Mixture Distributions and the Construction of Reference Ranges for Biochemical parameters

Shahid Kamal
Institute of Statistics
University of the Punjab
Q.A. Campus, Lahore

Clive J. Lawrence
Department of Mathematical Statistics & Operational Research
University of Exeter, Laver Building
North Park Road, Exeter EX4 4QE, UK

Vivian F. Trewin
Department of Pharmacy
Royal Devon and Exeter Hospital (Heavitree)
Gladstone Road, Exeter EX1 2ED, UK

Aisha Sekandar
Institute of Statistics
University of the Punjab
Q.A. Campus, Lahore

Abstract

The ability to make use of accurately documented, local and specific, reference ranges is an important part of clinical work. The most readily available, often the only, routinely collected data is that obtained from hospital admissions. A mixture distribution approach is employed to arrive at reference range appropriate to particular groups of patients. A classification rule is used to provide a spectrum of such ranges that allows for the lack of complete separation of the components of the mixture, a common feature when such models are fitted to data. The results of analyses are given for a range of biochemical parameters; in some examples, these are shown separately for males and females, and for those taking, and not taking, a diuretic drug – the most commonly occurring drug-group in the elderly.

1. Introduction

In the course of clinical decision making, reference (sometimes known as 'normal', or 'standard') ranges are used to assess the significance of the values of biochemical parameters. The use of the term 'normal' is somewhat misleading, as has been pointed out by Royston [1], in that a value within the range may be interpreted to indicate that "individuals falling within such a range are *clinically* normal". Unfortunately, the basis and the nature of the information upon which such ranges are calculated are not always available to the clinician. Many range are constructed from historical and ad-hoc data of unspecified, and sometimes unknown, origin. Another point of concern to the clinician is to know whether the range specified is appropriate for the assessment of an individual with specific characteristics (for example, of a certain age and with a particular drug regimen). The degree to which drug groups or their individual members influence reference ranges, whether this is patient specific, is valuable information to the diagnosing

clinician. Reference ranges of serum electrolytes, creatinine, hemoglobin, and liver enzymes tend to shift to lower values due to the relative shrinkage of organ capacity in the elderly, and so it is often the case that any apparent low values in this age group have no serious implications.

These issues are especially relevant to elderly populations, where the vast majority have at least one clinical condition, varying degrees of change in their body-system functions, and to whom significant drug regimens are commonly administered [2]. Shifts in pathological measurements caused by drug treatment will be specific to an individual, and any excessive deviations from drug-related ranges would alert physicians to search for other causes of these abnormal values. The identification of a healthy group of elderly individuals from which to collect information and then construct reference ranges can, for a variety of different reasons, be a difficult, almost impossible task. The most easily available information on the elderly is collected at hospital admission. Although this information will have a number of drawbacks, it does afford the possibility of sufficient case information to be able to construct reference ranges specific to a target group. Where information about a particular biochemical parameter is available for a group of such patients the set of measurements might, at the very least, be thought to have arisen from a mix of both 'healthy' and 'unhealthy' individuals. This paper considers one way in which this mix might be modeled in order to provide the clinician with an improved tool for his or her assessment.

A considerable body of previous work has addressed various aspects of the construction of reference ranges [1-8]. One of the earliest papers, by McPherson et al [3], demonstrated the need for a transformation of the scale of measurement for many biochemical parameters, before any reference range is constructed. Royston et al [1,6,8] in a series of papers have used low-order polynomial curves to produce age-specific ranges, have demonstrated the theoretical and practical advantages of estimating symmetric percentiles of an assumed underlying normal distribution, and have examined the various statistical properties of using the three-parameter log-normal distribution in the calculation of reference ranges for clinical measurements. More recently, Wright and Royston [14] have conducted a detailed, comparative study of statistical methods used in the determination of age-related reference intervals.

There is also relevant work to be found in other medical research areas where the purpose is to establish some kind of 'standard'. In the study of human growth, research is often focused upon the estimation of reference centiles of some characteristic of interest (birthweight, birthlength, etc., see for example Lawrence et al [9], Thompson and Theron [10], Cole and Green [11]), to provide the clinician with a framework in which to assess the measurement made on a new patient.

