Homeopathy (2009) 98, 26–34

doi:10.1016/j.homp.2008.11.012, available online at http://www.sciencedirect.com

ORIGINAL PAPER

Statistical analysis of six repertory rubrics after prospective assessment applying Bayes' theorem

ALB Rutten*[,a], CF Stolper, RFG Lugten and RWJM Barthels

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

Background: **After prospective assessment of six homeopathic symptoms we validated some rubrics of the homeopathic repertory using Bayesian theory. In this paper we introduce statistical arguments for introducing or discarding entries from the repertory.**
Methods: **4094 patients entered the prospective study and 4072 prescriptions were evaluated. After translating typeface into Likelihood Ratios (LRs), Confidence Intervals and the probability of existing repertory entries compared to our findings were calculated.**
Outcome: **Our assessment yielded 121 relevant results to validate existing repertory entries. Five symptoms could be compared with Kent's original repertory; they have about the same prevalence (range 3.9–6.5%) in the whole population, but the size of the corresponding repertory rubrics varies from 3 to 103 entries. LR assessment reduced the larger rubrics and supplemented the smaller ones. Our results do not correspond with 56% of the existing repertory entries regarding five symptom-rubrics. This result cannot be generalised for the whole repertory.** *Homeopathy* (2009) **98**, 26–34.

Keywords: Bayes; Repertory; Likelihood Ratio; Symptoms; Homeopathy

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 practical experience with patients improved by the medicine (medicine population). How accurate are these data? One problem is that we use some medicines very often and many others infrequently.[1] We have many patients with good results from *Natrum muriaticum* but few patients for *Glonoinum*. Memory tells us that many patients improved by *Glonoinum* and many patients improved by *Natrum muriaticum* had headaches, but can we tell which proportions of those populations had headaches?[2] Computers and the Internet enable us to update our repertories virtually every minute, but this is a great threat to the reliability of our repertories. We cannot continue to update repertories without consistent rules, preferably based on sound theory.

Bayes' philosophy was published in 1763 and deals with 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 who 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 probability[5]:

$$\text{Posterior odds} = \text{LR} \times \text{Prior odds}$$

where odds = chance/(1 − chance) and chance = odds/(1 + odds).

LR 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 remainder of the whole treated population. These are variables we can measure easily with information technology. 'Cured' means a treatment outcome which is considered a good result in homeopathic

*Correspondence: Lex Rutten, Aard 10, 4813 NN Breda, The Netherlands.
E-mail: lexrtn@concepts.nl
[a]Committee for Methods and Validation of the Dutch homeopathic doctors association (VHAN).
Received 30 May 2008; revised 22 October 2008; accepted 18 November 2008

practice, i.e. not only in terms of the presenting complaint should improve, but also other complaints and the patients' general wellbeing.

There is also a risk involved in using numbers; they easily give the impression of some absolute truth, without reflecting statistical uncertainties or even bias in the measurements. All numbers should be accompanied by their statistical probability or Confidence Interval (CI).[6] Using statistical probabilities to construct a repertory is new, and we should reach consensus about when to accept a new entry or discard an old entry in repertories. We have previously proposed some cut-off values for this.[7] This paper shows how such cut-off values could change the existing repertory. Bias has always existed and cannot fully be ecluded, but systematic research should minimise it.

Methods

We performed a prospective assessment of LR of repertory rubrics symptoms, evaluating six symptoms: 'Diarrhoea from anticipation', 'Fear of death', 'Grinding teeth during sleep', 'Herpes lips', 'Sensitive to injustice' and' Loquacity'.[8] Symptoms were agreed upon in a consensus meeting by all participating doctors. These symptoms were selected as being keynotes for different homeopathic medicines, but not for the same medicines. This, and semantic incomparability, should minimise interrelation of the symptoms. We selected symptoms with an estimated prevalence of more than 1% in the whole population and we chose a mixture of vague and more specific symptoms. Results were evaluated using the Glasgow Homoeopathic Hospital Outcome Scale (GHHOS).[9] The symptoms were recorded in all consecutive new patients older than two years attending one of 10 homeopathic practices in The Netherlands. The assessment lasted from June 2004 until December 2007.

All symptoms except 'Sensitivity to injustice' could be compared with rubrics in Kent's original repertory.[10] 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 ≥ 1.5–3.0 for plain type means that we regard a medicine as indicated if the prevalence of the symptom in the medicine population is between 1.5 and 3 times larger than in the remainder of the population. A possible translation from type into LR is shown in Table 1.

