Credal Classification for Dementia Screening*

Marco Zaffalon[1], Keith Wesnes[2], and Orlando Petrini[3]

[1] IDSIA, Galleria 2,
CH-6928 Manno, Switzerland
zaffalon@idsia.ch
[2] Cognitive Drug Research Ltd., CDR House, 24 Portman Road,
Reading RG30 1EA, UK
keithw@cdr.org.uk
[3] Pharmaton SA, Via ai Mulini,
CH-6934 Bioggio, Switzerland
petrini@lgn.boehringer-ingelheim.com

Abstract. Dementia is a very serious personal, medical and social problem. Early and accurate diagnoses seem to be the key to effectively cope with it. This paper presents a diagnostic tool that couples the most widely used computerized system of cognitive tests in dementia research, the Cognitive Drug Research system, with the naive credal classifier. Although the classifier is trained on an incomplete database, it provides unmatched predictive performance and reliability. The tool also proves to be very effective in discriminating between Alzheimer's disease and dementia with Lewy bodies, which is a problem on the frontier of research on dementia.

1 Introduction

Dementia is becoming recognized to be one of the leading causes for concern in the elderly, both from the perspective of their own quality of life and also the social issues concerning the hugely growing costs of caring for the rapidly growing population who require constant supervision and medical care [1, 14]. The most important type of dementia is Alzheimer's Disease (AD) accounting for approximately 50% of all types of dementia. Vascular Dementia (VD) has traditionally been considered the second most common cause of dementia (up to 20% of dementias, either alone, or in combination with AD). Dementia with Lewy Bodies (DLB) is an only recently described type of dementia which is becoming recognized to be the second most common single form of dementia accounting for over 20% of all dementias [14]. It has previously often been misdiagnosed as Alzheimer's disease, as well as confused with schizophrenia. One of the major problems confronting trials with DLB is the correct diagnose of the disorder [14, 23–25].

There is currently no cure for any dementia although three compounds from a class of drug called anticholinesterases (galanthamine, rivastigmine and donepezil)

have been registered in various countries for the mild symptomatic relief of AD. Current research is trying to early identify AD as the hope is that these compounds may prove more effective in treating the early stages of dementia (often termed "mild cognitive impairment") and preventing rather than just reducing symptoms [27]. The first clinical trial has just been completed showing that rivastigmine can dramatically improve cognitive function in DLB [13].

Three major problems confront research in this field. On one hand, it is questionable whether or not systems are sensitive enough to detect early stages of dementia. On the other, it must still be confirmed whether or not tests are capable of differentiating different types of dementia; and measures for assessing therapeutic response to treatment must still be unequivocally determined. In this paper we focus on the first two problems. By coupling the power of emerging classification tools and the diagnostic capabilities of a well-targeted system of cognitive tests, we propose an automated diagnostic system that deals successfully with both problems.

As far as cognitive tests, we rely upon the Cognitive Drug Research (CDR) computerized assessment system. This has been designed to provide a valid, reliable and sensitive tool for assessing cognitive functions in dementia [2, 16, 17, 20, 23–25, 28]. The system is the *most widely* used automated system in dementia research [16] (see Sect. 2.1). We use a database describing the actual health state and the past responses to the CDR system tests for about 3,400 patients (Sect. 2.3). The data were not collected for the purpose of data mining, so they are not completely accurate, presenting a substantial amount of missing values. Missing data are recognized to be a fundamental problem in the machine learning literature [4]; treating them properly is essential to draw reliable inferences.

These challenging issues motivated us to choose the classification model called *Naive Credal Classifier* [30, 31] (NCC, Sect. 2.2), a generalization of the well-known discrete *Naive Bayes Classifier* (or NBC [6]) to sets of probability distributions [26]. The NCC is currently the only classifier that takes into account the imprecision arising from prior uncertainty about the unknown distribution generating the data; and it is also robust to all the possible mechanisms of missingness. The robust Bayes classifier from Ramoni and Sebastiani presents similarities in the treatment of missing data [18], although it appears to adopt an overly conservative approach. Also, it neglects the former type of imprecision.

The characteristics of the new paradigm of credal classification enable the NCC to automatically do reliable classifications. This is the first application of credal classification to the field of dementia screening. We realize it by analyzing the predictive behavior of the NCC by an empirical study on the database, in Sect. 3. In doing this, we remarkably improve upon the diagnostic accuracy described in similar past work [12], with up to 95% correct predictions. We also show that the system is very effective in discriminating among dementias, even between the two types currently only hardly distinguishable, AD and DLB. Overall, we successfully deal with the problem of obtaining reliable inferences, which is fundamental for the application domain under consideration and it is also more critical given the incompleteness of the database.

