Updating the Homeopathic Algorithms: Handling Confirmation Bias

Lex Rutten[1] Rainer Schäferkordt[2] José E. Eizayaga[3]

[1] Independent Researcher, Breda, The Netherlands
[2] Wisshom, Deutscher Zentralverein homöopathischer Ärzte, Boizenburg, Germany
[3] Department of Homeopathy, Maimonides University, Buenos Aires, Argentina

Address for correspondence Lex Rutten MD, Independent Researcher, Aard 10, 4813NN Breda, The Netherlands (e-mail: praktijk@dokterrutten.nl).

Homeopathy

Abstract

Background Homeopathy has always used algorithms, such as giving more weight to peculiar symptoms and repertorisation of symptoms for differential diagnosis of medicines. However, repertory entries are flawed and homeopathic data are liable to heuristic bias. Modernising the homeopathic repertory with statistical tools, such as Bayes' theorem, should be accompanied by handling (confirmation) bias.

Methods After systematic collection of 731 'Best Chronic Homeopathic Cases' (BCHC), we analysed patterns in the frequency distribution of likelihood ratios (LRs). We did the same with an existing Bayesian repertory based on historical materia medica data of more uncertain quality. The frequency distributions are assessed with theoretical considerations, mathematical tools such as (exponential) transformations and differentiation, and expert knowledge.

Findings The frequency distributions of LRs both showed the same two patterns: the middle part of the frequency distribution showed a loglinear progression, but at both ends there was an increasing slope of the curve. The confirmation bias in the middle part of the LRs can be corrected mathematically with exponentiation (power calculations). Clinical expertise and differentiation of the curve indicate LR = 7 as an eligible maximum for the vast majority of symptoms. There was no clear difference between the BCHC and the historical materia medica data in this respect.

Conclusion It is possible to correct partly for confirmation bias in a repertorisation algorithm by a combination of theoretical consideration, expert knowledge and mathematics. We found a striking similarity between the BCHC and historical data regarding confirmation bias.

Keywords
- data collection
- homeopathy
- materia medica
- confirmation bias

Introduction

An algorithm is defined as "a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer" (https://languages.oup.com/google-dictionary-en/). Homeopathy is a medical method that has used algorithms from the beginning, the most well-known being Hahnemann's aphorism 153, stating that "the more striking, singular, uncommon and peculiar (characteristic) signs and symptoms" are the most important.[1] This algorithm can easily be embedded in a computer program because we can translate it into numbers and a formula: the prevalence of a peculiar symptom is low. Such symptoms are generally represented in the homeopathic repertory in small symptom rubrics. By increasing the weight of small rubrics in repertorisation software, these rubrics get more attention. Repertorisation is also

received
March 5, 2025
accepted after revision
May 12, 2025

Georg Thieme Verlag KG, Oswald-Hesse-Straße 50, 70469 Stuttgart, Germany

DOI https://doi.org/10.1055/a-2606-4041.
ISSN 1475-4916.

an algorithmic approach: a matrix of symptoms and corresponding medicines, combined with ordering of medicines according to the strength of the relationship with symptoms. A combination of symptoms comes closer to the 'totality' of the patient, but homeopathic practitioners realise that the complexity of this 'totality' is much more than repertorisation could express. This does not preclude the use of repertorisations provided we realise the different functions of materia medica and repertory. The repertory is an index to the materia medica, with less precision of the personal expression of each symptom, but more oversight of a combination of symptoms and a global differentiation between several eligible medicines. Used this way, a repertorisation can be an instrument for finding the 'totality', but not without knowledge of the materia medica.

The different functions of materia medica and repertory are not always respected. We often see repertorisations where the prescribed medicine was in the first position to illustrate why this particular medicine was prescribed. This is *post-hoc* reasoning, the repertorisation was adapted to an improper purpose. The *post-hoc* choice of symptoms that places the prescribed medicine at the first position is based on knowledge of both materia medica and repertory. Many repertory rubrics are ambiguous in semantic and functional respects. For instance, the emotion anger/irritability has many synonymous rubrics in the repertory containing different medicines. Even concrete physical symptoms can cause semantic confusion: for instance, Kent's repertory uses 'emaciation' and 'marasmus' for losing weight, in combination with 'ravenous/canine/excessive appetite'. The practitioner using Kent's repertory who has prescribed *Cina* will enter 'marasmus' in the repertorisation; the prescription *Psorinum* will induce the choice of 'emaciation'. Different repertories will have different examples of this problem.

Adequate use of the repertory requires much study and experience. Practitioners have to be familiar with the structure of the repertory and its pitfalls. They learn to manage the shortcomings of the repertory intuitively. Computerised repertories can handle some of the problems, for instance by de-selecting frequently used medicines ('large medicines'). On the other hand, computerisation also caused a new problem by the ability to easily add new information. This amplified a basic flaw of the classic repertory: entries based on absolute occurrence of symptoms instead of prevalence. Mending this requires registration of the prevalence of symptoms and Bayesian analysis. This new approach would be a good opportunity to improve algorithms in homeopathy further.

There are two stages in our search for eligible medicines during a consultation: firstly, finding eligible medicines based on the most important symptoms that characterise the patient. Secondly, distinguishing between eligible medicines by checking for confirmatory symptoms for separate medicines. Confirmation by checking for keynote symptoms of different medicines is an important element of the homeopathic consultation. Many homeopathic practitioners know the emotion caused by the patient saying, 'How do you know that, doctor?', related to the fact that keynote symptoms occur frequently in patients responding well to specific medicines. The 'clairvoyance' of the doctor has a statistical background.