If a symptom has LR = 1.5 for a certain medicine it will only slightly increase the probability that that medicine will work. If, say, the prior chance is 20% the posterior chance becomes 27% if the symptom is present. With a prior chance of 50% posterior chance becomes 60%.

Table 1 Repertory entries translated into LR values

Type	*LR*
Plain	1.5–2.9
Italics	3.0–5.9
Bold	≥6.0

In general peculiar symptoms are found in small rubrics, common symptoms in large rubrics. The rubric 'Diarrhoea from anticipation' in Kent's repertory contains three medicines. This is a relatively small rubric considering the prevalence of 4.4% in the whole population in our assessment. The symptom 'Fear of death' has a slightly smaller prevalence, but Kent's repertory rubric is much larger, 103 medicines. This could mean several things:

1. The symptom with the smaller rubric 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.

Points 2 and 3 are probably both valid. Point 1 can 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 with LR ≥ 1.5 will come up. If the rubric is -includes too many medicines some medicines in the rubric will have LR < 1.5, but the range between LR = 1 and LR = 1.5 remains uncertain.

At the end of data collection we evaluated 4072 prescriptions (4094 patients), but these included a large number of medicines. Chance could play an important role in these results. The 95% CI indicates if the LR might be equal to 1. We can also calculate the possibility that a medicine is rightly in a certain rubric and in the correct degree. If there are no patients with the symptom in a specific medicine population it is impossible to calculate LR, but we can estimate the chance that the prevalence of that symptom is above a certain value.

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) × (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 95% CI we used the program CAI (BMJ) (logarithmic transformation). The p-values are calculated using the Vassar stats website (http://faculty.vassar.edu/lowry/VassarStats.html).

To suggest adding or discarding of medicines we used different cut-off values. For adding a medicine to a rubric the criterion was $p > 0.70$, for discarding a medicine we used $p < 0.40$. These cut-off values are subjective reflect our caution about discarding existing entries. We were also cautious with entries based on small numbers. In the tables we mention only medicines which had more than one patient that responded well with the symptom present. Our suggestions are meant to initiate the discussion about this subject, not to make definitive changes.

Table 2 Results of the assessment of the symptom 'Diarrhoea from anticipation'

Diarrhoea from anticipation	*181*	a	c	b	d	*LR+*	*95% CI*
Prevalence = 4.4%	arg-n	12	14	169	3899	**11.1**	7.13–17.2
	bell	2	19	179	3894	**2.17**	
	calc	6	69	175	3844	**1.84**	
	caust	3	43	178	3870	**1.48**	
	cimic	2	5	179	3908	**6.52**	2.00–21.17
	elaps	2	6	179	3907	**5.71**	1.70–19.06
	gels	8	5	173	3908	**14.5**	9.19–22.78
	merc	4	50	177	3863	**1.69**	
	ph-ac	7	16	174	3897	**7.12**	3.76–13.39
	staph	3	29	178	3884	**2.14**	

The 95% CI is mentioned if it does not include 1.
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.

Results

Diarrhoea from anticipation

This symptom was defined as 'Diarrhoea from various sorts of anticipation, like before exams, doctors' appointment etc.'. In Kent's repertory the rubric contains: *Arg-n.*, *Gels.*, *Ph-ac*, all in italic. The LR assessment showed 181 (prevalence 4.4%) patients with this symptom. Some of the results are shown in Table 2.

The 95% CIs (above 6, our cut-off value for bold type) shows that *Gelsemium* and *Argentum nitricum* could be upgraded to bold type. We must beware, however, that LR could be exaggerated by confirmation bias, see Discussion. *Phosphoricum acidum* could be upgraded to bold type, but that is less certain. The chance that LR ≥ 6 is $p = 0.746$ (binomial approximation of normal distribution, one-tailed). According to our proposed cut-off values for probability bold type is rectified. The same applies to *Cimicifuga* ($p = 0.725$).

Elaps could be added to this rubric in italic ($p = 0.921$). The probability that *Belladonna* could be rightly added in plain type to this rubric is $p = 0.842$. Other additions in plain type could be *Calcarea carbonica* ($p = 0.775$), *Mercurius* ($p = 0.716$) and *Staphisagria* ($p = 0.842$). The chance that for *Causticum* LR > 1.5 is $p = 0.639$, but the chance that LR > 1 is $p = 0.857$, so 'diarrhoea from anticipation' could be a weak indication for *Causticum*.

