Critical appraisal

Gary Collins
EQUATOR Network, Centre for Statistics in Medicine
NDORMS, University of Oxford

EQUATOR Network – OUCAGS training course
25 October 2014

Objectives of this session

- To understand the importance of critical appraisal
- To understand some of the key issues to consider when appraising a research study

Why do we need to critically appraise?

- **It is an integral part of evidence-based medicine**

- **Because**
  - large swathes of published reports of clinical studies are
    - poorly conducted (poor methodology)
    - poorly reported
    - authors naturally place lots of 'spin' on their results

- **Peer review doesn't infer study quality**

PLOS | MEDICINE

Browse | For Authors | About Us | Search | advanced search

OPEN ACCESS

ESSAY

1,151,608 VIEWS | 1,413 CITATIONS | 13,400 SAVES | 10,475 SHARES

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding

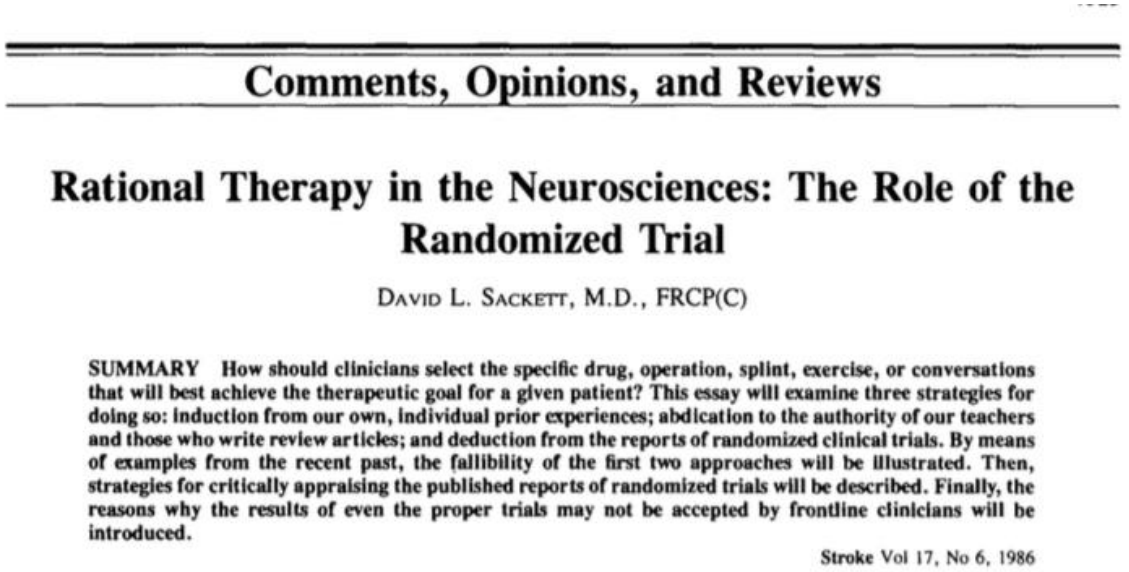
factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the

- **"There are only a handful of ways to do a study properly but a thousand ways to do it wrong."**



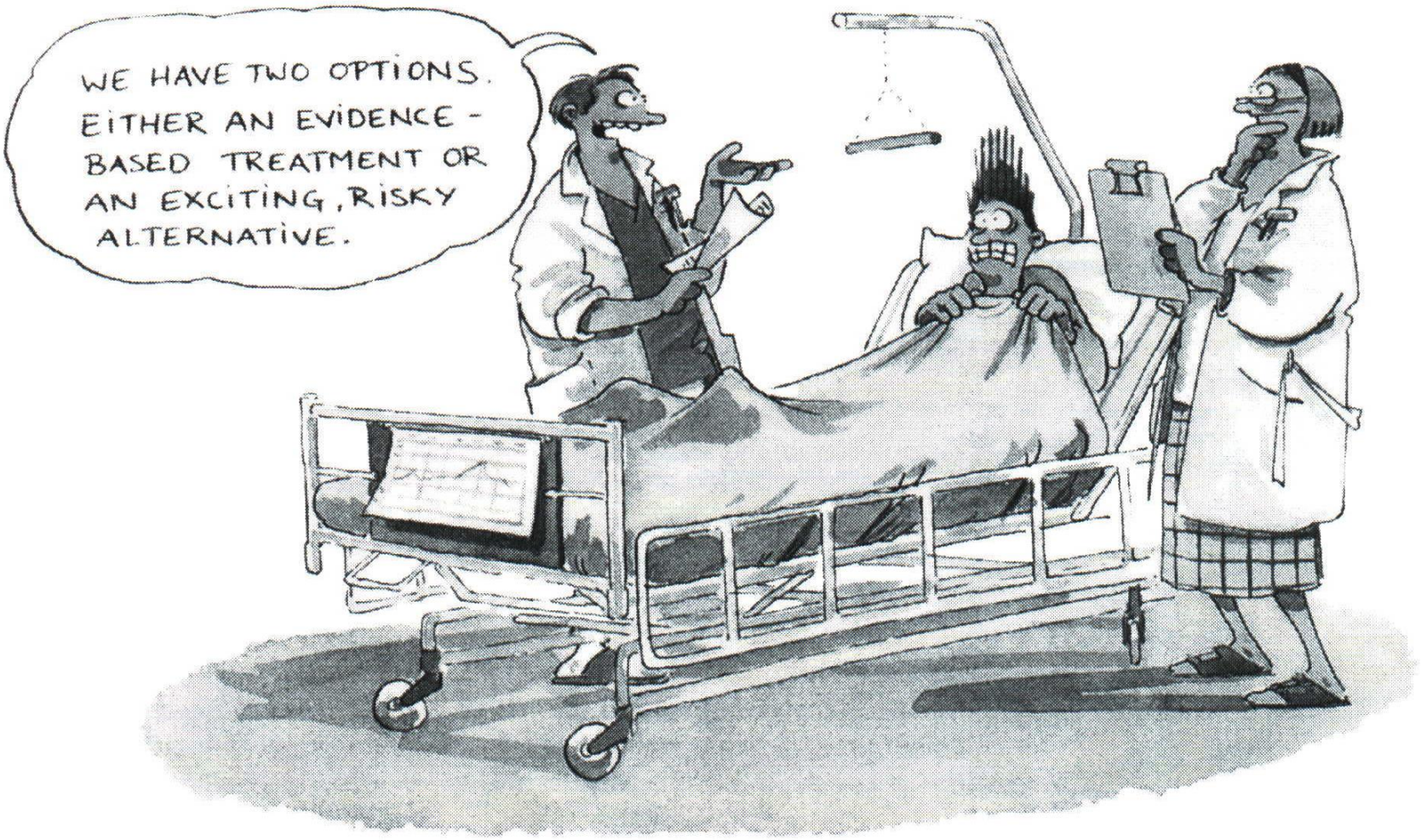
Comments, Opinions, and Reviews

Rational Therapy in the Neurosciences: The Role of the Randomized Trial

DAVID L. SACKETT, M.D., FRCP(C)

SUMMARY How should clinicians select the specific drug, operation, splint, exercise, or conversations that will best achieve the therapeutic goal for a given patient? This essay will examine three strategies for doing so: induction from our own, individual prior experiences; abdication to the authority of our teachers and those who write review articles; and deduction from the reports of randomized clinical trials. By means of examples from the recent past, the fallibility of the first two approaches will be illustrated. Then, strategies for critically appraising the published reports of randomized trials will be described. Finally, the reasons why the results of even the proper trials may not be accepted by frontline clinicians will be introduced.

