Truth Survival in Clinical Research: An Evidence-Based Requiem?

Thierry Poynard, MD, PhD; Mona Munteanu, MD; Vlad Ratziu, MD; Yves Benhamou, MD, PhD; Vincent Di Martino, MD; Julien Taieb, MD; and Pierre Opolon, MD

Purpose: **Factors associated with the survival of truth of clinical conclusions in the medical literature are unknown. The authors hypothesized that conclusions derived from studies using better methodology should have a longer half-life.**

Data Sources: **MEDLINE and hand searches of journals with studies on cirrhosis and hepatitis.**

Study Selection: **Original articles and meta-analyses published from 1945 to 1999 about cirrhosis or hepatitis in adults.**

Data Synthesis: **In 2000, 285 of 474 conclusions (60%) were still considered to be true, 91 (19%) were considered to be obsolete, and 98 (21%) were considered to be false. The half-life of truth was 45 years. The 20-year survival of conclusions derived from meta-analysis was lower (57% ± 10%) than that from nonrandomized studies (87% ± 2%) ($P < 0.001$) or randomized trials (85% ± 3%) ($P < 0.001$). The survival of conclusions was not different when studies of high methodologic quality were compared with those of low quality. In randomized trials, the 50-year survival rate was higher for 52 negative conclusions (68% ± 13%) than for 118 positive conclusions (14% ± 4%) ($P < 0.001$).**

Conclusions: **Contrary to the authors' hypothesis, conclusions based on recognized, good methodology had no clear survival advantage. To better convince clinicians of the long-term utility of evidence-based medicine, better prognostic factors should be developed.**

Ann Intern Med. 2002;136:888-895. **www.annals.org**

For author affiliations, see end of text.

Science progresses through a series of paradigms that are held to be true until they are replaced by a better approximation of reality (1). Since the development of the steam engine in the late 18th century, economists have recognized 50-year cycles during which critical technological innovation is introduced (2). In 1997, Hall and Platell (3) estimated the half-life of dogma relating to the practice of surgery. From their analysis of 260 abstracts published from 1935 to 1994, they estimated that the half-life of truth for clinical conclusions in the surgical literature was 45 years. We hypothesized that some factors should be related to this truth survival.

The first hypothesis was that conclusions derived from better methodology should have a longer half-life. If correct, this observation could be a validation of "good methodology," often called *evidence-based medicine* (4). Therefore, we compared the survival of conclusions from meta-analyses with those from isolated, randomized trials or nonrandomized studies. For the conclusions from randomized trials (isolated trials or meta-analyses), we also compared the survival rate on the basis of high versus low methodologic scores.

The second hypothesis was that survival of truth should be higher for negative conclusions than for positive conclusions. A negative conclusion has a better chance of survival because the only way it would not continue to be negative is if it were found to be false. A positive conclusion risks being found to be false or becoming obsolete. We also thought that publication of a negative conclusion in a reputable journal often indicated that a previous positive conclusion had been found to be false. We concluded, therefore, that this second-line analysis should be of higher quality.

To reduce heterogeneity in sampling and evaluation, we chose a single medical discipline—cirrhosis and hepatitis—and focused on two selective journals. We tried to categorize the conclusions into three groups: those that were true (referred to as "true" in this article), those that were not false but became obsolete (referred to as "obsolete"), and those that are now considered false (referred to as "false"). An example of an obsolete conclusion is the efficacy of immunoglobulins for preventing hepatitis A virus infection, since an effective vaccine is now available. An example of a false conclusion is the efficacy of corticosteroids for treating acute viral hepatitis.

METHODS

We identified original articles about cirrhosis or hepatitis in adults from 1945 to 1999. The articles were divided into eleven 5-year periods. Nonoriginal studies and studies involving children were excluded.

Selection of Nonrandomized Studies

In each 5-year period, we selected 20 nonrandomized articles published in two journals—10 from *Lancet* and 10 from *Gastroenterology*. We chose these journals because they have published clinical studies about hepatitis and cirrhosis since at least 1945, they are peer-reviewed and highly selective, and they have impact factors greater than 10. Articles from 1945 to 1985 were selected by a hand search. Because a true randomization was very difficult to organize, we selected by order of publication within each 5-year period. We selected the first article about cirrhosis or hepatitis that appeared within each 5-year period, then the last published article in the period, then the second article after the first, then the second-to-the-last article, and so on up to 20 articles. From 1985 to 1999, we used a PubMed electronic search and specified the following limits: "cirrhosis or hepatitis," "human," and "Lancet or Gastroenterology." Abstracts were downloaded by using a similar selection method, stratified into 5-year periods. We selected the first abstract listed on the first electronic page, then the first on the last electronic page, then the last on the second electronic page, then the last on the next-to-the-last electronic page, and so on up to 20 articles.