All 10 results of our assessment of this symptom differ from Kent's original repertory. One could, however, question our cut-off values indicating upgrading of three entries. No medicines should be removed from this small rubric, six could be added and one could possibly be added.

These results show that we have some significant LRs. But even without sufficient data to calculate LR we have indications about medicines that should not be in the rubric. If, say, zero out of 88 *Sulphur* cases have 'diarrhoea from anticipation' we can be sure that this symptom is not an indication for *Sulphur* ($p = 0.011$).

Table 3 Assessment of the symptom 'Fear of death'

Fear of death	*158*	a	c	*LR+*	*Expected prevalence (%)*	*p-Value*	*95% CI*
Prevalence = 3.9%	acon	4	6	**10.6**	23.4	0.939	4.87–22.93
	am-c	2	7	**5.82**	5.9	0.987	1.69–19.87
	anac	5	7	**11.1**	5.9	0.999	5.57–22.02
	arg-n	2	24	**2.01**	11.7	0.399	
	ars	6	21	**5.95**	23.4	0.549	**2.88**–12.2
	calc	4	71	**1.39**	23.4	<0.001	
	carc	4	39	**2.45**			0.95–6.28
	ign	3	30	**2.38**	5.9	0.872	
	kali-p	2	14	**3.27**	5.9	0.935	
	lac-c	2	6	**6.55**	23.4	0.720	1.95–21.88
	lach	4	38	**2.51**	11.7	0.446	0.97–6.42
	lyc	4	82	**1.21**	11.7	0.020	
	mag-c	2	17	**2.75**			
	nat-m	3	153	**0.49**	5.9	0.016	
	nux-v	2	38	**1.3**	11.7	0.138	
	phos	4	72	**1.37**	23.4	<0.001	
	puls	2	57	**0.88**	11.7	0.038	
	sep	6	87	**1.7**	5.9	0.691	
	sil	2	31	**1.58**			
	sulph	1	87	**0.29**	5.9	0.047	
	verat	2	4	**8.74**	11.7	0.976	2.78–27.27

The expected prevalence is the prevalence according to the existing entry in Kent's repertory. *p*-Value is the chance that the repertory-entry is correct.

With five patients with 'diarrhoea from anticipation' out of 156 *Natrum muriaticum* cases it is unlikely that the prevalence of 'Diarrhoea from anticipation' in the *Natrum muriaticum* population will exceed 6.6% ($p = 0.058$), which corresponds with LR ≥ 1.5. Using the traditional method for creating rubrics we would add *Nat-m.* because the symptom is repeatedly seen in connection with this medicine. According to our Bayesian method this medicine should not be represented in this rubric.

For the next tables we will not indicate the prevalence of the symptom in the rest-population (columns *b* and *d*), because that does not vary much. Instead we mention the expected prevalence of the symptom in the medicine population according to the existing repertory grading and the probability of this entry considering our results ($p = 0.156$ means probability is 15.6%). If the LR is significant (does not include 1) or nearly significant, we mention the 95% CI.

Fear of death

The symptom was defined as 'Strong fear of death, more than once a week'. The rubric 'Fear of death' contains 103 medicines in Kent's repertory. There were 158 patients in this assessment with this symptom. The prevalence of the symptom in the whole population is 3.9%. Therefore we need a prevalence of at least 5.9% (=1.5 × 3.9%) in the medicine population before entering the medicine in the rubric. If the prevalence in the medicine population is 11.7% (=3 × 3.9%) it should be entered in italic, and if the prevalence is over 23.4% (=6 × 3.9%) it should be entered in bold type. The results are shown in Table 3.

Lycopodium, *Natrum muriaticum*, *Pulsatilla* and *Sulphur* should not be in this rubric. There are also some medicines where we could not calculate LR because no patient had fear of death. Numbers were sufficient to state that it is highly unlikely that these medicines should be in this rubric. These medicines are *Calcarea phosphorica*, *Graphites*, *Medorrhinum*, *Mercurius*, *Phosphoricum acidum* and *Staphisagria* (see Table 4).

The bold entry of *Calcarea carbonica* is incorrect. Even plain type is questionable ($p = 0.540$), but above our cut-off value for discarding the entry. The same goes for *Phosphorus* ($p = 0.532$ for plain type). *Lycopodium* in plain is unlikely ($p = 0.397$), but it is just under our cut-off value for discarding entries ($p < 0.40$).

