the aircraft, and conducted a follow-up interview to ascertain vital status and to record any injuries within five minutes of impact with the ground, and again at 30 days. The primary outcome was the composite of death and major traumatic injury as measured by an Injury Severity Score greater than 15.9 Secondary outcomes included death and major injury at 30 days, as well as 30 day quality of life.

Statistical analysis

The primary efficacy analysis tested the hypothesis that parachute use is superior to the control in preventing death and major traumatic injury. We tabulated baseline characteristics of the two trial arms to examine for imbalance in variables, and tested for differences between outcomes using Student's t test (continuous variables) and Fisher's exact test (categorical variables). We also compared characteristics of individuals who were screened but chose not to enroll with individuals who enrolled.

For full details of the methods and statistical analysis, please see bmj.com.

Table 2 | Baseline characteristics of participants versus screened individuals. Values are numbers (percentages) unless stated otherwise

Characteristics	Participants	Screened	P value
Total	23	69	
Demographics			
Median (SD) age (years)	38.4 (9.7)	43.0 (14.9)	0.1
Women	10 (44)	32 (46)	
Men	13 (57)	37 (54)	
Ethnic group:			0.4
American Indian or Alaska Native	0 (0)	2 (3)	
East Asian or South Asian	8 (35)	13 (19)	
Black or African American	0 (0)	2 (3)	
More than one race	0 (0)	4 (6)	
White	15 (65)	48 (70)	
Mean (SD) height (cm)	171.7 (8.5)	171.2 (11.0)	0.8
Mean (SD) weight (kg)	75.2 (18.9)	73.5 (15.5)	0.7
Medical history			
Broken bones	9 (39)	26 (38)	0.9
Acrophobia	9 (39)	23 (33)	0.6
Parachute use	3 (13)	9 (13)	>0.9
Family history of parachute use	2 (8.7)	10 (15)	0.7
Frequent flier (average >4 flights per month)	4 (17)	14 (20)	>0.9
Flight			
International v domestic flight:			0.02
International	0 (0)	8 (21)	
Domestic	23 (100)	31 (80)	
Aircraft type:			<0.001
Jetliner	0 (0)	69 (100)	
Biplane	11 (48)	0 (0)	
Helicopter	12 (52)	0 (0)	
Mean (SD) velocity (km/h)	0 (0)	800 (124)	<0.001
Mean (SD) altitude (m)	0.6 (0.1)	9146 (2164)	<0.001

Table 3 | Event rates for primary and secondary endpoints. Values are numbers (percentages) unless stated otherwise

Endpoint	Parachute	Control	Mean difference (95% CI)	P value
On impact				
Death or major traumatic injury	0 (0)	0 (0)	0	>0.9
Mean (SD) Injury Severity Score	0 (0)	0 (0)	0	>0.9
30 days after impact				
Death or major traumatic injury	0 (0)	0 (0)	0	>0.9
Mean (SD) Injury Severity Score	0 (0)	0 (0)	0	>0.9
Health status				
Mean (SD) Short Form Health Survey score	43.9 (1.8)	44.0 (2.4)	0.1 (−2.0 to 2.2)	0.9
Mean (SD) physical health subscore	19.6 (0.7)	19.7 (0.5)	0.04 (−0.5 to 0.6)	0.9
Mean (SD) mental health subscore	24.3 (1.3)	24.3 (2.1)	0.08 (−1.6 to 1.8)	0.9

Results

Study population

A total of 92 individuals were screened for participation in the PARACHUTE trial. Among those screened, 69 (75%) were unwilling or ineligible to be randomised.

Baseline characteristics of enrolled participants were similar between the intervention and control arms (see bmj.com for details). In addition, participants in the study were similar to those screened but not enrolled with regard to most characteristics (table 2). However, participants were less likely to be on a jetliner, and instead were on a biplane or helicopter (0% *v* 100%; P<0.001), were at a lower mean altitude (0.6 m, SD 0.1 *v* 9146 m, SD 2164; P<0.001), and were traveling at a slower velocity (0 km/h, SD 0 *v* 800 km/h, SD 124; P<0.001).

The figures show representative jumps (additional jumps are shown in supplementary materials).