Selection of Randomized Trials

In each 5-year period, we tried to select 20 randomized trials about cirrhosis or hepatitis, 10 from *Lancet* and 10 from *Gastroenterology*. This was possible from 1970 to 1999. From 1945 to 1969, we selected all randomized trials that could be identified in any journal (range, 4 trials [1945 to 1950] to 20 trials [1965 to 1969]). From 1945 to 1982, we used the hand searching method previously described (5). We completed the random selection by hand searching articles from 1982 to 1985 and using PubMed (as described for nonrandomized studies) to search for articles from 1985 to 1999.

Selection of Meta-Analyses

From 1945 to 1992, we used a hand-searching method, as described in a systematic review of meta-analyses (6). Thereafter, we used PubMed and specified the following limits: "meta-analysis" and "cirrhosis or hepatitis." Because of the limited number of meta-analyses, we included all journals. All of the meta-analyses consisted solely of randomized trials.

Database Development and Observer Review

We obtained abstracts from all of the articles and selected the sentence from each abstract that seemed to best summarize the findings. These sentences were then copied to a database. Editing of these sentences was restricted to the rephrasing of outdated terminology and the elimination of redundant words.

Six hepatologists, called "observers," assessed a form that contained the selected conclusion sentences in a random order. Observers were full-time hepatologists with different clinical subspecialties (viral hepatitis, $n = 2$; HIV, $n = 1$; fibrosis, $n = 1$; alcoholic liver disease, $n = 1$; and transplantation, $n = 1$); worked in the same hospital; and were between 31 and 65 years of age. They had graduated from six different universities; three had worked in the United States, each in a different university. Observers were blinded to period, journal, authors, method (meta-analysis, randomized trial, or nonrandomized study), and the methodologic quality from which each conclusion was derived. They classified each conclusion into one of three categories: still true in 2000, obsolete but not false in 2000, or false in 2000.

Quality Assessment of Methodology and Consideration of Prognostic Factors

Independent of this study, the quality of the randomized trials was assessed by means of a scoring method (range, 2 to 14; mean, 12) that included 14 items (7, 8). Also independent of this study, the quality of the meta-analyses was assessed by means of a slightly modified version (6) of the scoring method established by Sacks and coworkers (9) and described in detail elsewhere (7). This scoring method (range, 0 to 54; mean, 27) included 27 items. We analyzed the meta-analyses that combined individual data as a separate category of research. In classic meta-analysis, the results of each trial are combined. In meta-analysis that combines individual data, the results for each patient are combined with the patient's prognostic factors, thus permitting better adjustment of the treatment effect on prognostic factors. Articles were rated as high quality when the score was greater or equal to the mean (12 for randomized trials and 27 for meta-analyses) and as low quality when the score was less than the mean. The methodologic quality of nonrandomized studies was classified as low because no specific scoring method was available.

In addition to methodologic quality, the following

Table 1. **Characteristics of Original Articles (*n* = 474)**

Year of Publication*	Method†			Disease‡			Journal§		
	Nonrandomized Studies	Randomized Trials	Meta-Analyses	Hepatitis	Portal Hypertension	Other	*Lancet*	*Gastroenterology*	Other
	←			*n*					→
1945–1964	80	30	0	48	23	39	43	41	26
1965–1979	60	60	0	39	25	56	40	49	31
1980–1999	80	80	84	103	74	67	77	90	77
Total	220	170	84	190	122	162	160	180	134

* Articles were selected every 5 years for a total of eleven 5-year periods. Between 1945 and 1964 (first 4 periods), articles were included by hand searching only. It was not possible to identify 20 randomized studies per period. Between 1965 and 1979, 20 randomized studies were identified for each period; no meta-analyses had been published. All meta-analyses were published between 1980 and 1999 (last 4 periods).
† For nonrandomized studies, only articles published in *Lancet* or *Gastroenterology* (20 every 5 years) were used. Between 1980 and 1999, 20 randomized trials were identified in *Lancet* or *Gastroenterology* for each 5-year period. From 1945 to 1979, because of the shortage of randomized trials, we searched for all randomized trials in any journal; 33 randomized trials were identified in *Lancet* or *Gastroenterology* and 57 in other journals. For meta-analyses, we systematically reviewed all journals. Of the 84 meta-analyses identified, 7 were published in *Lancet* or *Gastroenterology* and 77 in other journals.
‡ Disease was divided into hepatitis, portal hypertension (varices, hemorrhage, ascites), and other (alcoholic liver disease, primary biliary cirrhosis, and other disorders).
§ Of the 134 included articles published in other journals, 57 were randomized trials, none were nonrandomized studies, and 77 were meta-analyses. Of the 340 included articles published in *Lancet* or *Gastroenterology*, 113 were randomized trials, 220 were nonrandomized studies, and 7 were meta-analyses.
‖ Before 1965, PubMed was not available, and all randomized trials were found by hand searching.

factors were considered: negative or positive conclusion, type of disease (hepatitis, portal hypertension, alcoholic liver disease, primary biliary cirrhosis, or miscellaneous), domain of clinical research (therapeutic, diagnostic, or cognitive study [cognitive studies were defined as explanatory studies that did not assess treatment or diagnostic tests]), journal of publication (*Lancet*, *Gastroenterology*, or other), and specialty (medicine or surgery).

