

Repertory corrections based on validation of symptom-rubrics by Likelihood Ratio assessment

Lex (ALB) Rutten, Erik (CF) Stolper, Roland (RFG) Lugten, Rob (RWJM) Barthels*
August 2006

Abstract

Background: After two years of prospective assessment of six homeopathic symptoms (planned duration three years) we can validate some rubrics of the homeopathic repertory using solid bayesian scientific theory. In this article we try to introduce statistical reasoning in the process of introducing or discarding entries in/from the repertory.

Methods: At this moment 2506 patients entered the prospective study and 2266 prescriptions are evaluated. LRs, confidence intervals and the probability of repertory entries compared to our findings were calculated.

Outcome: The size of the assessed repertory rubrics varies more than expected; large rubrics contain superfluous entries, small rubrics are incomplete. According to our translation of present entries in the repertory into LR more than 50% of the outcome of this assessment did not correspond with the existing repertory entries. This outcome cannot be generalised for the whole repertory. We should reach consensus about criteria for adding or deleting entries in the repertory.

Introduction

The homeopathic repertory should indicate which medicines are indicated if a symptom is present, and how strong the indication is. The sources and methods of the homeopathic repertories are unclear. Entries of a homeopathic medicine in repertory-rubrics are based on occurrence in provings and in practice experience with patients cured by the medicine (medicine population). How accurate are these data? One problem is that we use some medicines very often and many other medicines infrequently.¹ We have many patients cured by *Natrium muriaticum* and few patients cured by *Glonoinum*. Our memory tells us that many patients cured by *Glonoinum* and many patients cured by *Natrium muriaticum* had headaches, but can we tell which proportions of those populations had headaches?² Are the data in the repertory the same as in prospective research?

Bayes' philosophy was published in 1763 and handles about predictions from experience in the past.³ It tells us that chances of success with a medicine increase if a symptom is frequently present in patients that are cured by that medicine, more frequently than in other patients.⁴ This is expressed by the Likelihood Ratio (LR). If the symptom 'loquacity' occurs four times more frequently in Lachesis patients than in other patients LR=4.

Bayes' philosophy is expressed in a formula that is derived from the mathematical rule of conditional probability⁵:

Posterior odds = LR * Prior odds

Where odds = chance / (1-chance) and chance is odds / (1+odds)

This formula enables us to estimate the increase of chances that a medicine will cure.

* Committee for Methods and Validation of the Dutch homeopathic doctors association (VHAN)

Methods

Likelihood ratio is the frequency (prevalence) of a symptom in the population cured by a certain medicine divided by the frequency of the same symptom in the rest-population. The Committee of Methods and Validation of the Dutch homeopathic doctors association is performing the first prospective assessment of LR of homeopathic symptoms, evaluating six symptoms: 'Diarrhoea from anticipation', 'Fear of death', 'Grinding teeth during sleep', 'Herpes lips', 'Sensitive to injustice' and 'Loquacity'.⁶ The choice of these symptoms was rather arbitrary. The different symptoms should be keynotes for different homeopathic medicines, not for the same medicine. We selected symptoms with an estimated prevalence of more than 1% in the whole population and we chose a mixture of vague and less vague symptoms. Results were evaluated using the GHHOS scale.⁷

To make a comparison with the existing entries of Kent's repertory we have to translate type (expressing importance of the symptom related to that medicine) into numbers. Such a translation is arbitrary, A cut-off value like $LR > 1.5$ for plain type means that we regard a medicine as indicated if the prevalence of the symptom in the medicine-population is at least 1.5 times larger than in the rest-population. A possible translation from type into LR could be like Table 1:

Table 1: Repertory entries translated into LR values

Type	LR
Plain	1.5-3.0
<i>Italics</i>	3.0-6.0
Bold	> 6.0

In general peculiar symptoms are related to small rubrics, common symptoms to large rubrics. The rubric 'Diarrhoea from anticipation' in Kent's repertory contains three medicines. This is a relatively small rubric considering the prevalence of 4% in the whole population in our assessment. The symptom 'Fear of death' has the same prevalence, but Kent's repertory-rubric is much larger, 103 medicines. This could mean several things:

1. The symptom is related to a small number of homeopathic medicines; in that case we expect the LRs to be high.
2. The rubric 'Diarrhoea from anticipation' is incomplete.
3. The rubric 'Fear of death' is over-complete.