The entry of *Anacardium* should be upgraded to bold type ($p = 0.960$, the *p*-value in Table 3 is the probability that LR > 1.5). *Veratrum* might be upgraded, but numbers are small.

Table 4 probabilities that medicines are indicated if the symptom 'Fear of death' is present, taking LR > 1.5 as cut-off value

Medicine	N	*Fear*	p
calc-p	28	0	0.182
graph	27	0	0.194
med	24	0	0.232
merc	54	0	0.037
ph-ac	23	0	0.247
staph	32	0	0.143

Carcinosinum ($p = 0.842$) and *Magnesium carbonicum* ($p = 0.902$) could be added. *Silicea* might be considered ($p = 0.692$).

For this symptom 17 out of 27 results differ from the original rubric. Three medicines should be added, 10 discarded, another four entries are in the wrong grade. The difference between rubric-size of 'diarrhoea from anticipation' and 'fear of death' is much less than in the original repertory. We also see lower LR values in the larger rubric, which might be consistent with the assumption that 'diarrhoea from anticipation' is a 'stronger' symptom, but also consistent with the confirmation bias hypothesis. If our existing knowledge comprises few medicines for a specific symptom we tend to choose fewer medicines if the symptom is present.

Grinding teeth during sleep

The symptom was defined as 'Grinding teeth during sleep, more than once a week'. In Kent's original repertory this rubric contains 36 medicines. There were 219 patients with this symptom and the prevalence in the whole population was 5.3%. Results are shown in Table 5.

Arsenicum album is in bold type in the original repertory. In our assessment there was only one *Arsenicum* patient out of 27 with grinding teeth during sleep. The chance that this medicine should be in this rubric, even in plain type, is $p = 0.356$, making this entry questionable.

Ammonium muriaticum ($p = 0.812$) and *Cenchris* ($p = 0.861$) might be added in bold type. *Cocculus* might be added in italic ($p = 0.861$).

Belladonna may be over-rated with bold type; italic is more probable ($p = 0.962$). There are good arguments to replace *Calcarea carbonica* in this rubric by *Calcarea muriatica* in italic ($p = 0.751$) and *Calcarea phosphorica* in plain type ($p = 0.933$).

Ignatia and *Mercurius* should be in this rubric, but possibly in plain type; chances are $p = 0.882$ and $p = 0.938$, respectively. *Staphisagria* could be added to this rubric ($p = 0.752$), as well as *Carcinosinum* ($p = 0.981$), *Pulsatilla* ($p = 0.672$) and *Phosphoricum acidum* ($p = 0.725$). *Argentum nitricum* is possibly indicated, but below our cut-off value ($p = 0.658$) for adding new entries.

The LR value for *Carcinosinum-con-cuprum* has to be handled with care. These prescriptions were made by two doctors only and confirmation bias is possible, see Discussion.

Out of 19 assessments of this symptom 11 did not correspond with the existing entries; one medicine (*Calcarea carbonica*) should be discarded (possibly also *Arsenicum album*), 10 added. Another three entries should be downgraded.

Herpes lips

The symptom was defined as 'Herpes lips, more than six times a year'. In Kent's original repertory this rubric contains 34 medicines. There were 205 patients with this symptom. The prevalence of the symptom in the whole population was 5%, see Table 6.

Table 5 Assessment of the symptom ' Grinding teeth during sleep'

Grinding teeth	*219*	a	c	*LR+*	*Expected prevalence (%)*	p-*Value*	*95% CI*
Prevalence = 5.3%	am-m	2	3	**7.54**			2.56–22.28
	arg-n	2	24	**1.44**			
	bell	6	15	**5.46**	31.8	0.468	2.75–10.90
	calc	3	72	**0.74**	7.95	0.147	
	calc-m	2	9	**3.42**			0.97–12.10
	calc-p	4	24	**2.7**			1.08–6.77
	carc	7	36	**3.11**			1.56–6.22
	car-c-c	2	7	**4.18**			1.23–14.34
	cench	2	2	**9.42**			3.52–25.39
	cocc	3	10	**4.36**			1.61–11.90
	ign	4	29	**2.29**	15.9	0.359	0.91–5.81
	merc	7	47	**2.47**	15.9	0.345	1.23–5.01
	ph-ac	2	21	**1.63**			
	psor	2	12	**2.69**	7.95	0.901	
	puls	5	54	**1.6**			
	sep	8	85	**1.63**	7.95	0.664	
	staph	3	29	**1.76**			
	thuj	2	22	**1.56**	7.95	0.703	

If the expected prevalence is not mentioned the medicine is not entered in Kent's repertory. *p*-Value is the chance that the repertory-entry is correct.

