Homeopathy (2007) 96, 4–16

doi:10.1016/j.homp.2006.11.005, available online at http://www.sciencedirect.com

ORIGINAL PAPER

A systematic review of the quality of homeopathic pathogenetic trials published from 1945 to 1995

F Dantas[1,2,*], P Fisher[3], H Walach[4], F Wieland, DP Rastogi[5], H Teixeira[1], D Koster, JP Jansen, J Eizayaga[6], MEP Alvarez[7], M Marim, P Belon and LLM Weckx[2]

[1]*Universidade Federal de Uberlândia, Brazil*
[2]*Universidade Federal de São Paulo, Brazil*
[3]*Royal London Homoeopathic Hospital, UK*
[4]*University of Northampton and Samueli Institute, European Office, UK*
[5]*Central Council of Research in Homeopathy, India*
[6]*Universidad Maimonides, Argentina*
[7]*Escuela Nacional de Medicina y Homeopatia, Mexico*

Background: **The quality of information gathered from homeopathic pathogenetic trials (HPTs), also known as 'provings', is fundamental to homeopathy. We systematically reviewed HPTs published in six languages (English, German, Spanish, French, Portuguese and Dutch) from 1945 to 1995, to assess their quality in terms of the validity of the information they provide.**

Methods: **The literature was comprehensively searched, only published reports of HPTs were included. Information was extracted by two reviewers per trial using a form with 87 items. Information on: medicines, volunteers, ethical aspects, blinding, randomization, use of placebo, adverse effects, assessments, presentation of data and number of claimed findings were recorded. Methodological quality was assessed by an index including indicators of internal and external validity, personal judgement and comments of reviewers for each study.**

Results: **156 HPTs on 143 medicines, involving 2815 volunteers, produced 20,538 pathogenetic effects (median 6.5 per volunteer). There was wide variation in methods and results. Sample size (median 15, range 1–103) and trial duration (mean 34 days) were very variable. Most studies had design flaws, particularly absence of proper randomization, blinding, placebo control and criteria for analysis of outcomes. Mean methodological score was 5.6 (range 4–16). More symptoms were reported from HPTs of poor quality than from better ones. In 56% of trials volunteers took placebo. Pathogenetic effects were claimed in 98% of publications. On average about 84% of volunteers receiving active treatment developed symptoms. The quality of reports was in general poor, and much important information was not available.**

Conclusions: **The HPTs were generally of low methodological quality. There is a high incidence of pathogenetic effects in publications and volunteers but this could be attributable to design flaws. Homeopathic medicines, tested in HPTs, appear safe. The central question of whether homeopathic medicines in high dilutions can provoke effects in healthy volunteers has not yet been definitively answered, because of**

*Correspondence: Flávio Dantas, Department of Clinical Medicine, Federal University of Uberlândia and Federal University of São Paulo, Brazil.
E-mail: dantasfla@gmail.com
Received 23 March 2006; revised 26 June 2006; accepted 6 November 2006

methodological weaknesses of the reports. Improvement of the method and reporting of results of HPTs are required.
References: **References to all included RCTs are available on-line at www.sciencedirect.com/homp/** *Homeopathy* (2007) **96,** 4–16.

Keywords: homeopathic pathogenetic trial; proving; systematic review; homeopathy; human volunteers; evidence-based medicine; methods

Introduction

Homeopathy is a controversial therapy, dating from the European Enlightenment. Its founder, Samuel Hahnemann, even wrote the Enlightenment ideal, rationality, into the title of his magnum opus the Organon der rationellen Heilkunde. He strongly held the enlightenment view that knowledge is not innate, but comes only from observation guided by reason, insisting that: 'The pure, characteristic, curative virtues of medicines cannot be apprehended by specious a priori sophistry, or from the smell, taste or appearance of the medicine, or from chemical analysis.'[1]

It was from such considerations that he developed 'provings', an infelicitous translation of the German Prüfung. In 19th century English 'prove' had the sense of try, test the qualities of, find out by experience.[2] Provings are more accurately known as *homeopathic pathogenetic trials* (HPT).[3] Hahnemann tested several toxic substances used as medicines in his time (eg arsenic and mercury) in healthy volunteers to identify their effects in body and mind, initially in substantial doses but later in high dilutions, subsequently applying the results to clinical practice on the basis of *Similia similibus curentur*—Let like cure like. He called these tests ***Prüfungen***, translated into English as 'Proving'.

HPTs are unique to homeopathy. Their purpose is to test a substance at a non-toxic level, on healthy volunteers to determine the symptoms it provokes and the type of person who may be sensitive to it and which, according to the Similarity Principle, it may be used to treat. HPTs have certain similarities to phase I trials for new pharmaceutical products: they are conducted on healthy volunteers, but there are key differences. Most importantly, the doses used in HPTs are too small to risk serious adverse effects, and the data collected are mostly qualitative,[3] phase I trials are intended mostly to provide quantitative pharmacological and pharmacokinetic data. We adopted the following definition: HPTs are clinical trials designed to investigate the effects of the exposure of human volunteers, in good health, to potentially toxic or pathogenetic substances, diluted and serially agitated according to homeopathic pharmacopoeial methods, with a view to providing data to inform their use as homeopathic medicines.

For Hahnemann, a true materia medica should be a collection of the authentic, pure, reliable effects of simple medicinal substances in themselves, where all conjecture, everything merely asserted or entirely fabricated, should be completely excluded. He tested 99 substances in quasi-experimental studies (one-group pretest–posttest design) and published the results.[4] To minimize bias, he recommended the selection of trustworthy and conscientious human healthy volunteers (usually friends and sympathizers of homeopathy), use of only one medicine in its purest form and in moderate dose, close supervision of the subjects and some rules for controlling confounding variables as diet, life style, ingestion of medicines and consumption of alcohol and coffee.[1]

The function of HPTs

HPTs are one of the sources of information on homeopathic materia medica, others are the observation of toxic effects of substances in humans and clinical experience. The validity and reliability of information gathered from HPTs are therefore fundamental for the success of homeopathic practice and clinical research. In fact, many important symptoms used in homeopathic prescribing cannot be traced to HPTs[5,6]. Possible explanations for this include that the theory linking symptoms detected in healthy volunteers to those treated in the sick is wrong, or that the methods used in HPTs are inadequate to detect them, specially the chronic ones.

