Fulminant hepatic failure secondary to acetaminophen poisoning: A systematic review and meta-analysis of prognostic criteria determining the need for liver transplantation

Benoit Bailey, MD, MSc, FRCPC; Devendra K. Amre, MBBS, PhD; Pierre Gaudreault, MD, FRCPC

Objectives: **To summarize and compare different prognostic criteria used to determine need for liver transplantation in patients with fulminant hepatic failure secondary to acetaminophen poisoning.**

Data Sources: **Studies published in the literature that investigated criteria for hepatic transplantation secondary to acetaminophen-induced liver failure as identified by a preestablished MEDLINE strategy (1966 through October 2001).**

Study Selection: **Studies were included if 2 × 2 tables could be reconstructed and if they did not assume that patients undergoing transplantation would have eventually died had they not received the transplant.**

Data Extraction: **Relevant articles were reviewed by two authors independently. Discrepancies or disagreements, if any, on the inclusion or exclusion of studies were resolved by consulting the third author.**

Data Synthesis: **King's criteria (pH < 7.30 or prothrombin time of >100 secs plus creatinine of >300 μmol/L plus encephalopathy grade of ≥3) were evaluated in nine studies, pH < 7.30 in four, prothrombin time of >100 secs in three, prothrombin time of >100 secs plus creatinine of >300 μmol/L plus encephalopathy grade of ≥3 in three, creatinine of >300 μmol/L in two, and one each for increase in prothrombin time day 4, factor V of <10%, Acute Physiology and Chronic Health Evaluation (APACHE) II score of >15, and Gc-globulin of <100 mg/L. King's criteria were more sensitive than pH: 69% (95% confidence interval, 63–75) vs. 57% (95% confidence interval, 44–68). Their specificities were, however, comparable: 92% (95% confidence interval, 81–97) vs. 89% (95% confidence interval, 62–97). APACHE II score of >15 had the highest positive likelihood ratio (16.4) and the lowest negative likelihood ratio (0.19) but was evaluated in only one study. The accuracy measures of all other criteria were lower than that of King's criteria or pH < 7.30.**

Conclusions: **Presently, available criteria are not very sensitive and may miss patients requiring transplantation. Future studies should further evaluate the efficacy of the APACHE II criteria. (Crit Care Med 2003; 31:299–305)**

Key Words: acetaminophen; liver failure; liver transplantation; prognosis; sensitivity; specificity; meta-analysis

After acetaminophen poisoning, some patients will develop fulminant hepatic failure. Although many patients will recover fully with appropriate medical management, in some cases, mortality will be high unless liver transplantation is performed. Presently, it is not well known which patients will benefit from the latter procedure. Moreover, as liver transplantation is expensive, requires immunosuppression for life, depends largely on the availability of donors, and its success is not perfect, it is important that criteria be developed to determine its appropriate implementation.

Over the last 15 yrs, a number of criteria have been developed to identify those patients with fulminant hepatic failure secondary to acetaminophen poisoning who are most likely to benefit from a liver transplant. The mostly widely used criteria are the King's criteria (arterial pH < 7.30 after adequate fluid resuscitation or the combination of prothrombin time of >100 secs and creatinine of >3.3 mg/dL (300 μmol/L) and grade III or IV encephalopathy) (1). However, the use of other criteria such as serial prothrombin time (2–5), coagulation factor V and VIII/V ratio (3), factor VIII/V ratio of >30, or a factor V of <10% (4, 6), serum Gc protein (7, 8), and Acute Physiology and Chronic Health Evaluation (APACHE) II score (8) has been proposed. Although promising, reported measures of accuracy for these criteria have varied among studies. The purpose of the present meta-analysis was to systematically review studies investigating these criteria to summarize information across studies, to compare the efficacies of the different criteria, and to identify those that are most likely to benefit the patient and that could be implemented for practical use in a normal clinical setting.

METHODS

Retrieving the Literature. A MEDLINE search for the years 1966 to October 2001 was done with the help of Ovid by using the strategy, exp acetaminophen and exp liver failure or exp liver transplantation or exp mortality or exp prognosis or exp sensitivity and specificity or exp ROC curve or exp meta-analysis or accuracy.mp or criteria.mp or systematic review.mp. The abstracts of these articles were reviewed to identify potentially relevant articles that investigated criteria for hepatic transplantation secondary to acetaminophen-induced liver failure. Articles so identified were then fully reviewed. Case reports and animal studies were excluded. The bibliography of the relevant articles were further cross-checked to search for articles not referenced

From the Division of Emergency Medicine (BB), the Division of Clinical Pharmacology and Toxicology, Department of Pediatrics (BB, PG), and the Research Centre (BB, DKA), Hôpital Ste-Justine, Université de Montréal, Montréal, Québec.