Point 2 and point 3 are probably both valid. Point 1 should be proven in due time, after assessment of a substantial number of symptoms. All three points can be proven by LR assessment. If the rubric is incomplete new medicines will come up with $LR > 1.5$. If the rubric is over-complete a number of medicines in the rubric might have $LR < 1.5$, but the range between $LR = 1$ and $LR = 1.5$ remains uncertain.

At this moment we have 2266 evaluated prescriptions (2506 patients), but they are divided over a large number of medicines. Chance could play an important role in these results. The 95% Confidence Interval (95%CI) indicates if LR could be one, and therefore useless. We can also calculate the possibility that a medicine is rightfully in a certain rubric in that degree. According to the translation of LR into type we proposed above bold type should be used if the prevalence of the symptom is at least six times the prevalence in the rest-population. The prevalence of a symptom in the rest-population is nearly the same as the prevalence in the whole population. So, if the prevalence of the symptom is 4% in the whole population, it should be 24% in the medicine population if the entry in the rubric is in bold type.

For medicines not mentioned in the rubric, but with possibly meaningful LR, we calculated the possibility that LR could be higher than 1.5. For these medicines we did not mention expected prevalence in the tables.

We used exact binomial calculations (one-tailed) to calculate P-values, and calculations via binomial approximation of normal distribution if (number of patients cured by the medicine)x(expected prevalence of symptom)>5. The prevalence of a symptom in a medicine population is calculated as $a/(a+c)$, where a is the number of patients cured by the medicine with the symptom and c is the number of patients cured by the medicine without the symptom. For calculating CI we used the software program CAI from BMJ (logarithmic transformation).

Diarrhoea from anticipation

In Kent's repertory the rubric contains: *Arg-n.*, *Gels.*, *ph-ac*, all medicines in Italics.

The LR-assessment has shown 101 patients with this symptom so far. Some of the results are in Table 2

Table 2: Results of the assessment of the symptom 'Diarrhoea from anticipation'. The 95% Confidence Interval (95% CI) is mentioned if it does not include 1.

Diarrhoea from anticipation		a	c	b	d	LR+	95% CI
	101						
prevalence=4%	<i>arg-n</i>	4	6	97	2399	10,29	4,70 - 22,54
	<i>calc</i>	3	45	98	2360	1,57	
	<i>gels</i>	2	3	99	2402	10,11	3,40 - 30,08
	<i>ign</i>	2	16	99	2389	2,79	
	<i>nat-m</i>	4	88	97	2317	1,08	
	<i>ph-ac</i>	4	9	97	2396	7,91	3,42 - 18,29

a = the medicine cured and the symptom is present

b = rest-population with the symptom

c = the medicine cured and the symptom is not present

d = rest-population without the symptom

The 95% confidence interval shows that *Argentum-nitricum* could be upgraded to bold, but that is not absolutely certain; chances are 93% ($p=0.93$). $LR < 6$ is still possible according to the 95% confidence interval. The upgrading of *Gelsemium* ($p=0.907$) and *Phosphoricum-acidum* ($p=0.818$) to bold type is slightly less certain.

Natrium muriaticum is not in the repertory-rubric. According to our assessment it is indeed probably not indicated by this symptom. For LR to be > 1.5 the prevalence in the *Natrium muriaticum* population should exceed 6% (= 1.5 times the prevalence in the whole population). Probability that LR could be higher than 1.5 is 0.326 (binomial approximation of normal distribution, one-tailed).

Probability that *Calcareo carbonica* could be rightfully entered in this rubric is 0.675; for *Ignatia* this is 0.910.