Historically, HPTs have been methodologically innovative. The first double-blind placebo controlled homeopathic 'proving' was conducted in 1835, and was one of the first double-blind placebo controlled trials in the history of medicine.[7] In 1895 the suggestion of including a pre-observation 'run-in' period to prepare the volunteer was made,[8] and one of the earliest multicentre double-blind clinical trials was an HPT of Belladonna conducted by Bellows, published in 1906.[9] Early in the 20th century, Hughes[10] critically analysed Hahnemann's conclusions. A modern analysis of Hahnemann's guidelines found many flaws, which could not have been anticipated by Hahnemann, all likely to lead to an over-estimation of pathogenetic effects.[3] A review of HPTs published in the UK by two of us (FD, PF) included 45 studies, and showed a great variability in terms of the medicines tested, methodology, volunteers, sample size and outcome.[11]

We report an exploratory systematic review[12] of HPTs published in six languages (Dutch, English, French, German, Portuguese, Spanish), in the five

decades from 1945. We focus particularly on the characteristics which account for the observed differences in the reported effect sizes; we also discuss the findings, and their implications for improving the design and conduct of HPTs.

Methods

The protocol was based on that of the previous study of HPTs conducted in the UK. The reviewers were homeopathic doctors or researchers with experience of conducting randomized clinical trials or HPTs.

Search strategy and study selection

Trials were sought by manual searching of books and journals, scanning reference lists and expert knowledge, by authors for their respective languages, and searches of bibliographic databases (HOMINFORM—British Homoeopathic Library, HOMEOINDEX, MEDLINE, Pre-MEDLINE). Only published reports of trials using diluted and potentized homeopathic medicines with the aim of detecting changes in at least one healthy human volunteer resulting from exposure to a specific homeopathic medicine were included. Printed publications of all kinds were included: books, proceedings of congresses and homeopathic meetings (regional, national and international), and journal articles published from 1945 to 1995 in six languages. The decision on inclusion was made by the authors after referring to the review protocol. Only written information in the public domain was included, private reports of HPTs by homeopathic companies, for instance, were excluded. Repeated publications, translations of HPT publications done before 1945 and papers dealing only with theoretical or methodological aspects of HPTs and not reporting any experimental results were excluded, as well as trials in which only mother tinctures were used.

Study design

A draft data extraction form, similar to that used in the previous study, was developed by the first author, and sent to the reviewers for comment. A second form was then designed and sent to all reviewers for a pilot study using three reports of HPTs originally published in English. These three reports were selected because of their methodological differences: one was a self-experiment, another a randomized double-blind placebo controlled trial using different dilutions of the same medicine and the third a randomized double-blind placebo controlled trial of one medicine in a single dilution. Suggestions and comments were again incorporated to prepare the final version of the data extraction form, which comprised 86 closed questions and a final open question for methodological criticisms.

Procedures

The form was designed to collect relevant information on the setting, population, design, outcomes, assessment and interpretation of results in HPT reports. For each report bibliographic details, description of the setting, tested substance, method of preparation, volunteers included in the study, details of the study design, assessment of outcomes, presentation and interpretation of results and the reviewer's overall personal appraisal were recorded. Withdrawal rates, study methodology, presence of adverse effects, percentage of responsive volunteers and number of claims per HPT were also extracted. Each HPT report was independently analysed by two reviewers per language.

For each medicine the name, dilution(s), method of dilution, presentation, dose, frequency per day, repetition of doses, total duration of the trial, number of active treatment periods and duration per volunteer, source of the drug, method of preparation and responsibility for preparation was recorded. For study population we extracted the initial and final number, ethnic origin, sex, age, occupation, number of control volunteers, inclusion criteria, exclusion criteria, assessment of health status prior to admission, training of volunteers, personality traits, physical characteristics, informed consent, method of recruitment. The study method was assessed in terms of approval of protocol by an Ethical Committee, direction/coordination, randomization, sequence generation of subjects in the trial, allocation concealment, masking (blindness) of volunteers and of supervisor, use of placebo, pre-trial observation ('run-in') period with or without placebo, placebo distinguishable from verum, placebo potentized, comparative group, crossover, washout period (post-treatment observation), management of adverse effects, rules for stopping medicine, rationale and source of the medicine. Assessment was evaluated in terms of: use of symptom diary, type of diary, initial interview (case-taking/collection of previous symptoms), follow-up interview, use of laboratory investigations, use of psychological tests, withdrawal/dropout of volunteers, reason for withdrawal, withdrawal due to severe adverse effects, presence of adverse effects, pre-defined categories for assessment of the attributes of a symptom.

For the presentation of results we extracted information on the frequency of symptoms, their chronology, character, location, duration, onset, intensity, modalities, presence of concomitant symptoms, description of complete symptoms, analytical presentation, inclusion of prior symptoms that improved during the trial, detailed report of individual volunteers, claimed percentage of sensitive volunteers, use of symptom tables and charts. The authors' interpretation of the results was reviewed in terms of pre-defined criteria for including symptoms (such as time after taking the medicine, peculiarity, intensity and duration of the symptom, etc), use of descriptive

statistics, use of statistical tests and presence and number of significant findings claimed. Each reviewer made a personal judgement on the study in 4 multiple choice questions, with a final open questions asking for methodological criticisms.

Pathogenetic effects

We defined pathogenetic effects ('proving symptoms') as all changes in clinical events and laboratory findings reported by volunteers during an HPT and recorded in the final report. In other words, the findings claimed by authors to be compared by practitioners with the symptoms of their patients in order to prescribe a homeopathic medicine. The overall incidence of pathogenetic effects in each trial was calculated by dividing the number of volunteers who had at least one reported pathogenetic effect by the total number of volunteers taking the medicine and who contributed symptoms or signs. The incidence of pathogenetic effects per volunteer was defined as the total number of findings claimed in the trial divided by the total number of subjects using the medicine and included in its final pathogenetic description. We counted as one pathogenetic effect a piece of information which could be included in an homeopathic repertory as an independent subheading. For instance, boring headache ameliorated by pressure counted as one claim.

The results were entered on a database and reviewers' discrepancies were noted and discussed to reach consensus. Unresolved disagreements were adjudicated by the first author.

Assessment of methodological quality

An improved version of the Methodological Quality Index (MQI) for HPTs, used in the previous study, was developed. It is based on key components of methodological quality including internal and external validity items. The MQI includes aspects such as randomization, inclusion and exclusion criteria, blinding and criteria for selection of pathogenetic effects, with values ranging from 1 to 4 for each component, giving a range from 4 to 16. Scores were divided into 4 methodological classes, where class I is the worst and class IV is the best, with arbitrary cutoff points ($\leqslant 6$ for Class I; 7–10 for Class II; 11–13 for Class III; >14 for Class IV). Reviewers' judgements on validity, reliability and clinical applicability of study findings were transformed into a numerical scale, where 0 means the lowest confidence (Table 1).

Graphical and statistical procedures

We used descriptive statistics, charts and visual data plots to present as much raw data as possible due to the exploratory character of the review. Spearman correlation coefficients (r_s) were used to verify relationships between validity and reliability of information from HPTs, including association between MQI and subjective judgements by reviewers. Kappa statistics were used to evaluate agreement between reviewers on judging methodological quality components and to estimate the disagreement on global judgements of quality.