Stroke Vol 17, No 6, 1986


WE HAVE TWO OPTIONS. EITHER AN EVIDENCE-BASED TREATMENT OR AN EXCITING, RISKY ALTERNATIVE.

Considerations when appraising a published study

- **Does the study address a clearly focused question?**
  - PICO
- **Did the study use valid methods to address this question?**
- **What are the results?**
- **Are the results useful?**

Study aim and design

- **Does the study have a clear aim?**
  - do these match any pre-specified outcomes?
- **Has the study provided sufficient information on what is known on the topic?**
- **Is the study design appropriate for the research question?**
- **Is there a protocol available**
  - with a protocol we can check what the investigators planned against what they actually did

Sample size

- **Was there a sample size calculation?**
  - if an RCT has no sample size calculation we have no idea if the achieved sample size is sufficient. We also don't know what treatment effect they were attempting to seek.

- **Is it clear what assumptions the sample size calculation was based on?**
  - is there sufficient information to allow replication?

- **Has the study achieved the planned sample size?**
  - have the authors mentioned this in the article?

Randomization

- **Goal of randomization is to create comparable (balanced) groups with respect to known and unknown prognostic factors to allow an unbiased comparison**

- **How were the participants randomised to the intervention?**
  - Have they mentioned any prognostic factors and whether this has been included in the randomization procedure?

- **Do the randomised groups look similar at baseline (usually Table 1)?**
  - Be wary of those that do statistical tests for differences in baseline characteristics!

Continuous pralidoxime infusion versus repeated bolus injection to treat organophosphorus pesticide poisoning: a randomised controlled trial

Kirti S Pawar, Ramesh R Bhoite, Chandrakant P Pillay, Sujata C Chavan, Dhananjay S Malshikare, Saraswati G Garad

Lancet 2006; 368: 2136–41

See Comment page 2110

Giriraj Hospital and Intensive

Summary

Background The role of oximes for the treatment of organophosphorus pesticide poisoning has not been conclusively established. We aimed to assess the effectiveness of a constant pralidoxime infusion compared with repeated bolus doses to treat patients with moderately severe poisoning from organophosphorus pesticides.

Pawar et al, Lancet 2006

	Control group (n=100)	Study group (n=100)
Men	52	57
Oral route of consumption	94	98
Ingestion of diethyl pesticides*	41	23
Ingestion of dimethyl pesticides†	59	77
Intubated during resuscitation	69	63
Median (IQR) age (years)	29 (22–35)	28 (22–33)
Median (IQR) time between ingestion and admission (min)	112·5 (60·0–150·0)	120·0 (90·0–142·5)
Median (IQR) quantity of poison consumed (mL)	15 (10–20)	15 (15–20)
Median (IQR) Glasgow coma score	10 (8–12)	10 (10–12)
Median (IQR) serum butyrylcholinesterase activity (IU/L)‡	808·0(534·8–911·0)§	866·0 (751·8–939·0)¶
Mean (SD) pulse (bpm)	50·5 (7·85)	50·8 (9·35)
Mean (SD) systolic blood pressure (mm Hg)	110·2 (14·72)	116·2 (14·79)
Mean (SD) diastolic blood pressure (mm Hg)	70·3 (10·78)	74·5 (10·77)

Data are count (percentage) for categorical variables; mean (standard deviation) for normally distributed continuous variables, median (interquartile range—25th to 75th percentile) for other continuous variables. *Chlorpyrifos (40 control, 22 study), quinalphos (1, 1). †Dimethoate (45, 65), monocrotophos (5, 6), methyl parathion (6, 5), malathion (2, 2), fenitrothion (1 control). ‡Normal range: 2710–11 510 IU/L. §n=93. ¶n=83.

Table 1: **Baseline demographic and clinical characteristics at admission**

Problems with imbalance

- **Potential loss of credibility**
  - raises questions on the randomization procedure (and concealment)
- **Analysis requires adjustment**
  - possible problems, depends on magnitude of imbalance
  - not always satisfactory
- **Complicates the interpretation of the results**
  - leaves an air of uncertainty

COMET Trial, Lancet 2001

Effect of low-dose mobile versus traditional epidural techniques on mode of delivery: a randomised controlled trial

*Comparative Obstetric Mobile Epidural Trial (COMET) Study Group UK**

Lancet 2001; **358:** 19–23

equator network

COMET Trial, Lancet 2001

Characteristic	Traditional epidural (n=388)	Combined spinal epidural (n=335)	Low-dose infusion (n=331)
Age			
⩽19	1 (0·3%)	69 (21%)	73 (22%)
20–24	12 (3%)	110 (33%)	100 (30%)
25–29	131 (34%)	111 (33%)	117 (35%)
30–34	208 (54%)	17 (5%)	16 (5%)
⩾35	36 (9%)	28 (8%)	25 (8%)
Ethnic group			
White	322 (83%)	301 (90%)	298 (90%)
Asian	43 (11%)	24 (7%)	26 (8%)
Other	23 (6%)	10 (3%)	7 (2%)

trials experts, separate from the study team, to exclude bias. The programme included minimisation to balance maternal age and ethnicity but unfortunately an error in the programme that assigned women according to age resulted in a severe imbalance in distribution between the groups for age and ethnic origin (table 1). Both we and the funding body immediately commissioned two independent groups of national experts in clinical trials analysis who recommended repeat recruitment of a further complete sample. This second sample was

Outcomes

- **Do the outcomes (primary and secondary) match what was recorded in the trial registry and the protocol?**

- **All outcomes should be clearly described**
  - including timing of measurement

