Homeopathy (2011) 100, 175–182

doi:10.1016/j.homp.2010.11.005, available online at http://www.sciencedirect.com

CLINICAL

Comparison of effectiveness of frequently and infrequently used homeopathic medicines

ALB Rutten*

Commissie Methode en Validering VHAN (Dutch Association of Homeopathic Physicians), The Netherlands

Rationale: **Patients treated with homeopathy may respond to infrequently used or even 'new' medicines. But does the introduction of an unlimited number of (new) medicines improve homeopathy? Do new medicines solve old problems?**
Methods: **1. Consensus meetings to evaluate best cases. 2. Patient outcome study in 10 Dutch practices.**
Results: **Good cases are scarce for many medicines, random variance is an important source of uncertainty. 50 Medicines are responsible for 72% of all successful prescriptions. There is no difference in effectiveness of frequently and less frequently used medicines. Confirmation bias is found for a few well-known symptom-medicine combinations.**
Conclusion: **'New' and infrequently medicines are as effective as 'old' frequently used medicines. Improving the use of frequently used medicines is more effective in improving results than seldom used medicines. Large numbers are required and old mistakes should be avoided developing new medicines. A research agenda for improving the use of homeopathic medicines is imperative.** *Homeopathy* (2011) **100**, 175–182.

Keywords: Homeopathy; Bayes' theorem; Effectiveness; Causal relationship; Confirmation bias

Introduction

It is a common experience that apparently well-indicated homeopathic medicines do not, or only temporarily, work. Sometimes infrequently used medicines succeeded after several well-known medicines failed. It is intriguing to find and use 'new' medicines, but do they increase the success of homeopathy? Is a new medicine that worked after several old medicines failed a better medicine with a higher success rate? Does the success rate after our first, second, or subsequent prescriptions increase by adding new medicines? New medicines could suffer from the same or other shortcomings as the old ones.

Causes of failure are potentially diverse; most relevant here are inaccuracies in the Materia Medica and Repertories and an insufficient number of medicines. This paper does not question if new medicines work or not. Probably many new and effective medicines are to be found, but we already have hundreds of infrequently used medicines in the existing Materia Medica. Are the existing medicines optimally used: only prescribed when they should be and not prescribed when they should not be? Is each new medicine a useful addition, just because patients improve after it? We have insufficient scientific guidelines to discover new medicines that could optimise homeopathy. There is a need for a strategy to improve homeopathy by introducing, and validating the indications for new medicines.

What if errors in the Materia Medica and Repertories are the most important source of failures? If the same errors occur with new medicines much of the effort will be in vain. Such errors include:

- Misinterpreting causal relationship between medicine and effect
- Random variance
- Confirmation bias
- Systematic mistakes.

*Correspondence: Lex Rutten, Aard 10, 4813 NN Breda, The Netherlands.
E-mail: lexrtn@concepts.nl, http://www.dokterrutten.nl/inc.html
Received 13 April 2010; revised 15 September 2010; accepted 4 November 2010

Causal relationship

Many successful treatments are in fact spontaneous recoveries or due to context effects instead of the prescribed

medicine. Assessment of the case by the patient and his doctor is a poor benchmark for establishing the relationship between improvement and the prescribed medicine. Probably many databases of 'cured' cases for specific medicines are contaminated by 'cures' attributable to causes other than the prescribed medicine. This decreases the strength of the indicating symptoms in these databases.[1] According to Walach *et al.* context effects of Conventional and Alternative Medicine (CAM) could be higher than of conventional medicine.[2] But Nuhn *et al.* found no confirmation for this higher context effect for homeopathy.[3]

Causal relationship between cure and medicine could be established by proper case descriptions,[4] but scoring systems might help. The Glasgow Homeopathic Hospital Outcome Scale (GHHOS), later referred to as Outcome in Relation to Impact on Daily Living (ORIDL), measures patient-rated outcomes.[5] Improvement on this scale does not indicate causal relationship, but 'homeopathic improvement' might indicate a causal relationship. Homeopathic improvement implies improved wellbeing and other complaints besides improvement of the presented complaint; this is reflected by a GHHOS/ORIDL score >1. Peer review could be applied to assess if improvement could be due to other factors than the prescribed medicine. Sustained improvement might also suggest a causal relationship.

Random variance

Much information about our medicines is based on few cases or homeopathic pathogenetic trials (HPTs) with few participants. Such information can be misleading. This can be explained by basic statistics. Intuitively we know that it is less probable to throw six with a dice four times in a row than heads with a coin four times in a row. This is a simple chance calculations, shown in (Table 1).