Confirmation is also a delicate instrument that can easily result in confirmation bias: preference for information that confirms an existing idea. If you have an angry impression of the patient in front of you, symptoms that confirm 'angry medicines' will be checked first. Confirmation bias also influences our knowledge about medicines, in our own experience and in shared experience in the materia medica. Our first experience with a cure by a specific medicine and its symptoms 'primes' us for those symptoms as indications for the medicine.[2] Shared experience with the same medicine is also likely to cause confirmation bias in our materia medica. Confirmation bias results in more awareness of specific symptoms that confirm an eligible medicine and possibly less if the medicine is not preferred. In statistical terms: the prevalence is over-estimated in the medicine population and under-estimated in the remainder of the population. Both factors increase the LR {LR = (prevalence in the medicine population)/(prevalence in the remainder of the population)} of the symptom–medicine relationship. A long-term assessment of six keynote symptoms showed that a long follow-up results in reduction of confirmation bias and resulted in LRs between 2 and 4 at the end despite very high LRs at the beginning of the assessment.[3] This gives us some indication of the LRs we can expect in the case of keynote symptoms.

Algorithms have pros and cons; the homeopathic algorithms are not different in that respect. Homeopathic practitioners have, by training and experience, knowledge of both materia medica and repertory and know how to use the repertory algorithm and avoid its pitfalls. Nevertheless, they will profit from a better repertory — especially inexperienced practitioners. Modern science and technology can help by technical solutions, such as thesaurus and refined algorithms embedded in continuously updatable computerised repertories.

Over-reliance on algorithms in homeopathy can undermine clinical judgment and even be counter-productive. Algorithm quality firstly depends on the quality of the inputted data: flawed or biased input results in poor output. Algorithm output also depends on correct statistical analysis and interpretation of data. Bias can occur in the whole process between the individual patient's presentation, via processing of aggregated data, up to the interface that guides the practitioner. The complexity of this whole process makes optimising the effectiveness of homeopathy a challenge. It is therefore obligatory that algorithms are explained and discussed. Some examples are the process of constituting a treatment algorithm for coronavirus disease 2019/viral infections,[4] the 'Best Chronic Homeopathic Cases' (BCHC) repertory,[5] and a Bayesian repertory of materia medica data.[6]

In this paper, we describe a process that resulted in an algorithm for handling extremely high likelihood ratios (LRs) in collections of best cases and in existing materia medica. Such high LRs tend to dominate repertorisations, resulting in

a preference for specific medicines that is not in accordance with clinical expertise. We demonstrate some methods that can be used to understand and correct for bias underlying this problem.

Materials and Methods

Materials

The materials underlying this paper are data from two sources:

1. BCHC[5]: a relatively small dataset of an ongoing collection of selected 'best cases' containing (at the time of this analysis) the aggregated data from a total of 731 de-identified patients, constituting a repertory of 41 medicines and 561 symptoms.
2. Bayesian repertory of materia medica data[6]: data of a repertory based on historical materia medica data, here referred to as 'materia medica data'. Sources are: drug provings, poisonings and clinical experiences. It is a large database with a broad diversity between frequently and infrequently used medicines, 1,185 in total, with approximately 775,000 symptoms in total.

The two databases (BCHC data and 'materia medica data') represent the existing variety of knowledge bases in homeopathy; the 'materia medica data' are mostly collected in a period where quality of data was not an explicit issue. The knowledge about validity of different sources is limited. Data are produced by many doctors with various levels of methodological knowledge. Therefore, the quality of data is mostly unknown.

The BCHC data are collected with a more structured method, considering quality aspects such as causality, heuristic bias and statistics, and training of practitioners in bias reduction. This method of data collection started in 1997 with consensus meetings (Materia Medica Validation) twice a year between 1997 and 2004, where on average 20 experienced Dutch doctors discussed their best chronic cases of pre-selected medicines with at least one year of follow up and other criteria such as one medicine causing the cure.[7] All meetings were structured following Delphi consensus procedures. About half of all cases were rejected during the consensus meeting, because there was insufficient certainty about quality of the case, especially about the certainty that one specific medicine caused the improvement and possible other causes of improvement. In total, 310 cases were collected this way. This was the start of continuing data collection by a group of doctors discussing and developing relevant methodological aspects. One of the outcomes of these discussions was the application of modified Naranjo criteria.[8] The remaining cases were later submitted by six doctors originating from this group and a group of three doctors from Argentina with similar training. To give an indication about case selection: for one data collection (author L.R.), about 17,000 cases out of 40 years of homeopathic practice were checked and 1.5% selected.

At the time of this analysis (January 2025) the BCHC repertory was based on at least five cases of each of the 41 most prescribed medicines. Doctors contributing cases had informed patients that their anonymised data could be used for research purposes if no objection. All patients received usual treatment. The BCHC data collection contains aggregated, de-identified, records only.

The historical materia medica repertory contains a much larger number of symptoms and variety of medicines (see above):

- 510 'small' medicines with <100 symptoms each.
- 259 'medium' medicines with 100 to 400 symptoms each.
- 416 'large' medicines with >400 symptoms each.

For analysis and comparability, the wordings of the symptoms in the materia medica dataset were synchronised with the BCHC dataset. This resulted in the following numbers of LRs:

- Small medicines: 31,969 LRs.
- Medium medicines: 59,198 LRs.
- Large medicines: 288,430 LRs.