These results show that we have some significant LRs. On the other hand we have indications about medicines that should not be entered in the rubric. With such a large number of *Natrium muriaticum* cases it is unlikely that the prevalence of 'Diarrhoea from anticipation' in the *Natrium muriaticum* population will exceed 6%. According to Bayes we should not enter this medicine in this rubric. But what should we do if the medicine were in the rubric?

For the next tables we will not indicate the prevalence of the symptom in the rest-population (fields b and d), because that does not vary much. Instead we mention the expected prevalence of the symptom in the medicine population according to repertory grading and the probability

considering our results (P=0.156 means probability is 15.6%). If the LR is significant (does not include 1), we mention the 95% CI.

Fear of death

The rubric 'Fear of death' contains 103 medicines in Kent's repertory. There are also 103 patients in this assessment with this symptom. The prevalence of the symptom in the whole population is 4%. Therefore we expect a prevalence of 6% in the medicine-population before entering the medicine in the rubric. If the prevalence in the medicine-population is 12% it can be entered in Italics, and if the prevalence is over 24% it should be entered in bold type. The results are in Table 3.

Table 3: Assessment of the symptom 'Fear of death'. The expected prevalence is the prevalence according to the existing entry in Kent's repertory.

fear of death	a	c	LR+	expected prevalence(%)	P-value	95% CI
prevalence=4% acon	4	2	16.84	24	0.996	9.26 to 30.61
anac	3	4	10.71	6	0.999	4.46 to 25.74
ars	4	13	5.92	24	0.612	2.46 to 14.24
arg-n	2	8	4.94	12	0.891	1.41 to 17.33
calc	3	45	1.54	24	0.003	
lach	2	21	2.14	12	0.467	
lyc	2	36	1.29	12	0.150	
nat-m	0	92	<0.1	6	0.014	
phos	2	30	1.53	24	0.016	
sep	4	52	1.77	6	0.755	

The entry of *Anacardium* might be upgraded, possibly to bold type (p=0.938, the p-value in Table 3 is the probability that LR>1.5). The bold entry of *Calcarea carbonica* is incorrect, plain type could be correct (p=0.675), the same goes for Phosphorus (p=0.699). *Lycopodium* still might be possible, but plain, not in Italics (p=0.599). *Natrium muriaticum* should not be in this rubric.

Grinding teeth during sleep

In Kent's original repertory this rubric contains 36 medicines. There were 138 patients with this symptom and the prevalence in the whole population was 5.5%

Table 4: Assessment of the symptom 'Grinding teeth during sleep'. If the expected prevalence is not mentioned the medicine is not entered in Kent's repertory.

grinding teeth	a	c	LR+	expected prevalence(%)	P-value	95% CI
prevalence=5.5% bell	2	7	4.08	33	0.385	1.19 to 14.00
calc	2	46	0.75	8.25	0.231	
calc-p	3	9	4.62		0.987	1.71 to 12.48
carc	3	22	2.21		0.853	
carc-c-c	2	3	7.36		0.995	2.48 to 21.79
ign	3	15	3.07	16.5	0.590	
merc	5	25	3.10	16.5	0.626	1.37 to 7.02
sep	3	53	0.97	8.25	0.311	

Belladonna is over-rated with bold type; Italics is more probable (p=0.826). There are good arguments to replace *Calcarea carbonica* in this rubric by *Calcarea phosphorica*, maybe even in Italics (p=0.878). The LR value for *Carcinosinum con cuprum* has to be handled with care. These prescriptions were very recent and their evaluation could be (confirmation) biased (see

Discussion). *Ignatia* and *Mercurius* should be in this rubric, but possibly in plain type; chances are $p=0.944$ and $p=0.967$ respectively. *Sepia* should not be in this rubric.

Herpes lips

In Kent's original repertory this rubric contains 34 medicines. There were 129 patients with this symptom. The prevalence of the symptom in the whole population was 5%, see Table 5.