Outcomes

There was no significant difference in the rate of death or major traumatic injury between the treatment and control arms within five minutes of ground impact or at 30 days (0% for parachute *v* 0% for control; P>0.9 for both time points, table 3). Health status at 30 days as measured by the Short Form Health Survey was similar between groups (43.9, SD 1.8 for parachute *v* 44.0, SD 2.4 for control; P=0.90).

Discussion

We have performed the first randomised clinical trial evaluating the efficacy of parachutes for preventing death or major traumatic injury among individuals jumping from aircraft. Our groundbreaking study found no difference in the primary outcome between the treatment and control arms.

A minor caveat is that event rate was substantially lower in this study than was anticipated, which could have somewhat underpowered our ability to detect clinically meaningful differences. Although randomised participants had similar characteristics compared with those screened, they could have been at lower risk of death or major trauma because they jumped from an average altitude of 0.6 m on aircraft moving at an average of 0 km/h.

Limitations of this study

The study has several limitations. First, our findings might not be generalisable to the use of parachutes in aircraft travelling at a higher altitude or velocity. Previous theoretical work supporting the use of parachutes could reduce the feasibility of conducting future trials in these higher risk settings.[16] Second, the individuals screened but not enrolled in the study were limited to passengers unfortunate enough to be seated near study investigators during commercial flights, and might not be representative of all aircraft passengers. The participants who ultimately did enroll agreed with the knowledge that the aircraft were stationary and on the ground.

The PARACHUTE trial satirically highlights some of the limitations of randomised controlled trials. Studies evaluating devices entrenched in clinical practice face the difficult task of ensuring that patients with the greatest expected benefit from treatment are included. Overcoming such a hurdle requires extreme commitment on the part of the investigators, clinicians, and patients. Stronger efforts could be made to ensure that definitive trials are conducted before new treatments become inculcated into routine practice. In addition, comparisons of study participants and non-participants should be reported consistently to facilitate the assessment of study generalisability.[14] Finally, there could be instances where clinical beliefs justifiably prevent a true randomised evaluation of a treatment from being conducted.

Conclusion

Parachute use compared with a backpack control did not reduce death or major traumatic injury when used by participants jumping from aircraft. This largely resulted from our ability to only recruit participants jumping from stationary aircraft on the ground. Individual judgment should be exercised when applying these findings at higher altitudes.

BMJ OPINION Robert W Yeh and colleagues

We jumped from aircraft without parachutes (and lived to tell the tale)

"Would you be willing to jump out of this plane without a parachute?" For the last year we've posed this question, mid-flight, to dozens of unsuspecting travellers seated on commercial aeroplanes.

Why would we set out to ask such a ridiculous question? Some background may be in order. In 2003, Smith and Pell published a tongue-in-cheek systematic review which concluded that there were no randomised clinical trials (RCTs) evaluating the effectiveness of parachutes in preventing major trauma related to "gravitational challenge." They argued that the "most radical protagonists of evidence based medicine" should volunteer to participate in a randomised, double blind trial of the parachute. In the two decades since the appearance of this seminal work in *The BMJ* Christmas issue, the parachute has been the paragon of biological plausibility. The saviour of anecdote. The arch-nemesis of evidence based medicine.

> We conducted the trial to illustrate the perils of interpreting trials outside of context

The PARACHUTE trial is our satirical attempt at bringing the parachute, as well as the almighty RCT, back down to earth.

That no one would ever jump out of an aeroplane without a parachute has often been used to argue that randomising people to either a potentially life saving medical intervention or a control would be inappropriate, and that the efficacy of such an intervention should be discerned from clinical judgment alone. We disagree, for the most part. We believe that randomisation is critical to evaluating the benefits and harms of the vast majority of modern therapies, most of which are unlikely to be nearly as effective at achieving their end goal as parachutes are at preventing injury among people jumping from aircraft.

However, RCTs are vulnerable to pre-existing beliefs about standard of care, whether or not these beliefs are justified. Our attempts to recruit in-flight passengers to our ambitious trial were first met with quizzical looks and incredulity, predictably followed by a firm, "No, I would not jump without a parachute."