Address requests for and reprints to: Benoit Bailey, MD, MSc, FRCPC, Hôpital Ste-Justine, 3175 Chemin Côte Ste-Catherine, Montréal, Quebec H3T 1C5, Canada. E-mail: baileyb@med.umontreal.ca

DOI: 10.1097/01.CCM.0000034674.51554.4C

Table 1. Studies excluded from the present meta-analysis

Study	Setting	Criteria Studied	Reason for Exclusion	Type of Study	Patients, n
Harrison et al., 1990 (2)	London	Increase PT on day 4; peak PT at ≥ 180 secs; increase PT on day 4 + peak PT at ≥ 180 secs; increase PT on day 4 or peak PT at ≥ 180 secs	Transplant included as died	Retro	150
Pereira et al., 1992 (3)	London	King's[a]; pH < 7.30; increase PT on day 4 or peak PT at > 180 secs; factor V ≤ 10%; factor VIII/V > 30; factor V ≤ 10% + encephalopathy ≥ 3	Transplant included as died	Retro	22
Mutimer et al., 1994 (22)	Birmingham, UK	pH < 7.30; PT > 100 secs + creatinine > 300 μmol/L + encephalopathy grade ≥ 3	Transplant included as died (?)	Retro	92
Lee et al., 1995 (7)	London	King's; Gc-globulin < 34 μg/mL	2 × 2 tables impossible	Retro	47
Bradberry et al., 1995 (6)	Birmingham, UK	Factor V ≤ 10%; factor VIII/V ≥ 30	2 × 2 tables impossible	Retro	22
Anand et al., 1997 (5)	Birmingham, UK	King's	2 × 2 table impossible	Retro	120

PT, prothrombin time; Retro, retrospective.
[a]King's criteria are pH < 7.30 or a combination of PT > 100 secs + creatinine > 300 μmol/L + encephalopathy grade ≥ 3.

in MEDLINE. In addition, a number of medical toxicology and hepatology textbooks were also reviewed (9–19).

Selection of Studies, Inclusion and Exclusion Criteria, and Data Extraction. All relevant articles were reviewed by two authors independently. Discrepancies or disagreements, if any, on the inclusion or exclusion of studies were resolved by consulting the third author. Studies were included in the meta-analysis only if it was possible to extract the data and reconstitute them in the form of 2 × 2 tables. Some studies assumed that patients undergoing transplantation would have eventually died had they not received the transplant and included them in this category (mortality positive). These studies were excluded from the analysis. The small number of relevant studies precluded the exclusion or inclusion of studies on the basis of methodologic quality. Whenever possible, the raw data were used to construct the 2 × 2 tables. When raw data were unavailable, the tables were constructed using given measures of sensitivity and specificity. The 2 × 2 tables represented patients who either died or survived when they met or did not meet the criteria under study.

Quality Assessment. We evaluated the methodologic quality of the included studies by applying the criteria for assessing design-related bias recently described by Lijmer et al. (1999) (20). In most cases, these biases tend to lead to an inflation of the accuracy of the test or criteria under study. We examined each study for the following potential biases: 1) spectrum bias, which could occur when diagnostic accuracy is examined by comparing test results among patients known to have disease and among a group of normal patients (case control study) as opposed to a clinical population covering the spectrum of disease; 2) verification bias, which could exist when the decision to perform the reference test is based on the result of the test under examination (20); 3) selection bias, which could occur when patients are not recruited consecutively and randomly; and 4) investigator bias, which could occur if investigators are not blinded to the results of the study and reference test. In addition to potential limitations in the study design, we examined each study with regard to the methods utilized for data collection and reporting. Insufficient details could hinder interpretation, replication, validation, and generalization of findings. With regard to the present review, we evaluated whether there was adequate description on the nature of data collection (prospective or retrospective), clear definitions for cutoff values for the test criteria, information on treatment with N-acetylcysteine, and information on delay between ingestion and admission or ingestion and treatment (population description). Each of the latter was examined separately, and studies not providing the respective details were classified as not sufficient.

Meta-analysis. The meta-analysis approach that uses linear regression techniques to combine data from independent studies evaluating similar diagnostic tests/criteria and as described by Moses and Shapiro (21) was used. To create the summary receiver operating characteristic (SROC) curve, we first calculated the true-positive rate (TPR) and false-positive rate (FPR) from each individual study from the reconstructed 2 × 2 tables. They were then converted to their logistic transform (log [TPR/1-TPR] and log [FPR/1-FPR]). The sum (S) and the difference (D) of these logistic transforms were calculated for each study and a regression line fitted to these points, with D as the dependent variable and S as the independent variable (D = a + bS). If the co-efficient was not significant, the values of sensitivity and specificity for constructing the SROC curve were then calculated as follows: sensitivity = 1/(1 + 1/DOR × (1 − specificity/specificity)). If the co-efficient was significant, the values of sensitivity and specificity required to construct the SROC curve were then calculated as: sensitivity = $1/(1 + 1/e^{a/(1-b)} \times (1 - \text{specificity}/\text{specificity})^{(1+b)/(1-b)}$. For both the above-mentioned formulae, one needs to provide values of specificity to obtain corresponding values for sensitivity. We considered values of specificity in the range from 0.5 to 1.0. The resulting values were than plotted in SROC space to obtain the SROC curve. We took into account the differences in sample sizes among the studies by weighting each observation by the reciprocal of the variance of D and performing weighted regression. As these weights are related to the size of D, unweighted regression was also done. To further compare the accuracy between the King's criteria and pH < 7.30, we calculated the Q value from the SROC curves obtained for each of these criteria. This value represents the intersection point of the SROC curve with a diagonal from the left upper corner to the right lower corner of the ROC space and is a global measure of the sensitivity and the specificity of individual criteria. Sensitivity and specificity are equal on this diagonal line. A higher Q value determines a higher sensitivity and specificity, and thus, a test with a higher value will represent the more accurate test. To enable comparison with other criteria besides King's and pH < 7.30, we also calculated the positive and negative likelihood ratios for these latter criteria from their respective SROC equations.