2 Methods

2.1 The CDR system

The CDR system is a computer based system, the patient responding by using a simple response box containing just a "yes" and "no" button for each test. It takes between 25 and 40 minutes to administer the tests of the system, depending on the level of dementia shown by the patients.

Results from the CDR system have encouraged the *International Group on Dementia Drug Guidelines* to issue a position paper on assessing function in future clinical trials [10]. The working group concluded that existing testing procedures (e.g. the Alzheimer's disease assessment scale) do not properly identify all of the cognitive deficits suffered by AD patients, particularly attentional deficits, and have recommended that automated procedures should be used alongside more traditional ones to support them in this and to ultimately determine whether they should supersede traditional methods [10]. The CDR system has shown sensitivity in identifying mild cognitive impairment [17, 27], has been shown capable of differentiating various types of dementia (AD, DLB, VD, Huntington's Chorea; [2, 16, 22–25]), of measuring therapeutic response to a variety of medications in both AD [1, 9, 15, 19, 21] and DLB [13], and has shown superior sensitivity in identifying AD and Huntington's disease to all of the most widely used non-automated procedures [16].

The CDR system is the most widely used automated system in clinical research worldwide. It is used in almost every European country, North and South America, Russia, South Africa, India, Australia and New Zealand. It is used in hospitals, universities, private and government research facilities to study the effects of new medicines, ageing, disease, trauma, dementia as well as other factors such as mobile phones, altitude and so on.

2.2 Credal Classification

Classification is a technique concerned with allocating new objects to a finite set of previously defined groups (classes) on the basis of observations on several characteristics of the objects, called *attributes* or *features* [7]. In the present application, each test of the CDR system is an attribute of the problem the values of which are the possible outcomes of the test. The class is the variable the values of which are the possible states of dementia (including the state "no dementia").

Credal classification [30] is a new way of conceiving the task of prediction that generalizes the common notion of classification: a credal classifier is defined as a function that maps an instance of a set of features to a *set of states* of a categorical class variable (a common classifier maps an instance of the features to a single class). Classifiers are usually inferred from a database. A credal classifier is a basis from which imprecision in the data can be taken into account, as generated by unobserved or rare events, small sample sizes and missing data. As a consequence, for a given state of the attributes, imprecision in the input may

prevent a single output class from being obtained; then the result of a credal classifier is a set of classes, all of which are candidates to be the correct category. That is, a credal classifier recognizes that the available knowledge may not suffice to isolate a single class and thus gives rise to a set of alternatives. Reliability is thus a concept intrinsic in the definition of credal classification.

It is easy to use a credal classifier for diagnostic purposes: in the present case, the vector of responses to the CDR tests for a patient is mapped to a set of possible states of dementia by the function represented by the classifier. Credal classification is realized in this paper by the naive credal classifier [30]. The NCC is a special type of credal network [8, 3] that generalizes the NBC. It maintains the good properties of the naive Bayes classifier [5], but it relaxes the assumption that the model probabilities be precise. Most importantly, it can easily and robustly be inferred from possibly small and incomplete data sets [31]. In particular, an incomplete data set is regarded as a collection of complete data sets that arises by considering all the possible replacements of the missing data with admissible values [29]. The NCC is inferred by considering the collection as a set of possible samples, so that the NCC inferences are robust to each possible mechanism of missing data. This is equivalent to considering many probability distributions be plausible given the data and to treat all of them together as a set of possible distributions, as it is common in the field of *imprecise probabilities* [26]. A set of distributions gives rise to a set of output classes in general, given an instance of the attributes, thus realizing a credal classifier.

2.3 The Database

The database is constituted by 3,385 records. Each record stores the responses to the CDR system tests for a patient. The results are expressed by either continuous or integer numbers. Each record also reports the actual health state of the patient, which is classified into 5 categories: normal, AD, to undergo Coronary Bypass Surgery (CBS), DLB and VD. Table 1 shows the percentages of patients in each class.

Table 1. Percentual distribution of the classes in the database.