Results

Number and distribution

156 HPTs were included in the analysis, in which 143 medicines were tested by 2815 volunteers (1169 male, 857 female). In total 20.538 pathogenetic effects were reported. 116 HPTs were published in journals, 13 in congress proceedings, 11 were books and 16 monographs or academic dissertations. Authorship was relatively concentrated: 15 authors contributed to 52% of studies.

There was great heterogeneity among studies regarding methods and outcomes description. An increasing number of HPTs were published across the decades, particularly in the last decade. Table 2 shows included publications by language in the period from 1945 to 1995.

Quality of reporting

The quality of reports was in general poor, and much important information for methodological analysis and

Table 1 Methodological Quality Index for HPTs

Component	*Score*			
	1	2	3	4
Randomization	Not stated	Only stated, no details	Description of sequence generation or allocation concealment	Description of sequence generation *and* allocation concealment
Blinding	Not stated	Single blind	Double-blind without verification	Double-blind with post-trial verification
Inclusion and exclusion criteria	Not stated	One partially stated	One clearly stated or both partially stated	Clearly stated
Criteria for selection of effects	Not stated	At least one defined	2 to 4 defined	More than 4 defined

Table 2 Number of included HPTs per decade and language

Time	Language						Total
	English	German	Dutch	French	Spanish	Portuguese	
1945–1955	9	0	0	1	1	0	11
1956–1965	11	1	0	0	0	0	12
1966–1975	16	3	0	0	0	0	19
1976–1985	16	9	1	2	0	0	28
1986–1995	32	20	16	8	6	4	86
Total	84	33	17	11	7	4	156

reproducibility of HPTs was omitted from many reports. Complete description of the source of medicines was given in only 17 publications and of these only 7 gave adequate information on the preparation of the medicine. 57% of the reports did not state the age and 34% the gender of volunteers. Ethnicity data was generally absent from the reports. Little information on volunteers' characteristics was reported, with a few exceptions. Some studies of apparently good design did not describe adequately their methods and outcomes. For instance, the use (or not) of homeopathic case-taking at the beginning of HPTs was described only in 31% of reports, this suggests poor reporting.

Ethical aspects

Informed consent was obtained from volunteers in 19 studies (the first in 1980), approval by Ethical Review Committees was mentioned in 7 publications. Commercial sponsorship was not mentioned in any paper, some publications mentioned the supplier of the tested medicine. Conflicts of interest were not declared in any case.

Settings and responsibility

HPTs were conducted mainly in India (36 studies) and United Kingdom (30), followed by Germany (17), The Netherlands (17), Austria (16), France (13), United States (12), Mexico (9), Brazil (2), New Zealand (2), Norway (1) and Argentina (1). 9.6% of studies were multicentric, generally in two centres. Most HPTs were done in homeopathic teaching or research centres, under the supervision of medical doctors, with students or sympathizers of homeopathy as volunteers.

Volunteers

Age range was 5–76 years, most studies were done in young adults. 28 studies (18%) involved more women than men, this trend increased in the later decades and correlates with larger number of pathogenetic effects ($r = 0.26$; $p = 0.014$). Volunteers underwent laboratory tests in 22 (14%) of HPTs. Criteria used for the definition of healthy volunteers mostly followed Hahnemann's guidelines.

Medicines and rationale

65 publications tested medicines in single dilutions and 91 in different dilutions. 30c was the most frequently used dilution (66 trials) followed by 6c (33) and 6 × (32). HPTs using single dilutions above 12c ($n = 32$) generated on average 6.2 effects/trial, those below 6x ($n = 12$) 8.9 effects. We could not compare effects obtained from different dilutions of the same medicine in the same trial because of inadequacies of reporting. 51% of published HPTs tested new homeopathic medicines. Plants were the most common source of medicines (75), followed by animal (29), mineral (18), chemicals (14) and pharmaceutical drugs (11). Two publications studied energy sources (sun light and solar eclipse ray) and one reported an unidentified coded substance. Fig. 1 shows the rationale for selection of the substances included in our sample of HPTs

Study design

There was a large variation in methods. Quasi-experimental designs, without control groups were the most common type of study, particularly before–after studies, followed by trials using placebo parallel group (36% of the sample). 17 uncontrolled 'dream provings' were published, all in Dutch. There is a trend in the later decades to larger sample size and randomized placebo controlled designs. Of the placebo controlled HPTs 14 were of crossover design. Sample size was small (median 15, range 1–103). 7 HPTs involved a single volunteer and 3 two volunteers. Median trial duration, reported in 63% of the sample, was 44 days (range 1–540). Placebo was taken by 769 volunteers in 56% of trials. Use of placebo control by volunteers was highly variable (range 0–100%, median 22). A placebo run-in phase preceded 16% of trials, a pre-observation period without placebo was reported in 14% of trials, 3% had two run-in phases with and without placebo. Placebo was described as indistinguishable from verum in 21% of reports.

Pathogenetic effects

Only 3 reports did not claim pathogenetic effects. The number of effects per publication varied from 0 to 1100 (median 88). A mean of 153 effects were reported

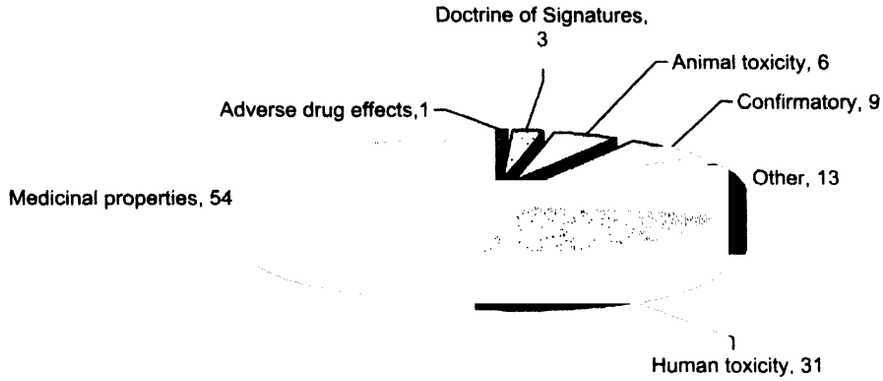

Fig. 1 Rationale for HPTs (%)

per trial. Each volunteer produced a median of 6.5 pathogenetic effects. The incidence of symptoms in volunteers was high with, on average, about 84% of volunteers receiving verum describing at least one symptom. The nature of the effects was very diverse, but many seemed to be common symptoms, associated with placebo in clinical trials or reported as everyday symptoms, but we did not systematically assess this. The frequency of symptoms in the sample was reported in 40% of studies and the chronology in 23%. In 52% of studies there was no mention to duration of symptoms and in 62% none of time of onset of symptoms. In general pathogenetic effects seem to occur mainly in the first week after taking the medicine and to be of short duration (up to 7 days).