In this paper a methodology involving mixtures of distributions is employed to construct a spectrum of reference ranges for a parameter; obtained by fitting a univariate distribution model to a reduced population. The population reduction occurs as a result of the application, successively, of a classification rule designed to identify a core of values from which to derive the reference value. In

any study of this kind a balance needs to be struck between basing the reference values upon known pharmacological effects, the use of data from a very specific group of individuals, where inference to a wider population is not really justified, and the use of a more heterogeneous set of data, such as that from hospital admissions.

2. Method

The population P is assumed to comprise of a mixture of k sub-populations $\{P_1..P_k\}$. The proportion of each sub-population may or may not be known. Let $c_1,..,c_k$ denote these proportions, where $\sum_1^k c_i = 1$ and $c_i \geq 0$ for $i = 1,..,k$. The probability distribution function of a measurement Y in P can be represented in the mixture form $F\left(y;\underset{-}{\Theta}\right) = \sum_{i=1}^k c_i F_i\left(y;\underset{-i}{\theta}\right)$, where $F_i\left(y;\underset{-i}{\theta}\right)$ and $f_i\left(y;\underset{-i}{\theta}\right)$ denote, respectively. The distribution function and probability density function of Y in P_i, and, $\underset{-}{\Theta} \equiv \left(\underset{-}{c};\underset{-}{\theta}..;\underset{-k}{\theta}\right)$ denotes the set of unknown parameters to be estimated.

The possible presence of mis-recorded or other unusually high or low values, quite likely for the examples considered in this paper, makes it sensible to assume measurement Y to be 'censored' to the interval (L,U). Suppose that in a sample of N measurements, n_1 are observed in $(L=B_0 \leq Y < B_1),..,n_M$ are observed in $(B_{M-1} \leq Y < B_M = U)$, where $N = \sum_{m=1}^M n_m$. Then Prob$\{Y \leq y\}$ = D(y,L)/D(U,L), where $D(a,b) = \sum_{k=1}^K c_k\left[F_k(a;\underset{-k}{\theta}) - F_k(b;\underset{-k}{\theta})\right]$, and Prob$\{B_{m-1} \leq Y < B_m\}$=D($B_m,B_{m-1}$)/D(U, L)=$G_m$,say. The log-likelihood function is given by ln L=constant $+\sum_{m=1}^M n_m\left[\ln D(B_m,B_{m-1}) - \ln D(U,L)\right]$. This function can be maximised, using standard numerical optimisation techniques, to obtain the ML estimator $\hat{\underset{-}{\Theta}}$ of $\underset{-}{\Theta}$. Estimates of the variances (and covariances) of the estimator $\hat{\underset{-}{\Theta}}$ can be obtained from the matrix $\left[E\left\{\frac{\partial \ln L}{\partial \Theta_r} \cdot \frac{\partial \ln L}{\partial \Theta_s}\right\}\right]^{-1}$, where since $E\left\{\frac{\partial \ln L}{\partial \Theta_r}\right\} = 0$ for all r, it can be shown that

$$E\left\{\frac{\partial \ln L}{\partial \Theta_r} \cdot \frac{\partial \ln L}{\partial \Theta_s}\right\} = N \sum_{m=1}^M G_m \frac{\partial G_m}{\partial \Theta_r} \frac{\partial G_m}{\partial \Theta_s}.$$

These estimated characteristics can be used to partition population P into two groups $P_A(p)$ and $P_R(p)$, using the classification rule:

Assign population element 'y' to $P_A(p)$ if the estimated conditional probability

$$\text{Prob}\left\{\text{element } \varepsilon \ P_j | Y = y\right\} = \frac{\hat{f}_j(y)\hat{c}_j}{\sum\limits_{k=1}^{K} \hat{f}_k(y)\hat{c}_k} \text{ is } \geq p \text{ for some threshold value } 0 \leq p \leq 1;$$

otherwise, assign to $P_R(p)$. In this paper sub-population P_j was chosen to be the largest estimated component of the mixture distribution. In all our analyses, this has proved pharmacologically satisfactory. At the lower end of the scale 'p' $P_A(0)$ will comprise the original population, and as 'p' increases the rule successively restricts $P_A(p)$. The nature of the subsequent change in reference range for successive 'p' provides the clinician with an additional information in the assessment of a new patient value.

