# An advisory system for the treatment of kidney stones

Judy Hardy*/Alistair Armitage,

School of Computing

Napier University

a.armitage@napier.ac.uk

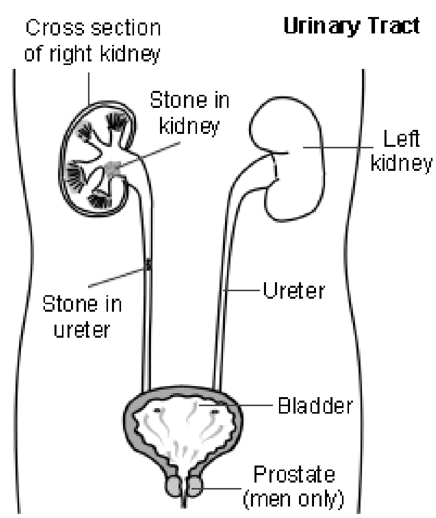
*Currently at the Edinburgh Parallel Computing Centre

## Contents



## 1. What are kidney stones?

The kidneys act as filters, concentrating waste into urine. Sometimes, chemicals deposit out. Stones are small accretions, generally of calcium (70-80%), but occasionally other materials, such as uric acid. They form in the kidney, and occasionally get stuck in the ureter. About 3 in 20 men, and 1 in 20 women will develop kidney stones, many of which are passed naturally.
Small stones (<4 mm) may go undetected.




### Diagnosis

Classic symptoms of acute renal colic:

- severe pain in groin, back, stomach.
- fever, nausea and vomiting.
- elevated blood pressure.
- urgent need to urinate; blood in urine.

Standard diagnosis techniques include:

- X-ray or intravenous pyelogram (IVP).
- Ultrasound.

# 2. Treatment: non-surgical

–Wait and see (conservative).
–Endoscopy: pass an endoscope up the ureter, with an attachment to:

- – Catch the stone in a basket.
- – Blast it with ultrasound or electrically induced shock waves.

– Chemolysis. Non-calcium stones can often be dissolved.
– Extracorporeal Shock Wave Lithotripsy (ESWL).

## ESWL

Commonest outpatient treatment. Highly focussed ultrasonic impulses from outside of the body are focussed onto the stone to shatter it. The small fragments are passed in the urine (but it may take some time for this to happen).

## Surgical

'Keyhole': Percutaneous Nephrolithotomy (PCN), uses local anaesthesia. Large incision: Open Nephrolithotomy uses general anaesthesia. Cut a chunk out: Partial Nephrectomy. Cut the whole kidney out: Complete Nephrectomy.

## Treatment issues

Many factors affect treatment:

- – Stone location, size, composition.
- – Degree of pain.
- – Presence of infection.

For instance: an upper ureteral stone is probably best treated by ESWL (the fragments will probably pass in the urine). But a kidney stone in the lower calyx may best be treated by PCN.

## Some statistics

63% of patients are male. Mean age is 52 years ($\pm$ 28). 28% of patients have more than one stone. 30% of stones are located in the ureter. Single stones: avg size 10 mm. (95% are between 3 and 30 mm.) Multiple stones: avg size 20 mm. 95% are between 5 and 80 mm.

## Summary of the medical bit

Large variety of possible treatments. Different degrees of invasiveness: from out-patient to surgery under anaesthetic. Different success rates: stones re-occur in many cases.