But what if we provided assurances that the planes were stationary and on the ground, and that the jump would be just a couple of feet? It was at this point that our study took off. We set out in two groups, one at Katama Airfield on Martha's Vineyard and the other at the Yankee Air Museum in Ann Arbor. One by one, our study subjects jumped from either a small biplane or a helicopter, randomised to either a backpack equipped with a parachute or a look-a-like control. As promised, both aircraft were parked safely on terra firma. The matchup was, unsurprisingly, a draw, with no injuries in either group. In the first ever RCT of parachutes, the topline conclusion was clear: parachutes did not reduce death or major traumatic injury among people jumping from aircraft.

But topline results from RCTs often fail to reveal the full story. We conducted the PARACHUTE trial to illustrate the perils of interpreting trials outside of context. When strong beliefs about the standard of care exist, often only low risk patients are enrolled in a trial, which can unsalvageably bias the results, akin to jumping from an aircraft without a parachute. Assuming that the findings of such a trial are generalisable to the broader population may produce disastrous consequences.

Before you jump to the conclusion that we're suggesting we jettison RCTs from clinical research, let us clarify that that is not our intention. In an ideal world, new interventions would always be carefully evaluated through rigorous RCTs before widespread adoption. But when pre-existing convictions about an untested intervention affect the population enrolled, even a well conducted RCT can provide misleading results. Without careful attention to context, extrapolating findings from such an RCT to the patient in front of us may be, well, a leap too far.

Robert W Yeh is associate professor of medicine at Harvard Medical School

Dhruv Kazi is associate director of the Richard A and Susan F Smith Center for Outcomes Research in Cardiology at Beth Israel Deaconess Medical Center, Boston, Massachusetts

Linda R Valsdottir is clinical research coordinator at the Richard A and Susan F Smith Center for Outcomes Research in Cardiology at Beth Israel Deaconess Medical Center

Brahmajee K Nallamothu is professor of medicine at the University of Michigan, Ann Arbor

Fig 1 | *Self experimentation by an emeritus neurologist. Urica ferox* is a tree nettle growing >2 m tall, endemic to New Zealand (A), with trichomes that contain a putative neurotoxin (B). During a collection expedition and subsequent sample preparation an emeritus neurologist exposed himself to the fluid in the trichomes (C) and recorded the experience in detail (D)

Feats of self experimentation

Gareth Parry and **Eric Buenz** explore the storied history of scientists using themselves as guinea pigs

Self experimentation is a rich medical tradition, leading to remarkable scientific advances but also erroneous conclusions and, sometimes, death.

Recently we explored the properties of *Urtica ferox*, a stinging nettle endemic to New Zealand (fig 1). During a collection expedition a 71 year old emeritus neurologist indulged in inadvertent, and subsequently deliberate, self experimentation. His notes of the evolving neurological manifestations after exposure provide clues to the toxin's mechanism of action that would be difficult to draw without self experimentation (box).

Self experimentation through the years

Historically, self experimentation was an important part of the scientific process, allowing medical advances that would have been hard to achieve otherwise because no sane human would agree to be a research participant and no ethical review board in its right mind would approve the experiment. For example:

- Hooke calling Newton's bluff on distorted vision through inserting a blunt needle between the eyeball and the orbit. He shamed Newton into performing the procedure and self experimenting with the retinal perception of light.[1]
- Carrion's self inoculation with blood from a wart on a patient with verruga peruana, to establish the link between these chronic skin lesions and the acute febrile illness Oroya fever, caused by the bacterium *Bartonella bacilliformis*. Carrion died for this advance.[2]
- Head's examination of cutaneous innervation and reinnervation by transecting the nerves in his own arm and studying the return of function through regeneration.[3]
- Haldane's personal explorations of hyperoxia, causing seizures resulting in vertebral fractures.[4]
- Forssman's success at self cardiac catheterisation, after lying to the nurse about who the research participant was and initially strapping her to the table so that she couldn't stop him from initiating the procedure—which ultimately resulted in a Nobel prize.[5]

Below we outline four famous instances of self experimentation that have led to notable medical advances.

Gareth J Parry, old research professor
Eric J Buenz, young research professor,
Nelson Marlborough Institute of Technology,
Nelson, New Zealand
eric.buenz@nmit.ac.nz

GJP wrote the first draft of the manuscript. EJB added the humour and eloquence. GJP replied, "Who are you calling an old curmudgeon?! I'm not old." After discussion it was agreed that GJP technically was an old curmudgeon. All authors approved the final manuscript.