Studies evaluating other criteria besides King's and pH < 7.30 were few. It was not possible to describe them adequately by SROC curves. The FPRs and TPRs from the individual studies were then simply pooled to obtain summary estimates, and likelihood ratios were calculated from these measures.

RESULTS

From the initial search of the MEDLINE database (1966 to October 2001), a total of 502 articles were retrieved. On

Table 2. Raw data (2×2 tables) and accuracy estimates for each individual study included in the present analysis

Study and Setting	Criteria Studied[a]	Inclusion Criteria	Mortality Criteria −	Mortality Criteria +	Sensitivity	Specificity	+ LR	− LR
O'Grady et al., 1989 (1) London	King's; pH < 7.30; PT > 100 secs; creatinine > 300 μmol/L; PT > 100 secs + creatinine > 300 μmol/L + encephalopathy grade ≥ 3	PT > 32 secs	31/37	12/84	0.72	0.92	9	0.3
			21/22	22/99	0.49	0.99	49	0.52
			34/60	9/61	0.79	0.67	2.4	0.31
			30/54	13/67	0.70	0.69	2.3	0.43
			10/15	12/90	0.45	0.94	7.5	0.58
O'Grady et al., 1991 (24) London	King's	FHF	19/23	7/37	0.73	0.88	6.1	0.31
Donalson et al., 1993 (23) Toronto	King's	FHF as Trey/Davidson (32)	4/4	0/10	1	1	—	—
Schiødt et al., 1996 (8) Copenhagen	King's; Gc-globulin < 100 mg/L	FHF as Trey/Davidson (32)	6/10	5/8	0.54	0.43	0.95	1.07
			3/3	7/15	0.3	1	—	0.7
Izumi et al., 1996 (4) London	King's; pH < 7.30; PT > 100 secs; factor V < 10%	FHF as Trey/Davidson (32)	24/26	11/55	0.69	0.96	17.3	0.32
			18/20	17/61	0.52	0.95	10.4	0.51
			24/44	11/42	0.69	0.61	1.8	0.79
			29/51	6/30	0.83	0.52	1.7	0.33
Anand et al., 1997 (5) Birmingham, UK	pH < 7.30; PT > 100 secs; creatinine > 300 μmol/L; PT > 100 sec + creatinine > 300 μmol/L + encephalopathy grade ≥ 3; increase PT on day 4	FHF as Trey/Davidson (32)	24/31	15/41	0.62	0.79	3.0	0.48
			33/60	15/60	0.69	0.62	1.8	0.5
			30/61	15/58	0.67	0.58	1.6	0.57
			19/24	26/63	0.42	0.88	3.5	0.66
			12/19	26/63	0.41	0.9	4.1	0.66
Gow et al., 1999 (28) Australia	King's	Severe hepatotoxicity	0/7	0/11	—	0.61	—	—
Mitchell et al., 1998 (27) London	King's; APACHE II score > 15	Standardized admission criteria	12/16	4/80	0.75	0.95	15	0.26
			13/20	3/83	0.81	0.92	10.1	0.21
Bernal et al., 1998 (26) London	King's	Standardized admission criteria	51/56	28/424	0.65	0.99	0.65	0.35
Shakil et al., 2000 (25) Pittsburgh	King's; pH < 7.30; PT > 100 secs + creatinine > 300 μmol/L + encephalopathy grade ≥ 3	ALF as O'Grady et al. (33)	17/21	4/19	0.81	0.79	0.24	0.8
			9/13	1/5	0.9	0.5	1.8	0.2
			8/8	3/14	0.73	1	—	0.27

+ LR and − LR, positive and negative likelihood ratios; PT, prothrombin time; FHF, fulminant hepatic failure; APACHE, Acute Physiology and Chronic Health Evaluation; ALF, acute liver failure.
[a]King's criteria are pH < 7.30 or a combination of PT > 100 secs + creatinine > 300 μmol/L + encephalopathy grade ≥ 3.