This classification rule is based upon the optimum decision rule for a Bayesian classifier with rejection, first introduced by Chow[12,13], that assigns each population element to that class with maximum aposteriori probability. To allow the acceptance into $P_A(p)$ of population elements 'closer-to' $P_R(p)$, $k \neq j$, Chow's procedure needs to be modified to the form given above. The reason for this being that in general verification systems, of which this is an example, the classes (or sub-populations $P_1..P_k$) may be quite close together. When this occurs, estimation of certain of the parameters in the mixture model will be, with even quite large sample size N, subject to poor precision. This could give rise to a considerable amount of anomalous allocation if the criterion of the maximum a posterior probability is used.

Having established, for a given value of p, the sub-population $P_A(p)$, a single-component model $F_A(x; \underset{-}{\theta})$ is then fitted and the reference range $R_{100(1-\alpha)\%}(p)$ estimated for specified significance level $100(1-\alpha)\%$. These calculations produce a spectrum of reference values for the different values of p. At p=0, $P_A(0)$ comprises all the population elements and is identical to P. As p increases, elements not "close to" P_j are removed. A plot of $R_{100(1-\alpha)\%}(p)$ against p provides both a summary of the particular measurement under study and a visual aid to the clinical interpretation of the significance of a measurement for any new patient.

3. Results

This method of reference range construction is illustrated using hospital admissions data from an Exeter, UK Study [5,7]. In each case the mixture model components $\{Fi\left(y; \underset{-i}{\theta}\right)\}$ were chosen to be Normal $N\left(\mu_i, \sigma_i^2\right)$ on a logarithmic scale, with k = 3. The choice of three sub-populations was found to be satisfactory, or more than so (k = 2 being sufficient), for all the biochemical parameters considered. Maximum Likelihood estimates of the model parameters for k=3 and for sub-populations $P_A(0.75)$ are given in Table 1 together with the estimated

95% reference ranges $\left(\hat{R}_{L,95\%}, \hat{R}_{U,95\%}\right)$ derived from them, and locally used ranges $\left(L_{local}, U_{local}\right)$.

An important objective of reference range determination is to try to produce information for the clinician as relevant as possible to a particular patient group. To illustrate this process summary results from a selection of such analyses are given in Table 2 and the accompanying Figures 1 & 2. Table 2(a) and Figure 1 show the results of analyses applied to a reduced starting population for each of bilirubin, blood-sugar, and serum sodium. In each case, the population was reduced by removing any patient whose drug regimen or disease condition may have affected that particular biochemical parameter. Analyses were carried out for thyroxine and urea to determine separate ranges for males and females, and for creatinine a breakdown into those administered a diuretic drug and those not; the details are given in Table 2(b) and Figure 2.

The results highlight a number of important issues in the construction of reference values. Requiring the estimated conditional probability of belonging to the 'target' sub-population to be greater than p=0.75 gives rise to an approximate dichotomy of the set of biochemical parameters. Those which use 70+% of the original population elements (namely, bilirubin, gamma glutamyl transpeptidase, aspartate aminotransferase, alkaline phosphatase, thyroxine, Red blood cells (count), and mean cell volume), and those which use approximately 40-50% (blood-sugar, haemoglobin, White blood cells (count), platelets, sodium, potassium, urea, and creatinine) to establish the reference range. These general reductions are what might be expected with the use of hospital admission data, where the methodology is trying to remove the effects of disease, etc. Another important factor is that the data are from an elderly population (≥ 65 years of age), and will reflect the complicated changes that take place in body-system performance as age increases. Locally-used ranges are often determined from healthy groups of much younger people. For example in Fig 2(b) it is seen that for both males and females, despite some convergence to the locally-used range as we increase 'p' and restrict the population on which the range calculation is based, the estimated range remains significantly wider, particularly at the upper limit. In the case of creatinine, Fig 2(c), both the diuretic and non-diuretic groups produce ranges with higher mean values for values of 'p' up to 0.85.