Table 5: Assessment of the symptom 'Herpes lips'

herpes lips	129	a	c	LR+	expected prevalence(%)	P-value	95% CI
prevalence=5%							
carc		2	23	1.56		0.712	
lach		2	21	1.70	7.5	0.754	
lyc		3	35	1.55		0.754	
merc		2	28	1.30		0.682	
nat-m		11	81	2.45	30	<0.001	1.37 to 4.38
sep		7	49	2.51	30	0.003	1.23 to 5.18
staph		3	14	3.49		0.966	1.23 to 9.87
sulph		3	52	1.06	7.5	0.401	

This rubric gives low LR values, but there are few entries that should not be there at all, like Sulphur. Natrium muriaticum and Sepia should be downgraded, even Italics are improbable; $p=0.251$ for Natrium muriaticum and $p=0.367$ for Sepia.

Some medicines might be added: Carcininum ($p=0.712$), Lycopodium ($p=0.754$).

Staphisagria should be there, at least plain ($p=0.966$), maybe in Italics ($p=0.756$).

Sensitive to injustice

This rubric does not exist in the original repertory of Kent. In RADAR-Synthesis (v 8.140) it contains: **Caust.**, Cupr., Dros., *Ign.*, Merc., Nux-v., Sep., **Staph.**, Verat. There were 250 patients with this symptom and the prevalence was 10% in the whole population, see table 6:

Table 6: Assessment of the symptom 'Sensitive to injustice'

injustice	250	a	c	LR+	expected prevalence(%)	P-value	95% CI
prevalence=10%							
anac		2	5	2.88		0.926	
aur		2	5	2.88		0.926	
bell		3	6	3.37		0.966	1.33 to 8.55
carc		5	20	2.03		0.838	
carc-c-c		3	2	6.08		0.973	2.94 to 12.55
caust		9	11	4.64	60	0.128	2.82 to 7.65
chel		2	1	6.73		0.939	2.30 to 15.11
cocc		2	3	4.03		0.973	1.37 to 11.88
ign		2	16	1.11	15	0.480	
kali-bi		2	9	1.83		0.779	
merc		5	25	1.68		0.711	
nux-v		2	19	0.95	15	0.370	
ph-ac		3	10	2.33		0.882	
sep		6	50	1.08	15	0.246	
staph		2	15	1.18	60	<0.001	

Anacardium and *Aurum* could be related to sensitivity to injustice. The same goes for *Chelidonium* and *Cocculus*. *Carcinosinum*, *Belladonna*, *Kalium bichromicum*, *Mercurius* and *Phosphoricum acidum* are reasonable options for this rubric, more so than *Ignatia* and *Nux vomica*. The bold entry for *Causticum* seems too much ($p=0.128$), Italics is better ($p=0.952$). This however, could be depending on cut-off value for the symptom (see discussion). The

bold entry for *Staphisagria* is unlikely, even in *Italics* it would be improbable (p=0.084). For *Carcinosinum con cuprum* see above.

Loquacity

In Kent's original repertory this rubric contains 97 medicines. See table 7. There were 172 patients with this symptom and the prevalence in the whole population was 7%, see Table 7:

Table 7: Assessment of the symptom 'Loquacity'

loquacity	172	a	c	LR+	expected prevalence(%)	P-value	95% CI
prevalence=7%	caust	2	18	1.46	10.5	0.648	
	hyos	5	4	8.31	42	0.877	4.55 to 15.17
	lach	6	17	3.90	42	0.089	1.93 to 7.89
	lyc	4	34	1.55		0.631	
	med	2	10	2.45		0.876	
	merc	3	27	1.47		0.612	
	nat-m	2	90	0.31	10.5	0.007	
	phos	2	30	0.91	21	0.236	
	sep	6	50	1.58		0.728	
	stram	2	6	3.67	42	0.275	
	sulph	3	52	0.79	10.5	0.159	

It seems contra-intuitive that *Lachesis* should not be in bold type; *Italics* is better (p=0.808). There could be a number of reasons, like:

- Prevalence of the symptom in the research-population may be high. The prospective setting causes this; people who are not obviously loquacious during consultation could confirm this symptom after questioning.⁸
- Subjectivity in assessing this symptom is high.
- 'Changing the subject frequently' is an important specification for this symptom regarding *Lachesis*; this specification is not made during this assessment.⁹

There are weak indications for *Lycopodium*, *Mercurius* and *Sepia* to be relevant for this symptom. *Medorrhinum* probably should be added. *Stramonium* seems to fit better in *Italics* (p=0.775). *Natrium muriaticum* and *Sulphur* could be removed from this rubric. *Phosphorus* could be removed also, even in plain type this medicine does not fit in (p=0.333).