Table 3. Assessment of the methodologic quality of studies included in the present meta-analysis according to the criteria suggested by Lijmer et al. (2001) (20)

Study	Spectrum Bias		Verification Bias		Investigator Bias		Selection Bias		Data Collection Bias		Study and Reference Test Bias		Population Description	
	Population	Case Control	Absent	Present	Blinded	Not	Consecutive	Not	Pros	Retro	Sufficient	Not	Sufficient	Not
O'Grady et al., 1989 (1)	+		+			+	+			+	+			+
O'Grady et al., 1991 (24)	+			+		+	+		+		+			+
Donalson et al., 1993 (23)	+			+		+	+			+	+			+
Schiødt et al., 1996 (8)	+		+			+	+			+	+		+	
Izumi et al., 1996 (4)	+		+			+	+			+	+			+
Anand et al., 1997 (5)	+		+			+		+[a]		+	+			+
Gow et al., 1999 (28)	+			+		+	+			+	+		+	
Mitchell et al., 1998 (27)	+		+			+	+		+		+			+
Bernal et al., 1998 (26)	+			+		+	+			+	+			+
Shakil et al., 2000 (25)	+		+			+	+			+	+			+

Pros, prospective; retro, retrospective

[a]Although entered consecutively, 21% of patients could not be analyzed because of missing data.

Table 4. Diagnostic test characteristics of criteria evaluated in more than one study

Criteria	No. of Studies Included	Pooled Sensitivity (95% CI)	Range Sensitivity	Pooled Specificity (95% CI)	Range Specificity
King's[a]	8[b]	0.69 (0.63–0.75)	0.55–1.0	0.92 (0.81–0.97)	0.43–1.0
pH < 7.30	4	0.57 (0.44–0.68)	0.49–0.90	0.89 (0.62–0.97)	0.50–0.99
PT > 100 secs	3	0.72 (0.63–0.79)	0.69–0.79	0.64 (0.57–0.70)	0.61–0.67
PT > 100 secs + creatinine > 300 μmol/L + encephalopathy grade ≥ 3	3	0.55 (0.44–0.66)	0.42–0.73	0.94 (0.90–0.98)	0.88–1.0
Creatinine > 300 μmol/L	2	0.68 (0.58–0.77)	0.67–0.70	0.64 (0.52–0.74)	0.58–0.69

CI, confidence interval; PT, prothrombin time.

[a]King's criteria are pH < 7.30 or a combination of PT > 100 secs + creatinine > 300 μmol/L + encephalopathy grade ≥ 3;

[b]one study (27) could not be included in the meta-analysis because the sensitivity could not be computed.

reviewing the abstracts, 33 articles were retained. On detailed review of these articles and after searching the bibliographies and other related information sources (textbooks), 15 articles were deemed relevant to the meta-analysis (1–8, 22–28). From among these 15 relevant articles, five were excluded: two because the 2 × 2 tables could not be reconstructed and three because the results assumed that subjects undergoing transplant would otherwise have died (Table 1). In one study that evaluated multiple criteria, the 2 × 2 tables could be reconstructed for all criteria except the King's criteria (5): the results pertaining to this criterion were also excluded from the meta-analysis. From the remaining nine studies (Table 2), data extraction was possible for all criteria. Results of each individual study included in the present analysis and their 2 × 2 tables are presented in Table 2.

As would be expected, none of the studies were free from all the potential biases and limitations described above (Table 3). Nevertheless, most of the studies were in general of good quality. All were free from spectrum bias and investigator bias, and the majority were free from selection and verification bias. Most of the studies, however, involved retrospective data collection and did not provide sufficient information on lag times between ingestion and admission or ingestion and treatment and whether any treatment was administered.

Estimates of sensitivity and specificity of the different criteria evaluated in more than one study are shown in Table 4. One study (28) was excluded from the calculation because the sensitivity could not be computed: there was no death. For criteria evaluated in only one study, the sensitivity and specificity were the following: increase in PT day 4, sensitivity = 0.41 and specificity = 0.90 (5); factor V < 10%, sensitivity = 0.83 and specificity = 0.52 (4); APACHE II score > 15, sensitivity = 0.81 and specificity = 0.92 (27); and Gc-globulin < 100 mg/L, sensitivity = 0.30 and specificity = 1.0 (8).

The SROC curve is plotted over the domain of TPR and FPR in Figure 1 for eight studies included in the meta-analysis that evaluated King's criteria. As mentioned above, the study by Gow et al. (28) was excluded from the SROC curve analysis because the sensitivity could not be computed. The SROC curve provides evidence on the individual contribution of each study to the regression analysis. The graph suggests that although six studies conform to the pattern of the curve, two studies behave differently. Study 3 (Donalson et al. (23)) is a small study that gives perfect results (100% sensitivity and specificity). Study 2 (Schiødt et al. (8)) is also small and does not provide good diagnostic information as seen by the fact that it falls on the diagonal line. These studies are outliers and probably responsible for the flatness and overdispersion of the SROC curve. We reanalyzed the regression after ex-