Statistical Analysis

A conclusion was considered to be true, obsolete, or false when three or more observers stated it to be so. When there was a split decision (3 to 3) about whether conclusions were true or not true (9 of 474 articles [1.9%]), the conclusion was considered to be true. When there was a split decision (3 to 3) about whether conclusions were obsolete or not obsolete (26 of 474 articles [5.5%]), the conclusion was considered to be obsolete. When the article was not classified as either true or obsolete, it was considered to be false.

Conclusions from older research are at greater risk for being refuted or becoming obsolete than are conclusions from more recent studies. Because the end points were highly time dependent, we used time-dependent analyses. The half-life was calculated according to the Kaplan–Meier method, using the censored time as the duration between the year of publication and the year 2000. The censored time is the time at risk for being refuted or found obsolete. For example, an article published in 1950 had a censored time of 50 years. First, we analyzed the truth survival: If the conclusion was assessed to be true, it was censored at 50 years. If the conclusion was assessed to be false or obsolete, it was considered a failure ("death of truth"). Second, we analyzed the "nonfalse" survival; true or obsolete conclusions were considered censored at 50 years. If the conclusion was false, it was considered a failure ("death of nonfalse"). The factors were compared by using the two-sided log-rank test and the multivariate proportional hazards regression analysis. Agreement among observers was analyzed by using κ statistics.

RESULTS

Characteristics of the 474 identified articles are given in **Table 1**. All nonrandomized studies were published in *Lancet* (50%) or *Gastroenterology* (50%); randomized trials were published in *Lancet* (29%), *Gastroenterology* (41%), or other journals (30%); and 92% of meta-analyses were published in other journals. Compared with the total number of articles published every year about hepatitis or cirrhosis, this sample represents less than 0.1% of nonrandomized studies, 100% of randomized trials from 1945 to 1974 but less than 5% thereafter, and 100% of meta-analyses (**Table 1**). Only six meta-analyses combined individual data. The main characteristics of articles published in *Lancet* and *Gastroenterology* did not differ from those of articles published in other journals (data not shown).

In 2000, 285 of 474 conclusions (60%) were still

Table 1—**Continued**

Type of Research			Articles Selected	Articles Identified by Using PubMed or by Hand Searching‖		
Therapeutic	Diagnostic	Cognitive		Nonrandomized Studies	Randomized Trials	Meta-Analyses
←			*n*			→
56	14	40	110	Unknown	30	0
66	18	36	120	32 524	118	0
172	31	41	244	80 898	1951	84
294	63	117	474	>100 000	2099	84

considered to be true, 91 were obsolete (19%), and 98 (21%) were false. The agreement among observers was excellent ($P < 0.001$ for all observers), with κ coefficients ranging from 0.67 to 0.88 for true conclusions and from 0.50 to 0.79 for obsolete conclusions.

Studies with true conclusions were more likely to be published more recently. Between 1980 and 1999, the percentage of true conclusions was 76%, compared with 43% between 1945 and 1979 (chi-square = 52). The half-life of truth was 45 years when expressed as the percentage of studies without false or obsolete conclusions and 55 years when expressed as the percentage of studies without false conclusions (**Figure 1**).

When methodology was considered without taking into account the publication period, there was a higher percentage of true conclusions in meta-analyses (82%) compared with randomized trials (62%) (chi-square = 10) and nonrandomized studies (50%) compared with randomized trials (chi-square = 6). When time-dependent analysis was used, there was no further difference between randomized trials and nonrandomized studies and a significantly lower 20-year survival in meta-analyses (57% ± 10%) compared with other studies (87% ± 2%) (**Figure 2**). In contrast, truth survival at 15 years was 100% for conclusions derived from the six meta-analyses that combined individual data.

Positive conclusions, conclusions derived from meta-analyses, and conclusions from studies of primary biliary cirrhosis were significantly associated with low truth survival in univariate and multivariate analysis (**Table 2**). For false conclusions, meta-analyses and studies of primary biliary cirrhosis were significantly associated with low truth survival in univariate and multivariate analysis. In randomized trials, the truth survival of 82 high-quality trials (50-year truth survival, 23% ± 19%) did not differ from that of 82 low-quality trials (22% ± 6%) (log-rank = 0.23). In meta-analyses, there was no significant difference between the 20-year truth survival of 47 high-quality meta-analyses (46% ± 14%) and that of 37 low-quality meta-analyses (69% ± 13%) (log-rank = 1.4).