	Normal	AD	CBS	DLB	VD
Percentage	67.4%	22.9%	2.4%	3.9%	3.4%

The tests are carried out in different ways according to the state of the patient. This leads to the presence of more features for the dementia patients than for the normal controls. Seven attributes are used when normal people are involved in the study (the number in parentheses is the percentage of missing values): delayed word recognition reaction time (6.4%), delayed word recognition sensitivity index (5.6%), digit vigilance false alarms (4.3%), digit vigilance reaction time (3.3%), digit vigilance accuracy (3.3%), choice reaction time (1.6%) and choice reaction time accuracy (1%). The choice reaction time is obtained

as follows: either the word "no" or the word "yes" is presented on the monitor and the patient is instructed to press the corresponding button as quickly as possible. There are 20 trials for which each stimulus word is chosen randomly with equal probability and there is a varying inter-stimulus interval. We do not describe here how all other measures are obtained. The interested reader can consult the related references.

The above attributes are used in the first analysis (Sect. 3.2). The second analysis (Sect. 3.3) is based on the 1,103 units of the database restricted to the dementia group. In the second analysis we have included 18 attributes (the number in parentheses is the percentage of missing values): digit vigilance reaction time (4.6%), numeric working memory reaction time (3.9%), numeric working memory original stimuli accuracy (3.9), numeric working memory new stimuli accuracy (3.9%), numeric working memory sensitivity index (3.9), digit vigilance accuracy (3.5%), picture recognition speed (2.3%), picture recognition original stimuli accuracy (2.3%), picture recognition new stimuli accuracy (2.3%), picture recognition sensitivity index (2.3%), delayed word recognition reaction time (2.1%), delayed word recognition original stimuli accuracy (2.1%), delayed word recognition new stimuli accuracy (2.1%), delayed word recognition sensitivity index (2.1%), choice reaction time (1.5%), choice reaction time accuracy (1.5%), simple reaction time (1.1%) and age (0.2%).

3 Experiments

3.1 Experimental Methodology

The database was randomly split into a learning set and a test set. The learning set was used to infer the classifier; its size being fixed at 50% of the database size. On the remaining 50% of cases, i.e. the test set, the true classes were hidden to the classifier in order to study its predictive accuracy, i.e. the relative number of correct guesses on a set of unseen units.

As far as the NCC, it can have different degrees of caution expressed by the real parameter s [31]. The parameter plays a role analogous to the weight given to the prior in Bayesian models. This study uses $s = 1$. Further, since the NCC assumes that the attributes are categorical, the database was initially discretized by MLC++ [11], default options. The discretization was made on the basis of the learning set only.

The empirical analysis of the NCC benefits from comparing it with the NBC [31]. Several NBCs are considered, related to the Bayesian prior distribution chosen to infer them. We consider four well-known so-called noninformative priors: Haldane, Uniform, Perks and Jeffreys. We also consider other three priors obtained by modifying some of the former. In this case the original priors are required to satisfy the structural constraints implied by the special classifier under consideration. These priors are called: Uniform$'$, Perks$'$ and Jeffreys$'$ (see [31] for a thorough explanation). The Bayesian classifiers were inferred by discarding, separately for each attribute, the missing values (this is possible since the NBC assumes independence of the attributes conditional on the class).

3.2 Detecting Dementia

In the first experiment the goal is to distinguish normal people from people in the dementia group. Dementias are clustered into one class so that the class variable is binary with values in: normal group (67.4%) and dementia group (32.6%). There are 7 attributes describing a patient, as reported in Sect. 2.3. The results of the experiment are shown in Tab. 2. Each row in the table refers to a different

Table 2. Results of the discrimination between normal people and people in the dementia group. A value in a cell is expressed as a percentual number ± its standard deviation.

	C_1%	N%	Ns%	S%
Haldane	94.77±0.57	92.59±0.64	72.22±3.89	9.68±0.72
Perks	94.77±0.57	92.29±0.65	69.14±4.02	9.68±0.72
Perks′	94.77±0.57	92.41±0.65	70.37±3.97	9.68±0.72
Uniform	94.77±0.57	92.41±0.65	70.37±3.97	9.68±0.72
Uniform′	94.77±0.57	92.47±0.65	70.99±3.95	9.68±0.72
Jeffreys	94.77±0.57	92.41±0.65	70.37±3.97	9.68±0.72
Jeffreys′	94.77±0.57	92.47±0.65	70.99±3.95	9.68±0.72

prior distribution for the NBC. The columns are described below. Bear in mind that the predictive accuracy is the relative number of correct guesses.