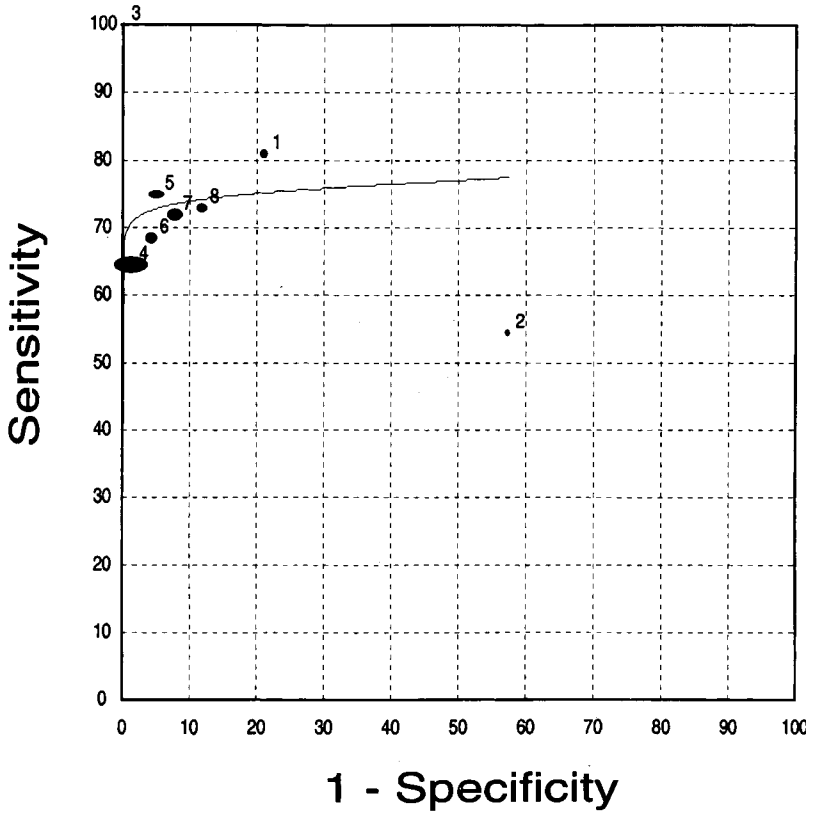

Figure 1. Preliminary summary receiver operating characteristic curve for King's criteria for the eight studies that were included in the summary receiver operating characteristic curve analysis (see text for details for the study inclusion/exclusion). *Ovals* and *circles*, studies. *Numbers*, the study's number.

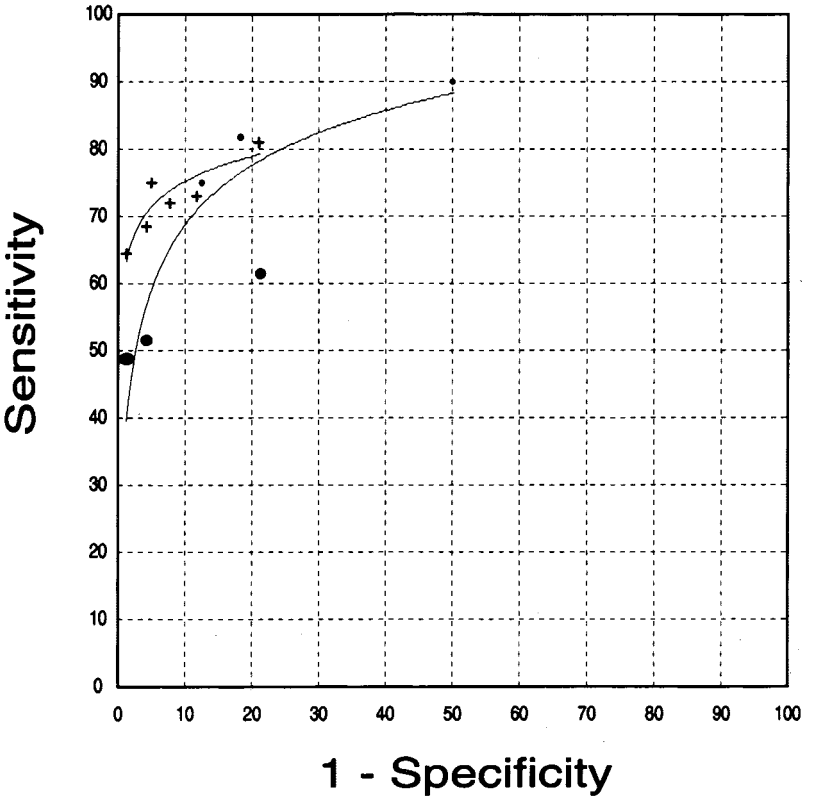

Figure 2. Summary receiver operating characteristic curves comparing King's criteria and pH < 7.30 criteria. Six studies evaluating King's criteria and four studies evaluating the pH < 7.30 criteria were included in the final analysis. +King's criteria studies; *oval* and *circles*, pH < 7.30 studies.

cluding these studies. Figure 2 shows the subsequent plot of the SROC curve obtained for these six studies. This plot is more typical of a SROC curve and is not adversely influenced by extreme values.