Cite this as: *BMJ* 2018;363:k5006

THE EMERITUS PROFESSOR'S NOTES AFTER EXPOSURE TO THE NEUROTOXIC STINGING NETTLE

Immediate, moderately severe, burning pain at the site of penetration spread over 5-10 seconds to involve an area 1 cm in diameter. The pain began to subside within five minutes and had resolved within 60 minutes. As the pain subsided paraesthesias appeared that were intense and annoying but not truly painful, and allodynia was noted in the affected area.

Paraesthesias were constant for 18 hours and then became intermittent (particularly triggered by cold) and resolved completely by 48 hours. Numbness developed within 30 minutes of onset of paraesthesias. At nadir, complete loss of cold thermal and light touch sensation was noted, and pin prick thresholds were increased, but hyperalgesia occurred when the threshold was exceeded. At 18 hours the numbness began to recede in severity and extent, and it resolved completely by 72 hours.

Urtica ferox contains several chemicals that may account for the acute pain but not for the evolving neurological features. The observed sequence of events suggests a capsaicin-like response with initial burning pain and paraesthesias, followed by numbness that persists for several days. Like capsaicin, the unidentified molecule in the *U ferox* extract may bind to a channel in the nerve terminal, causing initial depolarisation but then preventing repolarisation.

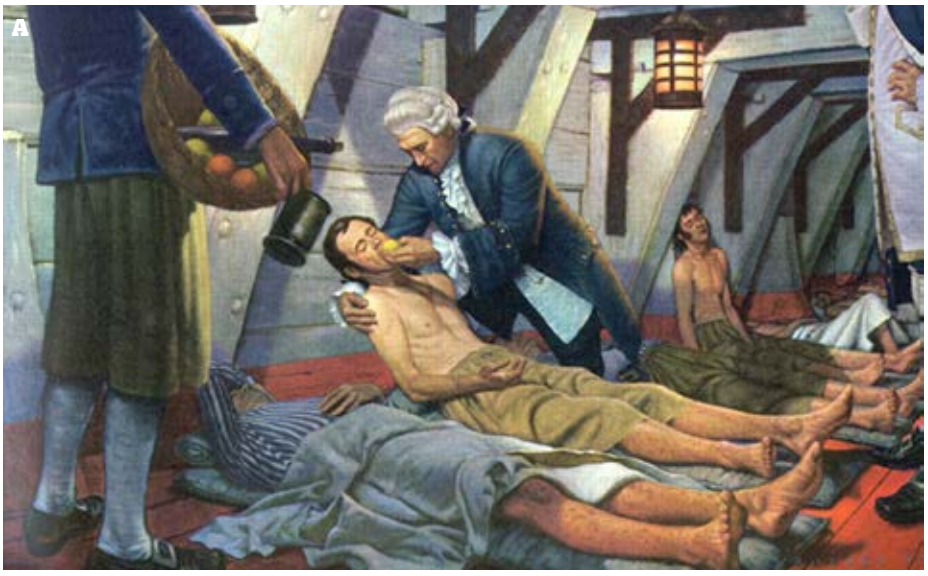


Fig 2 | *Historical examples of self experimentation.* (A) Throughout naval history scurvy was a predominant medical concern. Stark's dietary self experimentation set the foundation for supplementing sailors' diets with citrus. (B) *Conquerors of Yellow Fever* (oil on canvas), showing Cuban doctor Carlos Finlay (left) and US army surgeon Walter Reed (centre) observing as Lazear exposes Carroll to mosquitoes suspected of carrying yellow fever. (C) Through drinking a culture of bacteria isolated from patients with gastroenteritis Barry Marshall showed that *Helicobacter pylori* could serve as an aetiological agent for ulcers, and won a Nobel prize

It's unlikely that doctors and medical scientists are immune to ego clouding their clarity when deciding to self experiment

Stark's pudding and cheese diet

In naval exploration scurvy was a perpetual significant risk manifesting as fatigue, whole body pain, muscle atrophy, tooth loss, and death (fig 2A). Encouraged by John Pringle (the "father of military medicine"), the young William Stark devised dietary experiments involving weighing his food, water, and excrement to show that "a pleasant and varied diet was as healthful as simpler strict diets."[8]