Discussion

After doubling the amount of evaluated prescriptions compared to the previous data-analysis our conclusion that the repertory is flawed is confirmed.⁶ But certainty of LRs can still improve. In this article we try to introduce statistical reasoning in the process of introducing or discarding entries in/from the repertory. Our calculations concern the influence of chance on our results, not the influence of bias. The influence of chance will decrease as numbers become larger, but the influence of bias will remain the same for most variables. Confirmation bias could become less over time. In the first evaluation there were no *Causticum* patients without 'Sensitivity to injustice'¹⁰; later on people without this symptom, who started with another medicine, had better reactions on *Causticum*.¹¹ By comparing the last evaluation with former evaluations we noticed that *Carcinosinum con cuprum* was new in our database. To wash out confirmation bias we will close the inclusion of new patients a few months before the last evaluation.

We have, disregarding the values for *Carcinosinum con cuprum*, 56 LR values of interest at this stage. Of these 20 (35.7%) are additions to Kent's original repertory, mostly in the rubric

‘Sensitive to injustice’. Then 11 (19.6%) suggest removal and 10 (18%) downgrading of existing entries. This means 41 (73%) proposed adjustments. Downgrading of entries is somewhat disputable because of arbitrary cut-off values (see below). If we disregard downgrading there are 31 (55%) important changes to the existing repertory. Such figures depend on the chosen symptom rubrics. We need to assess much more symptoms before we can make a statement about the correctness of the existing repertory.

When should we introduce new entries to the repertory; and when should we remove old entries? We still must decide how much certainty is needed for our repertory-entries. Should this be 50%, 95%, or somewhere in between? If we demand 95% certainty for each entry many existing entries in the repertory should be deleted. But are our results valid all over the world? Wouldn't it be a waist if we delete medicines that are infrequently used because of too much uncertainty? In conventional medicine a two-tailed certainty of 95% is generally accepted, but this is linked to hypothesis testing. The Bayesian method is about (un)certainty; is a symptom linked to a medicine (probability) and is a cure more certain if a certain symptom is present (conditional probability)? In our case the relevance of research is decision making. If there is 78% chance that the LR of ‘Sensitive to injustice’ for Kalium bichromicum is at least 1.5, this information makes sense to a homeopathic physician. He uses this kind of information implicitly in daily practice, but is mainly referring to his own experience.

The LR values are related to thresholds for symptoms.⁶ In this assessment the prevalence of ‘Sensitive to injustice’ is 10%. This could be due to the technique of prospective research. With a higher threshold the prevalence of this symptom might be lower. LRs become higher in that case.

The threshold for different type is quite arbitrary. We could also take $LR > 4$ or 5 for bold type, or $LR > 1.3$ for plain type.

We disregard negative LR values for this moment. This should be evaluated later. We also disregarded the possibility of interaction between symptoms. Probably this will be hard to investigate, because of the large number of variables.¹²

There are several precautions we should take in interpreting our results. We already discussed confirmation bias and threshold values, but there may be other bias. On the other hand, the old repertory has become severely outdated. We have as yet no idea how much bias there is in the repertory, because the sources of the information are largely unknown. The results of LR research are well documented and can give indications for a gradually improving repertory.

A repertory based on prospective LR research gives new opportunities that should be considered carefully. We could choose to display all LRs, even the LRs that do not indicate a medicine. If you hesitate between Lycopodium and Natrium muriaticum regarding a patient that has fear of death, the information that for Lycopodium and this symptom $LR = 1.29$, and for Natrium muriaticum $LR < 0.1$. In an electronic repertory it would be possible to show such information on demand. It could also be possible that we discover that a certain LR value is different for children or for women. Such choices have important influence on the repertories. Who is going to make such choices?