All studies evaluating the pH < 7.30 criteria conformed to the pattern of the SROC curve. Figure 2 shows this SROC curve and its comparison with that obtained for King's criteria. Both criteria seem to have similar diagnostic capabilities, although King's criteria seem better as they lie slightly closer to the 1,1 corner of the space. However the Q values calculated from the curves were not different (King's criteria, Q = 0.61 [95% confidence interval, 0.55–0.67] and pH < 7.30, Q = 0.61 [95% confidence interval, 0.47–0.75]), suggesting that in terms of overall accuracy, there was not much difference between these two criteria.

Likelihood ratios for all criteria studied are presented in Table 5. Apache II score of >15 at admission had the highest positive likelihood ratio and the lowest likelihood ratio but was based on only one study. The accuracy estimates for the other criteria except King's criteria and pH < 7.30 were comparable.

DISCUSSION

The King's criteria are the best known and most widely used prognostic indicators to determine the need for liver transplantation in patients with fulminant hepatic failure secondary to acetaminophen poisoning (1). They were derived from a group of patients with acetaminophen poisoning treated from 1973 to 1985 and validated in another group of patients treated between 1986 and 1987. These criteria were described as having a sensitivity of 72% and a specificity of 92%, for an overall accuracy of 85%. However, since then, other studies have evaluated the King's criteria—in some cases in the same setting, King's College Hospital liver unit (3, 4, 7, 24, 26, 27), but in other cases, elsewhere (5, 8, 23, 25, 28). Subsequently, studies have proposed the use of other prognostic criteria such as pH < 7.30 (1, 3–5, 22, 25), creatinine of >300 μmol/L (1, 5), combination of PT of >100 sec plus creatinine of >300 μmol/L plus encephalopathy grade of ≥3 (1, 5, 22, 25), serial prothrombin time (2–5), coagulation factor V and VIII/V ratio (3), factor VIII/V ratio of >30 or a factor V of <10% (4, 6), serum Gc protein (7, 8), and APACHE II score (27). The goal of the present study was thus mainly to evaluate

Table 5. Likelihood ratios of each criteria studied

Criteria	No. of Studies Included	Positive Likelihood Ratio	Negative Likelihood Ratio
King's[a,b]	6	12.33	0.29
pH < 7.30[a]	4	7.44	0.48
PT > 100 secs + creatinine > 300 μmol/L + encephalopathy grade ≥ 3	3	7.30	0.48
PT > 100 secs	3	2.05	0.40
Creatinine > 300 μmol/L	2	1.91	0.50
APACHE II score > 15	1	16.4	0.19
Increase in PT day 4	1	4.1	0.66
Factor V < 10%	1	1.73	0.33
Gc-globulin < 100 mg/L	1	Infinity	0.70

PT, prothrombin time; APACHE, Acute Physiology and chronic Health Evaluation.

[a]Likelihood ratios on pooled measures of sensitivity and specificity based on the final summary receiver operating characteristic model;

[b]King's criteria are pH < 7.30 or a combination of PT > 100 secs + creatinine > 300 μmol/L + encephalopathy grade ≥ 3.

the performance of these criteria and compare their accuracies with those of the widely used King's criteria.

Overall in our meta-analysis, King's criteria had moderate sensitivity (69%; range, 55% to 100%) and high specificity (92%; range, 43% to 100%). However, the SROC analysis showed that King's criteria ability to distinguish between subjects requiring transplantation and those who did not seems limited, with a Q value of 0.61: a Q value of 1 reflects a perfect test, and a Q value of 0.5 reflects an uninformative test. In comparison, the pH criteria had similar accuracy (similar specificity and the same Q value), although its sensitivity was slightly lower. All the other criteria had much lower sensitivities and specificities.

An alternative method to evaluate the accuracies of diagnostic criteria is to compare their likelihood ratios (29). Likelihood ratios are semiquantitative measures of the performance of a diagnostic test that indicate how much a diagnostic procedure modifies the probability of disease (29): the greater the positive likelihood ratio and the lower the negative likelihood ratio, the better the criteria are. For the present study, an APACHE score of >15 criteria had the highest positive and lowest negative likelihood ratios. However, this criteria was evaluated in only one study. When evaluated on the first day of admission, the APACHE criteria had a sensitivity of 81% and a specificity of 92%, for an accuracy of 90%. In comparison with King's criteria, the APACHE criteria specificity was similar but had a higher sensitivity (Table 4). One major disadvantage of the APACHE criteria is that it is cumbersome to use. Nevertheless, initial results suggest that it is very accurate and should be evaluated in further studies. King's criteria and pH < 7.30 had the second and third best likelihood ratios. The low accuracies of the other criteria suggest that they may not be clinically very useful.