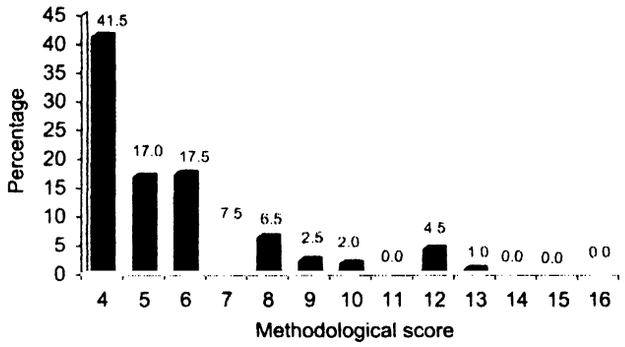

Fig. 2 Publications according to their methodological score (%)

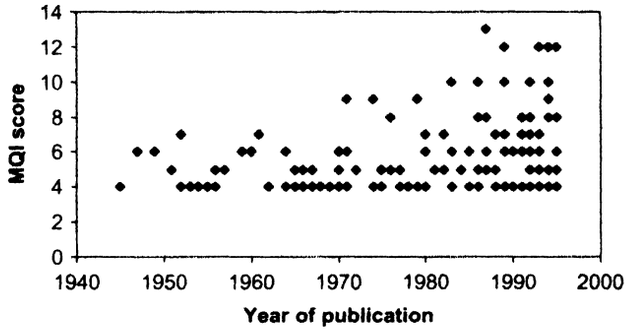

Fig. 3 MQI scores across time

Safety

Withdrawal was reported in 22% of publications, the incidence varied by quality of the study. In studies of class III (better quality) 10% of volunteers dropped out whereas in class II it was reported withdrawal of 18% and 6% in class I. In all classes an extremely low value of withdrawals due to severe adverse effects was reported (1–2%). 85% of reports did not mention how adverse effects were managed. Placebo and verum symptoms were superficially quite similar, but we could not properly compare due to insufficient information in reports. Withdrawal of volunteers taking placebo was also reported.

Methodological quality

Most HPTs were of low methodological quality, both according to the MQI and the reviewer's personal judgement. Mean methodological score was 5.6 (median 5) with possible range 4–16 and observed range 4–13. Fig. 2 shows the distribution of studies according to methodological score:

In all decades there were many low quality studies. The methodological quality showed a trend to improvement in the later decades, there was a positive and significant correlation between methodological classes and decades ($r_s = 0,218$; $P = 0,006$) (Fig. 3).

Most studies were of flawed design, mainly absence of proper randomization, blinding, placebo control and criteria for analysis of outcomes. There was a trend to increased use of randomization and masking in the last two decades of our period of study. Randomization was first described in 1961, but only 15 reports mentioned it, 9 of these were published in 1986–1995. Sequence generation was described in only 2 studies and it was difficult, from reading the reports, to clearly separate concealment of allocation from masking procedure. 'Blinding' of volunteers was reported in 53% of publications, 33% for investigators. Post-trial verification of blinding was not reported in any publication. Inclusion criteria were

not mentioned in 78% of reports, when mentioned they included a clinical history (94%) and laboratory findings (53%). Criteria for attributing symptoms to tested medicines were reported in 14% of our sample, such criteria include the presence of the symptom in more than one volunteer (33%), intensity and peculiarity (28%) (Fig. 4).

A comparison of better studies (MQI score 12-13) with studies with the lowest score (4), randomly selected and paired per year, showed that in low-quality studies there was no report of placebo, pre-observation period or judgement criteria for selection of pathogenetic effects but the mean number of pathogenetic effects was double with all volunteers reporting pathogenetic effects (Tables 3 and 4).

More symptoms per volunteer were reported from HPTs of lower MQI scores than from studies with higher MQI scores ($r_s = -0.204$; $P = 0.011$). Quality of publications in different languages was comparable, with the exception of those published in Dutch, which were of lower quality due to several so-called 'dream-provings' of poor quality. Better studies were positively and significantly correlated with larger numbers of volunteers ($r_s = 0.287$; $P<0.001$), and there was a significant difference in the number of volunteers in classes I and II (more in II, $P<0.001$, Tukey). Studies of longer duration tended to report more pathogenetic effects per volunteer ($r_s = 0.216$; $P = 0.031$) (Fig. 5).

MQI scores correlated with reviewers' judgment on reliability ($r_s = 0.375$; $P<0.001$). Reviewers' judgement showed internal consistency with correlations between clinical applicability and reliability ($r_s = 0.730$; $p<0.001$) and validity ($r_s = 0.869$; $P<0.001$). The reviewers overall considered 40% of the reports unreliable, yet 70% said they would apply the findings in practice. When asked if the reported symptoms could be reliably attributed to the medicine, the reviewers had serious reservations in 18% of HPTs, they rated 33% as possible, 38% probable and 11% certainly.