Examination of Table 1 indicates that the biochemical parameters divide into four broad groups: those where the estimated range falls within the locally-used one (gamma glutamyl transpeptidase, thyroxine, White blood cells, platelets, and potassium); those with a significant shift to a lower mean value (asparatate aminotransferase, haemoglobin, and Red blood cells); those with a higher mean value (alkaline phosphatase, mean cell volume, urea, and creatinine); and, bilirubin, blood-sugar, and sodium, where the calculated interval covers the local range. The reduced population analyses, carried out for this final group, maintain the same picture, except in the case of sodium, where the values are very close to hose of the local range, at p=0.75; see also Figure 1. The four main groups mirror very well expected changes with ageing, with a less active bone marrow

resulting in lower values of haemoglobin and Red blood cells, and declining renal function resulting in higher values of creatinine and to some degree urea, with osteoporotic changes accounting for higher values of alkaline phosphatase.

Although the analyses, in Table 2(b), for urea and creatinine, stratified by sex and diuretic-group respectively, indicate certain changes from those reported I Table 1, the ranges are still shifted to higher mean values compared to those of the local ranges. The case of thyroxine is more complicated, since the female group is seen, in Figure 2(a), to 'validate' the local range, whilst in the male group many more patient values are rejected, and an estimated range produced which is in agreement at the lower end but with a very significantly reduced upper boundary.

4. Conclusion

We have demonstrated how reference ranges can be determined from observational data that is inevitably going to consist of measurements from a collection of sub-populations in which the individuals are subject to a variety of different influences. For example, factors such as sex, age, and major drug-class regimen must be taken into account in the assessment of unusual or significant clinical measurements. At the same time the methodology needs to be flexible enough to be able to commend itself to simple clinical use, otherwise it will remain a somewhat academic exercise. The use of reference range plots such as we have illustrated enables the clinician to make a more informed interpretation of biochemical results that are presented. The ability to be able to focus attention onto groups of individuals with particular characteristics is important. This requires the availability of large, regularly updated databases, such as hospital-based sets of the kind used in this paper. The obvious disadvantage in using hospital admission data is that the patient set is, by definition, an ill population; although, not uniformly so across all biochemical parameters. However this drawback must be balanced against the need to have sufficient amount of data available, to properly focus the relevance of any reference range, in the way that the methodology in this paper has been able to do.

References

1. Albert, A. and Harris, E. K. (1987). Multivariate Interpretation of clinical Laboratory Data, Marcel Dekker Inc, New York.
2. Chow, C.K. (1957). 'An optimum character recognition system using decision functions', IRE Trans. Electronic Computers, 6, 247-254.
3. Chow, C.K. (1970). 'On optimum recognition error and reject tradeoff', IEEE Trans. Information Theory, 16 (1), 41-46.
4. Cole, T.J. and Green, P.J. (1992). 'Smoothing reference centile curves: the LMS method and penalised likelihood', Statistics in Medicine, 11, 1305-19.
5. Lawrence, C., Fryer, J. and Karlberg, P. (1989). 'Modelling of Reference Values for Size at Birth', Acta Paediatrica Scandinavica, 350, 55-69.

6. Lawrence, C.J. and Trewin, F.V. (1990). "Statistical Methods in the Construction of Pharmacological Reference Ranges and their use in the Identification of Possible Adverse Drug Reactions'. Statistica Applicata, 2(4) 333-349.

7. Lawrence, C.J. and Trewin, F.V. (1991). "the Construction of Biochemical Reference Ranges and the Identification of Possible Adverse Drug Reactions in the Elderly', Statistics in Medicine, 10, 831-837.

8. Leask, R., Andrews, G. and Caird, F. (1973). 'Normal values for sixteen blood constituents in the elderly', Age and Ageing 2, 14-23.

9. Mcpherson, K., Healy, M., Flynn, F., Piper, K. and Garcia-Webb, P. (1978). 'The effect of age, sex, and other factors on blood chemistry in health', Clinical Chimica Acta, 84, 373-397.

10. Royston, P. (1991). 'Constructing time-specific reference ranges', Statistics in Medicine, 10, 675-690.

11. Royston, P. (1992). "Estimation, Reference Ranges and Goodness-of-fit for the three-parameter log-normal distribution', Statistics in Medicine, 11, 897-912.

12. Royston, P. and Matthews, J. N. S. (1991). "Estimation of Reference Ranges from Normal Samles', Statistics in Medicine, 10, 691-695.

13. Thompson, M.L. and Theron G.B. (1990). 'Maximum Likelihood Estimation of Reference Centiles', Statistics in Medicine, 9, 539-548.