This is a small rubric and, like 'diarrhoea from anticipation', there are few entries that should not be there at all. *Sulphur*, *Natrum muriaticum*, *Rhus toxicodendron* and *Sepia* should be downgraded, even italic is improbable for *Rhus-t* and for *Sepia*. Italic would be more appropriate for *Natrum muriaticum* ($p = 0.596$). Remember that LR depends on the intensity of the symptom. Intuitively it is easy to understand that the indication for the medicine is stronger if the symptom occurs in a stronger degree than the mean in our assessment.

Borax might be upgraded to bold type ($p = 0.837$) according to our assessment, but numbers are small, so we prefer italic. Some medicines could be added in italic: *Aloe* ($p = 0.973$), *Baryta carbonica* ($p = 0.779$) and *Stramonium* ($p = 0.859$).

Some medicines could be added in plain type: *Bryonia* ($p = 0.932$), *China* ($p = 0.918$), *Gelsemium* ($p = 0.932$), *Lycopodium* ($p = 0.805$) and *Thuja* ($p = 0.733$). *Staphisagria* might even be added in italic ($p = 0.654$).

This rubric shows lower LR values than the other small rubric 'diarrhoea from anticipation', but these rubrics are not comparable. There are 9 values out of 17 that do not correspond with the original repertory, 9 medicines should be added. Another three entries should be downgraded.

Sensitive to injustice

The symptom was defined as 'Sensitivity to injustice, resulting in subsequent behaviour, like writing letters to papers and politicians, participating in protest groups, etc.'. This rubric does not exist in Kent's repertory. In RADAR-Synthesis (v 8.140) it contains: Calc., **Caust.**, Cupr., Dros., *Ign.*, Merc., Nux-v., Sep., **Staph.**, Verat. There were 379 patients with this symptom and the prevalence was 9.3% in the whole population, see Table 7:

None of the LR values we found indicate entries in bold type in this rubric. The bold entry for *Causticum* seems too much ($p = 0.016$), italic would be more appropriate ($p = 0.986$). This, however, could depend on cut-off value for the symptom (see Discussion).

The entries of *Calcarea carbonica* ($p = 0.161$), *Sepia* ($p = 0.051$) and *Staphisagria* ($p < 0.001$) in this rubric are

Table 6 Assessment of the symptom 'Herpes lips'

Herpes lips	*205*	a	c	*LR+*	*Expected prevalence (%)*	p-*Value*	*95% CI*
Prevalence = 5%	aloe	2	3	**8.06**			2.74–23.83
	bar-c	2	9	**3.66**			
	bor	2	3	**8.06**	7.5	0.996	
	bry	2	11	**3.09**			
	caust	6	40	**2.65**	7.5	0.946	1.27–5.68
	chin	2	12	**2.87**			
	gels	2	11	**3.09**			
	lach	4	38	**1.92**	7.5	0.795	
	lyc	8	78	**1.89**			
	nat-m	24	132	**3.35**	30	<0.001	2.26–4.98
	rhus-t	2	17	**2.11**	30	0.055	
	sep	10	83	**2.21**	30	<0.001	1.21–4.04
	sil	3	30	**1.83**	7.5	0.767	
	staph	5	27	**3.17**			1.41–7.20
	stram	2	7	**4.47**			1.31–15.33
	sulph	6	82	**1.37**	7.5	0.484	
	thuj	2	22	**1.67**			

Table 7 Assessment of the symptom 'Sensitive to injustice'

Injustice	*379*	a	c	*LR+*	*Expected prevalence (%)*	p-*Value*	*95% CI*
Prevalence = 10%	ambr	2	6	**2.71**			
	am-m	2	3	**4.34**			1.48–12.75
	anac	6	6	**5.47**			3.08–9.72
	aur	2	11	**1.67**			
	bell	4	17	**2.07**			
	bor	2	3	**4.34**			1.48–12.75
	calc	7	68	**1.01**	13.95	0.161	
	calc-m	2	9	**1.97**			
	carc	9	34	**2.29**			1.27–4.13
	car-c-c	4	5	**4.84**			2.32–10.12
	caust	18	28	**4.39**	55.8	0.016	3.02–6.38
	chin	2	12	**1.55**			
	cocc	5	8	**4.2**			2.10–8.40
	cupr	2	11	**1.67**	13.95	0.731	
	ign	6	27	**1.98**	27.9		0.95–4.11
	kali-bi	2	12	**1.55**			
	med	5	19	**2.27**			1.03–4.98
	merc	7	47	**1.41**	13.95	0.496	
	nat-m	15	141	**1.04**			
	nux-v	3	37	**0.81**	13.95	0.171	
	ph-ac	4	19	**1.89**			
	sep	7	86	**0.81**	13.95	0.051	
	staph	3	29	**1.01**	55.8	<0.001	

dubious, even in plain type. *Mercurius* ($p = 0.496$) is still possible.