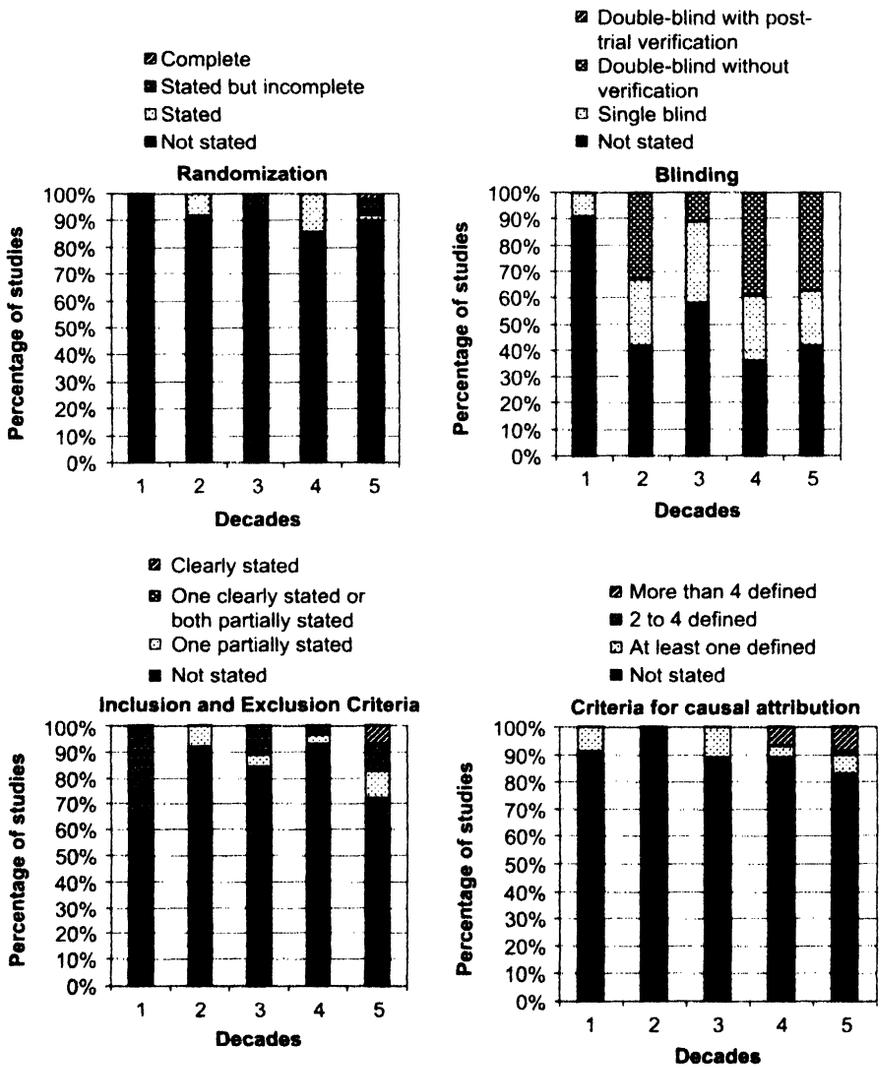

Fig. 4 Evolution of individual components of MQI per decade (%)

Table 3 Characteristics of HPTs with the best MQI scores 1986-1995

Medicine(year)	*Score*	*n*	*Effects per volunteer (mean)*	*Incidence in volunteers (%)*	*Type of study*	*Pre-observation*	*Use of placebo and ratio verum/ placebo*	*Number of criteria for pathogenetic judgement*
Cuprum (1987)	13	34	0	78	Parallel groups	Yes	Yes; 1,1	1
Propanolol (1989)	12	9	9,9	100	Parallel groups	No	Yes; 1,0	4
Belladonna (1993)	12	45	0	29	Crossover	No	Yes; 1,0	1
Nicotinamide (1994)	12	15	4,2	67	Parallel groups	Yes	Yes; 7,5	7
Fumaria (1994)	12	15	5,8	100	Parallel groups	Yes	Yes; 7,5	7
Myosotis (1994)	12	15	6,4	67	Parallel groups	Yes	Yes; 7,5	7
Veronica (1995)	12	17	6,7	100	Parallel groups	Yes	Yes; 7,5	5
Geranium (1995)	12	13	22,7	93	Parallel groups	Yes	Yes; 7,5	7

Table 4 Characteristics of HPTs with the lowest MQI scores 1986-1995

Medicine(year)	*Score*	*n*	*Effects per volunteer (mean)*	*Incidence in volunteers (%)*	*Type of study*	*Pre-observation*	*Use of placebo and ratio verum/ placebo*	*Number of criteria for pathogenetic judgement*
Vincetoxicum (1988)	4	6	4,5	100	Before-After	No	No	0
Heracleum (1989)	4	20	8,2	100	Before-After	No	No	0
Hyoscyamus (1993)	4	1	51	100	Before-After	No	No	0
Secale (1994)	4	21	4,6	100 (estimated)	Before-After	No	No	0
Viola odorata (1994)	4	6	1,8	100	Before-After	No	No	0
Carcinosinum (1994)	4	13	7,1	100	Before-After	No	No	0
Lac caprinum (1995)	4	2	18	100	Before-After	No	No	0
Gryllus (1995)	4	(?)	(?)	(?)	Before-After	No	No	0

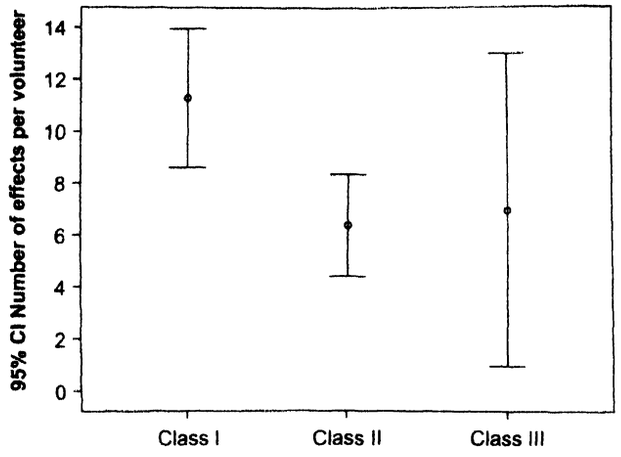

Fig. 5 Mean number of pathogenetic effects per volunteer by methodological classes

Reviewers

Data were extracted independently by 12 reviewers distributed in 11 pairs, each pair of reviewers read between 2 and 45. Agreement among reviewers varied from fair to good on different items of the MQI and personal judgment. For the pair which analysed most studies ($n = 45$) kappa was fair for global appraisal of the paper (0.30) and allocation concealment (0.32), moderate for sequence generation in randomization (0.49), good for exclusion criteria (0.65) and blinding (0.69) and very good for randomization (0.89) and inclusion criteria (1.0).

Discussion

Our exploratory meta-analysis examined methods and outcomes of HPTs published from 1945 to 1995. There was a great heterogeneity between studies in terms of substances tested, designs, volunteers, sample size and outcomes. This was reflected in great diversity in the incidence and types of reported effects. There was a clear association between the methodological quality of the trial and the number of effects reported: better trials produced fewer pathogenetic effects. Overall our analysis reveals methodological shortcomings which seriously compromise the validity, reliability and clinical applicability of the results.

HPTs play various roles for homeopathy. From an historical perspective they are a powerful evidence of the experimental nature of homeopathy since its inception. Their results have been disseminated and applied by homeopathic practitioners worldwide. For some homeopaths, HPTs are the very basis of homeopathy; others view them as marginal, since a large part of the homeopathic materia medica is build on toxicological sources and clinical confirmation gathered from practice. HPTs are not designed to 'prove' homeopathy, in the sense of proving that it works; but rather to identify the effects of substances in healthy human beings with a view to using them, on the basis of similarity, in clinical practice. They need to be replicated and then submitted to a rigorous process of systematic clinical verification, as suggested by Hering more than a hundred years ago.[13]

We included all published HPTs due to our original exploratory purpose for this systematic review. We also designed an index to assess the methodological quality of published trials, piloted in a previous publication, evaluating both internal and external validity components.[14]We did not include withdrawals and losses to follow-up in our index due to the unusual nature of HPTs: the whole point is to seek 'adverse effects' or pathogenetic changes, contrary to what is expected in conventional clinical trials.