The analogy between a coin, a dice and homeopathic symptoms is that some symptoms, for instance headache, occur frequently – like heads on a coin. Other symptoms occur less frequently, like six on a dice. Suppose headache occurs in 50% of the patients that respond well to medicine X. One in eight practitioners will see three patients in a row whose headache responds well to medicine X, by chance alone. This practitioner will think that headache is a strong indication for medicine X, while it is not. Or consider eight symptoms, each occurring in half of the patients responding well to medicine X. By mere chance one of theses eight symptoms will occur in three patients in a row. These examples show that we must be careful with frequently occurring symptoms if we have a small number of cases, say less than five. The chance that the prevalence of such symptoms in a population responding well to a specific medicine is over- or under-estimated is considerable.

The use of many symptoms to recognise a specific homeopathic medicine is also a source of error. Suppose the population responding well to medicine X has 10 symptoms each with a prevalence of 5%. Of two patients responding well to X, one of these 10 symptoms will occur in one patient by mere chance. But now this symptom has a prevalence of 50% in your small population. The formula is simple: if the number of questions you ask is one divided by the prevalence, and if this is the same for every symptom (in this case 1/0.05 = 20), you have one spurious symptom by random variance.

Confirmation bias

The calculations on random variance show when experience could be further misleading because of confirmation bias. Suppose that half of the population responding well to medicine X is chilly. As in the coin tossing case, there is a chance of 1:16 that a doctor will see four successive cases of chilly patients responding well to medicine X. In the experience of this doctor all patients responding well to medicine X are chilly. So, if the next patient is not chilly this doctor will hesitate to prescribe X. If this doctor is a teacher and is asked about his experience with medicine X he will say that these patients are chilly, subsequently all his pupils will hesitate to prescribe X to patients that are not chilly. Much of our knowledge about homeopathic medicines is based on a small number of cases and the influence of chance is inevitable. McKenzie states that the impact of confirmation bias tends to be larger in familiar situations.[6] Familiarity is greater if a medicine is linked to a small number of symptoms and could be greater still if a medicine is described on theoretical grounds or on signature or symbolic grounds.

Systematic mistakes

According to Bayes' theorem a symptom is an indication for a specific medicine if it occurs more frequently in the population responding well to the medicine than in the remainder of the population. The most important systematic mistake in homeopathy is that symptoms are added to Materia Medica and Repertories based on absolute occurrence instead of prevalence.

This happens with clinical experience, but also in HPTs. In an HPT in India seven out of 22 (31.8%) subjects had headache. Based on this figure one could conclude that headache is an indication for this medicine. However, this study had a large control group and 486 out of 1000 (48.6%) of this control group had headache.[7] Headache is therefore not an indication for the medicine, but in the existing methodology of translating HPTs into Materia Medica and Repertory the symptom would be entered in second (italic) grade. The main problem of frequently

Table 1 Chances for repetitions

	1 Time	*2 Times in a row*	*3 Times in a row*	*4 Times in a row*
Six with a dice	1 in 6	1 in 6 × 6 = 1 in 36	1 in 6^3 = 1 in 216	1 in 6^4 = 1 in 1296
Heads with a coin	1 in 2	1 in 2 × 2 = 1 in 4	1 in 2^3 = 1 in 8	1 in 2^4 = 1 in 16

used existing medicines is such incorrect additions to the repertory. Frei showed that using Bayes' theorem with Polarity Analysis increased effectiveness of the first prescription in a trial from 27% to 47%.[8,9] Computerised data-gathering could improve the success of existing medicines. Prospective research on six homeopathic symptoms showed that possibly half of the entries in the repertories is misleading.[10]

Towards a rational strategy

Infrequently used new medicines suffer from unreliable data because of random variance due to small numbers. New medicines should fill a real gap, not a gap caused by improper use of existing medicines. The problems mentioned above should be avoided in introducing new medicines, and the information on which their indications are based should be accessible. Homeopathic medicines are increasingly regulated, registration of medicines is expensive and worthwhile only for medicines which are sufficiently used.

We need data about existing practice to develop a rational strategy for improving homeopathic prescribing. Numerous new medicines have been introduced in recent years. How many medicines do we use now? How effective are these medicines? Evaluation of the effectiveness of frequently used medicines is hampered by false information (adding symptoms based on absolute occurrence), but are infrequently used medicines therefore more effective? What is the influence of misinterpreting causal relationship, random variance and confirmation bias? We attempted to elucidate these issues in two projects in the Netherlands which validated practice experience:

- Consensus meetings – Materia Medica Validation (MMV)
- Prospective assessment of six homeopathic symptoms (LR-project).