- C_1% is the accuracy of the NCC on the subset of instances where it is possible to provide a single class according to the NCC.
- N% is the accuracy of the NBC on the entire test set.
- Ns% is the accuracy of the NBC on the subset of instances for which the NCC outputs more than one class.
- S% is the percentage of instances for which the NCC outputs more than one class.

Discussion. When the credal classifier isolates a single class, it has a very high accuracy of prediction (C_1%). In about 10% of cases (S%), it suggests that there is not enough knowledge to isolate a single class and it outputs both, thus not giving any judgment. Note that on this subset *all* the Bayesian models have a much worse prediction (Ns%) than that of the NCC. For this reason the accuracy of the NBCs on the entire test set (N%) is worse than that of the NCC.

The NCC is thus able to isolate a subset of units where robust predictions are possible (despite the missing data). Note that the NBCs realize non-random predictions on the subset of instances where the NCC does not provide any judgment: the NBCs achieve about 70% accuracy that is greater than the 50% accuracy that would be obtained by randomly guessing (recall that the class is binary). This effect is due to the finiteness of the sample and does not mean that the NBCs should be applied on the subset related to S% [31]. Instead, this

may suggest that the data remarkably violate the assumption of independence between attributes conditional on the class—that is made both by the NCC and the NBC—and that it might be worth trying more structured credal classifiers.

3.3 Discriminating among Dementias

In the second analysis the goal is to assign a diseased patient to the correct disease type. The class takes values in the set of 4 dementias reported in Sect. 2.3, which also reports the 18 attributes used. The results are in Tab. 3. As

Table 3. Results of the discrimination among dementias.

	C_1%	Cs%	N%	Ns%	S%	Z
Haldane	94.05±1.01	98.28±1.20	91.79±1.17	83.62±3.40	21.64±1.76	2.16±0.43
Perks	94.05±1.01	98.43±1.10	89.76±1.30	75.59±3.81	23.22±1.81	2.31±0.66
Perks′	94.05±1.01	98.43±1.10	90.86±1.23	80.31±3.53	23.22±1.81	2.31±0.66
Uniform	94.05±1.01	98.43±1.10	90.86±1.23	80.31±3.53	23.22±1.81	2.31±0.66
Uniform′	94.05±1.01	98.43±1.10	91.22±1.21	81.89±3.42	23.22±1.81	2.31±0.66
Jeffreys	94.05±1.01	98.43±1.10	89.58±1.31	74.80±3.85	23.22±1.81	2.31±0.66
Jeffreys′	94.05±1.01	98.43±1.10	91.04±1.22	81.10±3.47	23.22±1.81	2.31±0.66

before, the rows of the table refer to different NBCs. (The statistics related to Haldane's prior were computed on the 98% of units only, because, since it is an improper prior, the NBC classification was undefined in the remaining 2% of cases.) There are two more columns with respect to Tab. 2:

- Cs% is the empirical probability that the actual class belongs to the set of classes proposed by the NCC. This measure is computed when the output set of the NCC is made by two classes at least;
- Z is the average number of classes proposed by the NCC. This measure is computed when the output set of the NCC is made by two classes at least.

Discussion. Again, we see that the performance of the credal classifier is very good when it isolates a single class (C_1%). When it outputs more than one class, there are about 2 on average (Z); and the probability that the actual class belongs to such a set is very high (Cs%). This happens about 1 time out of 4 (S%). The values of S% are larger than those in Tab. 2 because now the learning set size is about 1/3 of the learning set of the first experiment and because there are more attributes; larger collections of data, or less missing values, would quickly decrease this type of indeterminacy. The considerations related to the columns N% and Ns% are similar to the case of the preceding experiment.

In order to better analyze the capability of the credal classifier to assign a patient to the actual class, we represent the *confusion matrix* in Tab. 4. We restrict attention to the 76.88% of units where the NCC suggested only one class. The cells of the confusion matrix report the empirical probabilities of the

pairs of random variables representing the predicted class and the actual class, respectively, as computed from the test set: e.g. the empirical joint probability that the class predicted by the NCC is AD and the actual class is AD is 0.7571.

The confusion matrix shows that the NCC performance to assign patients to actual classes is excellent: for instance given that the NCC has predicted Alzheimer's disease, there is only a small probability that the actual disease is VD and zero probability that it is DBL or CBS. This outcome is similar for the other dementias. Most importantly, the confusion matrix shows that the NCC discriminates between AD and DLB very well. This is a very important result for research on dementias: DLB has frequently been misdiagnosed with AD. Here we show that the CDR system is capable of distinguishing them and that it is even possible to do this automatically by means of the NCC.