14. Wright, E.M. and Royston, P. (1997). 'A Comparison of Statistical Methods for Age-related Reference Intervals', Journal of the Royal Statistical Society A, 160, 47-69.

Table 1: Estimated reference ranges $\left(\hat{R}_{L,95\%}, \hat{R}_{U,95\%}\right)$, calculated using $N(\hat{\mu}_*, \hat{\sigma}_*)$, the fitted distribution for sub-population $P_A(p = 0.75)$, and locally used ranges (L_{local}, U_{local}). The population proportion $P_A(p = 0.75) / P_A(p = 0)$ is shown.

Body System-	biochemical parameter	Population proportion $P_A(p = 0.75) / P_A(p = 0)$	$\hat{R}_{L,95\%}, \hat{R}_{U,95\%}$	L_{local}, U_{local}	$\hat{\mu}_*$	$\hat{\sigma}_*$
Hepatic -	bilirubin (μmol/litre)	3202/3347 ~ 96%	2.9, 32.3	5, 17	2.28	0.611
	gamma glutamyl transpeptidase (iu/litre)	2526/3225 ~ 78%	5.6, 87.7	0, 100	3.10	0.703
	aspartate aminotransferase (iu/litre)	2409/3317 ~ 73%	11.7, 39.6	20, 57	3.08	0.302
	alkaline phosphatase (mmol/litre)	2480/3316 ~ 75%	56.1, 166.9	25, 130	4.57	0.278
Blood -	blood sugar (mmol/litre)	1608/3105 ~ 52%	3.1, 8.7	4.2, 6.6	1.64	0.267
	thyroxine (nmol/litre)	1898/2245 ~ 85%	65.8, 157.4	58, 170	4.62	0.223
	haemoglobin (g/100ml)	1844/3612 ~ 51%	12.2, 16.6	13, 18	2.66	0.075
	Red blood cells (x10^{12}/litre)	2899/3552 ~ 82%	3.6, 5.7	4.5, 6.5	1.51	0.117
	White blood cells (x10^9/litre)	2183/3606 ~ 61%	4.7, 10.6	4, 11	1.95	0.205
	mean cell volume (fl)	2973/3526 ~ 84%	80.5, 102.4	76, 96	4.51	0.061
	platelets (x10^9/litre)	1838/3210 ~ 57%	211, 382	150, 400	5.65	0.151
	sodium: (mmol/litre)	1985/3604 ~ 55%	122.5, 150.8	132, 144	4.91	0.053
	potassium: (mmol/litre)	1543/3594 ~ 43%	3.4, 5.0	3.5, 5.5	1.30	0.150
Urogenital -	urea: (mmol/litre)	2219/3331 ~ 67%	3.0, 11.6	2.1, 6.0	1.78	0.342
	creatinine (μmol/litre)	1698/3446 ~ 49%	76.9, 130.9	45, 120	4.61	0.136

Table 2a: Maximum Likelihood Estimates for the single-component (*with standard errors*) and three-component models, in each case natural log scales have been used.

Biochemical parameter	$\hat{c}_1$	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{c}_2$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{c}_3$	$\hat{\mu}_3$	$\hat{\sigma}_3$	Population proportion $P_A(p=0.75)/P_A(p=0)$	$\hat{\mu}_*$	$\hat{\sigma}_*$
Bilirubin (μmol/litre)	0.86	2.13	0.544	0.13	3.33	2.300	0.01	2.14	0.054	3094/3258 ~ 95%	2.22	0.506
											(0.0093)	(0.0071)
										=> Reference range	[3.4 - 24.8]	
Blood sugar (mmol/litre)	0.91	1.82	0.217	0.06	2.32	0.088	0.04	1.61	0.039	1578/1830 ~ 86%	1.81	0.232
											(0.0059)	(0.0045)
										=> Reference range	[3.9- 9.6]	
Sodium (mmol/litre)	0.71	4.93	0.023	0.24	4.90	0.059	0.05	4.87	0.002	1358/1804 ~ 75%	4.93	0.0189
											(0.0005)	(0.0004)
										=> Reference range	[133.7- 144.0]	

Shahid Kamal, Clive J. Lawrence, Vivian F. Trewin, Aisha Sekandar

Table 2b: Maximum Likelihood Estimates for the single-component (with estimated standard errors) and three-component models; in each case natural log scales have been used.