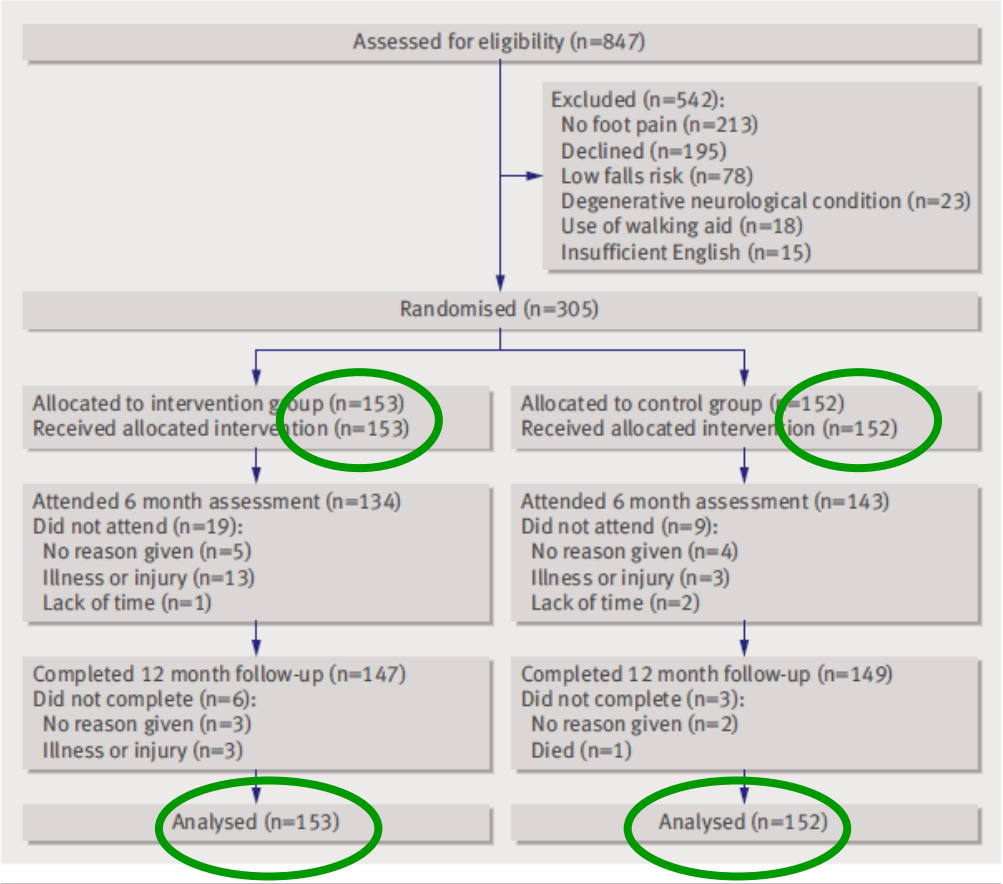
- **Are the conclusions based on the primary outcome?**
  - not uncommon to see authors switch primary and secondary outcomes (or bring into unspecified outcomes)

Missing data

- **Does the study mention anything about missing data? Have all participants been accounted for?**
  - Rarely do studies have complete data on all participants
  - What did they do?
    - Often unclear (i.e. they don't mention anything)
    - More than often those within any missing data will have been omitted ('ignored') from the analysis (so called 'complete-case' analysis)
    - Is there something special about those with missing data?
  - If the data are based on a survey, what is the response rate?
    - how low a response rate are you willing to accept, before generalisability becomes an issue?

- **For an RCT have they done an ITT**
  - Excluding those with missing data violates ITT
    - Further problems if the balance of exclusions differ across the groups
      - Implication for sample size calculations

Flow of participants, analysis population



Flow of participants through study

Spink MJ, et al. Effectiveness of a multifaceted podiatry intervention to prevent falls in community dwelling older people with disabling foot pain: randomised controlled trial. *BMJ* 2011;342:d3411.

How are the results presented?

- **How was the primary outcome analysed?**
  - Did they actually compare treatments (between arm comparison)?
  - Or did they report only a within-arm comparison?

THERAPEUTICS

BJD British Journal of Dermatology

A cosmetic 'anti-ageing' product improves photoaged skin: a double-blind, randomized controlled trial

R.E.B. Watson, S. Ogden, L.F. Cotterell, J.J. Bowden, J.Y. Bastrilles, S.P. Long* and C.E.M. Griffiths

Dermatological Sciences Research Group, School of Translational Medicine, Faculty of Medical and Human Sciences, The University of Manchester, Oxford Road, Manchester M13 9PT, U.K.

**Alliance Boots Ltd, Nottingham NG2 3AA, U.K.*

Note added after online publication:

Since the publication of this article online on 28 April 2009, the authors wish to recognize the following changes to the article:

Title: Effects of a cosmetic 'anti-ageing' product on photoaged skin

Conflicts of Interest: This study was funded by Alliance Boots Ltd.

Anti-ageing trial

- **60 patients randomised into two arms**
- **They reported significance tests within each arm**
  - Active treatment group; P=0.013
  - Control group; P=0.11

  } No effect size

  - Interpreted as 'an over-the-counter cosmetic anti-ageing product resulted in significant clinical improvement in facial wrinkles'
- **However, the two arms were NOT compared directly**
- **Therefore their conclusion is incorrect**

BJD British Journal of Dermatology

A cosmetic 'anti-ageing' product improves photoaged skin: a double-blind, randomized controlled trial

R.E.B. Watson, S. Ogden, L.F. Cotterell, J.J. Bowden, J.Y. Bastrilles, S.P. Long* and C.E.M. Griffiths

Dermatological Sciences Research Group, School of Translational Medicine, Faculty of Medical and Human Sciences, The University of Manchester, Oxford Road, Manchester M13 9PT, U.K.

**Alliance Boots Ltd, Nottingham NG2 3AA, U.K.*

Note added after online publication:

Since the publication of this article online on 28 April 2009, the authors wish to recognize the following changes to the article:

Title: Effects of a cosmetic 'anti-ageing' product on photoaged skin

Conflicts of Interest: This study was funded by Alliance Boots Ltd.

Summary

Methods For the patch test, commercially available test product and its vehicle were applied occluded for 12-days to photoaged forearm skin (n = 10) prior to biopsy and immunohistochemical assessment of fibrillin-1; all-*trans* retinoic acid (RA) was used as a positive control. Sixty photoaged subjects were recruited to the RCT (test product, n = 30 vs. vehicle, n = 30; once daily for 6-months; face & hands) with clinical assessments performed at recruitment and following 1-, 3- & 6-months of use. Twenty-eight subjects had skin biopsies (dorsal wrist) at baseline and at 6 months of treatment for immunohistochemical assessment of fibrillin-1 (test product, n = 15; vehicle, n = 13). All subjects received test product for a further 6-months. Final clinical assessments were performed at the end of this open period; 27 subjects received test product for 12-months.

Results In the 12-day patch test assay, we observed significant immunohistological deposition of fibrillin-1 in skin treated by test product and RA as compared to untreated baseline (P = 0·005 and 0·015 respectively). In the clinical RCT, at 6 months, compared to baseline assessment, 43% of subjects on test product had an improvement in facial wrinkles (P = 0·013), whereas only 22% of subjects using vehicle had clinical improvement (P = ns). Between group comparison of test product and vehicle was non-significant (P = 0·10). After 12 months, there was a significant benefit of test product over that projected for vehicle (70% vs. 33% of subjects improving; combined Wilcoxon rank tests, P = 0·026). There was significant deposition of fibrillin-1 in skin treated for 6 months with test product (mean ± SE; vehicle, 1·84 ± 0·23; test product, 2·57 ± 0·19; P = 0·019).