Most HPTs were suggested by the known properties of medicinal substances rather than their toxic properties, there were more trials of new medicines than confirmatory HPTs of established homeopathic medicines. It is difficult to understand the rationale for HPTs of sunlight, edible bird's nest and human foods (natural or processed), for instance, even more difficult to accept that these substances in high dilutions can yield a large number of pathogenetic effects. So-called 'dream provings' scored low in the MQI and were responsible for the low mean scores in Dutch publications. We included them because they met our definition, involving 'exposure' to homeopathic medicines. In three cases this did not mean ingestion, but sleeping with the medicine under the pillow, without previous validation of this new form of exposure. In retrospect it is debatable whether they can be classified as HPTs, and the word exposure in definition should perhaps be changed to ingestion. Pathogenetic effects reported from studies of low methodological quality should be excluded from homeopathic materia medica and repertories, since they generate large numbers of unreliable symptoms. Consideration should be given to the rationale for pathogenetic trials, the main rationale should presumably be human toxicity, particularly when the toxic effects simulate a disease or syndrome of interest. Toxic sources to human beings, such as rejected substances in pharmacological phase I clinical trials, could be good candidates to HPTs.

Several factors may account for the great variability in the results. Among others the settings, inclusion and exclusion criteria for volunteers, differences in duration, study design and use of placebo, style of supervision and criteria for selection of pathogenetic effects as well as details of blinding and randomization, and, last but probably important, the assumptions by investigators and volunteers that homeopathic medicines must cause pathogenetic effects in most volunteers (which we could not directly assess).

The causal attribution of changes in healthy volunteers after an intervention is very complex, and may be influenced by many factors. In the absence of adequate control, clinical studies usually yield results

favouring investigators' assumptions.[15] Most of the HPTs were conducted by a small number of investigators. The enthusiasm of supervisors and the subjective appraisal of pathogenetic effects could lead, in poorly controlled designs, to overestimate of symptoms in the publication. Self-observation and daily recording may also result in an increased recall of changes ('Hawthorne effect'). The use of poor designs, with multiple endpoints, probably inflated the number of effects reported in these studies.

Conditioning and expectancy is of great importance in the reporting of symptoms, as was demonstrated in medical students in an experiment where they were conditioned to expect sedative or stimulant effects but received only placebo in blue or pink capsules[16] or where the type of effect caffeine was expected to have on psychomotor performance predicted the type of placebo response displayed.[17] Volunteers' behaviour pattern has also been shown to influence the reporting of subjective symptoms after placebo.[18] Most of the reviewed HPTs were done in the context of courses, the volunteers were students of homeopathy. In this context at least two such factors could bias the outcome towards increased reporting of symptoms: students, presumably believers in homeopathy and in the potential of HPTs to produce valid symptoms, have a dependent relationship with their teachers/investigators and on the other hand supervising investigators expecting useful information from students during the trial.

Placebo control

Many investigators seemed to have taken for granted that every substance must elicit symptoms and for this reason felt it unnecessary to use placebo as a control or failed to consider symptoms experienced by volunteers taking placebo. Symptoms due to placebo that could be classified as adverse drug reactions in apparently healthy people are well documented, particularly in phase I clinical trial reports[19] or surveys with healthy university students and hospital staff.[20] On average seven body and mental changes were reported by healthy medical students responding to a survey on symptoms they experienced in the last week.[21] These findings point out to the importance of proper controls in pathogenetic research, with particular attention to intraindividual control, to prevent incorrect attribution of symptoms to the medicines. This may require the use of cross-over designs, or at least run-in periods.

Strictly speaking, some of the HPTs which used placebo were not placebo-controlled: they used placebo as an instrument to increase awareness or uncertainty of volunteers. Some investigators who used it this way progressively abandoned the use of placebo, perhaps because they came to view it as an 'unnecessary waste' of volunteers.

Ethnicity and gender could be related to differential production of pathogenetic effects. For instance it appears that the prevalence of "salt sensitivity" in blacks is higher than in whites.[22] Cultural heterogeneity needs also to be explored since many complaints seem to be diverse among different countries.[23] Female gender is considered a risk factor for the development of adverse drug reactions,[24] and in newly marketed drugs suspected adverse reactions are recorded more often in women than in men.[25]

Quality of study and symptoms

Our results show more effects per volunteer when the methodological rigour of the trial is low. It is clear that, on the whole, HPTs have hitherto greatly overestimated the incidence of effects. From a theoretical perspective this is consistent with the flaws in Hahnemann's original directions for conducting HPTs.[3] Empirically it is confirmed by two trials of Hydrogenium[26,27] which used different designs and controls (MQI scores 12 and 4, respectively): 5000% more pathogenetic effects per volunteer were reported in the HPT of lower methodological quality. Our findings are also consistent with other studies showing that estimates of treatment effects are exaggerated in trials of poor methodological quality[28] and that most claimed research findings are false due to bias, study power, small sample and effect sizes, greater number and lesser preselection of tested relationships and greater flexibility in designs, definitions, outcomes, and other methodological shortcomings.[29]

Only a small number of trials used a pre-observation run-in period with or without placebo, in general they did not present the symptoms collected during this period or how they differed from the reported pathogenetic effects. A landmark study published in 1964 showed that placebos tend to accentuate pre-treatment symptoms in some healthy volunteers or patients and to elicit symptoms not previously present in others.[30] It also showed that adverse effects from placebo were more evident in healthy subjects than in patients, pre-treatment symptoms of some patients were relieved by placebo and that the incidence of symptoms was higher in females than in males.

The studies we reviewed claim, on average, that homeopathic dilutions can elicit at least one symptom in 84% of volunteers. We are sceptical of this claim: if it were true one would expect many more undesirable effects of homeopathic medicines in clinical practice. Hahnemann noted that individuals have different susceptibility to homeopathic medicines, and recommended to test a new medicine if no symptoms appeared in volunteers. The occurrence of withdrawals due to adverse effects was very low, but in the context of a HPT pathogenetic changes are the main outcome of the study—adverse effects are expected! There should be an agreed definition of when a pathogenetic symptom becomes an adverse effect, to be reported as

such. Our data confirm that homeopathic medicines in high dilutions are safe[31] although many volunteers reported mild, transient mental and somatic changes.

The use of quality indices to analyse published papers is controversial,[32,33] they sometimes have to be developed specifically for the topic to be evaluated.[34] For the sake of balance and completeness, we used indicators for internal and external validity plus subjective appraisal and open criticisms. The good agreement observed among reviewers in the components of MQI, together with the generally good correlation between MQI findings and subjective judgments, confirms the validity and reliability of the MQI as an instrument.