| Biochemical parameter	$\hat{c}_1$	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{c}_2$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{c}_3$	$\hat{\mu}_3$	$\hat{\sigma}_3$	Population proportion PA(p=0.75)/PA(p=0)	$\hat{\mu}_*$	$\hat{\sigma}_*$
Urea (mmol/litre) -												
male:	0.65	1.89	0.362	0.28	2.43	0.368	0.08	3.24	0.010	723/1389 ~ 52%	1.72	0.329
											(0.0275)	(0.0075)
										=> Reference range	[2.93 -	10.63]
female:	0.85	1.91	0.413	0.14	2.85	0.405	0.01	3.06	0.130	1579/1946 ~ 81%	1.86	0.392
											(0.010)	(0.0075)
										=> Reference range	[2.97 -	13.82]
Thyroxine (nmol/litre) -												
male:	0.54	4.61	0.174	0.39	4.35	0.220	0.07	3.82	0.181	311/859 ~ 36%	4.74	0.113
											(0.0092)	(0.0052)
										=> Reference range	[91.5 -	142.2]
female:	0.92	4.62	0.243	0.07	4.01	0.150	0.01	4.72	0.000	1265/1378 ~ 92%	4.64	0.275
											(0.0078)	(0.0060)
										=> Reference range	[58.0 -	177.8]
Creatinine (µmol/litre) -												
√ diuretic:	0.65	4.65	0.296	0.34	5.11	0.440	0.01	4.61	0.010	824/1577 ~ 52%	4.50	0.259
											(0.0092)	(0.0071)
										=> Reference range	[53.9 -	148.8]
x no diuretic:	0.78	4.52	0.246	0.14	5.22	1.986	0.09	4.96	0.180	1458/1873 ~ 78%	4.48	0.248
											(0.0065)	(0.0047)
										=> Reference range	[54.3 -	143.3]

(a)

(b)