Conclusion An over-the-counter cosmetic 'anti-ageing' product demonstrated clear benefit over vehicle in fibrillin-1 deposition over a 6-month trial period. There was a corresponding but non-significant trend towards clinical improvement in facial wrinkles. Clinical improvements in the treated group were increased after a further

Correspondence
R.E.B. Watson.
E-mail: rachel.watson@manchester.ac.uk

Accepted for publication
3 April 2009

Key words
fibrillin-1, patch-test assay, randomized controlled trial, wrinkles

Conflicts of interest
S.P.L. is employed by Alliance Boots Ltd., the manufacturer of the commercially available preparation tested in this study.

DOI 10.1111/j.1365-2133.2009.09216.x

Summary

Background Very few over-the-counter cosmetic 'anti-ageing' products have been subjected to a rigorous double-blind, vehicle-controlled trial of efficacy. Previously we have shown that application of a cosmetic 'anti-ageing' product to photoaged skin under occlusion for 12 days can stimulate the deposition of fibrillin-1. This observation infers potential to repair and perhaps clinically improve photoaged skin.

Objective We examined another similar over-the-counter cosmetic 'anti-ageing' product using both the patch test assay and a 6-month double-blind, randomized controlled trial (RCT), with a further 6-month open phase to assess clinical efficacy in photoaged skin.

Methods For the patch test, a commercially available test product and its vehicle were applied occluded for 12 days to photoaged forearm skin ($n = 10$) prior to biopsy and immunohistochemical assessment of fibrillin-1; all-*trans* retinoic acid (RA) was used as a positive control. Sixty photoaged subjects were recruited to the RCT (test product, $n = 30$ vs. vehicle, $n = 30$; once daily for 6 months, face and hands) with clinical assessments performed at recruitment and following 1, 3 and 6 months of use. Twenty-eight volunteers had skin biopsies (dorsal wrist) at baseline and at 6 months treatment for immunohistochemical assessment of fibrillin-1 (test product, $n = 15$; vehicle, $n = 13$). All volunteers received the test product for a further 6 months. Final clinical assessments were performed at the end of this open period.

Results In the 12-day patch test assay, we observed significant immunohistological deposition of fibrillin-1 in skin treated with the test product and RA compared with the untreated baseline ($P = 0·005$ and $0·015$, respectively). In the clinical RCT, at 6 months, the test product produced statistically significant improvement in facial wrinkles as compared to baseline assessment ($P = 0·013$), whereas vehicle-treated skin was not significantly improved ($P = 0·11$). After 12 months, there was a significant benefit of the test product over that projected for the vehicle (70% vs. 33% of subjects improving; combined Wilcoxon rank tests, $P = 0·026$). There was significant deposition of fibrillin-1 in skin treated for 6 months with the test product [(mean ± SE) vehicle $1·84 \pm 0·23$; test product $2·57 \pm 0·19$; ANCOVA $P = 0·019$).

Conclusions In a double-blind RCT, an over-the-counter cosmetic 'anti-ageing' product resulted in significant clinical improvement in facial wrinkles, which was associated with fibrillin-1 deposition in treated skin. This study demonstrates that a cosmetic product can produce significant improvement in the appearance of wrinkles and further supports the use of fibrillin-1 as a robust biomarker for the repair of photoaged dermis.

All tissues, regardless of body site, are subject to intrinsic ageing, the result of the passage of time. Few clinically apparent changes occur in intrinsically aged skin until the individual is over 70 years of age at which point fine wrinkles become apparent.[1] Skin, more than any other organ, is also subject to environmental influences which can lead to extrinsic ageing. One such environmental factor is chronic exposure to sunlight which results in phenotypic changes termed photoageing—inevitably a combination of intrinsic ageing and photodamage. By comparison with intrinsic ageing, photoaged skin is rough, dyspigmented and exhibits both fine and deep wrinkles.[2,3] Histological examination of intrinsically aged skin reveals atrophy of the dermal extracellular matrix (ECM), with reduced levels of collagen and elastin.[4] Photoaged skin has a different ECM morphology with solar elastosis—the deposition of dystrophic elastic fibres in the dermis—being a prominent histological feature.[5] Photoaged dermis contains significantly reduced levels of collagen types I and III,[6] fewer anchoring

fibrils at the dermal–epidermal junction (DEJ; collagen VII)[7] and loss of the fibrillin-rich microfibrillar architecture in the papillary dermis.[8] These remodelled ageing phenotypes are thought in part to be due to increased cutaneous expression of matrix metalloproteinases (MMPs).[9–11]

Topical retinoids are used as the clinical, evidence-based 'gold standard' for the treatment of photoaged skin.[12] Numerous studies have shown the reparative effects of topical application of all-*trans* retinoic acid (RA), which includes the partial restoration of collagens I, III[13] and VII[14] and restoration of the fibrillin-rich microfibrillar network.[15] These ECM changes, together with reduced MMP expression may in part explain the clinical improvement of photoaged skin produced by topical retinoids.[16–18] We showed previously, in a 12-day occluded patch test assay, that a specific cosmetic 'anti-ageing' product also has the ability to stimulate the accumulation of fibrillin-1.[19]

Although prescription retinoids can affect these significant clinical and histological changes in photoaged skin there is scant evidence that any of the plethora of cosmetic 'anti-ageing' products can produce similar effects. We firstly examined whether another, similar cosmetic 'anti-ageing' product can induce accumulation of fibrillin-1 in photoaged skin using the patch test protocol. We then investigated the same product in a rigorous double-blind, randomized controlled trial (RCT) to ascertain whether or not its use results in a clinically detectable benefit.

Methods

Test products

A commercially available product provided by Alliance Boots Ltd (No7 Protect & Perfect Intense Beauty Serum™; Alliance Boots Ltd, Nottingham, UK) was investigated in these studies, together with a vehicle formulation. The product is a water in silicone emulsion with glycerine and other emollients and a complex of 'anti-ageing' ingredients comprising natural extracts and peptides: sodium ascorbyl phosphate, *Panax ginseng*, *Morus alba*, *Lupinus alba*, tocopherol, palmitoyl oligopeptide, palmitoyl tetrapeptide-7, *Medicago sativa* and retinyl palmitate. The vehicle was of identical composition, but without the complex of 'anti-ageing' ingredients.