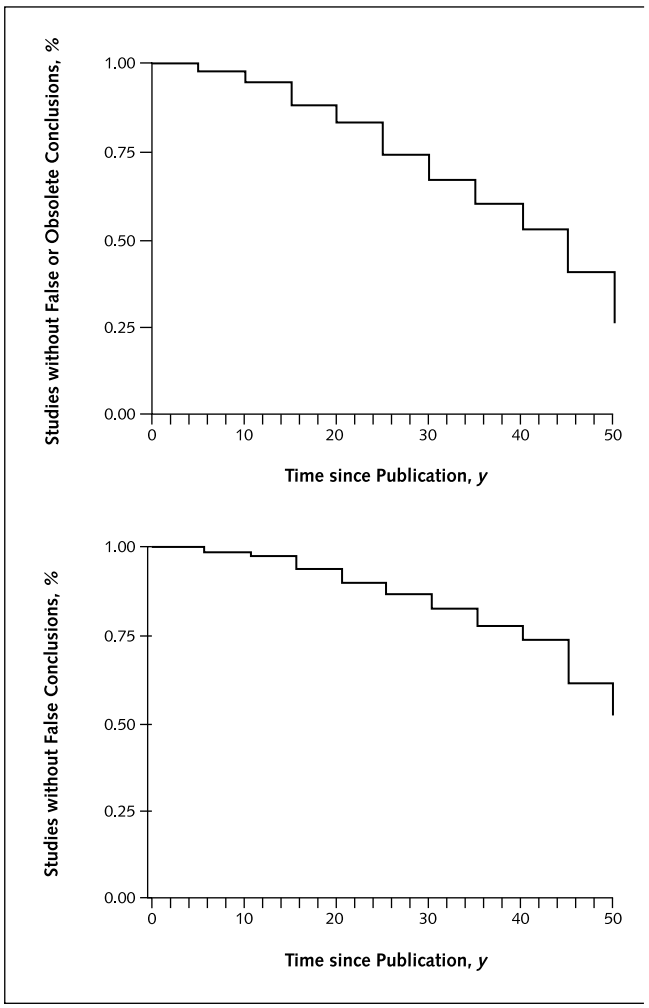
Negative conclusions were more frequent in therapeutic studies (29% vs. 14% in diagnostic or cognitive studies; $P < 0.001$), in randomized trials (31% vs. 16% in nonrandomized studies; $P < 0.001$), in meta-analyses (27% vs. 16% in nonrandomized studies; $P = 0.02$), in studies of alcoholic liver disease (43% vs. 26% for studies in other types of diseases; $P = 0.007$), and in surgical studies (43% vs. 22% in medical studies; $P = 0.02$). They were less frequent in studies from *Lancet* than in those from other journals (15% vs. 27%; $P = 0.002$). There was no difference in the number of negative conclusions in articles published before or after 1980 (23% vs. 23%; $P > 0.02$). The 50-year truth survival rate was higher for 110 negative conclusions (51% ± 12%) than for 364 positive conclusions (23% ± 3%) ($P = 0.02$). In randomized trials, the 50-year survival rate was higher for 52 negative conclusions (68% ± 13%) than for 118 positive conclusions (14% ± 4%) ($P < 0.001$). Contrary to randomized trials, the survival of a negative conclusion (34% ± 10%) was not higher than that of a positive conclusion (56% ± 25%) in nonrandomized therapeutic studies.

Finally, one example of initial false-negative conclusions obtained by evidence-based methodology derived from trials and meta-analysis is the treatment of hepatitis B virus infection with α-interferon (**Table 3**) (10–20).

Discussion

In the clinical literature on cirrhosis and hepatitis, we found the same half-life of truth (45 years) as that previously described in the surgical literature (3). One

Figure 1. **Truth survival in original articles and meta-analyses on hepatitis and cirrhosis.**