Both data sets apply the same Bayesian approach regarding the prevalence of symptoms, in the whole database and in sub-sets confined to specific medicines. The historical materia medica dataset is much larger regarding the number of symptoms and medicines and serves as a reference for the BCHC dataset with less statistical uncertainty because of the large sample size.

Methods

The data were evaluated with a combination of methods.

Description

Graphical representation of frequency distributions of LRs illustrate how they are distributed, enabling theoretical considerations about the causes of the distributions. Mathematical transformation of the LRs and/or axes of the graph show if the data are linear, loglinear or otherwise distributed.

Theoretical Considerations

LR is not a linear measure because it expresses the relationship between prior and posterior odds instead of chance. Not all homeopathic symptoms are equally valued; this could explain further deviation from linearity. Psychological and sociological mechanisms could be responsible for bias.

Mathematical Tools

If there is a systematic development in (part of) the data, this could be expressed as a formula but also compensated by a formula. In this case, we used mathematical transformation by exponentiation (power calculation by Excel)[a] to compensate for confirmation bias and improve the spread of the data. Such transformations of data are also used in multivariate analysis, for instance Discriminant Analysis, to show a clearer difference between variables. Mathematical tools, such as differentiation, can be used to describe curves in a distribution.

[a] For instance: exponentiation of 4 by 0.5 (noted as $4^{0.5}$) results in 2. Power 0.5 is the same as square root.

Table 1 Translation of chance into odds and *vice versa* with a hypothetical prior for different likelihood ratios

Prior chance (%)	Prior odds	LR	Posterior odds	Posterior chance (%)
10.0	0.111	2	0.222	18.2
10.0	0.111	4	0.444	30.8
10.0	0.111	8	0.889	47.1
10.0	0.111	16	1.778	64.0
10.0	0.111	32	3.556	78.0
10.0	0.111	64	7.111	87.7
10.0	0.111	128	14.222	93.4
10.0	0.111	256	28.444	96.6

Abbreviation: LR, likelihood ratio.

Prior Outcomes and Comparisons

Previous prognostic factor research (PFR) projects give an indication of the range of LRs we can expect in comparable symptoms. One example is the long-term prospective evaluation of six 'keynote symptoms' mentioned in the Introduction.[3] A keynote symptom is a relatively common symptom with a relatively high prevalence in a specific medicine population. The two databases allowed for comparison of two homeopathy data sets of different origins: one with undefined and one with defined quality.

Expert Knowledge

Homeopathic practitioners have experiential and intuitive knowledge about the relative importance of many symptoms. They also have experience with several kinds of bias and have learned to use 'small medicines', medicines with few symptoms that are infrequently used.

Findings

The Bayesian formula with LR is intuitively the easiest to understand, but it calculates odds instead of chance. ►**Table 1** shows how chance relates to odds for different LRs, given a hypothetical prior chance of 10%. This table shows that LR = 16 will increase the chance that the medicine will work from 10 to 64%, a relatively high certainty for one symptom. Two symptoms with LR = 16 will have a combined certainty of LR = 256, resulting in 96.6% certainty of effect. Experienced practitioners will expect this kind of certainty only from very peculiar symptoms, with very low prevalence in the whole population. Three keynote symptoms are regarded as a sufficient indication for a reliable prescription by many practitioners.

Data collections often show some unexpected outcomes. ►**Table 2** shows three symptoms of the BCHC database with very high LRs (290.4 and 165.7) and their prevalence in the whole research population, and two slightly more prevalent symptoms ('Forsaken feeling and 'Desire for ice-cream') with much lower LRs. The prevalence of 0.3% up to 0.5% is lower than expected for relatively common symptoms. For instance, a European survey of 'non-restorative sleep' showed a prevalence of 10.8% (95% confidence interval, 10.4 to 11.2%).[9] This could be explained by the retrospective character of this data collection. The symptom was not checked in every patient, for instance in cases where the symptom did not confirm the prescribed medicine. On the other hand, the symptom could be readily noticed if the symptom confirmed the medicine. This causes over-estimation of the symptom in the medicine population and under-estimation in the remainder of the population, possibly resulting in extreme LRs.

►**Table 2** also shows two symptoms that are known as keynote symptoms, 'Forsaken feeling' for *Pulsatilla (Puls)* and 'Desire for ice-cream' for *Phosphorus (Phos)*. Both symptoms have LR = 4.6, which is consistent with values of keynote symptoms in previous research.[3] A set of three keynote symptoms with LR = 4, would result in a combined LR = $4 \times 4 \times 4 = 64$ and, according to ►**Table 1**, an increased probability of cure from 10 to 87.7%.

A frequency distribution of 1,683 LRs > 1 of the three most frequently used medicines (*Lycopodium*, *Natrium muriaticum* and *Phosphorus*) in the BCHC database is shown in ►**Fig. 1**. The *y*-axis has a logarithmic scale to correct for the exponential increase of LRs demonstrated in ►**Table 1**. This works well, showing a linear rise of LRs until LR = 4. Nevertheless, the logarithmic transformation is no longer sufficient for linearity above LR = 4, the curve rises exponentially and becomes a nearly vertical line about LR = 12. The sharp inclination of LRs shows at both ends of the curve, for LRs > 1 and LR < 1.

