Reprinted from ANNALS OF INTERNAL MEDICINE Vol. 104; No. 6 June 1986
Reprinted in USA

Computer-Assisted Design of Studies Using Routine Clinical Data

Analyzing the Association of Prednisone and Cholesterol

ROBERT L. BLUM, M.D., Ph.D.; Stanford, California

To facilitate the analysis of routine, longitudinal, clinical data, we developed a computer program called the RX Study Module. Our prototype uses a small online knowledge base of medicine and biostatistics to help create and execute a detailed statistical study design. The program identifies possible confounding variables, selects methods for controlling them, creates a statistical model, determines patient eligibility criteria, and retrieves data from records. We used the program to examine the hypothesis that daily prednisone administration elevates serum cholesterol. Data from 49 patients with chronic rheumatologic disorders were analyzed from a database of 1787 patients. A regression model was fitted to each patient's record. Changes in cholesterol were significantly correlated ($p = 10^{-5}$) with changes in prednisone after a lag of at least 1 week and after recorded confounders were controlled: Δcholesterol $= 18.4 \log_e$(prednisone). Routinely collected patient data may become an important resource for generating and studying new medical hypotheses.

SINCE 1970 when large-scale integrated circuits were first marketed, the capabilities of computers and electronic memories have increased a thousand-fold while their costs have dropped by the same factor. These technologic developments have permitted the increasingly widespread collection of routine, longitudinal clinical data in electronic form. Motivation for recording data electronically comes not only from physicians and other practitioners who desire rapid access to records for patient management and clinical investigation, but also from third-party insurers, hospital administrators, and government agencies. In response to these two factors—the decreasing cost of computing equipment and the increasing demand for detailed and accurate clinical data—a number of software systems have emerged for recording clinical data sequentially as it is routinely gathered in ambulatory clinics. Widely known examples include COSTAR (1, 2), TMR (3), RMIS (4, 5), STOR (6), ARAMIS (7), MEDLOG (8), and CLINFO (9).

Recognizing the growing value of these databases for medical studies, we started work in 1978 on a software package called the RX Study Module. This program is intended to assist a clinical investigator with designing and analyzing a database of routine clinical data. The Study Module facilitates these tasks by assisting with the selection and control of confounding variables, assisting with the design of the mathematical model of the effect of interest, and automating the execution of the study design on a statistical package (10-14).

We will show how the Study Module works using one clinical hypothesis of interest. The study design includes three innovative aspects. First, the study design itself is assembled semi-automatically with the assistance of an online knowledge base of clinical medicine and biostatistics. Second, the resulting study design uses all the relevant data in long-term patient records, taking account of the detailed time relationships among the data. Third, the study design uses a two-stage regression method that enables it to quantitate the effects of drugs within individual patient records and to compare these effects across patients. These methods have not previously been applied to routinely collected, longitudinal clinical data. To demonstrate the methods used by the Study Module we will show the steps used to study the association of prednisone with subsequent increases in serum cholesterol.

The Effect of Prednisone on Cholesterol

Beyond serving as a demonstration of how the basic design of a study may be assisted by computer, the effect of chronic, daily prednisone administration on serum cholesterol is of considerable clinical importance. Prednisone is a potent steroid commonly administered for various chronic disorders. Hypercholesterolemia is a well-established risk factor for atherogenesis and subsequent vascular occlusion. If patients who are chronically taking steroids have elevated serum cholesterol levels they will be at increased risk for accelerated arteriosclerosis, myocardial infarction, and stroke. Indeed, the evidence supporting this relationship is abundant.

Over the past 35 years there have been several published reports showing that chronically administered steroids elevate serum lipoproteins with concomitant increases in serum cholesterol. The first report of this effect in humans appeared in 1950 (15). In 1962, Moran (16), studying the effects of daily, high-dose cortisone in rabbits, found severe hyperlipoproteinemia as well as fatty changes in the liver, kidney, and other organs. Reaven and associates (17) reported that corticosteroid administration in rats increased both plasma triglycerides and cholesterol, caused by accelerated hepatic lipoprotein synthesis. Zimmerman and colleagues (18), in a recent prospective study of 12 patients, found a 17% increase in cholesterol, mainly due to a 68% elevation in high-density lipoprotein cholesterol.