The latter project provides a large database of clinical experience. New medicines and infrequently used medicines in the existing Materia Medica both have few symptoms to rely on. Our prospective database could show differences between frequently and infrequently prescribed medicines.

Methods

The Dutch Committee for Methods and Validation organised two projects to validate homeopathic data: consensus meetings to evaluate best cases of specific medicines (MMV) and a prospective assessment of the relationship between six homeopathic symptoms and treatment outcome (LR-project). The two projects were closely related because the participants of the prospective research were recruited from the MMV participants, and MMV meetings were also used for training.

Consensus meetings

Twice a year between 1997 and 2007 Dutch doctors were invited to present their best cases. Each meeting was attended by 10–20 doctors presenting with best cases with two medicines. Cases were required to have a follow-up of at least 1 year and a result according to GHHOS/ORIDL of +3 or 4. The participants were warned of methodological pitfalls including causal relationship and confirmation bias. During the meetings all cases were assessed by peers for the scale of the effect and the causal relationship between medicine and effect.

Prospective assessment of symptoms

This project has previously been described in this journal.[11] The LR-project, an observational study, was conducted from June 2004 until December 2007 and included all consecutive new patients older than 2 years. The goal was to assess the relationship between symptoms and successful outcome. The observers were 10 Dutch medically qualified doctors with more than 10 years experience in homeopathy, all were extensively trained in assessing cases and confirmation bias during the MMV. Assessment of symptoms and results were discussed in consensus meetings before and during the project. Cases were regarded successful if they had a GHHOS ≥ 2 and a follow-up period longer than 3 months. GHHOS ≥ 2 was rated only if a causal relationship was suspected based on clinical judgement. If the effect declined within 1 year the GHHOS score was adjusted accordingly. Six symptoms were recorded: 'diarrhoea from anticipation', 'fear of death', 'recurrent herpes of the lips', 'grinding teeth during sleep', 'sensitivity to injustice' and 'loquacity'. Loss to follow-up was not included in the analysis. The population responding well to a specific medicine (GHHOS ≥ 2) was used to calculate the prevalence of the symptom in the 'medicine-population'; the remainder of the population served as control.

The data were analysed with Excel® spreadsheets, the statistical programs Epi-info® and SPSS®. Confidence intervals (CIs) for this paper were calculated with Confidence Interval Analyser (BMJ), Wilson method or log method for likelihood ratios (LRs).

Results

Consensus meetings

In the MMV meetings 24 medicines were validated. In most practices, even after more than 10 years experience, cases with result GHHOS 3–4 and a follow-up longer than 1 year are scarce. The number of cases studied in each meeting varied between five (for *Naja*) and 23 (for *Sulphur*). One of the most striking results of these meetings was that the absence of an expected symptom cannot be regarded as an absolute contra-indication for a medicine. In other words: there were many cases without the symptoms considered characteristic. Some examples are: 'sensitive to injustice' occurred in 40% (4/10) of the best *Causticum* cases, 'loquacity' in 44% (7/16) of the best *Lachesis* cases and 'fear of dark' in 42% (4/12) of the best *Stramonium* cases.

During the validation of *Sulphur* 23 cases from 15 doctors were evaluated. Only one of the 23 patients (4.3%) had

the symptom 'fear of death'. According to the existing methodology *Sulphur* should be added to the corresponding repertory rubric because of absolute occurrence. The doctor who presented this case had two *Sulphur* cases. For this doctor 'fear of death' was present in 50% of the *Sulphur* cases, and therefore an important indication for *Sulphur*. But for the group of 15 practitioners the symptom 'fear of death' would not be an indication for or against *Sulphur*, given the prevalence in the general population of about 4%.

Prospective research

4094 patients were included in the LR-project; 2752 (67.2%) female. Mean age was 39.6 (standard deviation 21, range 3–95). The number of evaluated prescriptions was 4072, the follow-up after each prescription was at least 3 months, maximum 3.5 years.