Entries in italic could be justified for: *Ammonium muriaticum* ($p = 0.864$), *Anacardium* ($p = 0.974$), *Borax* ($p = 0.864$), *Carcinosinum-con-cuprum* ($p = 0.864$), *Cocculus* ($p = 0.875$).

Plain type entries are justified for: *Ambra* ($p = 0.912$), *Aurum* ($p = 0.731$), *Belladonna* ($p = 0.841$), *Calcarea muriatica* ($p = 0.938$), *China* ($p = 0.691$), *Ignatia* ($p = 0.832$), *Kalium bichromicum* ($p = 0.691$), *Medorrhinum* ($p = 0.893$) and *Phosphoricum acidum* ($p = 0.790$).

For *Carcinosinum-con-cuprum* see above and the Discussion. We could not compare the outcome of our assessment with the original repertory, because the symptom is not there.

Loquacity

The symptoms was defined as 'Loquacity that is conspicuous during consultation, or mentioned by others as one of the most important characteristics; also indications like the most talkative of the class'. In Kent's original repertory this rubric contains 97 medicines. There were 267 patients with this symptom and the prevalence in the whole population was 6.5%, see Table 8.

It seems counter intuitive that *Lachesis* should not be in bold type but our analysis suggests that italic would be more appropriate ($p = 0.993$). There could be a number of reasons, like

- Prevalence of the symptom in the research population may be high because of the prospective setting; people who are not obviously loquacious during consultation might become so on questioning.[11]
- Subjectivity in assessing this symptom is high.
- 'Changing the subject frequently' is an important qualification for this symptom regarding *Lachesis*; this qualification was not included in our assessment.[12]

Our assessment shows some entries which are probably incorrect in this rubric: *Calcarea carbonica* ($p = 0.031$), *Causticum* ($p = 0.332$), *Natrum muriaticum* ($p = 0.002$), *Nux-vomica* ($p = 0.238$), *Phosphorus* – not even in plain type ($p = 0.129$) – and *Sulphur* ($p = 0.071$).

Other medicines that should be discarded because there were no patients with the symptom and sufficient numbers of good responders are *Aurum*, *Cocculus*, *Cuprum* and *Gelsemium* (see Table 9).

One medicine, *Veratrum album*, might be upgraded to bold type ($p = 0.834$), but numbers are small. Maintaining italic seems wiser. For the same reason we don't advise adding *Digitalis* (LR = 10.3) in bold type, but in italic ($p = 0.901$).

Our findings suggest downgrading the bold type entries to italic for *Lachesis* ($p = 0.993$) and *Stramonium* ($p = 0.751$). For *Hyoscyamus* ($p = 0.945$) italic seems to fit better.

Upgrading to italic could be justified for: *Ambra* ($p = 0.948$) and *Tarentula* ($p = 0.808$).

Medicines that could be added in plain type are as follows: *Calcarea muriatica* ($p = 0.916$), *Medorrhinum* ($p = 0.799$) and *Tuberculinum* ($p = 0.780$). *Lycopodium* ($p = 0.657$) is questionable.

Some entries illustrate the problem with our cut-off values. The chance that *Belladonna* is rightly in italic is $p = 0.609$. The chance that for Sepia LR > 1.5 is $p = 0.421$, but the chance that for LR > 1 is $p = 0.849$, so loquacity might be a weak indication for Sepia.

For this symptom 16 out of 26 assessments are different from the original entries; 10 medicines should be discarded, 6 added. One medicine (Sepia) is uncertain.