Clinical studies within the past 10 years have focused predominantly on patients who have undergone renal

▶ From the Department of Computer Science, Stanford University; Stanford, California.

transplantation and received long-term steroid administration to prevent graft rejection. Several investigators have reported substantial increases in serum cholesterol as well as dramatic increases in the incidence of myocardial infarction and stroke (19-23). In 1974 Ibels and associates (21) observed that the incidence of coronary thrombosis in such a population was 25 times greater than that expected in a control population, and the incidence of cerebral thrombosis was 300 times greater. Similarly, Stern and associates (23) saw several cases of myocardial infarction and marked hypercholesterolemia in relatively young female patients with rheumatologic disorders after chronic steroid administration.

Despite the fact that these and other studies involved prospective study design and data collection, several issues have not been resolved, namely the magnitude of the effect, its dose and response relationship, and its relation to clinical setting. Furthermore, previous studies have not quantitatively examined the steroid-to-cholesterol relationship in the presence of the nephrotic syndrome to determine whether the effects of steroids and the nephrotic syndrome are additive or interactive. By making use of multiple, sequential observations of cholesterol over various prednisone dosages over a number of years, we can address these quantitative issues.

ARAMIS: A Time-Oriented Clinical Database of Rheumatology

The study was done using the subset of the American Rheumatism Association Medical Information System (ARAMIS) database of rheumatology collected by the Stanford University Immunology Division. (The complete ARAMIS database comprises information from 11 health care institutions [7, 24]. The clinical attributes in the database have been reported by the American Rheumatism Association [25].) The Stanford database was begun in 1970 and includes the records of 1787 patients with a broad spectrum of rheumatologic diagnoses, including a relative abundance of severely ill patients referred to the university.

Many patients, particularly those with systemic lupus erythematosus, were chronically taking prednisone at differing doses depending on the severity of their disease. Occasionally, their serum cholesterol levels would be determined, either because it was deemed relevant or simply because cholesterol levels happened to be part of a multi-test chemistry panel. From 1971 through 1982 serum cholesterol levels were determined using an SMA 12/60 from Technicon Instrument Division (Tustin, California) using the Liebermann-Burchard reagent method. This colorimetric method is linear to a cholesterol concentration of 500 mg/dL.

Hazards of Making Causal Inferences from Non-Protocol, Non-Randomized Data

The portion of a patient record in Table 1 shows the format of the data analyzed by the Study Module and will also be used to show the difficulties the Study Module attempts to handle in drawing inferences from non-protocol data. (The data are from the record of Patient 26, Table 2.) Table 1 shows values for prednisone and cholesterol recorded over time in one patient with severe rheumatoid arthritis. The first column of the table shows the number of the clinic visit, and the second column shows the number of days elapsed since the first visit. Although it appears that the cholesterol tends to rise after an increase in the dosage of prednisone and falls with a decrease in prednisone, there are many problems in quantitating this time-lagged association and in deciding whether it is spurious.

Table 1. Data from Record of One Patient with Rheumatoid Arthritis*

Visit	Day	Prednisone	Cholesterol†
		mg/d	*mg/dL*
1	0	7.5	. . .
2	11	7.5	. . .
3	18	7.5	. . .
4	25	7.5	. . .
5	32	7.5	170
6	39	7.5	. . .
7	53	7.5	175
8	60	7.5	180
9	67	7.5	. . .
10	81	7.5	. . .
11	88	30.0	. . .
12	95	30.0	. . .
13	109	30.0	215
14	123	30.0	. . .
15	137	25.0	230
16	165	17.5	250
17	193	15.0	. . .
18	221	12.5	. . .
19	256	8.5	. . .
20	305	8.5	. . .
21	340	8.5	. . .
22	443	7.5	. . .
23	641	6.5	170
24	774	7.0	. . .
25	865	7.0	. . .

* Patient 26, Table 2.
† Cholesterol level measured only when clinically indicated.

First, even though we have several sequential values of prednisone and cholesterol in this record, the time intervals between clinic visits are irregular, and the intervals between prednisone administration and measurement of cholesterol are irregular. Prednisone is only given sporadically, if at all, to most patients, and the time of measurement of cholesterol is arbitrary.

Second, although we may assume, at least for periods less than 1 month or so, that patients are taking prednisone at a constant dose until the next clinic visit, this is certainly not guaranteed. Prescriptions may not have been filled right away, and patients may have discontinued taking the drug themselves or taken the medication erratically.