Table 4. Confusion matrix. The boldface values are percentages related to correct predictions.

		Predicted class			
		AD	CBS	DLB	VD
Actual class	AD	**75.71**	0.71	0.24	0.00
	CBS	0.00	**8.10**	0.00	0.00
	DLB	0.00	0.48	**5.95**	1.90
	VD	0.95	0.00	1.67	**4.29**

4 Conclusions

Cognitive tests for dementias are getting more and more important, as early diagnosis seems to be the basis for coping successfully with the diseases. This paper shows that coupling targeted cognitive tests such as the CDR computerized system with a reliable classifier such as the NCC, enables very accurate diagnoses to be automatically made. A particularly successful feature of the overall system described in this paper is the capability to discriminate between Alzheimer's disease and dementia with Lewy bodies. This is particularly important, as non-automated and non-computerized diagnoses often have problems in detecting the subtle differences in symptoms linked to the two disease types.

Diagnoses have also been shown to be robust to a non-negligible number of missing values in the database. This result is due to the powerful characteristics of credal classification, which is applied to such an important domain for the first time. The NCC enabled the difficult problem of the incomplete database to be dealt with easily and robustly.

However, the imprecision resulting from the missing data did result in less precise inferences. Future data sets with less missing values will reduce the indeterminacies in the predicted classes, and thus enable the demonstrated predictive capability of the method to be realized in all instances.

Acknowledgements

Thanks to L. M. Gambardella and C. Lepori for their kind attention and support. M. Zaffalon was partially supported by SUPSI DIE under CTI grant # 4217.1.

References

1. H. Allain, E. Neuman, M. Malbezin, V. Salzman, Guez D., K Wesnes, and J. M. Gandon. Bridging study of S12024 in 53 in-patients with Alzheimer's disease. *J. Am. Geriatr. Soc.*, 45:125–126, 1997.
2. G. A. Ayre, A. Sahgal, I. G. McKeith, C. G. Ballard, K. Lowery, C. Pincock, M. P. Walker, and K. Wesnes. Distinct profiles of neuropsychological impairment in dementia with Lewy bodies and Alzheimer's disease. *Neurology*. In press.
3. F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.
4. P. Domingos. Machine learning. In W. Klosgen and J. Zytkow, editors, *Handbook of data mining and knowledge discovery*. Oxford University Press, New York. To appear.
5. P. Domingos and M. Pazzani. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2/3):103–130, 1997.
6. R. O. Duda and P. E. Hart. *Pattern classification and scene analysis*. Wiley, New York, 1973.
7. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley, 2001. 2nd edition.
8. E. Fagiuoli and M. Zaffalon. 2U: an exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106(1):77–107, 1998.
9. T. D. Fakouhi, Jhee S. S., J. J. Sramek, C. Benes, P. Schwartz, G. Hantsburger, R. Herting, E. A. Swabb, and N. R. Cutler. Evaluation of cycloserine in the treatment of Alzheimer's disease. *J. Geriatr. Psychiatry Neurol.*, 8:226–230, 1995.
10. S. Ferris, U. Lucca, R. Mohs, B. Dubois, K. Wesnes, H. Erzigkeit, D. Geldmacher, and N. Bodick. Objective psychometric tests in clinical trials of dementia drugs. *Alzheimer Disease and Associated Disorders*, 11(3):34–38, 1997. Position paper from the International Working Group on Harmonisation of Dementia Drug Guidelines.
11. R. Kohavi, G. John, R. Long, D. Manley, and K. Pfleger. MLC++: a machine learning library in C++. In *Tools with Artificial Intelligence*, pages 740–743. IEEE Computer Society Press, 1994.
12. S. Mani, M. B. Dick, M. J. Pazzani, E. L. Teng, D. Kempler, and I. M. Taussig. Refinement of neuro-psychological tests for dementia screening in a cross cultural population using machine learning. In W. Horn, Y. Shahar, G. Lindberg, S. Andreassen, and J. Wyatt, editors, *Lecture Notes in Computer Science*, volume 1620, pages 326–335. Springer, 1999. Proc. of the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making, AIMDM'99, Aalborg, Denmark.
13. I. McKeith, T. Del Ser, F. Spano, K. Wesnes, R. Anand, A. Cicin-Sain, R. Ferrera, and R. Spiegel. Efficacy of rivastigmine in dementia with Lewy bodies: results of a randomised placebo-controlled international study. *Lancet*, 356:2031–2036, 2000.
14. I. G. McKeith and G. A. Ayre. Consensus criteria for the clinical diagnosis of dementia with Lewy bodies. In K. Iqbal, B. Winblad, T. Nishimura, M. Takeda, and H. M. Wisnieswski, editors, *Alzheimer's Disease: Biology, Diagnosis and Therapeutics*, pages 167–178. Wiley, 1997.