Given the heterogeneity of designs and low quality of reporting, we were unable to compare either different trials with the same medicine or results from different dilutions of the same medicine in the same trial. Our study was limited by the poor information quality of most of the HPTs. It is difficult to rule out publication bias, certainly 'negative' HPTs were very rare. The quality of the publication of HPTs would be enhanced if editors of homeopathic journals, as well as publishers of homeopathic repertories, agreed on minimal requirements for reporting such trials. These should include adequate description of the setting and of the tested medicine (source, mode of preparation, posology), demographic data, inclusion and exclusion criteria, study design, criteria for selection of pathogenetic changes, compliance, safety data and ethical aspects. Structured reporting of HPTs might also be helpful, allowing easy extraction of the main points. Additionally it would be helpful to establish a public registration system for HPTs at inception, as suggested by the World Health Organization and the International Committee of Medical Journal Editors for clinical trials. Sponsorship of HPTs by homeopathic companies should be declared as should authors' conflicts of interest.

The validity of HPTs

Our results show the evidence for the occurrence of pathogenetic effects in HPTs is contaminated. Homeopathic practitioners expect useful information from HPTs. Yet we do not even know the pattern of symptoms produced in HPTs. Do they occur in most volunteers or only in a minority? To resolve this initially it would suffice to do randomized trials with good intraindividual placebo control, with clear criteria for inclusion/exclusion and attribution of causality. However, if only a small minority of volunteers manifest changes (and the better designed studies suggest that this is the case) then new designs are required, perhaps drawing on experience for the detection of type II, or idiosyncratic, adverse drug reactions, with much larger samples and more qualitative detail of the reported symptoms. Alternatively, one could screen to find apparently sensitive volunteers and then conduct series of randomized, double-blinded, placebo-controlled single case trials with multiple crosss-over on these volunteers using different dilutions and exploring not only the occurrence of symptoms but the quality of the symptoms. One possible objection to this procedure is that sensitive volunteers may progressively loose their sensitivity in the course of such a procedure.

In any case new HPTs should be conducted by competent investigators and include a definition of a healthy volunteer and an assessment of health status. The populations should be described, the methods should minimize bias, suggestion and the incorrect attribution of spontaneous or unrelated changes to the medicine, clear instructions for volunteers and supervisors, sensitive and valid outcomes measurements, and of course conform to ethical standards for human experimentation.

Qualitative as well as quantitative evaluation may be required if we are to understand what happens to healthy volunteers taking homeopathic medicines compared to placebo. The use of an exclusively quantitative statistical analysis probably leads to an underestimate of pathogenetic effects but on the other hand inadequate use of control and failure to use placebo symptoms as a comparator within each individual leads to false-positive results. 'Rare, strange and peculiar' or idiosyncratic symptoms are believed to be of crucial importance in homeopathic prescribing, yet they occur in only a few or no volunteers in small HPTs. Qualitative criteria to discriminate verum from placebo effects in HPTs need to be validated. In HPTs of traditional parallel group design such idiosyncratic effects would be drowned in statistical 'noise' arising from spontaneous, incidental or irrelevant sources. We need methods to detect them. The answers to these problems are not yet clear, and the lessons that can be learnt from traditional trials few. They are important areas for methodological development.

Recent HPTs

Many new HPTs have been published since 1995, with innovations in design and conceptions. For instance, the possibility of non-local effects of homeopathic medicines in HPTs was raised in a study by Walach, but a rival hypothesis could not be eliminated.[35] On the other hand two randomized double-blind HPTs, with double-crossover, post-trial blinding verification and using three progressive filters to select pathogenetic changes, incorporated only 5% of noticed changes as possibly pathogenetic.[36] Comparison between outcomes of new HPTs with the classical literature has been done for Plumbum[37] and Belladonna,[38] but the results are disputable. Some modern HPTs have gone back to Hahnemann's original method of collecting qualitatively refined data, through the method of close and daily monitoring of volunteers for subtle changes, adding masking

procedures in parallel groups,[39] but failed to confirm blinding after the trial. Post-trial verification of blinding is very important in HPTs to confirm that volunteers and investigators were unaware of whether they were taking verum or placebo.

On the evidence we have reviewed it is not possible to answer to the main questions posed in HPTs: do homeopathic medicines in high dilution, cause changes in healthy volunteers? If they do, how can we discriminate the effects due to the substance tested from incidental effects? If appropriate, rigorous and well-designed research gives a negative answer to the first question, we should relegate HPTs of highly diluted homeopathic medicines to be purely historical and expunge information deriving from them from the homeopathic database. But if high-quality research shows that they can produce specific effects, we will need to refine the methodology of HPTs in order to clearly identify effects attributable to the substance tested. Findings from hormesis,[40] could offer new ways to better understand the stimulatory action of substances in ultra high dilutions in healthy human beings.

Hahnemann's original conception of homeopathy had a strong basis in ethical principles of respect for persons, beneficence, non-maleficence and justice. His Prüfungen were a revolutionary experimental method, far ahead of their time. They are human experiments and like any such experiment must be reviewed by a competent Ethics Review Committee. Our systematic review has shown that most HPTs lack adequate control and analysis, the results of such studies are unreliable and potentially harmful to patients treated, in good faith, by homeopaths. More positively, if the method can be vindicated, there is great scope for improving clinical practice and research if doctors prescribe homeopathic medicines based on evidence from well-done HPTs.

Our study highlights the need for methodological improvements to ensure that HPTs are rigorous and that their results can be trusted. We hope it will stimulate a close monitoring and comparison of methodological quality of HPTs done after 1995. We will be happy to make our data extraction form available to others who may be interested in undertaking this task. It is imperative to develop a consensus on minimal requirements for reporting HPTs. We need a *pure* homeopathic materia medica, with valid and reliable information from HPTs, to get better results in our clinical practice and research. As evidence accumulates for the efficacy and safety of homeopathy from rigorous clinical trials, there is an increasing need to investigate and develop valid methodologies for the experimental pillar of homeopathy– the homeopathic pathogenetic trial.

Contributors

F Dantas conceived the study and developed analysis form together with P Fisher, did literature searches, reviewed half of the publications and wrote the draft and final version of the paper. All authors contributed to the final draft. Extraction of data of HPTs in different languages were done by P Fisher, DP Rastogi, D Koster, MEP Alvarez, J Eizayaga, F Wieland, H Walach, H Teixeira, JP Jansen, M Marim and P Belon. They were also responsible for identification of eligible studies. LLM Weckx advised on quality assessment, data interpretation and presentation of results.

Conflict of interest statement

We declare that we have no conflicts of interest

Acknowledgements

CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil); Royal London Homoeopathic Hospital; Programa de Pós-Graduação em Otorrinolaringologia e Cirurgia de Cabeça e Pescoço da Universidade Federal de São Paulo; Cecília Lomônaco de Paula and Gleice Margarete de Souza Conceição for statistical advise; Mary Gooch for help in literature search; David Riley for early cooperation on HPTs done in USA.