In 1769 he was experimenting with a diet solely of honey puddings and Cheshire cheese when he experienced symptoms of scurvy. On Boxing Day he broke the restricted diet and consumed half a pint of blackcurrants, noting some improvement in symptoms. He considered introducing fresh fruit and vegetables but returned to his pudding diet. On 23 February 1770 he died, with symptoms of severe scurvy and other ailments. His meticulous record keeping ultimately provided key clues to the importance of an antiscorbutic substance in the diet, later identified as vitamin C.

Ffirth, Carroll, and Lazear investigate yellow fever

Historically endemic in Africa and most of the Americas, yellow fever was a major impediment to tropical exploration. In the 19th century the miasma theory of disease causation prevailed, but a postulated alternative was an infectious agent.

In the early 1800s Stubbins Ffirth, a Philadelphia medical student, devised a series of imaginative experiments to prove that yellow fever was not infectious[9]: first, he smeared vomit from an infected participant into incisions on his own arms and poured it into his eyes; second, he fried vomit and inhaled its fumes; third, he drank vomit directly from the mouth of an infected patient; and, finally, he smeared blood, saliva, and urine into cuts on his arms. Having still not contracted yellow fever he triumphantly, but erroneously, announced the results of his experiments as proof that the disease was not infectious.

In the 1890s Jesse Lazear was working in Cuba during a yellow fever epidemic (fig 2B).[10] To show that yellow fever was an infectious agent transmitted by mosquitoes he exposed James Carroll and (debatably[11]) himself to bites from *Aedes aegypti*, the mosquito suspected of being the vector. Lazear contracted the disease and died, but Carroll recovered—going on to identify the vector for yellow fever virus and facilitate public health strategies to prevent infection.

Marshall drinks a bacteria brew

Into the late 20th century peptic ulceration was treated with dietary manipulation, antacid medicines, and even surgery. The concept that bacteria could survive in the stomach and cause the condition attracted derision.

Barry Marshall, an Australian physician, identified a bacterium, *Helicobacter pylori*, in the stomach contents of several patients with peptic ulcers.[12] Having tried unsuccessfully to infect pigs with the bacterium he decided, without seeking ethics approval, to drink a culture of the patients' bacteria. Three days later he developed nausea, vomiting, and putrid breath; gastric biopsy showed marked gastritis. His wife insisted that he immediately take antibiotics, and he rapidly improved. Subsequent research established *H pylori*'s role in gastric and duodenal ulcers, and in 2005 Marshall was awarded the Nobel Prize for Physiology and Medicine (fig 2C).

Bier and Hildebrandt pioneer spinal anaesthesia

In 1898 August Bier, a German surgeon, had his assistant August Hildebrandt perform a lumbar puncture on him, attempting administration of intrathecal cocaine.[13] Spinal fluid was removed, but the experiment was aborted because of defective equipment. The next day the tables were turned, and Bier administered the cocaine to Hildebrandt. He tested the effects of the anaesthesia on Hildebrandt by thrusting a pin into his thigh to the bone, smashing a hammer against his shins, stubbing out a cigar on his skin, tugging on his testicles, and plucking his pubic hairs.

The tests all demonstrated successful anaesthesia; and ushered in the era of spinal anaesthesia.

. . . and in the future

We should acknowledge the value of self experimentation and bring it out from the shadows. However, the business world has long appreciated the risks associated with ego driven decisions[15] and "importantitis"[16]: it's unlikely that doctors and medical scientists are immune to ego clouding their clarity when deciding to self experiment.

A comprehensive review showed that only around 2% of documented efforts are made by women.[17] While 89% of self experiments ultimately support the hypothesis being tested,[17] the yellow fever and scurvy experiments show that the participant may be correct—but still dead. The sex bias in self experimentation could be the result of evolutional unimportance in males, a greater tendency towards narcissism in males,[19] or simply historical gender disparities in science and medicine.[20]

Self experimentation is not without its hazards, but it often leads to useful and timesaving medical advances, as in our investigations into stinging nettles. If an old curmudgeon would like to inject himself with a neurotoxin for the sake of science, describing the process in exquisite detail, he should be thanked and presented with an award—ideally in large type, suitable for framing.