Third, and perhaps most important, it is never possible to prove that an apparent drug effect is actually due to the drug. Unless treatments are randomly assigned to patients there is little possibility of controlling for unknown or unrecorded confounding variables (26). (Even in a randomized trial all one can say is that a demonstrated effect is unlikely to have been caused by a maldistribution of important variables.) The best one can do in a non-randomized study is to attempt to control for recorded

Table 2. Effects of Prednisone on Cholesterol in 34 Patients Without the Nephrotic Syndrome*

Patient	b_i	*seb*	*p*	w_i	*nvarx*	n_i	$pred_{max}$	$chol_{min,\ max}$	Diagnosis†
1	16.8	9.9	0.112	0.054	36.38	16	60	112-325	SLE
2	11.8	5.8	0.074	0.052	31.24	11	48	135-225	SLE
3	30.9	7.3	0.008	0.050	30.43	7	20	203-315	SLE
4	4.9	7.8	0.554	0.048	24.45	8	50	73-195	PM
5	30.1	3.4	0.000	0.047	23.26	8	60	160-290	SLE
6	15.8	16.0	0.368	0.046	22.06	7	20	156-355	SS
7	−5.2	9.9	0.616	0.045	21.72	7	30	175-260	SS
8	22.3	10.2	0.079	0.045	20.89	7	45	145-261	SLE
9	1.7	14.8	0.912	0.043	18.99	7	15	135-260	SLE
10	33.6	9.3	0.006	0.041	15.38	11	60	190-344	SLE
11	−3.4	6.6	0.611	0.038	13.26	15	60	173-277	WEG
12	15.0	6.9	0.161	0.034	13.05	4	10	140-195	RA
13	50.3	9.5	0.034	0.034	13.01	4	30	145-285	DERM
14	12.2	10.5	0.307	0.035	12.49	6	20	165-230	SLE
15	30.3	30.9	0.431	0.032	11.57	4	60	130-245	SLE
16	8.8	10.5	0.464	0.030	9.46	5	30	125-207	SS
17	8.4	12.1	0.560	0.027	8.56	4	5	150-200	SLE
18	35.1	13.4	0.059	0.029	8.43	6	30	185-300	ART
19	−2.9	16.2	0.862	0.029	8.31	8	60	185-300	SLE
20	26.1	14.5	0.146	0.027	7.61	6	60	145-245	SLE
21	19.8	5.3	0.013	0.027	7.47	7	6	180-220	RA
22	7.5	34.4	0.848	0.024	6.90	4	30	150-357	ART
23	34.1	25.5	0.199	0.026	6.29	19	40	200-360	ART
24	16.9	29.5	0.625	0.022	6.23	4	60	172-250	SLE
25	−22.4	11.8	0.154	0.018	4.45	5	50	118-260	DERM
26	44.4	14.5	0.037	0.016	3.55	6	30	170-250	RA
27	57.0	26.4	0.097	0.016	3.49	6	20	215-330	FIBR
28	83.1	59.3	0.296	0.012	2.92	4	60	180-445	SLE
29	25.9	20.2	0.270	0.013	2.81	6	40	115-180	SLE
30	12.4	7.9	0.257	0.010	2.25	4	20	260-280	FIBR
31	−4.3	6.4	0.551	0.010	2.13	5	8	195-210	RA
32	84.2	4.1	0.002	0.007	1.43	4	35	145-220	ART
33	29.5	11.6	0.026	0.007	1.26	14	30	170-290	SLE
34	−42.2	50.2	0.462	0.005	1.04	5	15	155-285	SLE

* b_i = regression coefficient of Δcholesterol on $\Delta\log_e$ (prednisone); *seb* = standard error of b_i; *p* = *p* value of b_i; w_i = weight assigned to b_i by the inverse variance method; *nvarx* = n_i times the variance of $\log_e$ (prednisone); n_i = number of observations in the *i*th patient's data matrix; $pred_{max}$ = the maximum recorded dose of prednisone as mg/d; $chol_{min,\ max}$ = minimum and maximum serum cholesterol level as mg/dL.

† SLE = systemic lupus erythematosus; PM = polymyositis; SS = systemic sclerosis; WEG = Wegener's granulomatosis; RA = rheumatoid arthritis; DERM = dermatomyositis; ART = arteritis; FIBR = myofibrosis or fibrositis.

confounding variables. Then, if a time-lagged association still remains, one may seek further confirmation of the effect in subsequent studies, particularly in randomized controlled trials.

With several hundred clinical variables in the ARAMIS database, one can attempt to adjust for many of the known confounding variables that have been recorded. But even the measured, confounding variables will have been recorded at irregular intervals, so that adjusting for them may be difficult. Again, one can never claim that a perceived association is actually due to a causal relationship rather than being spurious.