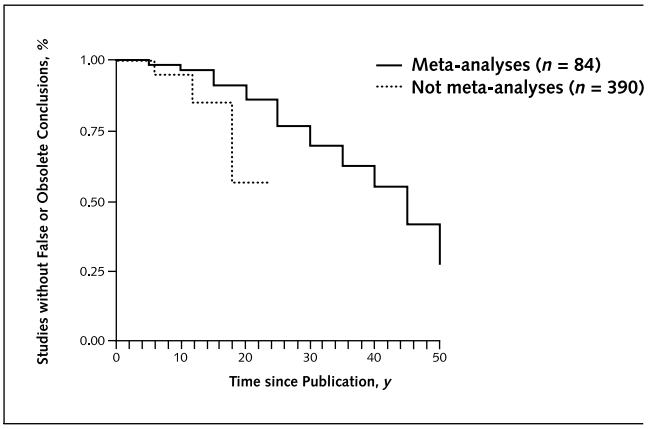


The top panel shows the percentage of studies without false or obsolete conclusions. At 50 years, the mean survival (± SE) was 26% ± 4%. The bottom panel shows the percentage of studies without false conclusions. At 50 years, the mean survival (± SE) was 53% ± 5%.

could state, as did Blau (21) of Hall and Platell's research, that our analysis proves half of Jackson's aphorism: "It takes 50 years to get a wrong idea out of medicine, and 100 years a right one into medicine." In fact, we observed that half of the nontrue conclusions in 2000 were obsolete rather than false. Many conclusions that were true at the time of their publication became obsolete following new discoveries. For example, research on the effectiveness of immunoglobulins for the prevention of hepatitis A was published in 1960 but was obsolete in 2000 because of the efficacy of the newly developed vaccine. As knowledge increases over time, earlier information becomes false or at least obsolete, no matter how good the science and methodology originally seemed. In our study, the percentage of false conclusions was limited to approximately 20%. This reassuring observation may be explained in part by the selection of *Lancet* and *Gastroenterology,* which are two very selective peer-reviewed journals.

We did not analyze a random sample of original articles but, rather, a sample of articles published in two selective journals. Of the total number of articles published, the nonrandomized studies that we analyzed represented less than 0.1% (220 of more than 100 000 published) and the randomized trials represented less than 5% (170 of 2099 published) (**Table 1**). Therefore, the limitation of our method is that we cannot extrapolate on the survival of articles published in less selective journals. However, we were able to assess truth survival in excellent scientific studies without interjournal variability. Our methodology could be used to better compare the selectivity of journals. Our sampling method was not a true randomization, which would have been very difficult to organize. A bias is impossible for meta-analyses, because all available meta-analyses were included. For randomized studies, bias is not likely because all randomized trials between 1945 and 1979 and 70% between 1980 and 1999 were included. For nonrandomized studies, 46% between 1945 and 1964, 9% between 1965 and 1979, and 9% between 1980 and 1999 were included. There was no significant difference

Figure 2. **Truth survival in meta-analyses.**



Twenty-year mean survival (±SE) was lower in meta-analyses (57% ± 10%) than in other studies (87% ± 2%).

Table 2. **Factors Associated with Truth Survival***

Factor	Truth Survival—Studies without False or Obsolete Conclusions					Not False Survival—Studies without False Conclusions				
	Mean ± SE, %	Log-Rank Test	*P* Value	Regression Coefficient	*P* Value	Mean ± SE, %	Log-Rank Test	*P* Value	Regression Coefficient	*P* Value
Conclusion										
Negative (n = 110)	51 ± 12	5.1	0.02	0.5	0.009	53 ± 12	0.8	>0.2	1.4	0.18
Positive (n = 364)	23 ± 3					28 ± 23				
Method†										
Meta-analysis (n = 84)	57 ± 10	7.0	0.03	8.0	<0.001	78 ± 9	2.6	0.11	8.5	<0.001
Randomized trial (n = 170)	85 ± 3					91 ± 3				
Nonrandomized study (n = 220)	87 ± 2					92 ± 2				
Methodologic quality										
High (n = 129)	21 ± 17	1.1	0.30	1.0	>0.2	28 ± 23	0.58	>0.2	1.0	>0.2
Low (n = 345)	27 ± 4					53 ± 5				
Disease										
Hepatitis (n = 190)	35 ± 5					58 ± 6				
Portal hypertension (n = 122)	16 ± 8					58 ± 11				
Alcohol (n = 30)	15 ± 13					17 ± 15				
Primary biliary cirrhosis (n = 29)	0 ± 13	5.2	0.02	2.3	0.004	0 ± 20	6.4	0.01	2.9	0.003
Other (n = 103)	23 ± 7					60 ± 8				
Type of research										
Therapeutic (n = 294)	25 ± 5					52 ± 6				
Diagnostic (n = 63)	22 ± 7	2.5	0.11	1.5	0.06	44 ± 11	3.7	0.06	1.9	0.02
Cognitive (n = 117)	31 ± 7					57 ± 8				
Journal										
Gastroenterology (n = 185)	21 ± 6	0.2	>0.2	1.1	>0.2	54 ± 7				
Lancet (n = 161)	29 ± 6					49 ± 7	2.3	0.13	1.7	0.03
Other (n = 128)	31 ± 9					57 ± 10				
Specialty										
Medicine (n = 451)	28 ± 4	0.1	>0.2	1.32	>0.2	53 ± 5	1.2	>0.2	0.6	>0.2
Surgery (n = 23)	0 ± 10					45 ± 19				