Discussion

We took the liberty of suggesting some corrections to the repertory before the discussion about cut-off values had

Table 8 Assessment of the symptom 'Loquacity'

Loquacity	*267*	a	c	*LR+*	*Expected prevalence (%)*	p-*Value*	*95% CI*
Prevalence = 7%	ambr	3	5	**5.8**	9.75	0.995	2.36–14.35
	anac	2	10	**2.57**	9.75	0.895	
	bell	4	17	**2.95**	19.5	0.609	1.22–7.20
	calc	2	73	**0.4**	9.75	0.031	
	calc-m	2	9	**2.8**			
	caust	3	43	**1**	9.75	0.332	
	cimic	2	5	**4.41**	19.5	0.860	1.36–14.34
	dig	2	1	**10.3**			4.60–23.16
	hyos	5	10	**5.19**	39	0.421	2.52–10.75
	lach	14	28	**5.34**	39	0.271	3.43–8.35
	lyc	9	77	**1.63**			
	med	3	21	**1.93**			
	nat-m	4	152	**0.38**	9.75	0.002	
	nux-v	2	38	**0.76**	9.75	0.238	
	phos	4	72	**0.8**	19.5	0.001	
	saccha	2	5	**4.41**			1.36–14.34
	sep	8	85	**1.33**			
	stram	2	7	**3.43**	39	0.248	1.01–11.73
	sulph	4	84	**0.69**	9.75	0.071	
	tarent	2	6	**3.85**	9.75	0.964	1.16–12.91
	tub	2	14	**1.92**			
	verat	3	3	**7.74**	19.5	0.928	3.46–17.43

even started. We used different cut-off values for entering new medicines ($p > 0.70$) than for discarding existing entries ($p < 0.40$). Our examples might help to make this discussion more concrete. The reliability of our repertories is seriously threatened if we allow new entries without considering Bayes' theorem. Our work shows that we should also assess the existing entries.

Our final results confirm our conclusion from a previous analysis, that the repertory is flawed.[8] In this paper we tried to introduce statistical reasoning in the process of introducing or discarding entries from the repertory as we previously proposed.

Three years of research yielded 121 relevant results: 48 significant values for LR, 73 other values with sufficient probability to validate repertory entries. Our results differ in 75% of 99 medicine-entries for five rubrics (excluding 'sensitive to injustice') in Kent's repertory, but if we disregard upgrading or downgrading of entries 56% of the medicines are wrongly entered in these five repertory rubrics or missing from them. For these five rubrics 21 results suggest removal and 34 adding of medicines. Most additions are to the small rubrics, most removals are from the large rubrics. These figures represent only five rubrics. We need to assess many more symptoms before we can make a statement about the correctness of the existing repertory. Note that we compared our results only with the original repertory, not with later versions.

We found discrepancies between our results and Kent's original repertory, as expected because the Kent's did not consider the influence of chance. But should the repertory be adjusted towards our results? There are uncertainties regarding semantics; doctors have their own interpretation of the existing rubrics and these interpretations might differ from ours. It is uncertain if the results of this group in The Netherlands are reproducible all over the world. There is also significant variance within our research group. And there is bias.

The most important bias is confirmation bias. Our prescriptions are based on existing knowledge and sometimes the absence of a symptom will prevent us from prescribing the related medicine. This is known for symptom–medicine relations like 'fear of dark' – *Stramonium*, 'loquacity' – *Lachesis* and 'sensitive to injustice' – *Causticum*. Confirmation bias could become less over time. In the first evaluation of the LR project there were no *Causticum* patients without 'Sensitivity to injustice'[13]; later on people without this symptom, who started with another medicine, had better reactions on *Causticum*.[14] Figure 1 shows how the proportions of *Causticum* patients with and without

Table 9 probabilities that medicines are indicated if the symptom 'Loquacity' is present, taking LR > 1.5 as cut-off value

Medicine	N	*Loquacity*	p
Aur	13	0	0.264
Cocc	13	0	0.264
Cupr	13	0	0.264
Gels	13	0	0.264

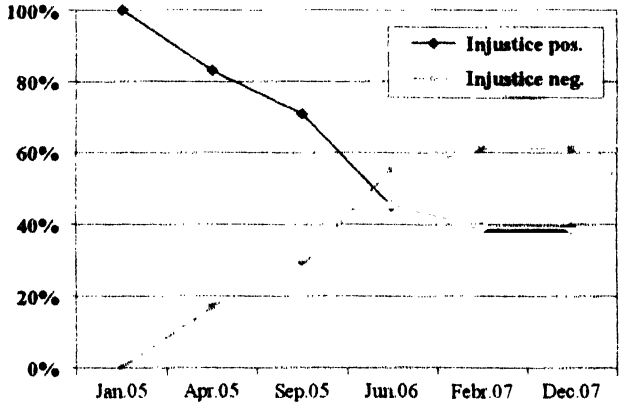

Figure 1 The influence of confirmation bias on the prevalence of the symptom 'Sensitive to injustice' in the population responding well to *Causticum*.