Key opinion leaders' guide to spinning a disappointing clinical trial result

Adam Hartley and colleagues present a playbook for commenting on research with disappointing results

The onward march of medical science is marred by a great deal of shuffling on the spot, since most wonderful ideas turn out to not work. To maintain a perception of an advancing tide of discovery and excitement, drug and device companies selectively promote specialty experts who can be relied on to provide an upbeat view in all situations. Internally, companies call them "key opinion leaders."

When key opinion leaders are asked to comment on disappointing trial results in news reports or at conferences, we have observed that they seem curiously unable to recognise that the treatment doesn't work. They prefer to argue that the trial design was wrong, drawing from a set of stereotyped criticisms.

Using cardiology as an example, we have systematically analysed the excuses they provide to compose the Panellists' Playbook, an anthropological classification that will be useful not only for readers but for key opinion leaders in need of inspiration (or backbone).

Compiling the playbook

We reviewed five years of reports from the world's three largest cardiology annual scientific congresses—European Society of Cardiology, American Heart Association, and American College of Cardiology—examining remarks made to large news organisations that maintain a database: Medscape and MedPage Today.[1 2] These conferences each cover all of the many subspecialties of cardiology and have tens of thousands of attendees. Medscape and Medpage Today are widely read, providing comprehensive media coverage of these congresses.

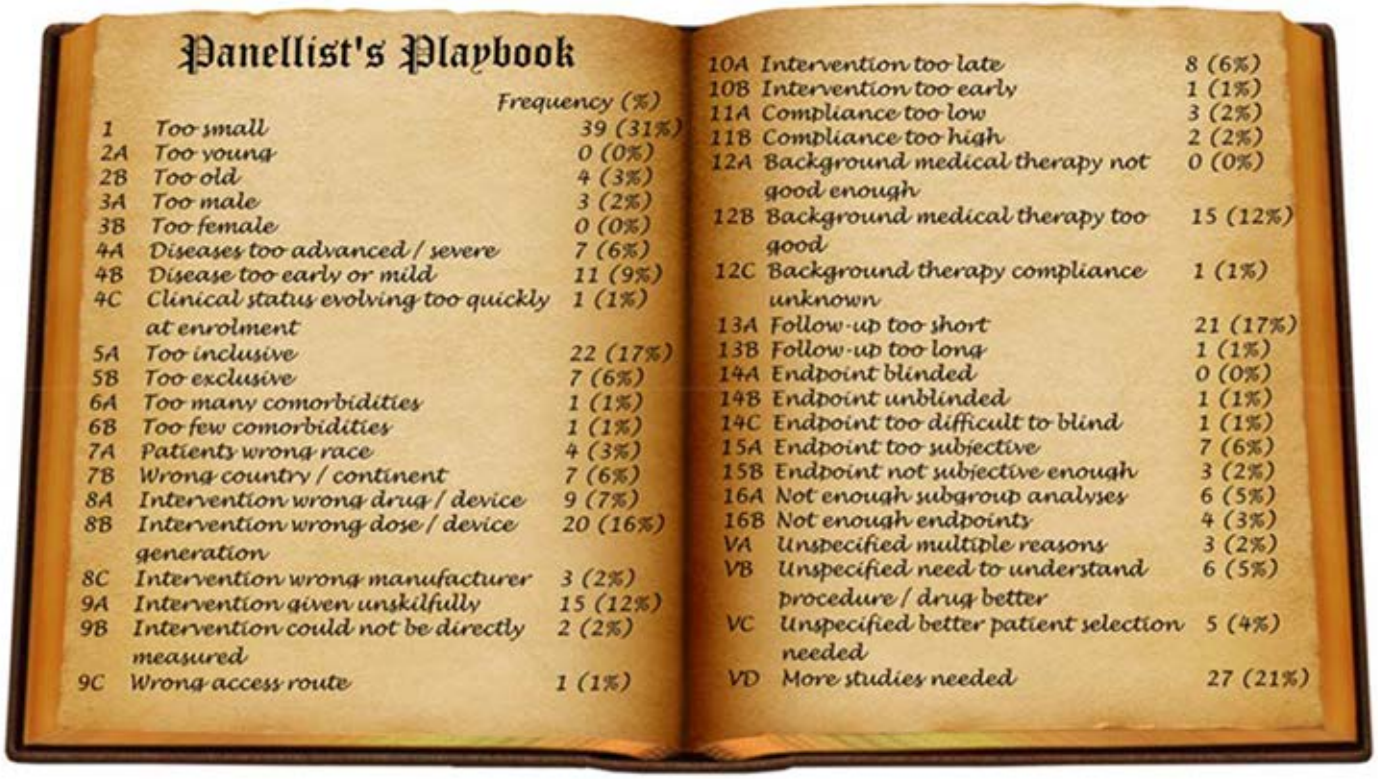
We categorised a trial as having negative findings if a primary endpoint was not met. Two of us (AH and MS) independently identified the trial reports and extracted the explanations provided to journalists by key opinion leaders. We defined key opinion leaders as specialty experts who provided quotes or statements for publication, regardless of whether they reported any current financial associations with industry. We then combined these independent analyses, resolved discrepancies, and formulated categories.