The much larger database of historical materia medica data shows a difference between small and large medicines (►**Figs. 2**, **3**). For small medicines, the graph (►**Fig. 2**) shows a hardly noticeable second acceleration of LR values at the right end, which is clearer in large medicines. Since the BCHC

Table 2 Variation in prevalence of relatively common symptoms in a contemporary collection of best cases, resulting in strong variation in likelihood ratios

Symptom	Prevalence (%)	Medicine	LR
01. Mind-SADNESS-Waking, on	0.3	*Calc p*	290.4
18. Sleep-UNREFRESHING	0.4	*Hep*	290.4
24. Modalities-PREGNANCY-during-agg	0.5	*Cocc*	165.7
01. Mind-FORSAKEN FEELING	2.2	*Puls*	4.6
10. Food-DESIRE FOR-Ice-cream	2.1	*Phos*	4.6

Abbreviations: Calc-p, Calcarea phosphorica; Cocc, Cocculus; Hep, Hepar sulphuris; LR, likelihood ratio; Puls, Pulsatilla; Phos, Phosphorus.

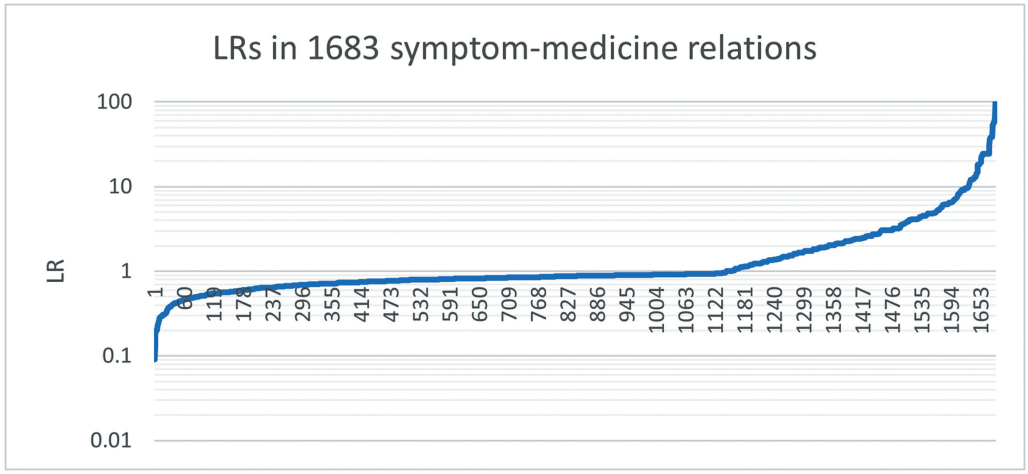

Fig. 1 Frequency distribution of 1,683 LRs of a dataset of best cases (BCHC). The *x*-axis represents the ordering of LRs from lowest to highest.

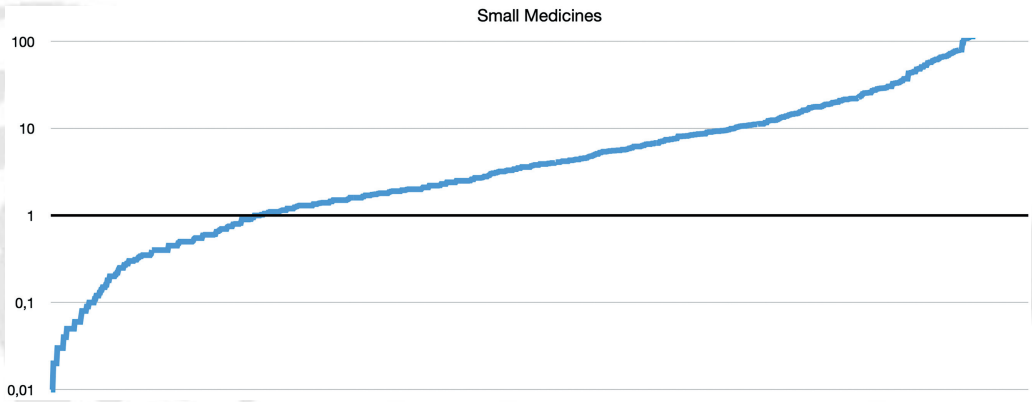

Fig. 2 Frequency distribution of 31,969 LRs of small medicines in Materia medica data, ordered from lowest to highest.

database comprises only large medicines, we did not evaluate this difference. ►**Figs. 1** and **3** are very similar for LRs > 1.

In the graphs of large medicines of both the smaller BCHC and the larger materia medica database, we see straight lines in the middle and at the ends, and between them a curve. For clarity, we will concentrate on LRs > 1 because contra-indications, indicated by LR < 1, are still a hardly explored territory in homeopathy. Therefore, clinical expertise, used here to estimate optimal corrections, is scarce regarding the use of contra-indication. There are two stages in our correction of LRs: power correction for relatively low LRs and a cut-off value for maximum LR based on the form of the curve.

Power Correction

First, we concentrate on the straight line in the middle expressing an exponential but steady rise of LRs. Confirmation

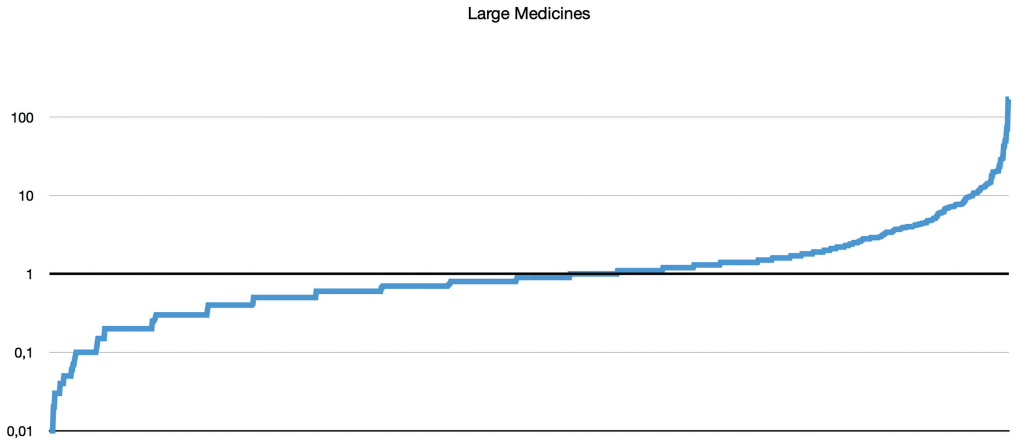

Fig. 3 Frequency distribution of 288,430 LRs of large medicines in Materia medica data, ordered from lowest to highest.