'Sensitivity to injustice' increased over time then stabilized We also noted his kind of development for 'grinding teeth at night' – *Mercurius*. The 'wash out' of confirmation bias took about two years. We stopped the inclusion of new prescriptions three months before the last evaluation to be more sure. This is not sufficient for medicines that were not prescribed in the first 18 months. *Carcinosinum-con-cuprum* appeared for the first time in May 2006 in the database. For the symptoms 'grinding teeth during sleep' and *Mercurius* and 'sensitive to injustice' and *Causticum* we found a similar tendency, see Figure 1, but the wash out of confirmation bias might have been incomplete for these medicine–symptom combinations.

Confirmation bias could also cause over-estimation of LR's in small symptom-rubrics. At the moment a limited number of medicines are considered if the patient has 'Diarrhoea from anticipation'. If the number of possible medicines grows the prevalence of this symptom will be spread out over more medicines and decrease for each medicine.

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 69% 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 mainly based on his own experience.

The LR values are related to thresholds for symptoms.[7] In this assessment the prevalence of 'Sensitive to injustice' is 9.3%. 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.

There are several precautions we should take in interpreting our results. We have already discussed confirmation bias and threshold values, but there may be other biases. On the other hand, Kent's 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. LR close to zero indicates a contra-indication for that medicine. In an electronic repertory it is possible to show such information on demand. We might also discover that a certain LR value is different for children or for women or for certain age categories. We also have to choose between displaying the entries as 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 rather than geographical factors.[8] 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

Our LR assessment confirms that frequently used medicines are over represented in the repertory, but not systematically. We also found that the size of repertory rubrics varies much more than expected from the prevalence of symptoms. Large rubrics tend to have superfluous entries where in small rubrics medicines are missing. In this assessment 56% 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.

Conflicts of interest

No conflicts of interest declared.

Acknowledgements

The doctors participating in this research 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.

The research was sponsored by KVHN (Royal Dutch patient's organisation for homeopathy), the Louise van Eeghen foundation, SFWOH (foundation for homeopathic research), the Blackie Foundation Trust and VHAN (Dutch homeopathic doctor's association).

The sponsors had no influence on study design, the collection, analysis and interpretation of data; on the writing of the manuscript; and on the decision to submit the manuscript for publication.

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 ALB, 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. *Philos Trans R Soc London* 1763; **53**: 370–418.
4 Gill CJ, Sabin L, Schmid CH. Why clinicians are natural Bayesians. *BMJ* 2005; **330**: 1080–1083.
5 Woodworth GG. *Biostatistics: a Bayesian introduction*. Wiley, 2004.
6 Altman DG, Machin D, Bryant TN, Gardner MJ (eds). *Statistics with confidence*. BMJ Books, ISBN 0-7279-1375-1; 2000.
7 Rutten ALB, Stolper CF, Lugten RF, Barthels RJ. New repertory, new considerations. *Homeopathy* 2008; **97**: 16–21.
8 Rutten ALB, Stolper CF, Lugten RF, Barthels RJ. A Bayesian perspective on the reliability of homeopathic repertories. *Homeopathy* 2006; **95**: 88–93.
9 Rutten ALB, Stolper CF, Lugten RF, Barthels RJ. 'Cure' as the gold standard for likelihood ratio assessment: theoretical considerations. *Homeopathy* 2004; **93**: 78–83.
10 Kent JT. *Repertory of the homoeopathic materia medica*. Enriched 6th American edn. New Delhi: World Homoeopathic Links, 1982.
11 Rutten ALB, Stolper CF, Lugten RF, Barthels RJ. Repertory and the symptom loquacity: some results from a pilot-study. *Homeopathy* 2004; **93**: 190–192.
12 Rutten ALB, Stolper CF, Lugten RF, Barthels RJ. Assessing likelihood ratio of clinical symptoms: handling vagueness. *Homeopathy* 2003; **92**: 182–186.
13 Rutten Lex, Stolper Erik, Lugten Roland, Barthels Rob. Onderzoek likelihood ratio: stand van zaken januari 2005. *SSC* 2005; **35**(2): 17–22.
14 Rutten Lex, Stolper Erik, Lugten Roland, Barthels Rob. Onderzoek likelihood ratio: april 2005. *SSC* 2005; **35**(2): 23–24.