The number of results with GHHOS +2–4 recorded in different practices varied between 29% and 79%, see Figure 1. These figures are probably not due to differences in competence between doctors, or doctors using different medicines. During MMV meetings and consensus meeting it appeared that interpretation of result is rather subjective. The subjectivity in assessing causal relationship (established by clinical judgement) could also be a source of variance. Results in early stages were not handled equally by all doctors: some recorded results after 1 month. This could lead to underestimation of result if the patient did not return after the second consultation. There was also much variance between the number of medicines each doctor used, between 35 and 295 different medicines, there is no clear explanation for this, see Figure 2. There is a possible moderate negative correlation between the number of prescribed medicines and the success rate suggesting that prescribing more medicines is associated with poorer results, the Pearson correlation is -0.581 ($p = 0.078$, two-sided). However, no firm conclusions can be drawn from this correlation because of the subjectivity in estimating results.

During the research period 421 different medicines were prescribed. Of all prescribed medicines 75 were responsible for more than four successful prescriptions per medicine. For 90 medicines two to four successes were recorded. One successful prescription was recorded for

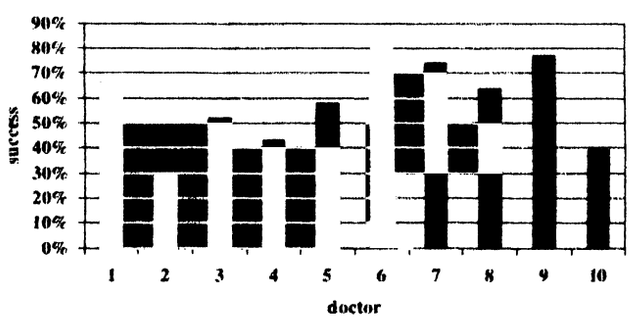

Figure 1 Success rates as recorded by 10 doctors during a patient outcome study assessing the relationship between six symptoms and specific medicines.

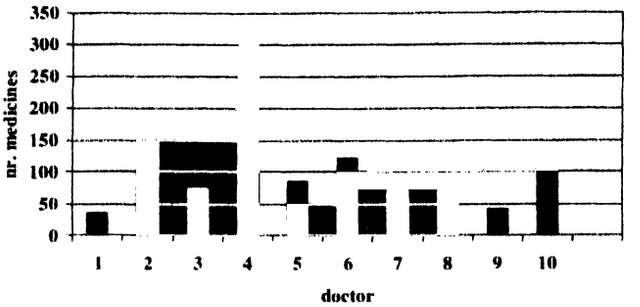

Figure 2 The number of different medicines used by 10 doctors.

128 medicines, and the same number of medicines rendered no result, see Figure 3.

The 10 most frequently prescribed medicines were responsible for 38% of all successful prescriptions, the second 10 most frequently prescribed medicines added 15% of all successes. The 50 most prescribed medicines were responsible for 72% of all successful prescriptions, see Figure 4.

The overall success rate (GHHOS 2–4) of all 4072 prescriptions was 50.8%. Table 2 shows the success rates of the 20 most prescribed medicines. Apart from less success with *Thuja* and *Nitricum acidum* there are no clear differences between the medicines. The success rate of the 20 most prescribed medicines is 1080/2019 = 53.5% (95% CI 51.3–55.7).

Table 3 shows the success rate of the least frequently prescribed medicines. The once and twice prescribed medicines seem less successful than the 20 most prescribed medicines.

The success rate of subsequent groups of 10 medicines in descending order of use was calculated for the 200 most prescribed medicines and is shown in Figure 5. There is no clear difference between frequently prescribed medicines and seldom prescribed medicines and no discernible trend.

Natrum muriaticum was the most frequently prescribed medicine with 248 prescriptions and 63% success. We investigated what medicines were prescribed after *Nat-m.*

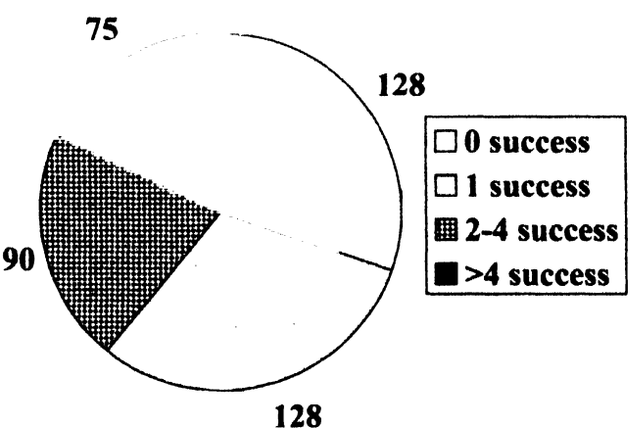

Figure 3 Number of successful prescriptions of 421 different medicines.