In vivo patch test study

Ten healthy but photoaged volunteers were recruited (four men, six women; age range 61–76 years) and subjected to an extended 12-day patch test assay.[19] Test substances (vehicle and test product 20 μL) were applied separately to the extensor photoaged aspect of the forearm under standard 6-mm diameter Finn chambers (Scanpore, Tuulsula, Finland). In addition, an area was left untreated but was occluded to provide a baseline control sample. Test products were applied to clean skin on days 1, 4 and 8 of the assay. RA (0·025%; Retin-A® cream; Janssen-Cilag Ltd, Beerse, Belgium; 20 μL) was applied to an untreated site on day 8 and left *in situ* for 4 days to avoid potential complications of irritancy caused by extended occlusion. On day 12, the Finn chambers were removed and 3-mm punch biopsies were obtained under 1% lignocaine local anaesthesia from each test site. Biopsied tissue was embedded in optimal cutting temperature compound (Tissue-Tek®; Miles Laboratories, Elkhart, IN, U.S.A.), snap frozen in liquid nitrogen and stored at −70 °C prior to immunohistochemical analyses. The North Manchester Local Research Ethics Committee approved the study and all subjects gave written, informed consent.

Slide preparation

Frozen sections were prepared at a thickness of 10 μm (OTF cryostat; Bright Instruments Ltd, Cambridge, U.K.) and mounted onto gelatin-coated slides prior to histological analysis.

Immunohistochemistry

Immunohistochemistry was performed as previously described[19] to identify a panel of ECM molecules or remodelling enzymes in frozen sections from the 12-day patch test assay and from the RCT. Primary antibodies were applied overnight at 4 °C. These were: mouse antihuman fibrillin-1 (clone 11C1.3; Neomarkers, Union City, CA, U.S.A.) diluted 1 : 100; rat antihuman procollagen-1 (pCI) (clone M-58; Chemicon International Inc., Temecula, CA, U.S.A.) diluted 1 : 1000; or mouse antihuman MMP-1 (Oncogene Research Products, Boston, MA, U.S.A.) diluted 1 : 100. Negative controls were by incubation of isotype sera at the appropriate concentration or omission of the primary antibody. Sections were washed in TBS prior to incubation with the appropriate biotinylated secondary antibody for 30 min. Antibody staining was visualized using a well-characterized immunoperoxidase reaction (Vecta-Stain® Elite ABC system; Vector Laboratories, Burlingame, CA, U.S.A.) utilizing Vector SG® as chromogen. Following light counterstaining with nuclear fast red, sections were serially dehydrated and permanently mounted. Stained sections were randomized, blinded and examined on a Nikon OPTIPHOT microscope (Tokyo, Japan). The degree of immunostaining for fibrillin-1 and pCI was assessed as previously described.[8,15,19] In brief, a five-point semiquantitative scale was used where 0 = no staining and 4 = maximal staining within the experiment. The numbers of epidermal keratinocytes positive for MMP-1 were quantified per high-power field (hpf; ×400). Four sections (including control) were examined per subject, per site, per treatment and the average score calculated.

Randomized controlled trial

Sixty healthy but photoaged volunteers were recruited to this study (11 men, 49 women; age range 45–80 years). All test products were supplied in identically packaged, coded containers so that the investigators and subjects were unaware as to the treatment. Subjects were randomly allocated to self-treatment with either the vehicle formulation or the test product as described by a randomization programme (StatsDirect

Ltd, Altrincham, U.K.) and instructed on the use of their allotted cream—daily evening application to the entire face and dorsa of the hands, including the wrists and extensor forearm, for 6 months. Clinical assessments of the skin of the face and dorsal hands were performed for all participants at baseline and following 1, 3 and 6 months of product use. The following four parameters were assessed at each visit: fine lines and wrinkles, dyspigmentation, overall clinical grade of photoageing and tactile roughness. The degree of fine lines and wrinkles, dyspigmentation and the overall level of photoageing were scored according to the well-characterized Griffiths photonumeric scale for photoaged skin.[20] The scale ranges from 0 to 8, where 0 represents no evidence of photoageing and 8 represents the most severe photoageing. Pigment was assessed on a similar 0–8 scale, where 0 denotes a uniform coloration of the skin with absence of photoageing-related colour change and 8 represents severe dyspigmentation. Likewise, tactile roughness was scored on the treated areas from 0 to 8, where 0 represents totally smooth skin with no rough patches and 8 represents very roughened skin.

In addition, 28 subjects provided 3-mm skin biopsies from the dorsal wrist at the beginning and end of the 6-month study period (vehicle formulation, $n = 13$; test product, $n = 15$). These biopsies were evaluated for the expression of fibrillin-1 in the papillary dermis, as previously described.[19] All the subjects were monitored for the occurrence of serious adverse events up to, and including, 28 days after their involvement with this study. The Salford and Trafford Local Research Ethics Committee approved the study and all the subjects gave written, informed consent.

Statistical analyses

In vivo patch test study

Differences in the amount of fibrillin-1 immunostaining produced by the vehicle and test product were assessed for significance using the repeated measures analysis of variance (ANOVA). Results were considered significant if $P < 0{\cdot}05$ (95% confidence level) and were calculated using SPSS+ v 11.5 software (SPSS Inc., Chicago, IL, U.S.A.).

Randomized controlled trial

Clinical assessment. Analysis of covariance (ANCOVA), using the baseline as covariate, was used to assess the 6-month data. Linear regression analysis was used to extrapolate the vehicle response to 12 months, thus allowing direct comparison with the test product, validity confirmed by the Monte Carlo simulation.[21] As all volunteers used the test product in the final 6 months, the 12-month clinical assessment data were analysed using a combination of Wilcoxon's matched pairs signed rank and rank sum tests, to give an overall P-value.

In-use biopsy samples. ANCOVA was performed, using the baseline as covariate to assess changes in the deposition of fibrillin-1 in the papillary dermis, with significance taken at the 95% confidence level (SAS 9.1; SAS Institute Inc., Cary, NC, U.S.A.).

Results

In vivo patch test study

RA (the clinical 'gold standard') produced a significant deposition of fibrillin-1 in the papillary dermis compared with that observed at baseline ($P = 0{\cdot}015$). Application of the vehicle, following the 12-day patch test assay, produced little effect on fibrillin-1 deposition ($P > 0{\cdot}05$). However, application of the test product resulted in a significant deposition of fibrillin-1, the accumulation being at a similar level to that observed using RA (mean ± SE) (baseline 1·27 ± 0·11; vehicle formulation 1·70 ± 0·17; test product 2·64 ± 0·22, $P = 0{\cdot}005$; RA 2·51 ± 0·28, $P = 0{\cdot}015$; Fig. 1e). As in previous studies,[15] treatment with RA had little effect on deposition of pCI or on the expression of MMP-1 in the epidermis (Table 1).