## The Scottish Lithotriptor Centre

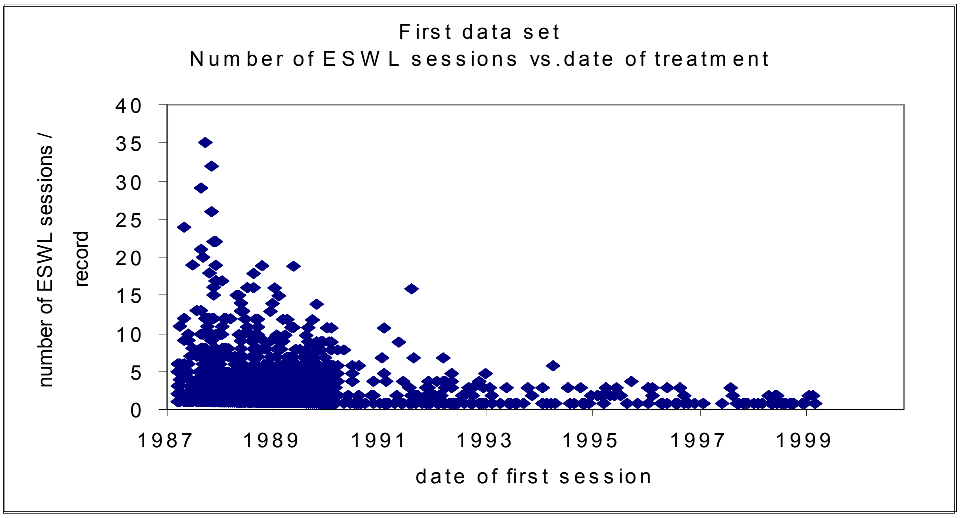
Based at the Western General in Edinburgh. Treatment of large numbers of patients, mostly by ESWL. Also records details of outcomes, follow-up. >10,000 patient records. Some on paper, but many in Excel.

# 3. Patient Records

Patient details. Stone details (position, size, single/multiple). Treatment (no. of shocks, kV used, date). Outcome (fragmented? Size of fragments). Follow-up (Stone free? Complications). Stone re-growth (date, size). Other records: x-ray, IVP.

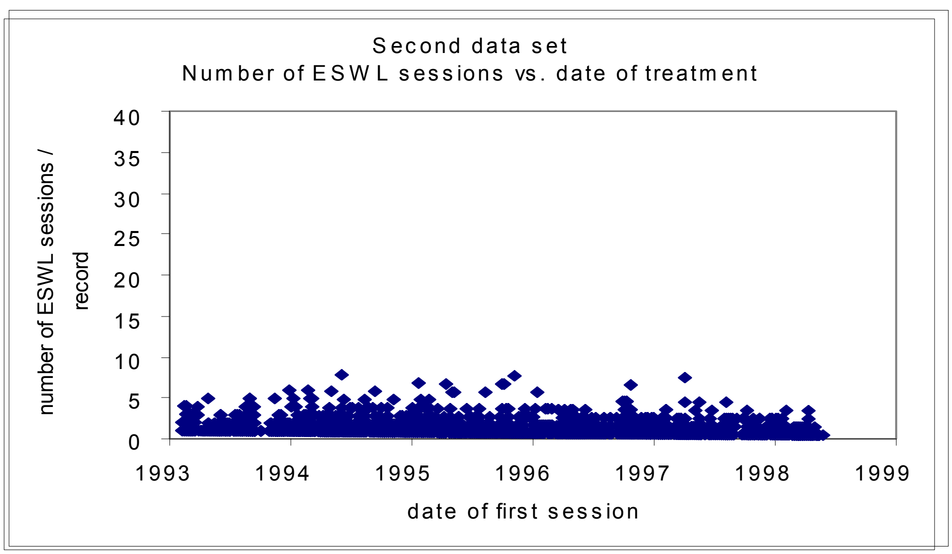
## Initial Statistical Analysis

It was known that there were problems with data: errors (e.g dates in wrong format), omissions (no outcome; when do you say somebody is stone free?). Initial study to look at distribution of age, stone size, number of treatment sessions.




### Time dependence

For ESWL, the maximum number of treatment sessions varies with the date. Before 1993, patients could have as many as 15 sessions (11% had 5 or more sessions). After 1993, they tended never to have more than two or three sessions of ESWL. (<1%) It turns out this is because the 'regime' had changed. Doctors learned that if it didn't work within a few sessions, it was better to try something else.