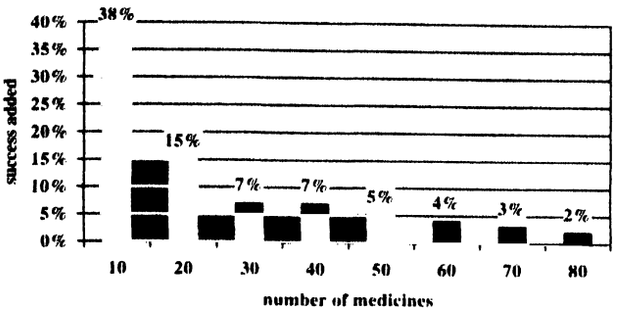

Figure 4 Success vs frequency of prescriptions. The 10 most prescribed medicines account for 38% of all successes, the second 10 added 15%, and so on.

failed (92 cases). For these 92 cases there were 51 successful prescriptions of other medicines, 28 (55%) of these were among the 50 most prescribed medicines.

To investigate confirmation bias we compared results in different stages. Some results after 10 months differed from results at the conclusion of the study (3.5 years). After longer follow-up more medicines were prescribed without characteristic symptoms. For example the symptom, 'sensitivity to injustice' was present in every patient prescribed *Causticum* at the onset of the study but after 2 years the prevalence of the symptom was only 40% in the population responding well to *Causticum* (Figure 4). During that time the medicine *Causticum* was prescribed on other grounds to patients who were not sensitive to injustice. The LR+ developed towards 4.39 in the end. We found similar but less marked evolution for the combinations *Arsenicum* – 'fear of death', *Mercurius* – 'grinding teeth during sleep' and *Lachesis* – 'loquacity'. After 2 years the prevalence of these symptoms stabilises (Figure 6). The number of symptom-medicine combinations that indicated confirmation bias was small and confined to well-known symptom-medicine combinations.

Table 2 20 most frequently used medicines and their success rate. *N* = number of prescriptions of each medicine

	N	*Success %*	*95% CI*
Nat-m	248	62.9	56.7–68.7
Sep	177	52.5	45.2–59.8
Sulph	175	50.3	43.0–57.6
Lyc	161	53.4	45.7–61.0
Calc	143	52.4	44.3–60.5
Phos	127	59.8	51.1–68.0
Puls	105	56.2	46.6–65.3
Merc	94	57.4	47.4–67.0
Caust	88	52.3	42.0–62.4
Carc	79	54.4	43.5–65.0
Staph	73	43.8	33.0–55.2
Nux-v	70	57.1	45.5–68.1
Ign	67	49.3	37.7–60.9
Lach	65	64.6	52.5–75.1
Sil	65	50.8	38.9–62.5
Thuj	63	38.1	27.1–50.4
Calc-p	60	46.7	34.6–59.1
Arg-n	56	46.4	34.0–59.3
Graph	54	50.0	37.1–62.9
Nit-ac	49	38.9	26.4–52.8
	2019	53.5	51.3–55.7

Table 3 Once or twice used medicines and their success rate. *N* = number of different medicines

Times prescribed	N	*Success %*	*95% CI*
Once	181	45.0	37.7–52.0
Twice	59	47.5	38.7–56.4
Once and twice	240	45.8	40.3–51.5

Of 88 *Sulphur* cases only one (1.1%) had the symptom 'fear of death'. The prevalence of 'fear of death' in the whole population of 4094 patients was 3.9%, LR = 0.29 (95% CI 0.041–2.048). The symptom 'fear of death' is therefore a contra-indication for *Sulphur*. According to the prevalence of 'fear of death' of 4.4% in the 23 patients in MMV this symptom was not an indication for or against *Sulphur*. For one doctor attending MMV the symptom 'fear of death' was a strong indication for *Sulphur* because of a prevalence of 50% according to his experience.

Discussion

Theoretically there are many pitfalls that can hamper effectiveness of homeopathic prescribing, the most important ones being use of absolute rather than relative prevalence, spontaneous recovery, random variance and confirmation bias. Frequent use of a medicine reduces errors due to variance, but a new source of error is introduced by frequent use if we use absolute rather than relative prevalence. Outcome of daily practice should tell us the extent of such problems. Systematic recording and evaluation of cases in homeopathy is rather scarce, despite the fact that practice experience is one of the cornerstones of homeopathy. Our assessments of daily practice showed that there is much variance in prescribing and interpretation of results. Our retrospective consensus meetings showed that the number of cases with long-standing excellent results is limited.