Any excuse

We found comments on 321 trials from the 15 international scientific congresses held during 2013 to 2017. Of these trials, 127 (40%) had negative results and received a total of 438 remarks from key opinion leaders. Excuses were provided for 108 (85%), with a mean of 2.5 published excuses for each trial. We defined an excuse as any explanation given for a trial's result other than the treatment not working. Sources for all 272 separate excuses are available in the supplementary table on bmj.com.

We identified 17 themes for the published excuses. Most themes could have more than one possible type of excuse—for example, the theme "age" could have two excuse types, "patients too young" or "patients too old." We used the 17 themes to create 40 theoretically possible types of excuse and compiled the Panellists' Playbook, a categorisation that is useful both for us as readers and for current and future key opinion leaders.

Of the 40 theoretically possible types of excuse, 36 were used (figure). The most common excuse was "sample size too small," used for 39 trials, 31% of those with negative results. This was followed by: "more studies



Panellists' Playbook. An efficient standardised framework for busy key opinion leaders asked to comment on trials with negative results

are needed," "study population too inclusive," and "follow-up too short."

Although key opinion leaders fell over themselves to criticise trials for being inadequately sized, in only one case did the person quoted provide a calculation for the correct sample size.[3] It is not clear whether in the other 38 of 39 (97%) cases the experts were reluctant to divulge the fruits of their 30 seconds of calculation or that the excuse was simply the first thing that came into their head.

We considered some excuses vacuous and categorised them under excuse theme "V" to avoid dignifying them with a numeral. For example, the uninformative, "more studies are needed," offered for one fifth of negative trials, suggests to us that the key opinion leader simply does not like the result and wants another throw of the die.

Practical applications

The Panellists' Playbook provides a comprehensive approach to summarising, or even generating, cheery key opinion leader remarks in the face of disappointing results. With the help of the playbook, no intervention is too ineffective for an excuse.

Even if they lack ability to think deeply about the trial, opinion leaders can pluck items from the playbook for an effortless veneer of insight

Medical journalists often contact key opinion leaders for comment with short deadlines. Key opinion leaders generally feel responsible for helping exaggerate the perceived efficacy of the specialty's interventions; indeed, their cohort is the result of decades of relentless selection for this predisposition rather than a track record of prompt, clear, and dispassionate analysis of scientific data. Even if they lack the time, inclination, or ability to think deeply about the trial, they can pluck items from the Panellist's Playbook to provide an effortless veneer of insight.

In principle, some of the excuses reported could be correct. An intervention may theoretically be neutral or harmful in the patient group tested but beneficial in another. However, industry does not spend tens of millions of dollars without carefully choosing the characteristics of participants. If despite all that expertise, expense, and effort an intervention did not work, as individual physicians we should not delude ourselves that we can somehow select patients better. It is easy to slip into the belief that personal clinical practice is better than trials, but this is only because we lack randomisation, allocation concealment, and (most importantly) blinding of outcome assessment.