### Second data set

Detailed study done on a 'better' set of 2178 records from the period 1993 – 1999. Obviously incomplete records were discarded, leaving 2162 records. This still left a lot of records with special conditions (there were over 40 different entries in the 'special' field, many appearing on only one record). The data set was simplified by removing records with 'special' entries. This left 1557 records.

# 4. Predictive ability

The objective was to see if a method could be found, using statistical or neural techniques, to predict the outcome from the patient record. This could be used as an 'Advisory system' (Medical Doctors are unlikely to go for a system that told them what to do). This might take the form of a system that said:
"For these conditions, ESWL will result in a 95% chance of the patient being stone-free, PCN will give a 79% chance of being stone-free. "

## What to predict?

In principle, what we want to predict is whether the outcome is stone-free or not. A real system will make errors: predicting stone-free when that is not the outcome, and vice-versa. In practice, it may be important to correctly identify the patients that will be stone-free after ESWL, even if some are predicted not to be stone-free, who are.

## Practical Measures

- Classification accuracy: % of patients who are correctly classified.
- Sensitivity: % of stone-free patients that are predicted accurately.
- Specificity: % of not stone-free patients that are predicted accurately.

So, we might want to maximise Sensitivity at the expense of Specificity.

## Statistical Methods

With a number of input variables and a categorical output variable (stone-free/not stone-free) discriminant analysis and logistic regression can be used. Discriminant analysis doesn't work if input values have many state nominal values. In this study, the stone postion is of this form: upper calyx, lower calyx, upper ureter. So, logistic regression (LR) was used. Different sets of input variables were used to get different Logistic Regression networks: LR1, LR2..

## Logistic Regression

Extension of linear regression. Applicable for categorical output data where there are only two possible outcomes.

- e.g. stone free / not stone free.

Gives probability p of a given outcome:

$$\text{logit}(p) = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_kx_k$$

Same assumptions as linear regression i.e.

- Linear relationship between input and output variables. Independent input variables.
- Output has normal distribution and constant variance over range of input variables.

## Discussion of logistic regression

Pros:

- Standard technique, widely reported in medical literature.
- Coefficients can easily be interpreted and used by other researchers to check data.

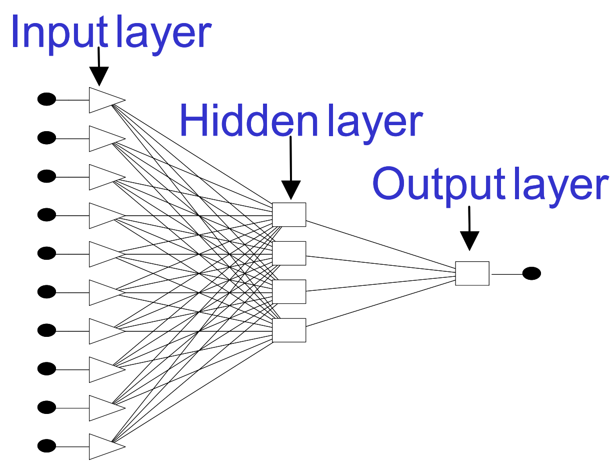
Cons:

- Medical problems are not often linear or independent.
- Underlying assumptions about distribution of data may not be valid.

Other statistical methods available e.g. discriminant analysis (if some of input data is ignored).

## 5.Neural Network

- Many highly connected simple processing units.
- E.g. Multilayer perceptron (MLP):
- Network trained by varying weights and thresholds to minimise overall error (e.g. backpropagation).




### Choice of Neural network

A large number of training examples are available, so supervised learning can be used. The application is a fairly standard "classification" type of problem. This still leaves a number of candidate networks, such as MultiLayer Perceptrons, Radial Basis Function networks, and so on. Used Trajan neural network simulator:

http://www.trajan-software.demon.co.uk/

### Intelligent Problem solver

The Trajan software has an Intelligent Problem Solver that automatically tries a number of candidate networks. From these candidate networks, a 3-layer Multi Layer Perceptron (MLP) with 7 input variables and 8 hidden nodes gave the best results. ½ of the samples were used for training, ¼ for verification and ¼ for validation.