Our prospective observational study was designed to assess the relationship between symptoms and successful prescriptions, not to test effectiveness of homeopathy or of individual prescribers; the latter would require more standardisation of assessing results. Because of subjectivity in interpretation of results we cannot conclude that doctors who use more medicines are better or worse

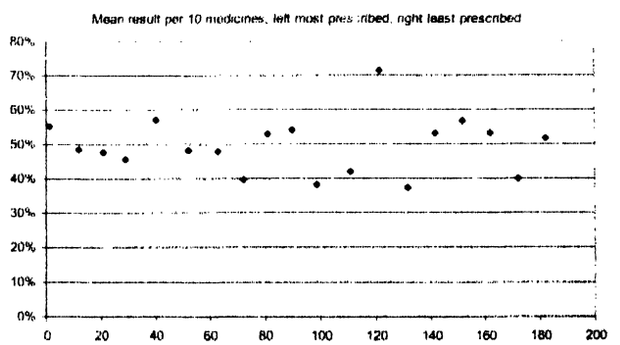

Figure 5 Success rate of medicines ranged by groups of 10 in order of frequency of use.

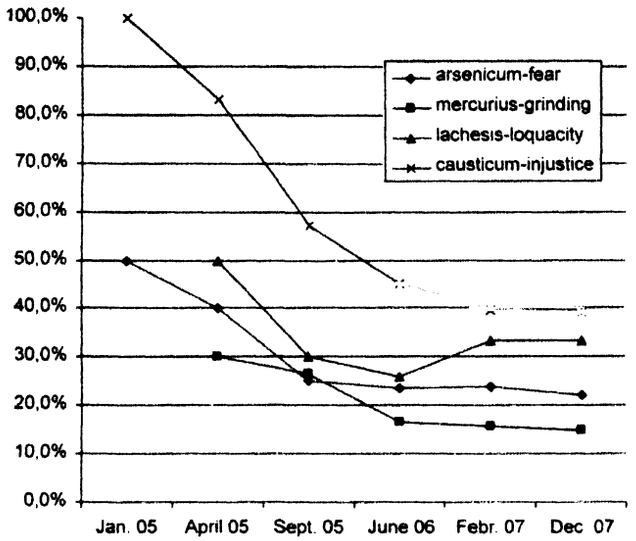

Figure 6 Decreasing score as a function of time of exaggerated success for a small number of symptom-medicine combinations. The exaggerated success is due to confirmation bias shortly after the first prescription.

prescribers than doctors who use few medicines. But the most frequently prescribed medicines were largely the same for all doctors and all doctors used infrequently prescribed medicines. Our results give some indication about the difference regarding effectiveness between frequently and infrequently prescribed medicines.

The number of different medicines was 421. There were no clear indications that frequently used medicines perform better than seldom used medicines, nor for the opposite. The 50 most prescribed medicines were responsible for 72% of all successful prescriptions. After *Natrium muriaticum* failed 55% of the subsequent successful prescriptions were with the 50 most prescribed medicines. This and Frei's results with Polarity Analysis (which amounts to Bayesian correction of the repertory) show that the use of the most prescribed medicines can be improved.

Both our retrospective and prospective research indicated that there is confirmation bias, especially in well-known combinations of symptoms and medicines. This is consistent with existing literature.[6] The observation that about 40% of all *Causticum* patients are sensitive to injustice is consistent with the outcome of MMV. In the LR-project 30 patients had strong sensitivity to injustice and received *Causticum*. In 12 out of these 30 cases (40%) *Causticum* had no effect. In other words if we rely too much on the symptom 'sensitive to injustice' in prescribing *Causticum* we miss 60% of potential responders to *Causticum* and still have 40% failures.

Our experience with the symptom 'fear of death' and *Sulphur* showed that due to random variance this symptom might be regarded as a strong indication for *Sulphur* by one out of 15 doctors, as no indication for or against *Sulphur* by 15 doctors, and as a contra-indication against *Sulphur* after prospective research in 4094 patients and 88 *Sulphur* cases. This example shows the influence of random variance. Even a consensus procedure of more than 20 best cases can yield misleading results.

Our prospective research does not permit conclusions about effectiveness of homeopathy as a method because loss to follow-up was not assessed. Our aim was to calculate LRs of symptoms in successful cases. Another limitation was the lack of independent observation of results, this was not feasible in constant daily monitoring of results in 10 practices. Despite these shortcomings we think that our results give some indication about comparative effectiveness of different medicines.