We also have to choose between displaying the entries like before – using typeface – or by numbers. The problem with numbers is that they are not as exact as they would seem to some and should be adapted to individual circumstances by the practitioner in each case. Like the prevalence of the symptom in your own practice related to the research population and the intensity of the symptom in each patient. Each practitioner has his own cut-off values; which seem to depend on personal factors rather than geographical.⁶ The practitioner needs thorough training to handle numbers.

The new repertory based on LR will also influence our method. Now we handle the shortcomings of the repertory intuitively. Some shortcomings of the repertory are solved and some not by LR. New problems might even arise, introducing a circular process of developing a new intuition about shortcomings before we can mend them. Some will fear that the art of case taking will be lost, but such fears existed also when Kent's repertory was first published. The art of case taking and judging the intensity and importance of each symptom will remain necessary. The art, however, is based on solid data; that was also the case one century ago. But the problem that retrospective data are inaccurate is now largely solved.

Conclusion

So far our LR assessment confirms what was expected: Frequently used medicines are overrated in the repertory and infrequently used medicines are underrated. But this occurs in a variable and unpredictable degree. We also found that the size of repertory rubrics varies much more than expected on prevalence of symptoms. Large rubrics tend to have superfluous entries where in small rubrics medicines are missing. In this assessment more than 50% of the results did not correspond with the existing repertory, but this fact should not be generalised for the whole repertory. We should reach consensus about criteria for entering or discarding entries in the repertory.

References

-
- ¹ Stolper CF, Rutten ALB, Lugten RFG, Barthels RJWMM. Improving homeopathic prescribing by applying epidemiological techniques: the role of likelihood ratio. *Homeopathy* 2002;**91**:230-238
 - ² Rutten A.L.B., Stolper CF, Lugten RF, Barthels RJ. Is assessment of likelihood ratio of homeopathic symptoms possible? A pilot study. *Homeopathy*. 2003;**92**:213-216
 - ³ Bayes, T. An Essay Toward Solving a Problem in the Doctrine of Chances, *Philosophical Transactions of the Royal Society of London* 1763;**53**, 370-418.
 - ⁴ Gill CJ, Sabin L, Schmid CH. Why clinicians are natural bayesians. *BMJ* 2005;330:1080-3
 - ⁵ Woodworth GG. *Biostatistics: A Bayesian Introduction*. Wiley. 2004
 - ⁶ Rutten A.L.B., Stolper CF, Lugten RF, Barthels RJ. A Bayesian perspective on the reliability of homeopathic repertories. *Homeopathy*. 2006;**95**:88-93
 - ⁷ Rutten ALB, Stolper CF, Lugten RF, Barthels RJ. 'Cure' as the gold standard for likelihood ratio assessment: theoretical considerations. *Homeopathy* 2004;**93**:78-83
 - ⁸ Rutten A.L.B., Stolper CF, Lugten RF, Barthels RJ. Repertory and the symptom loquacity: some results from a pilot-study. *Homeopathy*. 2004;**93**:190-192
 - ⁹ Rutten A.L.B., Stolper CF, Lugten RF, Barthels RJ. Assessing likelihood ratio of clinical symptoms: handling vagueness. *Homeopathy*. 2003;**92**:182-186
 - ¹⁰ Rutten Lex Stolper Erik, Lugten Roland, Barthels Rob. Onderzoek Likelihood Ratio: stand van zaken januari 2005. *SSC* 2005(2);35:17-22
 - ¹¹ Rutten Lex Stolper Erik, Lugten Roland, Barthels Rob. Onderzoek Likelihood Ratio: april 2005. *SSC* 2005(2);35:23-24
 - ¹² Bachmann LM, Puhan MA, Riet G ter, Bossuyt PM. Sample sizes of studies on diagnostic accuracy: literature survey. *BMJ*, doi:10.1136/BMJ.38793.637789.2F (24 april 2006)