Randomized controlled trial

Clinical assessment

At 6 months, the test product produced statistically significant improvement in facial wrinkles as compared to baseline assessment ($P = 0{\cdot}013$), whereas vehicle-treated skin was not significantly improved ($P = 0{\cdot}11$). 43% of the subjects who had received the test product showed an improvement in facial wrinkles compared with the baseline assessment, whereas only 22% of the subjects receiving the vehicle showed improvement compared with baseline (Fig. 2). Use of the test product produced a clinically significant improvement in facial wrinkles after 12 months of use, with a statistically significant between-groups benefit of test product vs. the vehicle (test product, 70% of subjects improving compared with vehicle, 33% improving; combined Wilcoxon rank tests, $P = 0{\cdot}026$) (Fig. 2). No benefits of the test product were seen for improvement in mottled dyspigmentation. Use of either formulation produced an improvement in skin texture over that recorded at baseline (vehicle, $P = 0{\cdot}001$; test product, $P = 0{\cdot}001$), but the test formulation did not perform significantly better than the vehicle (data not shown; $P = 0{\cdot}72$).

In-use biopsy samples

The test product produced a significant accumulation of fibrillin-1 in the papillary dermis of photoaged skin at 6 months compared with the vehicle (mean ± SE) (vehicle formulation 1·84 ± 0·23; test product 2·57 ± 0·19; ANCOVA, $P = 0{\cdot}019$; Fig. 3).

Discussion

We show here, for the first time, that a commercially available over-the-counter 'anti-ageing' product improves the

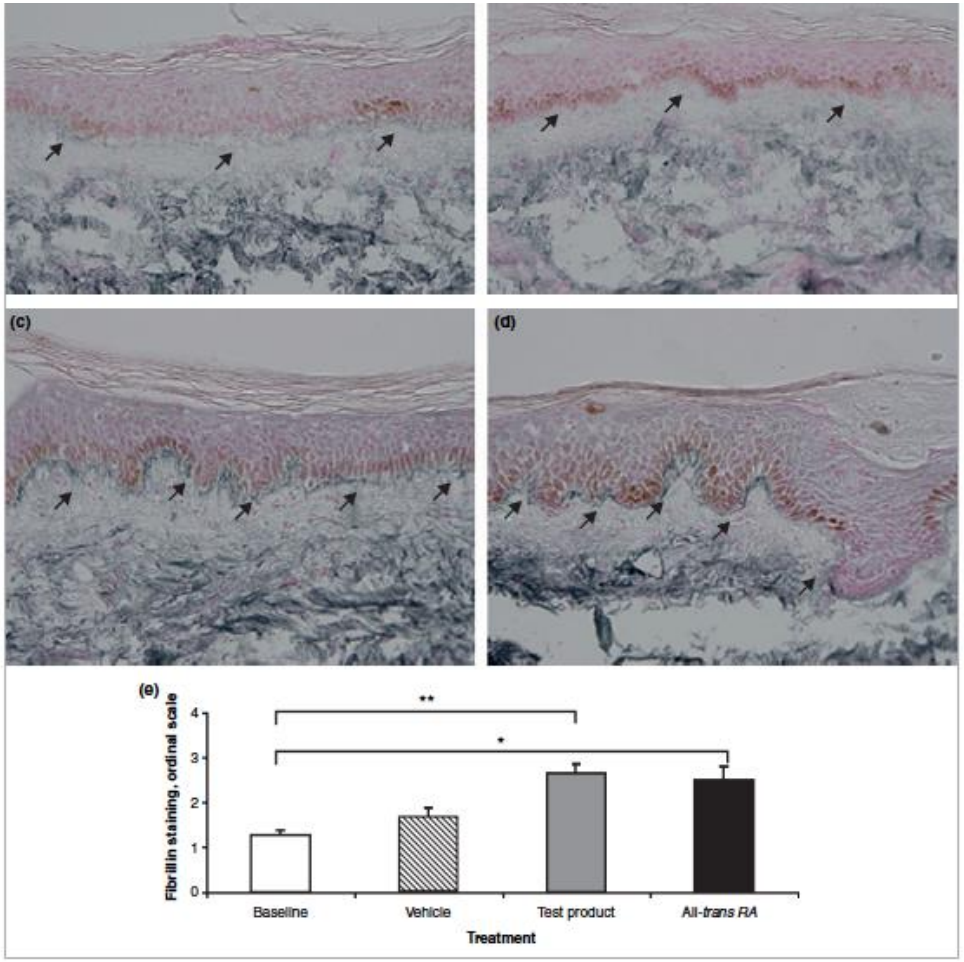


Fig 1. Fibrillin-rich microfibrils are deposited in test product- and all-trans retinoic acid (RA)-treated human skin in a short-term patch test assay. Representative photomicrographs showing fibrillin-rich microfibrils (arrows) in photoaged extensor forearm following the extended 12-day patch test protocol: (a) baseline; (b) vehicle; (c) test product. (d) The positive control (0·025% RA) was applied for 4 days to avoid harmful side-effects. Original magnification × 400. (e) Quantification of fibrillin-rich microfibrils following the occluded patch test assay with vehicle, test product or the positive control (RA), compared with the baseline. We identified significantly more fibrillin-rich microfibrils with both the test product (**P = 0·005) and RA (*P = 0·015).

appearance of facial wrinkles when used in the long term. This improvement is associated with restoration of fibrillin-1, the major component of fibrillin-rich microfibrils in product-

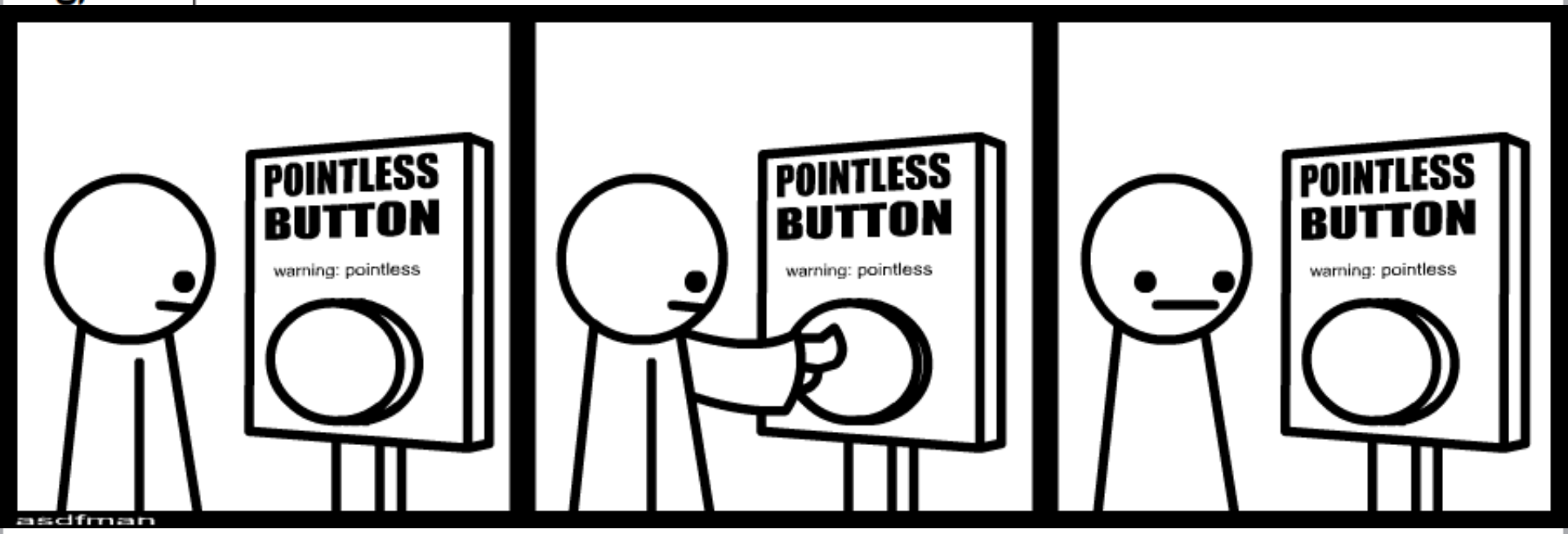
improvement in the appearance of photoaged facial skin. The trial was executed to the highest standards, with study creams coded and randomized at source and with the