Researchers planning a trial could also peruse the Panellists' Playbook for its exhaustive list of everything that might be said to undermine the results if negative. For example, they risk being criticised for having background therapy that adheres too well to guidelines. How to avoid this is less clear, since physicians and patients participating in trials tend to be more interested in prescribing and adhering to good medical therapy than are the general population. The only fix would be to seek out doctors or patients who are not interested in good care.

The extent to which the Panellists' Playbook will be used and whether it applies to other medical specialties is not yet clear. However, one thing is certain: more research is needed.

Adam Hartley, academic clinical fellow, National Heart and Lung Institute, Imperial College London
Mit Shah, registrar, Hammersmith Hospital, Imperial College NHS Healthcare Trust
Alexandra N Nowbar, clinical research fellow
Christopher Rajkumar, academic clinical fellow
James P Howard, clinical research fellow
Darrel P Francis, professor, National Heart and Lung Institute, Imperial College London
darrel@drfrancis.org.

Cite this as: *BMJ* 2018;363:k5207

Eyes down for appraisal bingo

Rachel Thomas and Kirsty Mozolowski present a game that's fun for all the faculty

Revalidation and regular appraisal are now an essential part of clinical life. Having known no different, current junior doctors accept regular evaluation as a compulsory part of training but more mature colleagues have found documenting their everyday practice to be challenging [personal correspondence to authors].

The authors have found that this reluctance is demonstrated in the limited language used during written feedback. Having more than 20 years combined of Annual Reviews of Competence Progress, we analysed the idioms most regularly used in our appraisals. There was a recognised homogeneity of phrases and, after further multidisciplinary analysis, we found that this finding transcended specialties and was universal in all disciplines at postgraduate level.

There was a recognised homogeneity of phrases and, after further multidisciplinary analysis, we found that this finding transcended specialties

This observation led us to look at new ways to increase engagement with the appraisal process. Previous studies have shown that game play can improve interaction with learning.[1 2] In addition, given that laughter offers health benefits across many domains,[3-7] we suggest that this multiplayer bingo game will bring much needed humour to the appraisal process.

Rachel AB Thomas, surgical registrar

Kirsty Mozolowski, surgical registrar, Department of Colorectal Surgery, Western General Hospital, Edinburgh

Cite this as: *BMJ* 2018;363:k5132

BINGO

Build good rapport	Communication skills	Reflected on...	Developed skills	On-going practice
Demonstrated areas for improvement	Appreciated complexity	Member of the team	Skills to build on	Well-liked
Comprehensive reflection	Positive opportunity		Proactive approach	Productive event
Highlighted areas of achievement	Outpatient skills	Diligent	On-going aspiration	Educational needs
Questions current practice	Current approach	Patient centred	Structured approach	Evidence base

BINGO

Research and audit	Sound knowledge	Questions current practice	Developed skills	On-going practice
Outpatient skills	Demonstrated insight	Reliability	Abreast of current research	Sound clinical skills
Ensures engagement	Member of the team		Progressing at a good pace	Well-liked
Reflected on...	Highlighted areas for improvement	Patient centred	Evidence base	Thorough
Demonstrates responsibility	Thoughtful decision making	Competent technical skills	Structured approach	Feedback

BINGO

Educational needs	Organisation	Sound knowledge	On-going aspirations	Evidence base
Outpatient skills	Questions current practice	Structured approach	Ensures engagement	Patient centred
Comprehensive reflection	Sound clinical skills		Current approach	Thorough assessment
Demonstrates responsibility	Highlighted areas of achievement	Member of the team	Thoughtful decision making	Recognises success
Positive opportunity	Leadership and management	Reliable	Meticulous record-taking	Appreciated complexity

BINGO

Builds good rapport	Ensures engagement	Displays leadership qualities	Competent technical skills	Patient centred
Demonstrates responsibility	Skills to build on	Member of the team	Appreciated complexity	Proactive approach
Acknowledged insight	Positive opportunity		Developed skills	Consistent note-taking
Highlighted areas of achievement	Feedback	Diligent	Conscientious	Considered Educational needs
Demonstrated areas for improvement	Current approach	Well-liked	Technical dexterity	Proficient growth of...

DUNCAN THOMAS

Simply cut out the cards and share with three friends. First one to match all the phrases during their appraisal wins!