bias depends on practical experience and knowledge acquired by training. The spread of such knowledge resembles the spread of a viral disease: one patient can transmit the disease to several others. The number of transmissions by one patient is expressed by R0 (R nought). The next transmission steps increase numbers exponentially. The transmission of information in medical practice develops in a similar fashion. Practitioners communicate with each other, spreading knowledge. Charismatic teachers with more contacts transmit their knowledge to more students. However, at some point, the acceptance of a symptom–medicine relationship 'goes viral' as we know from social media, and most practitioners identify the medicine with that specific symptom, resulting in uncontrollably high LRs.

The development of materia medica has resemblances and differences with the spread of experiential knowledge. It starts the same way, but the recording in writing adds another dynamic, for instance copying of information in later writings.

Estimation of R0 for confirmation bias is arbitrary and could be different for different data collections. Suppose R0 = 1.25, then an initial LR of, say, LR = 4 becomes LR = $4^{1.25}$ after the first transmission of this information. This could be corrected by a power transformation of LR with power = 0.8 (because $0.8 \times 1.25 = 1$). The influence of three different power corrections on original LRs—power = 0.8, power = 0.7 and power = 0.5 (or square root)—is shown in ►**Fig. 4**. The original (uncorrected) LR in the BCHC data is on the horizontal axis and the corrected LR is on the vertical axis. With the strongest power correction (power = 0.5), an LR = 14 on the horizontal axis corresponds to an exponentially corrected LR = 3.8 on the vertical axis. With the weakest power correction (power = 0.8), the same LR = 14 corresponds to an exponentially corrected LR = 8.5 on the vertical axis.

We chose a conservative power correction of 0.8 to see the effect in the number of LR values for LR > 1. With this power correction, 489 out of 539 LRs (90.7%) were below LR = 7 (50 remaining), whilst without correction 455 out of 539 (84.4%) were below LR = 7 (84 remaining). This correction included 34 (6.3%) more symptoms before the hard cut-off value of LR = 7.

The elegance of power correction is that it increases the number of LR values (symptoms) that differentiate between medicines: that is, the LRs that are not extremely high. In daily practice, we use the difference in LR values to select the most eligible medicines. The difference between LR = 2 and LR = 8 in ►**Table 1** is 18.9% more chance that the medicine will work; the difference between LR = 128 and LR = 256 is only 3.2% effectiveness. On the other hand, over-correction with power calculation reduces the difference between symptoms.

Cut-off for Maximum Likelihood Ratio

The interpretation of differences in LR value and the bias caused by high LR values are a strong argument to maximise LR, but the question is at *which* LR value. Looking at ►**Fig. 5**, the curvature of the uncorrected and power corrected LR begins at LR = 4; the curvature becomes sharper up to a certain point and then straightens almost to a vertical line at LR = 12. The position of the strongest curvature is hard to define, but this is the point where the LR values start to become uncontrollably high, resulting in single symptoms that dominate in the repertorisation.

Mathematics helps to explore the different parts of a curve. We can quantify the sharpness of the curve at each point by the angle of the tangent line. This tangent is approximated by the rise in LR value (*y*-axis) and the number of LR outcomes between two LR values (*x*-axis). The inverse value of this tangent (cotangent, shown in ►**Fig. 6** and in ►**Table 3**) equals the number of LRs between two subsequent LR values.

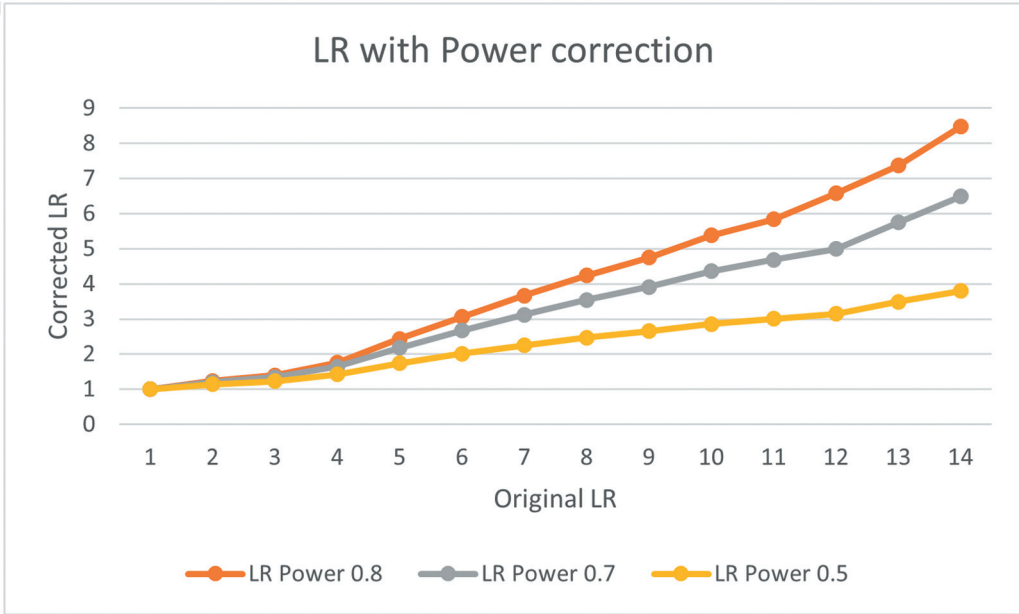

Fig. 4 Influence of different power corrections on uncorrected LRs in BCHC data. The uncorrected (original) LRs on the *x*-axis and the corrected LR on the *y*-axis.