ECM component	Baseline	Vehicle	Test product	0·05% RA
Fibrillin-1	1·27 ± 0·11	1·70 ± 0·17	2·64 ± 0·22**	2·51 ± 0·28*
pCI	2·65 ± 0·15	2·64 ± 0·15	2·70 ± 0·13	2·63 ± 0·21
MMP-1 (cells/hpf)	115·8 ± 7·9	140·3 ± 13·6	122·2 ± 14·1	143·7 ± 12·5

Table 1 Expression of extracellular matrix (ECM) molecules in photoaged skin following 12-day occlusion with a cosmetic 'anti-ageing' product

(a) 100

(c) 4



(b)

appearance of facial wrinkles. This improvement in the appearance of facial wrinkles became significant only after 12 months of daily product use comparing between groups.

the baseline, the test product did lead to a noticeable clinical improvement in facial wrinkles ($P = 0·013$) in 43% of treated individuals after 6 months, compared with only 22%

of those treated with the vehicle where there was no significant improvement in appearance ($P = 0{\cdot}11$). In a comparison between groups, this improvement was not statistically

was our belief that a combination of ingredients with activities known to address the multiple changes which occur in photoaged skin (degradation of collagen and elastin, the

Was there anything obviously missing?

There was also no sample size calculation and the outcomes were unclear

N t

We ea

how two groups.

contains the retinol ester, retinyl palmitate, together with natural plant extracts, peptides and lipopeptides and antioxidants. Other authors have shown evidence for the role of many of these cosmetic ingredients in protecting against mechanisms that lead to dermal degradation, such as increased MMP activity[22] and stimulating repair of dermal components.[23–25] It

References

1 Lavker RM, Zheng PS, Dong G. Aged skin: a study by light, transmission electron, and scanning electron microscopy. J Invest Dermatol 1987; 88:44s–51s.

The £19 anti-wrinkle cream set to cause a stampede at Boots after scientists claim it DOES actually work

By FIONA MACRAE FOR THE DAILY MAIL
UPDATED: 17:06, 3 May 2009

Two years ago, Boots produced a wrinkle cream so effective that stores sold out in a day.

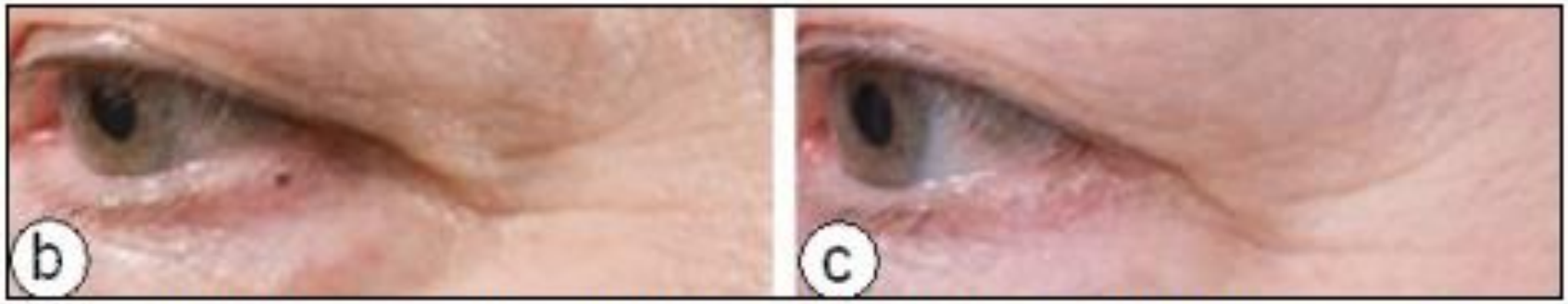
On Thursday, when its latest product hits the shelves, the high street chemist sets out to show customers it has done it again.

Aimed at mature skin with more wrinkles than its sister lotion, No7 Protect & Perfect Intense Beauty Serum is said to be twice as good as normal moisturisers at plumping up the skin and smoothing away wrinkles.

And Boots is offering proof. A year-long study at Manchester University found that 70 per cent of those who used the £19.75 cream for a year had fewer, finer lines, the British Journal of Dermatology reported.

I'm convinced…

It includes vitamin A and extract of lupin. And while the original Protect & Perfect Serum and the new cream share some ingredients, there are crucial differences. The original cream uses 'pentapeptides' - groups of five amino acids - to trigger the rejuvenation of skin.



Left, a volunteer's eye before applying Boots' Intense serum; right, the eye following a year's daily application of the cream

The Intense version, which has actually been on sale for 18 months under the name

How are the results presented?

- **How was the primary outcome analysed?**
  - Did they actually compare treatments (between arm comparison)?
  - Or did they report only a within-arm comparison?

- **Were relative or absolute risk differences presented?**
  - Results can appear to look more impressive if only relative risk differences are reported (Mayor BMJ 2002)
    - "if a disease kills two in every million people, a drug that reduces the death rate to one in a million would give a relative risk reduction of 50% which appears to be a major benefit. However, the absolute risk reduction would only be one in a million"

Statistical significance

- **Remember**
  - Statistical significance does not necessarily imply clinical significance
  - 'Not significant' does not equate to 'they are the same'
  - Statistical significance may be due to large sample size
  - Statistical insignificance may be due to small sample size

Systematic reviews/meta-analyses

- **Did the search include both published and unpublished material? English/non-English studies**
  - Were authors contacted for additional information not presented in the published article?
- **Were the criteria to select the studies appropriate?**
- **Has a risk of bias assessment been carried out?**
  - Were results of the risk of bias used in the analysis?