(c)
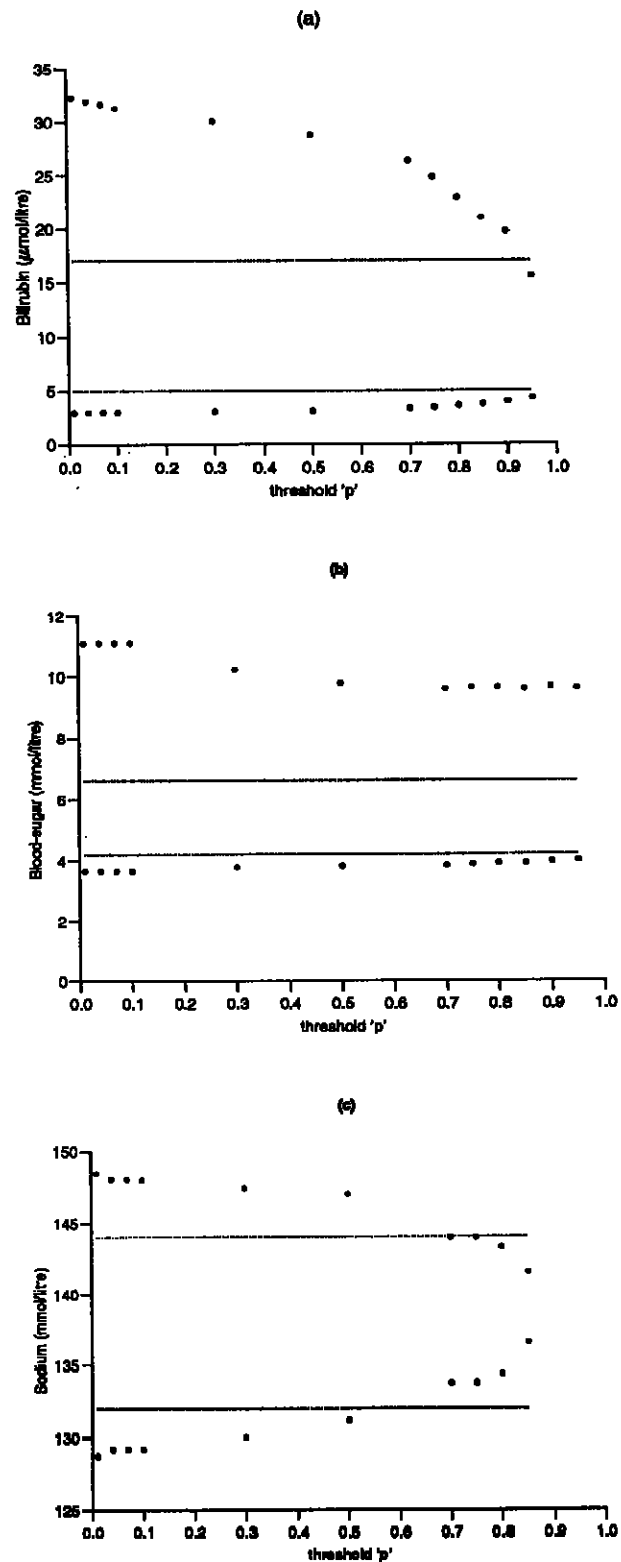
Figure 1: A graph of $\left\{\hat{R}_{L,95\%}(p), \hat{R}_{U,95\%}(p); p\right\}$ for (a) bilirubin, (b) blood sugar, and (c) sodium. In each case the straight-line denote, respectively, the lower and upper locally-used reference bounds.
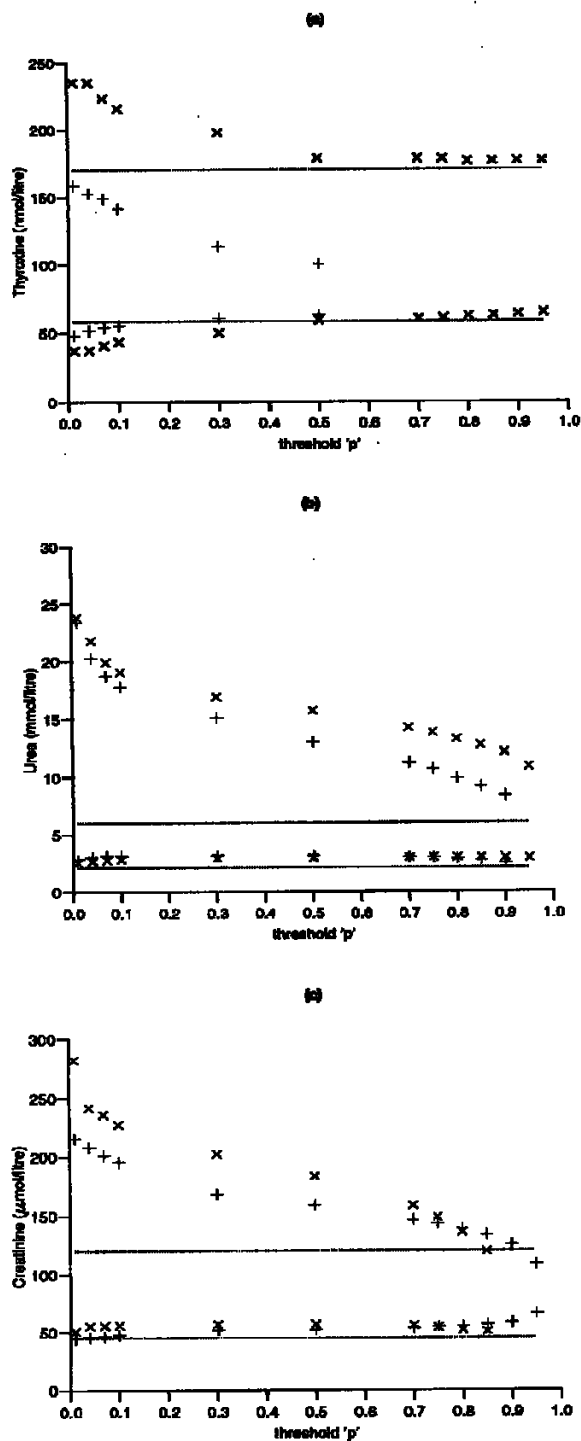
Figure 2: A graph of $\left\{\hat{R}_{L,95\%}(p), \hat{R}_{U,95\%}(p); p\right\}$ for (a) thyroxine and (b) urea, with separate plots shown for males [+] and females [x], and (c) creatinine, stratified by whether or not (+, x) the patient was receiving a diuretic drug. In each case the straight-line plots denote, respectively, the lower and upper locally-used reference bounds.