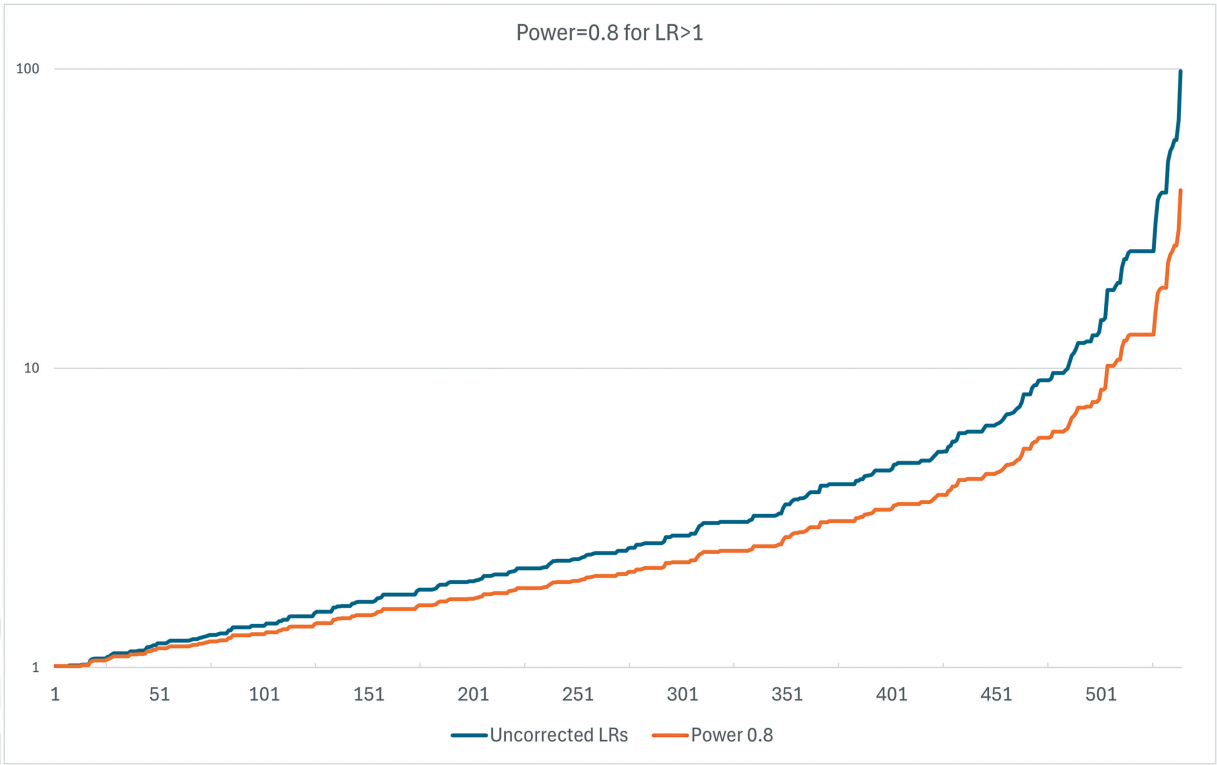

Fig. 5 Frequency distribution of 539 LR values in BCHC, with and without power correction of 0.8.

Cotangent is used here to avoid numbers < 1. The second derivative is the derivative of the first derivative and so on. A second derivative with value zero represents a straight line, but noise in real-world data causes a fluctuation around zero. The practical value of this procedure is that we see the relationship between two subsequent LR values and the number of symptoms between the two LR values. As the number of symptoms between two LR values decreases (frequency distribution becomes steeper), the chance of exaggerated LRs becomes higher.

We combined LR values > 1 of 37 medicines (3,467 LRs in total) of the best cases database ordered from low to high after power correction of 0.8. These LRs represent a combination of larger and smaller medicines. ►**Fig. 6** shows the LR distribution curves of the first and second derivatives as a function of original LR values. Theoretically, the second derivative being zero means that the line is straight. Noise in real-life data makes the curves less smooth but viewed more generally the distribution of LR values becomes a straight line when LR is 10 to 12.