Systematic reviews/meta-analyses

- **Did more than one person screen the studies, extract information and evaluate risk of bias from the primary studies**
  - Deciding which studies to include requires judgment
    - Mistakes (random errors)
    - Bias (systematic errors)

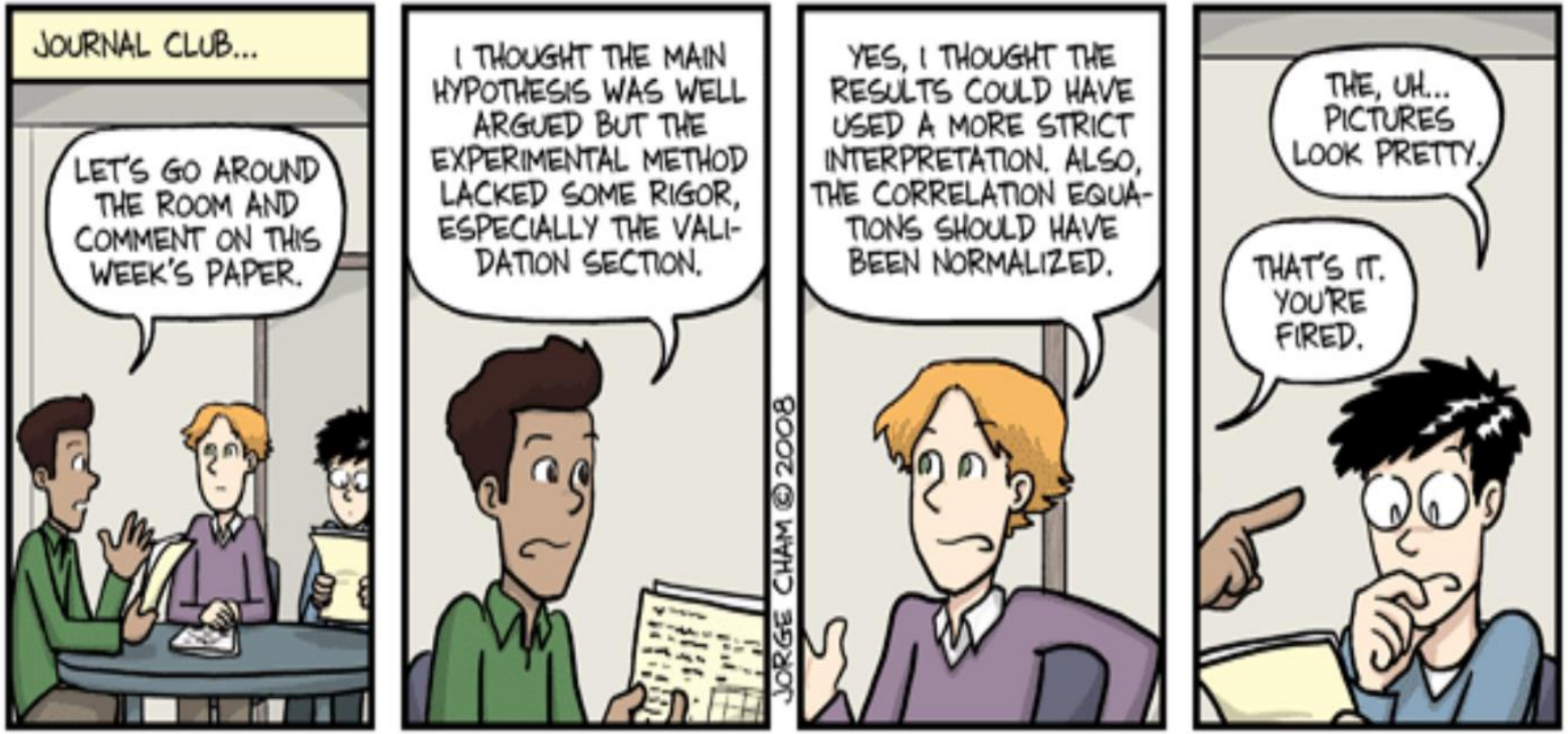
    } Multiple extractors reduces errors

- **Were studies sufficiently similar (homogenous) to meta-analyse?**

- **Were there a sufficient number of studies to meaningfully meta-analyse?**

Journal Clubs

- **Regular meetings to discuss and critically evaluate scientific articles**
  - shared learning experience
  - but it is only beneficial if there is discussion amongst the group
- **Usually, one person will present a summary of the article**
- **Good way to discuss issues ideas and gain understanding of current topics**
  - keep up-to-date with the literature


JOURNAL CLUB...
LET'S GO AROUND THE ROOM AND COMMENT ON THIS WEEK'S PAPER.
I THOUGHT THE MAIN HYPOTHESIS WAS WELL ARGUED BUT THE EXPERIMENTAL METHOD LACKED SOME RIGOR, ESPECIALLY THE VALIDATION SECTION.
YES, I THOUGHT THE RESULTS COULD HAVE USED A MORE STRICT INTERPRETATION. ALSO, THE CORRELATION EQUATIONS SHOULD HAVE BEEN NORMALIZED.
THE, UH... PICTURES LOOK PRETTY.
THAT'S IT. YOU'RE FIRED.
JORGE CHAM © 2008
WWW.PHDCOMICS.COM

Summary

- **Don't always believe what you read**
- **Are primary (and secondary) outcomes clearly specified**
- **Check sample size assumptions and whether the study recruited enough participants**
- **Check if all participants are accounted for**
  - In an RCT, are the two groups comparable
  - Have they done something sensible about missing data
- **Any evidence to suggest selective reporting**
  - i.e. didn't analyse pre-specified outcome, or placed emphasis on secondary outcome over primary outcome

Useful links & reading

- Critical Appraisal Skills Programme (CASP) www.casp-uk.net
- BMJ Statistics Notes
  - www-users.york.ac.uk/~mb55/pubs/pbstnote.htm
- How to read a research paper (Trish Greenhalgh BMJ series)
  - www.bmj.com/about-bmj/resources-readers/publications/how-read-paper
- Vickers AJ. Clinical appraisal: how to read a clinical research paper. Complementary Therapies in Medicine 1995; 3: 158-166.
- Heneghan C, Perera R. The essentials of critical appraisal of randomized controlled trials. The Foundation Years 2009; 5: 9-12.
- Schriger DL, Cooper RJ. Achieving graphical excellence: suggestions and methods for creating high-quality visual displays of experience data. Ann Emerg Med 2001; 37: 75-87.
- Stratton IM, Neil A. How to ensure your paper is rejected by the statistical reviewer. *Diab Med* 2005; 22: 371-373.
- Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005; 2(8): e124.
- Young JM, Solomon MJ. How to critically appraise an article. *Nat Clin Pract* 2009; 6: 82-91.