Prognostic indicators for death in fulminant liver failure are critical to determine who should need a liver transplantation. These indicators need to be accurate. To be accurate, they need to be highly specific to avoid performing liver transplantation in patients who would have otherwise survived without it (avoid false positives) and to be sensitive enough so as to not miss several patients who would survive from the procedure (avoid false negatives). An important characteristic of the prognostic indicator should be that it could be implemented early enough in the course of the disease such that suitable donors can be recruited before irreversible multiple organ failure precludes liver transplantation. Our results indicate that other criteria with higher sensitivity and specificity comparable with the King's criteria need to be developed. The APACHE score of >15 may adequately fulfill the above requirements and should be evaluated further.

We used meta-analytic methods to summarize results across different studies. To obtain valid results, appropriate inclusion and exclusion criteria were used. We excluded studies for which it was impossible to reconstruct the 2 × 2 tables (Table 1). We also excluded studies that considered patients who underwent liver transplantation as patients who would have otherwise died without it. The inclusion of these patients would have falsely increased the accuracy of the studied criteria. As the number of studies were limited, we could not statistically assess the contribution of interstudy variation in population characteristics, study design, methodology, and analysis to the variation in the accuracy measures observed. With regard to differences in population characteristics (population heterogeneity), most studies were done in one center among reasonably similar populations, and hence, resulting differences if any were likely to be minimal.

Furthermore, the evaluation of methodologic quality showed that most studies were largely free from bias, and there were no major differences within studies. Nevertheless, certain important features were of note. As mortality was the outcome evaluated in all the studies, factors that influence mortality could influence the accuracy of the criteria used. For example, the lag time between acetaminophen ingestion and presentation could be important. It is well known that early presentation within 24 hrs is an important determinant of survival (30). This variable was, however, not evaluated in most studies, and variation could influence the accuracy measures estimated. In addition, it is important to consider whether treatment measures were initiated for the studied populations. It is well established that patients administered n-acetylcysteine have a better prognosis (30). Here again, information on this variable was not available for most studies, and we were unable to account for possible differences among criteria-positive and criteria-negative groups. More studies need to be available for evaluating the contribution of these and other factors to the overall accuracies of selected criteria. Furthermore, some of the criteria could have been used by clinicians to determine the need for liver transplantation, and this could have potentially affected the outcome (mortality). The observation that the APACHE score had the highest likelihood ratio is

Presently, available criteria used to determine the need for liver transplantation in patients with fulminant hepatic failure secondary to acetaminophen poisoning, including King's criteria, are not very sensitive and may miss patients requiring transplantation.

interesting considering that these criteria were unlikely to have been used by clinicians to determine the need for liver transplantation.

Meta-analysis techniques for summarizing diagnostic test/criteria are presently in their development stages (31). The SROC analysis we have used does not provide the clinician with one unique joint summary estimate of sensitivity and specificity. However, it does provide a reasonable comparison between different criteria and also allows the investigator or clinician to decide on the choice of the most appropriate test suitable for his or her clinical setting. We have further supplemented the analysis by comparing the likelihood ratios for the different studies.

Presently, available criteria used to determine the need for liver transplantation in patients with fulminant hepatic failure secondary to acetaminophen poisoning, including King's criteria, are not very sensitive and may miss patients requiring transplantation. Future studies should further evaluate the efficacy of the APACHE II score criteria and develop newer and more efficient criteria. In the interim, King's criteria should probably be used as the "default" criteria, but their low sensitivity should be taken into consideration.

ACKNOWLEDGMENTS

We thank Dr. Jacques Lacroix and Dr. Fernando Alvarez for reviewing an early version of this article and Dr. Lawrence Joseph for critically evaluating the meta-analytic methods utilized.