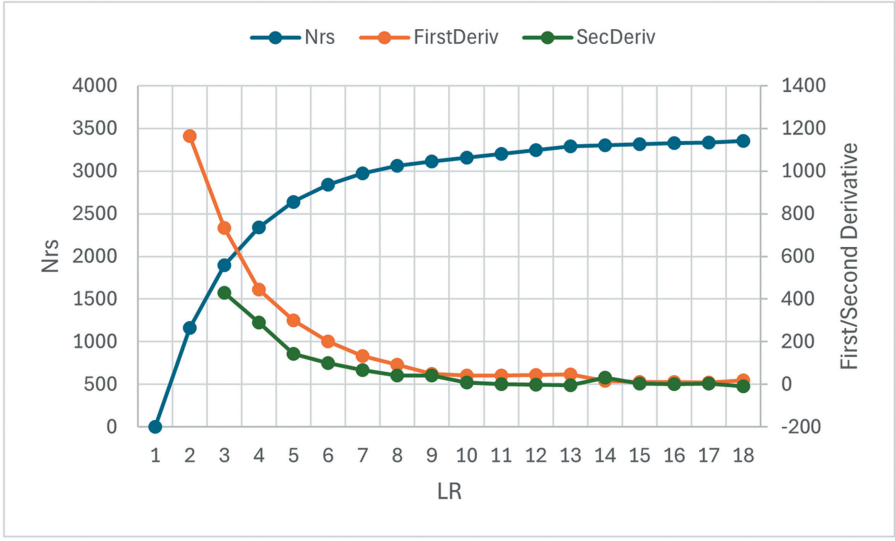

Fig. 6 Frequency distribution of LRs in BCHC, and the inverse (cotangent) of the first and second derivatives ('FirstDeriv' and 'SecDeriv').

Table 3 Materia medica data and their derivatives noted as cotangent

	Small			Medium			Large		
LR	Count	Deriv1	Deriv2	Count	Deriv2	Deriv3	Count	Deriv2	Deriv3
1	8,202	8,202	–	25,389	–	–	155,748	–	–
2	12,463	4,261	3,941	36,967	13,811	–	222,765	8,8731	–
3	15,216	2,753	1,508	42,740	5,805	8,006	245,836	43,946	44,785
4	17,219	2,003	750	46,171	2,342	3,463	257,050	11,857	32,089
5	18,721	1,502	501	48,325	1,277	1,065	263,809	4,455	7,402
6	19,884	1,163	339	49,900	579	698	268,264	2,304	2,151
7	20,871	987	176	51,040	435	144	271,466	1,253	1,051
8	21,687	816	171	51,918	262	173	273,797	871	382
9	22,359	672	144	52,640	156	106	275,612	516	355
10	22,941	582	90	53,229	133	23	277,109	318	198
11	23,429	488	94	53,702	116	17	278,291	315	3
12	23,851	422	66	54,137	38	78	279,287	186	129
13	24,282	431	−9	54,500	72	−34	280,111	172	14
14	24,601	319	112	54,806	57	15	280,831	104	68
15	24,930	329	−10	55,085	27	30	281,441	110	−6
16	25,199	269	60	55,322	42	−15	281,926	125	−15
17	25,433	234	35	55,524	35	7	282,385	26	99
18	25,662	229	5	55,702	24	11	282,835	9	17

Notes: The first derivative is not shown for the medium and large medicines because they are less informative. For 'Small medicines': arranged from lowest to highest the first LR with value LR = 1 or higher is the 8202nd LR value and the first with LR = 2 or higher is at position 12,463. Between LR = 1 and LR = 2, there are 12,463–8,202 = 4,261 LR values. This is the first derivative expressed as cotangent. The second derivative is 4,261–8,202 = 3,941. Likewise, there are 229 LR values between LR = 17 and LR = 18 (first derivative).

The larger amount of data allows a more detailed calculation of parts of the slope of the curve, especially for medium and large medicines. The evolution of the curves is more detailed in the third derivative. Smaller values in each derivative correspond with an acceleration of rise in LR values up to a point where the derivative oscillates (because of noise in the data) around a low value. The derivatives are shown in tabular form in ►**Table 3**.

►**Table 3** shows the calculated derivatives for the three groups of the materia medica data, the first for small medicines and second and third for medium and large medicines. The more gradual rise of LR values in small medicines is also visible in ►**Fig. 2**. The tabular representation of the derivatives of small, medium and large medicines show that the bend of the curve starts at LR = 4. The tabular representation of the third derivatives of medium and large medicines show that the curves become steeper at LR = 7.

Both databases show similar frequency distributions of LR values with a straight line at the lower LRs (after correction for exponentiality), then a slow curve before the second and steeper straighter line. The almost vertical straight line at the right side of the best cases graph represents a small number of LR values that seem 'out of control'. The more horizontal left line seems to represent 'normal' confirmation bias and can be controlled with power correction, which increases the number of LR values on the more horizontal straight line and reduces the number of LR values on the vertical line. This bend extends roughly between LR = 4 and LR = 12. Intuitively, it is understandable that we have to maximise LR at a value below LR = 12, because the influence of higher LR values on repertorisation is otherwise hard to control and affects few symptoms. Maximising LR at LR = 4 would exclude many LR values from being controlled. The area between LR = 4 and LR = 12 deserves further consideration.

The large database of historical materia medica data shows more detail of the curves, expressed in higher derivatives. The third derivative shows that at LR = 7 the curve becomes notably steeper. This second acceleration of the slope of the curve indicates that maximum LR can best be set at LR = 7. This value seems the best compromise between more discriminatory power of symptoms and excessive influence of a limited number of symptoms caused by confirmation bias.

Discussion

Algorithms, especially repertorisation and appreciation of peculiar symptoms, have always played a role in homeopathy, albeit unconsciously. Practitioners handled the shortcomings of these algorithms intuitively. Scientific methods

for improving the homeopathic instruments, materia medica and repertory, are based on data collection and algorithms derived from those data. Even if the practitioners were perfectly trained in data collection, we will have biased data that could harm the effectiveness of the algorithm. We must detect bias in the data by clinical expertise, inspection of data and statistical instruments. The correction of biased data should also be based on these same methods. We share these considerations for two reasons. Firstly, homeopathic instruments should not become 'black boxes'. Practitioners always had access to original sources providing all available knowledge. Computer repertories add calculation possibilities to the available data but that should be understandable to the user. Secondly, these mathematical aspects are important assets for homeopathy because they illustrate the systematic handling of practical experience. Practitioners should have some basic knowledge of them to understand why collection of cases with scientific rigor is vital for this method. The quality of the homeopathic instruments depends on the quality of the data. This requires some extra teaching of practitioners that is provided in a series of PFR-related papers in this journal.