Appendix A. Supplementary materials

Supplementary data, including a full listing of the reviewed HPTs, associated with this article can be found in the online version at doi:10.1016/j.homp.2006.11.005

References

1 Hahnemann S. *Organon of Medicine*. 6th edn. Los Angeles: J.P. Tarcher, 1982.
2 The Concise Oxford Dictionary of Current English., 9th edn. Oxford: Oxford University Press, 1990.
3 Dantas F. How can we get more reliable information from homoeopathic pathogenetic trials? *Br Hom J* 1996; **85**: 230–236.
4 Hahnemann S. *Materia Medica Pura*. 2v. London: Homoeopathic Publishing Co., 1936.
5 Campbell A. Lycopodium from provings. *Br Hom J* 1981; **70**: 94–99.
6 Campbell A. Rhus from provings. *Br Hom J* 1981; **70**: 179–182.
7 Stolberg M. Die Homöopathie auf dem Prüfstein. Der erste Doppelblindversouch der Medizingeschichte im Jahr 1835. *Münchener Medizinische Wochenschrift* 1996; **138**: 364–366.
8 The Medical Investigation Club of Baltimore. *A Pathogenetic Materia Medica*. Philadelphia: Boericke & Tafel, 1895.
9 Bellows HP. The Test Drug-Proving of the O. O. & L. Society: *A Reproving of Belladonna Being an Experimental Study of the*

Pathogenic Action of that Drug Upon the healthy Human Organism. Boston: The O. O. & L. Society, 1906.

10 Hughes R. *A Manual of Pharmacodinamics*, 6th edn. London: Leath and Ross, 1912.

11 Dantas F, Fisher P. A systematic review of homoeopathic pathogenetic trials ('provings') published in the United Kingdom from 1945 to 1995. in: Ernst E, Hahn EG editors. *Homoeopathy: A Critical Appraisal*. London: Butterworth-Heinemann, 1998. p. 69–97.

12 Anello C, Fleiss JL. Exploratory or analytic meta-analysis: should we distinguish between them? *J Clin Epidemiol* 1995; **48**: 109–116.

13 Hering C. *Guiding Symptoms of our Materia Medica*, vol 1. New Delhi: B. Jain Publishers, 1974.

14 Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001; **323**: 42–46.

15 Sackett DL. Bias in analytic research. *J Chron Dis* 1979; **32**: 51–63.

16 Blackwell B, Bloomfield SS, Buncher CR. Demonstration to medical students of placebo responses and non-drug factors. *Lancet* 1972; **1**: 1279–1282.

17 Fillmore M, Vogel-Sprott M. Expected effect of caffeine on motor performance predicts the type of response to placebo. *Psychopharmacology* 1992; **106**: 209–214.

18 Drici M, Raybaud F, De Lunardo C, Iacono P, Gustovic P. Influence of the behaviour pattern on the nocebo response of healthy volunteers. *Br J Clin Pharmacol* 1995; **39**: 204–206.

19 Sibille M, Deigat N, Olagnier V, Durand DV, Levrat R. Adverse events in phase one studies: a study in 430 health volunteers. *Eur J Clin Pharmacol* 1992; **42**: 389–393.

20 Reidenberg MM, Lowenthal DT. Adverse nondrug reactions. *New Engl J Med* 1968; **279**: 678–679.

21 Dantas F. Incidência de efeitos patogenéticos não-farmacológicos e triviais numa amostra de estudantes de medicina. *Rev Homeopatia* 2004; **69**: 5–10.

22 Flack JM, Ensrud KE, Mascioli S, Launer CA, Svendsen K, Elmer PJ, Grimm Jr. RH. Racial and ethnic modifiers of the salt-blood pressure response. *Hypertension* 1991; **17**(1 Suppl): I115–I121.

23 Payer L. *Medicine & Culture: Varieties of Treatment in the United States, England, West Germany, and France*. New York: First Owl Books, 1996.

24 Tran C, Knowles SR, Liu BA, Shear NH. Gender differences in adverse drug reactions. *J Clin Pharmacol* 1998; **38**: 1003–1009.

25 Martin RM, Biswas PN, Freemantle SN, Pearce GL, Mann RD. Age and sex distribution of suspected adverse drug reactions to newly marketed drugs in general practice in England: analysis of 48 cohort studies. *Br J Clin Pharmacol* 1998; **46**: 505–511.

26 Schroyens F, Cecchi M, Saetonne MF *et al*. Homeopathic proving of Hydrogen. Proceedings of 54th International Congress of LMHI, Milan; 1996.

27 Sherr J. *The Homeopathic Proving of Hydrogen*. West Malvern: Dynamis School, 1992.

28 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**: 408–412.

29 Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005; **2**(8): e124.

30 Green DM. Pre-existing conditions, placebo reactions and "side effects". *Ann Int Med* 1964; **60**: 255–265.

31 Dantas F, Rampes H. Do homeopathic medicines provoke adverse effects? A systematic review. *Br Hom J* 2000; **89**(Suppl 1): S35–S38.

32 Greenland S. Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994; **140**: 290–296.

33 Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; **282**: 1054–1060.

34 Bhandari M, Richards RR, Prague S, Schemitsch EH. Quality in the reporting of randomized trials in surgery: is the Jadad scale reliable? *Control Clin Trials* 2001; **22**: 687–688.

35 Walach H, Sherr J, Schneider R, Shabi R, Bond A, Rieberer G. Homeopathic proving symptoms: result of a local, non-local or placebo process? A blinded, placebo-controlled pilot study. *Homeopathy* 2004; **93**: 179–185.

36 Fisher P, Dantas F. Homeopathic pathogenetic trials of Acidum malicum and Acidum ascorbicum. *Br Hom J* 2001; **90**: 118–125.

37 Signorini A, Lubrano A, Manuele G, *et al*. Classical and new proving methodology: Provings of *Plumbum metallicum* and *Piper methysticum* and comparison with a classical proving of *Plumbum metallicum*. *Homeopathy* 2005; **94**: 164–174.

38 Brien S, Lewith G, Bryant T. Ultramolecular homeopathy has no observable clinical effects. A randomized, double-blind, placebo-controlled proving trial of Belladonna 30C. *Br J Clin Pharmacol*. 2003; **56**: 562–568.

39 Riley DS. Contemporary drug provings. *J Am Inst Hom* 1994; **87**: 161–165.

40 Calabrese EJ. Paradigm lost, paradigm found: the re-emergence of hormesis as a fundamental dose response model in the toxicological sciences. *Environ Pollut* 2005; **138**: 379–411.