* The correlation-squared coefficient was 0.13 ($P < 0.001$) for truth survival and 0.11 ($P < 0.001$) for the percentage of studies without false conclusions.
† Survival analysis was done at 20 years because no meta-analyses were published before 1980.

among characteristics when included studies were compared with nonincluded studies, both for randomized or nonrandomized designs. Another weakness of our study is the definition of the truth, which does not represent a consensus of worldwide expertise. Our six observers may have been biased because they worked in the same center. However, we chose experts who represented a wide range of ages, who graduated from different universities, and who had different clinical subspecialties. There was no single opinion; the κ statistic for interobserver agreement was far from 100%, although highly significant.

Negative conclusions had a very significant survival advantage. Only 2% were rated obsolete compared with 25% of positive conclusions. A rare example of a negative conclusion that was rated obsolete was, "There was no difference between prophylactic fragmented immunoglobulins and nonfragmented immunoglobulins in reducing the occurrence of hepatitis A" (22). For a positive conclusion, there was the risk for being false positive and for being obsolete. In randomized therapeutic trials, truth survival was 68% at 50 years for negative conclusions compared with 14% for positive conclusions. This prognostic value of negative conclusions was not observed in nonrandomized therapeutic studies. As expected, we found that some negative studies had been published to reveal previous false-positive conclusions. An example in cognitive studies is the article that concluded that hepatitis B virus was not responsible for primary biliary cirrhosis (23). This article was published 18 months after another article had suggested an association between the virus and this disease (24).

The period between 1960 and 1975 was particularly rich in true conclusions. Compared with other periods, there was no difference in the percentage of negative findings. Compared with earlier periods, the true conclusions between 1960 and 1975 were mainly represented by the discovery and usefulness of hepatitis B virus marker, randomized trials showing the efficacy of corticosteroids for treating autoimmune hepatitis, and randomized trials showing the lack of effectiveness of

Table 3. **Initial False-Negative Conclusions Obtained from Randomized Trials and Meta-Analysis of Trials Using α-Interferon To Treat Chronic Hepatitis B Virus Infection**

Year	Study (Reference)	Type of Trial	Conclusion on Efficacy of α-Interferon
1976	Greenberg et al. (10)	Nonrandomized	Effective
1976	Desmyter et al. (11)	Nonrandomized	Effective
1980	Weimar et al. (12)	Randomized	Not effective
1986	Burke (13)	Meta-analysis	Not effective
1987	Alexander et al. (14)	Randomized	Effective
1988	Hoofnagle et al. (15)	Randomized	Effective
1991	Serra et al. (16)	Meta-analysis	Effective
1993	Tinè et al. (17)	Meta-analysis	Effective
1993	Wong et al. (18)	Meta-analysis	Effective
1993	Ryff (19)	Meta-analysis	Effective
1994	Krogsgaard et al. (20)	Meta-analysis with individual data	Effective

portocaval shunts for preventing first bleeding episodes in patients with cirrhosis. The period between 1975 and 1979 had a high percentage of obsolete conclusions that were mainly associated with results of randomized trials showing the efficacy of immunoglobulins for preventing transfusion-related hepatitis.

The last and disappointing finding was that, contrary to our hypothesis, conclusions based on recognized good methodology had no clear survival advantage. A superficial analysis would have concluded that evidence-based medicine was a favorable prognostic factor. Unfortunately, when the aging phenomenon was taken into account, we did not observe a clear survival benefit for either randomized trials or meta-analyses.

Survival did not differ between randomized and nonrandomized studies, even when comparison was restricted to therapeutic studies. One optimistic explanation is that the selective journals accurately identified the "good" nonrandomized therapeutic studies, especially in gray zones of clinical practice (25). In addition, we observed no survival difference between high-quality and low-quality randomized trials. An optimistic explanation is that the selective journals we studied did not include very poor randomized trials. However, even after 1980, 25% of published randomized trials were of low methodologic quality. Another explanation could be a weakness in the scoring system used (7, 8). Our scoring system, however, was derived from the Chalmers scoring system and contained most of the items recently recommended by the CONSORT (Consolidated Standards of Reporting Trials) statement (26).

There was no survival difference between conclusions from meta-analyses and isolated randomized conclusions. Furthermore, the truth survival was lower for meta-analysis conclusions than for conclusions derived from non–meta-analyses. One explanation could be that low-quality meta-analyses increase the risk for false-positive or false-negative conclusions (27). However, we observed no survival difference between high-quality and low-quality meta-analyses. Another explanation could be a weakness of the scoring system used (6), although the system used in our study is widely recognized (28).