The spread of information about the relationship between symptoms and medicines is intuitively comparable to the spread of a viral infection that can be quantified by R0 (R nought). Using power correction appeared to correct LR values to a certain extent, resulting in more likely repertorisations. However, inspecting frequency distributions of LRs, we see a second element causing an even stronger rise in LR values. The symptom seems to 'go viral' as an indication for a specific medicine, resulting in extreme bias. We therefore propose a two-step correction for confirmation bias: a first correction by mathematical power calculations, and a maximum LR value before it tends to go viral.

The choice of the optimal power correction could be based on clinical weighing of advantages and disadvantages. A strong correction will result in a larger number of different LRs, but with less difference between LRs resulting in less distinction between medicines. Trying different power corrections, a value of 0.8 resulted in an acceptable difference between LRs, combined with 93% of LRs below $LR = 8$.

The transition between a moderate and a strong increase in LR is not an angle, but a curve. The cut-off value for LR must be somewhere in this curve. The calculation of mathematical derivatives helps us to find the beginning and the end of this curve. The choice of the optimal cut-off value in this curve is arbitrary. If we set the cut-off at the start of the bend at $LR = 4$, we lose valuable information about symptoms with LRs situated in the curve. Those are often keynote symptoms of less frequently used medicines. We would like to be aware of such medicines too, without over-emphasising them. If we use a cut-off at the end of the curve, say, $LR = 12$, smaller medicines might get more attention than desirable.

Despite the different characteristics of both datasets, the curves and the derivatives are quite similar. The large number of symptoms in the medium and large medicines in the historical materia medica data shows more detail of the curvature of the frequency distribution of LRs between $LR = 4$ and $LR = 12$. With a third derivative, a second acceleration of the steepness of the frequency distribution becomes visible at $LR = 7$. This reinforces the choice of $LR = 7$ as a maximum LR. This maximum $LR = 7$ includes 93.3% of all LR values of the BCHC database and (273,796/379,597) 72.1% of the materia medica database.

An improved algorithm aims at improving the effectiveness of the repertory, but not at simplifying its use. A good repertorisation still requires some ground rules, such as:

- Not trying to favour preferred medicines by selection of symptoms.
- Avoiding related symptoms.
- Ensuring the hierarchisation of symptoms.
- Using the absolutely necessary symptoms with maximum relevance to the case based on careful hierarchisation.

Limitations

There is still insufficient information to do similar exercises with small medicines and LRs < 1 (contra-indications). This thought experiment could help to structure new observations and experiments addressing small medicines and contra-indications.

The quality of best cases should be a subject for continuous further discussion and one of the main issues of PFR. The best cases mentioned here were selected after open discussions by doctors highly motivated to improve the quality of homeopathic case descriptions, but it would be too soon to define definitive quality criteria now. We need many cases to start a structured evaluation of useful and feasible quality criteria.

We do not know the prior chance that any specific homeopathic medicine will work. All prior and posterior chances mentioned here are hypothetical, and the applicability of Bayes' theorem is still limited to comparison of medicines. Our intention has been not to improve the predictability of absolute chance of cure but to improve the differentiation between medicines. Our choice of power correction was based on that goal only.

Conclusion

The availability of a new dataset of best cases allows for comparison with historical materia medica data and further analysis. We found a striking similarity between the frequency distributions of LRs of best cases and materia medica data. Based on previous PFR, LRs of homeopathic keynote symptoms range mostly between $LR = 2$ and $LR = 4$. Above $LR = 4$, symptoms tend to increasingly dominate repertorisations. Higher LR values are more likely caused by confirmation bias. Mathematical tools allow further analysis and suggest two-step handling of exaggeration of LR by confirmation bias: the first step by power correction, the second step by cut-off.

Conflict of Interest
None declared.

References

1 Hahnemann S. Organon of Medicine, 6th edition. B. Jain Publishers; 1921
2 Kahneman D. Thinking, Fast and Slow. Penguin Books; 2011
3 Rutten ALB, Stolper CF, Lugten RFG, Barthels RWJM. Statistical analysis of six repertory rubrics after prospective assessment applying Bayes' theorem. Homeopathy 2009;98:26–34
4 Miglani A, Manchanda RK, Rutten L. Prognostic factor research in homeopathy: insights from condition-confined studies. Homeopathy 2025; Jul 28. Doi: 10.1055/a-2562-8734. Online ahead of print
5 Eizayaga JE, Rutten L. The challenge of building a Bayesian homeopathic materia medica and repertory from best cases' data. Homeopathy: Accepted for publication
6 Schäferkordt R. Repertorium der Phänomene: Neukonzeption eines vollständig in der Materia medica referenzierten Repertoriums mit phänomenbasierter Struktur und Bayes-Gradierung. AHZ 2021;226:16–23
7 Stolper E, Rutten L, Lugten R, Barthels R. Materia medica validation and meta-analysis. Homoeopath Links 2004;16:186–192
8 Rutten LA. The importance of case histories for accepting and improving homeopathy. Complement Ther Med 2013;21: 565–570
9 Ohayon MM. Prevalence and correlates of nonrestorative sleep complaints. Arch Intern Med 2005;165:35–41