REFERENCES

1. O'Grady J, Alexander G, Hayllar K, et al: Early indicators of prognosis in fulminant hepatic failure. *Gastroenterology* 1989; 97:439–445
2. Harrison P, O'Grady J, Keays R, et al: Serial prothrombin time as prognostic indicator in paracetamol induced fulminant hepatic failure. *BMJ* 1990; 301:964–966
3. Pereira L, Langley P, Hayllar K, et al: Coagulation factor V and VIII/V ratio as predictors outcome in paracetamol induced fulminant hepatic failure: Relation to other prognostic indicators. *Gut* 1992; 33:98–102
4. Izumi S, Langley P, Wendon J, et al: Coagulation factor V levels as a prognostic indicator in fulminant hepatic failure. *Hepatology* 1996; 23:1507–1511
5. Anand A, Nightingale P, Neuberger J: Early indicators of prognosis in fulminant hepatic failure: An assessment of the King's criteria. *J Hepatology* 1997; 26:62–68
6. Bradberry S, Hart M, Bareford D, et al: Factor V and factor VIII: Ratio as prognostic indicators in paracetamol poisoning. *Lancet* 1995; 346:646–647
7. Lee W, Galbraith R, Watt G, et al: Predicting survival in fulminant hepatic failure using serum Gc protein concentrations. *Hepatol* 1995; 21:101–105
8. Schiødt F, Bondensen S, Petersen I, et al: Admission levels of Gc-globulin: Predictive value in fulminant hepatic failure. *Hepatol* 1996; 23:713–718
9. Delaney K: Hepatic principles. *In:* Golfrank's Toxicologic Emergencies. Goldfrank L, Flomenbaum N, Lewin N, et al (Ed). Stamford, CT, Appleton and Lange, 1998, pp 213–228
10. Smilkstein M. Acetaminophen. *In:* Golfrank's Toxicologic Emergencies. Goldfrank L, Flomenbaum N, Lewin N, et al (Ed). Stamford, CT, Appleton and Lange, 1998, pp 541–564
11. Meredith T, Jacobsen D, Haines J, et al: IPCS/EC Evaluation of Antidotes Series: Antidotes for Poisoning by Paracetamol. Cambridge, Cambridge Press, 1995, p 7
12. Zimmerman H: Chemical hepatic injury. *In:* Clinical Management of Poisoning and Drug Overdose. Haddad L, Shannon M, Winchester J (Eds). Philadelphia, WB Saunders, 1998, pp 149–174
13. Perry H, Shannon M. Acetaminophen. *In:* Clinical Management of Poisoning and Drug Overdose. Haddad L, Shannon M, Winchester J (Eds). Philadelphia, WB Saunders, 1998, pp 664–674
14. Ellenhorn M. Acetaminophen. *In:* Ellenhorn's Medical Toxicology. Ellenhorn M (Ed). Baltimore, Williams and Wilkins, 1997, pp 180–195
15. Anker A. Acetaminophen. *In:* Clinical Toxicology. Ford M, Delaney K, Ling L, et al (Eds). Philadelphia, WB Saunders, 2001, pp 265–274
16. Murphy N, Wendon J: Fulminant hepatic failure treatment. *In:* Evidence-Based Gastroenterology and Hepatology. McDonald J, Burroughs A, Feagan B (Eds). London, BMJ Books, 1999, pp 468–490
17. Rosen H, Martin P: Liver transplantation. *In:* Schiff's Disease of the Liver. Schiff E, Sorrell M, Maddrey W (Eds). Philadelphia, Lippincott-Raven, 1999, pp 1589
18. Stieber A, Gordon R, Galloway J: Orthotopic liver transplantation. *In:* Hepatology: A Textbook of Liver Disease. Zakim D, Boyer T (Eds). Philadelphia, WB Saunders, 1996, pp 1759
19. Sussman N: Fulminant hepatic failure. *In:* Hepatology: A Textbook of Liver Disease. Zakim D, Boyer T (Eds). Philadelphia, WB Saunders, 1996, pp 618
20. Lijmer J, Mol B, Heisterkamp S, et al: Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 2001; 282:1061–1066
21. Moses L, Shapiro D: Combining independent studies of a diagnostic test into a summary ROC curve: Data-analytic approach and some additional considerations. *Stat Med* 1993; 12: 1293–1316
22. Mutimer D, Ayres R, Neuberger J, et al: Serious paracetamol poisoning and the results of liver transplantation. *Gut* 1994; 35:809–814
23. Donalson B, Gopinath R, Wanless I, et al: The role of transjugular liver biopsy in fulminant liver failure: Relation to other prognostic indicators. *Hepatol* 1993; 18:1370–1374
24. O'Grady J, Wendon J, Potter D, et al: Liver transplantation after paracetamol overdose. *BMJ* 1991; 303:221–223
25. Shakil A, Kramer D, Mazariegos G, et al: Acute liver failure: Clinical features, outcome analysis, and applicability of prognostic criteria. *Liver Transpl* 2000; 6:163–169
26. Bernal W, Wendon J, Rela M, et al: Use and outcome of liver transplantation in acetaminophen-induced acute liver failure. *Hepatology* 1998; 27:1050–1055
27. Mitchell I, Bihari D, Chang R, et al: Earlier identification of patients at risk from acetaminophen-induced acute liver failure. *Crit Care Med* 1998; 26:279–284
28. Gow P, Smallwood R, Angus P: Paracetamol overdose in a liver transplantation centre: An 8-year experience. *J Gastroenterol Hepatol* 1999; 14:817–821
29. Halkin A, Reichman J, Schwaber M, et al: Likelihood ratios: Getting diagnostic testing into perspective. *Q J Med* 1998; 91:247–258
30. Makin A, Wendon J, Williams R: A 7-year experience of severe acetaminophen-induced hepatotoxicity (1987–1993). *Gastroenterology* 1995; 109:1907–1916
31. Dekks JJ: Systematic reviews of evaluation of diagnostic and screening tests. *In:* Systematic Reviews in Health Care. Egger M, Smith G, Altman D (Eds). London, BMJ Books, 2001, pp 248–282
32. Trey C, Davidson CS: The management of fulminant hepatic failure. *In:* Progress in Liver Disease. Popper H, Schaffner F (Eds). New York, Grune and Stratton, 1970, pp 282–298
33. O'Grady J, Schalm S, Williams R: Acute liver failure: Redefining the syndromes. *Lancet* 1993; 342:273–275