There is no indication that 'new' medicines perform better than 'old' medicines. Probably the same mechanisms of failure apply for old and new medicines. It would be a waste of resources if we failed to handle these problems.

Absolute or relative occurrence

The present method of entering symptoms in our Materia Medica and Repertories based on absolute occurrence is a serious source of error. The HPT cited in the Introduction, our MMV and our LR-project showed that absolute occurrence can be misleading. Our LR-project indicates that for six investigated symptoms about half of the entries in Kent's repertory does not correspond with systematic assessment. A symptom is indicative for a specific medicine only if the prevalence of the symptom in the medicine population is greater than the prevalence of the same symptom in the remainder of the population. It is therefore vital to know or to be able to estimate the prevalence of each symptom in the whole population. This can be achieved by systematic data collection. Another option could be an online panel as used for market research. Such a panel of a representative sample of our patient population could be addressed with interviews about all kinds of symptoms.

Random variance

In the prospective research there were only 75 medicines with more than four successful prescriptions. This database was gathered by a group of 10 experienced doctors over 3.5 years by meticulous registration of their work. According to the calculations above we should be careful about the interpretation of daily occurring symptoms if we have less than five cases. But our example of 'fear of death' and *Sulphur* shows that even 20 cases can be insufficient to reveal the true value of a symptom. Dealing with random variance means that we should always be aware of this problem; it should be mentioned when presenting cases or CIs should be reported. We need large numbers, and that can only be achieved by systematic data collection.

Causal relationship

After critical evaluation by peers during retrospective MMV each doctor had just a few good cases with likely causal relationship between medicine and effect, even for frequently used medicines. Our criteria for success in the prospective research project were less strict regarding GHHOS score and duration of improvement. The GHHOS score does not assess causal relationship and this should be

Table 4 Proposed adaptation of Naranjo's algorithm for homeopathy

	Yes	*No*	*Don't know*
1. Was the case similar to other cases with this medicine?	+1	0	0
2. Did the effect appear after administration of the medicine?	+1	−1	0
3. Did the effect after one dose subside after a period of time?	+1	0	0
4. Was the improvement resumed after repeated administration of the medicine?	+2	1	0
5. Was there an initial aggravation?	+1	0	0
6. Did the effect comprise more than the presented complaint, e.g. wellbeing and other complaints, i.e. GHHOS/ ORIDL score two or higher?	+2	0	0
7. Did the course of improvement follow Hering's rule?	+2	0	0
8. Did old symptoms reappear for a while in the course of the improvement?	+1	0	0
9. Are there alternate causes (other than the medicine) that solely could have caused the improvement?	−3	+1	0
10. Did the patient have the same response to other homeopathic medicines?	−1	+1	0
11. Was the effect confirmed by objective evidence?	+1	0	0

estimated by clinical judgement. The participants were well trained in assessing causal relationship, but the variance of assessed results between doctors could indicate an unclear threshold.

There are no clear rules for assessing causal relationship before entering clinical or HPT results into our Materia Medica and Repertories. For assessing causal relationship in homeopathy it might be useful to develop algorithms. Algorithms for establishing causal relationship, like Naranjo's algorithm, are already available for adverse drug reactions.[12,13] Such algorithms could be adapted for homeopathy, a concept for clinical cases is shown in Table 4. The weighting of each item and the best threshold value for this algorithm should be discussed and validated.

Confirmation bias

If we were to rely too much on the symptom 'sensitive to injustice' in prescribing *Causticum* we would miss 60% of potential 'cures' by *Causticum* and have 40% failures after using this medicine. The impact of confirmation bias could be variable. As stated above, the impact of confirmation bias might be quite limited, but larger in familiar situations. This seems to be confirmed in our research, the relationship between 'sensitivity to injustice' and the medicine *Causticum* is the most well known of the relations we assessed. The other combinations, 'fear of death' – *Arsenicum*, 'grinding teeth during sleep' – *Mercurius* and 'loquacity' – *Lachesis* are also well known and the only combinations where confirmation bias could be demonstrated in our prospective research. Confirmation bias might be a problem when only a few symptoms are known for a medicine. This requires further investigation. Our research showed that follow-up longer than 2 years could reduce confirmation bias.