What can be proposed to increase the survival of clinical conclusions? One new gold standard could be the Cochrane reviews, which appear to have greater methodologic rigor and are more frequently updated than systematic reviews or meta-analyses published in paper-based journals (29). In the field of hepatology, too few Cochrane reviews have been performed to test this hypothesis. Another proposal is to encourage meta-analyses combining individual data; we base this on our observation that truth survival was 100% in such meta-analyses. Even nonexhaustive meta-analyses that combine individual data could have a better prognostic value than exhaustive meta-analyses that do not use individual data. One example is the negative conclusion of a classic meta-analysis of the efficacy of corticosteroids for treating acute alcoholic hepatitis (30). In this overview, the biggest weight (four times more than the other trials) was given to a negative trial in which patients with severe disease were mixed with patients without severe disease. We recently gathered the individual data from this negative trial and the data from two more recent trials (31) and found that the effect of corticosteroids on survival in patients with severe alcoholic hepatitis was highly significant. Other controversies, such as the efficacy of ursodesoxycholic acid in the treatment of primary biliary cirrhosis (32, 33), could be resolved by combining individual data.

In conclusion, if viewed pessimistically, our survival analysis could represent a requiem for evidence-based medicine. An optimistic interpretation is that conclusions published in good clinical journals, whether from randomized trials or nonrandomized studies, have excellent truth survival because editors and reviewers have eliminated the bad science. A balanced interpretation is that improved prognostic factors should be found to better convince clinicians of the long-term utility of ev-

idence-based medicine. Meta-analysis combining individual data should be encouraged.

From Groupe Hospitalier Pitié-Salpêtrière, Paris, France.

Acknowledgment: The authors thank Solko Schalm, MD, for his support and helpful comments.

Grant Support: By the European Association for Study of the Liver and Association pour la Recherche sur Maladies Hépatiques Virales.

Requests for Single Reprints: Thierry Poynard, MD, PhD, Service d'Hépato-Gastroentérologie, Groupe Hospitalier Pitié-Salpêtrière, 47-83 Boulevard de l'Hôpital, 75651 Paris Cedex 13, France; e-mail, tpoynard@teaser.fr.

Current author addresses are available at www.annals.org.

References

1. **Polanyi M.** Science, Faith, and Society. London: Oxford Univ Pr; 1946.

2. **Kernick DP.** Lies, damned lies, and evidence-based medicine. Lancet. 1998; 351:1824. [PMID: 9635992]

3. **Hall JC, Platell C.** Half-life of truth in surgical literature [Letter]. Lancet. 1997;350:1752. [PMID: 9413475]

4. **Sackett DL, Rosenberg WM.** The need for evidence-based medicine. J R Soc Med. 1995;88:620-4. [PMID: 8544145]

5. **Poynard T, Conn HO.** The retrieval of randomized clinical trials in liver disease from the medical literature. A comparison of MEDLARS and manual methods. Control Clin Trials. 1985;6:271-9. [PMID: 3907971]

6. **Auperin A, Pignon JP, Poynard T.** Review article: critical review of meta-analyses of randomized clinical trials in hepatogastroenterology. Aliment Pharmacol Ther. 1997;11:215-25. [PMID: 9146758]

7. **Poynard T.** [Evaluation of the methodological quality of randomized therapeutic trials]. Presse Med. 1988;17:315-8. [PMID: 2966351]

8. **Bernard B, Lebrec D, Mathurin P, Opolon P, Poynard T.** Beta-adrenergic antagonists in the prevention of gastrointestinal rebleeding in patients with cirrhosis: a meta-analysis. Hepatology. 1997;25:63-70. [PMID: 8985266]

9. **Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC.** Meta-analyses of randomized controlled trials. N Engl J Med. 1987;316:450-5. [PMID: 3807986]

10. **Greenberg HB, Pollard RB, Lutwick LI, Gregory PB, Robinson WS, Merigan TC.** Effect of human leukocyte interferon on hepatitis B virus infection in patients with chronic active hepatitis. N Engl J Med. 1976;295:517-22. [PMID: 950957]

11. **Desmyter J, De Groote J, Desmet VJ, Billiau A, Ray MB, Bradburne AF, et al.** Administration of human fibroblast interferon in chronic hepatitis-B infection. Lancet. 1976;02:645-7. [PMID: 60513]

12. **Weimar W, Heijtink RA, ten Kate FJ, Schalm SW, Masurel N, Schellekens H, et al.** Double-blind study of leucocyte interferon administration in chronic HBsAg-positive hepatitis. Lancet. 1980;1:336-8. [PMID: 6101791]