15. E. Mohr, V. Knott, M. Sampson, K. Wesnes, R. Herting, and T. Mendis. Cognitive and quantified electroencephalographic correlates of cycloserine treatment in Alzheimer's disease. *Clinical Neuropsychopharmacology*, 18:23–38, 1995.
16. E. Mohr, D. Walker, C. Randolph, M. Sampson, and T. Mendis. The utility of clinical trial batteries in the measurement of Alzheimer's and Huntington's dementia. *International Psychogeriatrics*, 3:397–411, 1996.
17. C. G. Nicholl, S. Lynch, C. A. Kelly, L. White, L. Simpson, P. M. Simpson, K. Wesnes, and B. M. N. Pitt. The cognitive drug research computerised assessment system in the evaluation of early dementia—is speed of the essence? *International Journal of Geriatric Psychiatry*, 10:199–206, 1995.
18. M. Ramoni and P. Sebastiani. Robust Bayes classifiers. *Artificial Intelligence*, 125(1–2):209–226, 2001.
19. K. R. Siegfried. Pharmacodynamic and early clinical studies with velnacrine. *Acta Neurol. Scand.*, 149(10):26–28, 1993.
20. P. M. Simpson, D. J. Surmon, K. A. Wesnes, and G. R. Wilcock. The cognitive drug research computerised assessment system for demented patients: a validation study. *International Journal of Geriatric Psychiatry*, 6:95–102, 1991.
21. L. Templeton, A. Barker, K. Wesnes, and D. Wilkinson. A double-blind, placebo-controlled trial of intravenous flumazenil in Alzheimer's disease. *Human Psychopharmacology*, 14:239–245, 1999.
22. M. P. Walker, G. A. Ayre, C. H. Ashton, V. R. Marsh, K. Wesnes, E. K. Perry, J. T. O'Brien, I. G. McKeith, and C. G. Ballard. A psychophysiological investigation of fluctuating consciousness in neurodegenerative dementias. *Human Psychopharmacology*, 14:483–489, 1999.
23. M. P. Walker, G. A. Ayre, J. L. Cummings, K. Wesnes, I. G. McKeith, J. T. O'Brien, and C. G. Ballard. Quantifying fluctuation in dementia with Lewy bodies, Alzheimer's disease and vascular dementia. *Neurology*, 54:1616–1625, 2000.
24. M. P. Walker, G. A. Ayre, J. L. Cummings, K. Wesnes, I. G. McKeith, J. T. O'Brien, and C. G. Ballard. The clinician assessment of fluctuation and the one day fluctuation assessment scale. *British Journal of Psychiatry*, 177:252–256, 2000.
25. M. P. Walker, G. A. Ayre, E. K. Perry, K. Wesnes, I. G. McKeith, M. Tovee, J. A. Edwardson, and C. G. Ballard. Quantification and characterisation of fluctuating cognition in dementia with Lewy bodies and Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, 11:327–335, 2000.
26. P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.
27. K. Wesnes. Predicting, assessing, differentiating and treating the dementias: experience in MCI and various dementias using the CDR computerised cognitive assessment system. In B. Vellas and L. J. Fitten, editors, *Research and practice in Alzheimer's disease*, volume 3, pages 59–65. Serdi, Paris, 2000.
28. K. Wesnes, K. Hildebrand, and E. Mohr. Computerised cognitive assessment. In G. W. Wilcock, R. S. Bucks, and K. Rocked, editors, *Diagnosis and management of dementia: a manual for memory disorders teams*, pages 124–136. Oxford Univ. Press, Oxford, 1999.
29. M. Zaffalon. Exact credal treatment of missing data. *Journal of Statistical Planning and Inference*. To appear.
30. M. Zaffalon. The naive credal classifier. *Journal of Statistical Planning and Inference*. To appear.
31. M. Zaffalon. Statistical inference of the naive credal classifier. In G. de Cooman, T. Fine, and T. Seidenfeld, editors, *ISIPTA'01*, pages 384–393, The Netherlands, 2001. Shaker Publishing.