Evaluation

We can be sure that some failures of homeopathy are caused by systematic mistakes in collection and interpretation of data. The hypothesis that new medicines can be the solution should be validated. The number of potential medicines could be thousands, but the number of medicines actually used in practice is limited to hundreds. Most successes are achieved with 50 medicines. Failures arise when a medicine is prescribed when it should not be prescribed, or not prescribed when it should be. Frequently prescribed medicines are still frequently not prescribed when they should be, and their use could therefore be optimised.

The purpose of new medicines should be to fill a gap between existing medicines. It is improbable, and our findings confirm this, that new medicines will be less inappropriately used. If we improve the effectiveness of the 20 most prescribed medicines (responsible for half of our successes) by 20% the effectiveness of homeopathy could rise from, say, 50–60%. Improving 20 medicines with a combined use of 5% by the same amount would result in only 1% improvement of the effectiveness. Therefore improving the use of our most frequently prescribed medicines should have priority.

It doesn't make sense to neglect the precautions for existing medicines while investigating new medicines. Collecting a sufficient number of cases will be the greatest challenge. Thorough measures like making a research agenda with a preferred list of medicines, international co-operation and much more (and adequate) registration of cases seem inevitable. Improvement of success is hard to measure if variance between doctors in assessing results is as large as in our study. We need new studies designed for this purpose and the use of an algorithm to establish causal relationship.

Conclusion

Most cases in homeopathy are successfully treated by 50 medicines, but many cases require medicines outside this group. Many uncertainties still exist about the number and the optimal choice of these new or little used medicines. The existing Materia Medica and Repertories contain many mistakes, partly due to avoidable systematic shortcomings, like use of absolute instead of relative occurrence and confirmation bias. Random variance is another source of mistakes. 'New' medicines are not more effective than 'old' medicines. Homeopathy would benefit the most from improving the use of the most frequently prescribed medicines first. New medicines should be developed without the shortcomings of the old ones. The greatest challenge is to obtain sufficient numbers. A research agenda setting priorities is required. We need an instrument to assess causal relationship between medicine and improvement.

Conflict of interest

There were no competing interests.

Acknowledgements

Thanks to José Eizayaga for his comments. The doctors participating in the LR-project were Rob Barthels, Hetty Buitelaar, Paul Fruijtier, Gerard Jansen, Jean Pierre Jansen, Stan Jesmiatka, Christien Klein, Roland Lugten, René van der Reijden, Lex Rutten, Erik Stolper, Janny Verhey and Mechtild Wijdeveld.

References

1 Rutten ALB, Stolper CF, Lugten RF, Barthels RJ. 'Cure' as the gold standard for likelihood ratio assessment: theoretical considerations. *Homeopathy* 2004; **93**: 78–83.

2 Walach H, Falkenberg T, Fonnebo V, Lewith G, Jonas WB. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol* 2006; **6**: 29. 10.1186/1471-2288-6-29.

3 Nuhn T, Lüdtke R, Geraedts M. Placebo effect sizes in homeopathic compared to conventional drugs – a systematic review of randomised controlled trials. *Homeopathy* 2010; **99**: 76–82.

4 Kienle GS, Kiene H. Methodik der Einzelfallbeschreibung. *Der Merkurstab* 2009; **62**(3): 239–242.

5 Reilly D, Mercer SW, Bikker AP, Harrison T. Outcome related to impact on daily living: preliminary validation of the ORIDL instrument. *BMC Health Serv Res* 2007; **7**: 139. doi:10.1186/1472-6963-7-139.

6 McKenzie CR. Increased sensitivity to differentially diagnostic answers using familiar materials: implications for confirmation bias. *Mem Cognit* 2006; **34**(3): 577–588.

7 Coutinho N. Personal communication.

8 Frei H. Polarity analysis, a new approach to increase the precision of homeopathic prescriptions. *Homeopathy* 2009; **98**: 49–55.

9 Rutten ALB, Frei H. Opposite repertory-rubrics in Bayesian perspective. *Homeopathy* 2010; **99**: 113–118.

10 Wassenhoven M van. Towards an evidence-based repertory: clinical evaluation of Veratrum album. *Homeopathy* 2004; **93**: 71–77.

11 Rutten ALB, Stolper CF, Lugten RF, Barthels RJ. Statistical analysis of six repertory-rubrics after prospective assessment applying Bayes' theorem. *Homeopathy* 2009; **98**: 26–34.

12 Naranjo CA, Busto U, Sellers EM, *et al*. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981; **80**: 289–295.

13 Arulmani R, Rajendran SD, Suresh B. Adverse drug reaction monitoring in a secondary care hospital in South India. *Br J Clin Pharmacol* 2008; **65**(2): 210–216.