13. **Burke CA.** A statistical view of clinical trials in chronic hepatitis B. J Hepatol. 1986;3 Suppl 2:S261-7. [PMID: 2439574]

14. **Alexander GJ, Brahm J, Fagan EA, Smith HM, Daniels HM, Eddleston AL, et al.** Loss of HBsAg with interferon therapy in chronic hepatitis B virus infection. Lancet. 1987;2:66-9. [PMID: 2885573]

15. **Hoofnagle JH, Peters M, Mullen KD, Jones DB, Rustgi V, Di Bisceglie A, et al.** Randomized, controlled trial of recombinant human alpha-interferon in patients with chronic hepatitis B. Gastroenterology. 1988;95:1318-25. [PMID: 3049216]

16. **Serra MA, Rodrigo JM, del Olmo JA, Escudero A, Rodríguez F.** [A meta-analysis of controlled studies of the treatment of chronic hepatitis due to the hepatitis B virus]. Med Clin (Barc). 1991;97:681-6. [PMID: 1722862]

17. **Tinè F, Liberati A, Craxì A, Almasio P, Pagliaro L.** Interferon treatment in patients with chronic hepatitis B: a meta-analysis of the published literature. J Hepatol. 1993;18:154-62. [PMID: 7691924]

18. **Wong DK, Cheung AM, O'Rourke K, Naylor CD, Detsky AS, Heathcote J.** Effect of alpha-interferon treatment in patients with hepatitis B e antigen-positive chronic hepatitis B. A meta-analysis. Ann Intern Med. 1993;119:312-23. [PMID: 8328741]

19. **Ryff JC.** To treat or not to treat? The judicious use of interferon-alpha-2a for the treatment of chronic hepatitis B. J Hepatol. 1993;17 Suppl 3:S42-6. [PMID: 8509638]

20. **Krogsgaard K, Bindslev N, Christensen E, Craxi A, Schlichting P, Schalm S, et al.** The treatment effect of alpha interferon in chronic hepatitis B is independent of pre-treatment variables. Results based on individual patient data from 10 clinical controlled trials. European Concerted Action on Viral Hepatitis (Eurohep). J Hepatol. 1994;21:646-55. [PMID: 7814812]

21. **Blau JN.** Half-life of truth in medicine [Letter]. Lancet. 1998;351:376. [PMID: 9652656]

22. **Brachott D, Mosley JW, Lipschitz I, Kendrick MA, Sgouris JT.** Fragmented IgG for post-exposure prophylaxis of type A hepatitis. Transfusion. 1972;12:389-93. [PMID 4264692]

23. **Sama S, Aach R, Benz W, Hacker E, Kaplan M.** False-positive Australia-antigen particles in primary biliary cirrhosis. Detection by electron microscopy. Lancet. 1973;1:14-7. [PMID: 4118536]

24. **Krohn K, Finlayson ND, Jokelainen PT, Anderson KE, Prince AM.** Electron microscopical and immunological observations on the serum-hepatitis (S.H.) antigen in primary biliary cirrhosis. Lancet. 1970;2:379-83. [PMID: 4194689]

25. **Naylor CD.** Grey zones of clinical practice: some limits to evidence-based medicine. Lancet. 1995;345:840-2. [PMID: 7898234]

26. **Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al.** Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA. 1996;276:637-9. [PMID: 8773637]

27. **Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al.** Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998;352:609-13. [PMID: 9746022]

28. **Sacks HS, Reitman D, Pagano D, Kupelnick B.** Meta-analysis: an update. Mt Sinai J Med. 1996;63:216-24. [PMID: 8692168]

29. **Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, et al.** Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. JAMA. 1998;280:278-80. [PMID: 9676681]

30. **Christensen E, Gluud C.** Glucocorticoids are ineffective in alcoholic hepatitis: a meta-analysis adjusting for confounding variables. Gut. 1995;37:113-8. [PMID: 7672658]

31. **Mathurin P, Mendenhall C, Carithers RL, Ramond MJ, Maddrey WC, Garstride P, et al.** Corticosteroids improve short-term survival in patients with severe alcoholic hepatitis (AH): Individual data analysis of the last three randomized placebo controlled double blind trials of corticosteroids in severe AH. J Hepatol. 2002;36:480-7. [PMID: 11943418]

32. **Goulis J, Leandro G, Burroughs AK.** Randomised controlled trials of ursodeoxycholic-acid therapy for primary biliary cirrhosis: a meta-analysis. Lancet. 1999;354:1053-60. [PMID: 10509495]

33. **Poupon RE, Lindor KD, Cauch-Dudek K, Dickson ER, Poupon R, Heathcote EJ.** Combined analysis of randomized controlled trials of ursodeoxycholic acid in primary biliary cirrhosis. Gastroenterology. 1997;113:884-90. [PMID: 9287980]