### Neural Net discussion

- Pros:
  - Can model complex non-linear interactions.
  - No assumptions needed about data.
  - Multiple models and training algorithms exist.
- Cons:
  - Model development is empirical: no "best fit".
  - Overlearning problem requires operator skill and judgement.
  - Doesn't identify causal relationships between variables.

### Neural net implementation

- Input variables selected from:

| | |
|---|---|
| fragmentation | gender |
| number of shocks | stone location |
| max. voltage | stone size |
| number of sessions | stone side |
| stent / no stent | single / multiple stones |

Logistic regression done using training and test sets with same input variables. Initially the training was done to maximise the classification accuracy.

## 6. Results

| | LR1 | LR2 | LR3 | N NET |
|---|---|---|---|---|
| Classification Accuracy % | 54 | 53 | 55 | 75 |
| Specificity % | 49 | 46 | 52 | 74 |
| Sensitivity % | 62 | 64 | 59 | 78 |

| | NN1 | NN2 | NN3 | LR |
|---|---|---|---|---|
| Classification Accuracy % | 75 | 60 | 76 | 53 |
| Specificity % | 74 | 43 | 90 | 46 |
| Sensitivity % | 78 | 89 | 53 | 64 |

Training for sensitivity

### Discussion

The best Neural net out-performed the standard statistical techniques. The sensitivity could be increased to 89% at the expense of specificity, which dropped to 43% (the overall classification accuracy dropped slightly). The analysis of which input variables were significant produced unexpected results.

### Important Variables

A Wilks Generalised likelihood ratio test was performed. It was found that statistically significant variables were:

- Whether the stone fragmented. 1.40
- The number of shocks. 1.18
- Voltage used. 1.10
- Number of sessions. 1.09

### Conclusions

- Treatment details are important.
- Patient details are unimportant.
  - Most significant was gender 1.02

How does this compare?

| | Network 1 | Network 2 | Michaels et al | Cummings et al |
|---|---|---|---|---|
| Sensitivity % | 78 | 89 | 91 | 100 |
| Specificity % | 74 | 43 | 92 | 57 |
| Classification Accuracy % | 75 | 60 | 91 | 76 |

### Differences in studies

Previous table is a simplification: Michaels ('98) actually predicted stone-regrowth. Cummings predicted spontaneous passage of stones. Other differences: we had more records (~1500, as opposed to 96 for Michaels and 181 for Cummings). We identified some limitations in our data:

### Data limitations

Recorded stone-free rates lower than expected.

- 3-month stone-free rates should be used.

Incomplete input data.

- e.g. presence of infection not recorded.

May not be possible to model all data with single network.

## 7. Conclusions

Neural techniques have produced an improvement in performance when compared with traditional statistical techniques. Limitations appear to be mostly in the limitations of the input data (incomplete or ambiguous records). Analysis of the input data was necessary in order to use only 'sensible' data. Future data should be more complete. From the clinician's point of view, the most important finding was not the ability to predict, but the identification of which variables were important in the ability to predict. The identification of gaps and inadequacies in the recording of data should also result in improvements in the long term. For instance, the measurement of whether stones have re-grown should be done at least three months after the treatment.

### Acknowledgements

The assistance of Michael Esposito and David Tolley at the Department of Urology, Western General Hospital in Edinburgh is gratefully acknowledged.

### References

Michael EK, Niederberger CS, Golden RM, Brown B, Cha, L & Hong Y, "Use of a neural network to predict stone growth after shock wave Lithotripsy", *Urology*, 1998, **51**, 335-8.

Cummings JM, Boullier JA, Izenber SD, Kitchens DM & Kothandapani RV, "Prediction of spontaneous ureteral calculous passage by an artificial neural network". *J Urol*, 2000, 164, 326-328.