Finally, once an appropriate data set is assembled, it is not clear how it should be analyzed statistically. We will show how each of these issues is considered, in turn, and how an appropriate statistical study design, whose objective is to attenuate these problems, is automatically constructed and executed by the Study Module.

RX Study Module

The RX Study Module is a large software system developed on the SUMEX-AIM (Stanford University Medical Experiment on Artificial Intelligence in Medicine) computer facility. The main computer of the facility is a DEC 20/60 (Digital Equipment Corporation, Maynard, Massachusetts) running TOPS-20. The Study Module uses methods of symbolic reasoning usually referred to as artificial intelligence. To mediate this logical reasoning, the program uses machine-readable knowledge bases of clinical medicine and statistics. Both the program and the knowledge bases are written in INTERLISP (27), a dialect of LISP, a language suitable for knowledge manipulation (28). The program makes use of a statistical package called IDL (Interactive Data-Analysis Language) (29), also written in INTERLISP.

The user first types the names of the independent and dependent variables into the computer; for example, "prednisone" and "cholesterol" will initiate a study of the hypothesis that prednisone affects cholesterol. Using its machine-readable knowledge bases of clinical medicine and statistics, the computer program generates an epidemiologic study design for the hypothesis, which it analyzes on appropriate information from the database. The user can interact with the program to modify or override its design decisions.

The flow of information through the Study Module is shown in Figure 1. The program transforms the entered hypothesis into a machine-readable description of a com-

prehensive statistical study design, which may then be executed by a statistical package. The second output of the program is a logical query directed to the clinical database, which specifies the criteria for selecting patients for the study. The resulting set of data produced by the database serves as the input to the statistical package. The results of the analysis are then passed back to the Study Module for interpretation.

IDENTIFICATION OF CONFOUNDING VARIABLES

The first step of the program is the identification of variables recorded in the database that may possibly confound the relationship of interest (that is, "does prednisone elevate cholesterol?"). To identify a set of confounding variables, the Study Module uses an online medical knowledge base that contains descriptions of clinical variables including diseases, syndromes, laboratory tests, and drugs (11).

Figure 2 shows the kind of information that the Study Module has available for determining a set of possible confounding variables for a given study. The figure shows the causal links connecting glomerulonephritis to the nephrotic syndrome and to prednisone. These links mean that glomerulonephritis occasionally causes the nephrotic syndrome, and that glomerulonephritis may induce a physician to treat his or her patient with prednisone. We also see links connecting the nephrotic syndrome to cholesterol and to prednisone, meaning that the nephrotic syndrome may elevate cholesterol and that it is occasionally treated by prednisone. The links from diabetic ketoacidosis signify that it too may elevate serum cholesterol, and furthermore, that it may cause a reduction in prednisone dosage, because it is a relative contraindication to

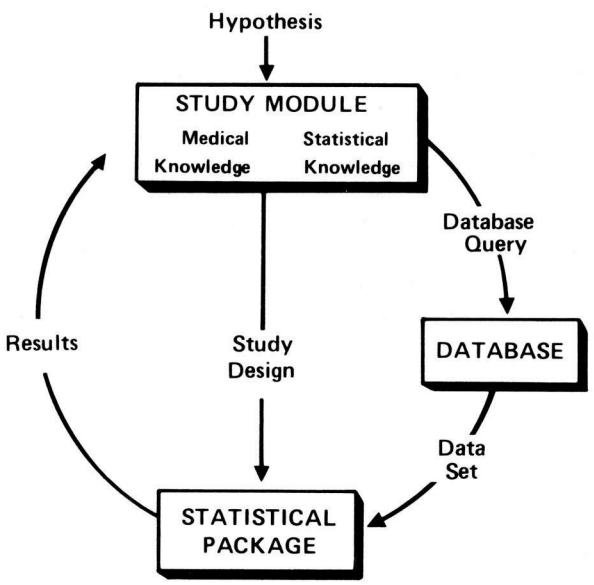

Figure 1. Study module information flow. A hypothesis is entered by the researcher into the Study Module. The Study Module uses its medical and statistical knowledge bases to formulate a study design, which is passed to the statistical package. The Study Module also formulates criteria for patient selection passed to the database as a query. The resulting data set is analyzed by the statistical package, and the results are then returned to the Study Module for interpretation.

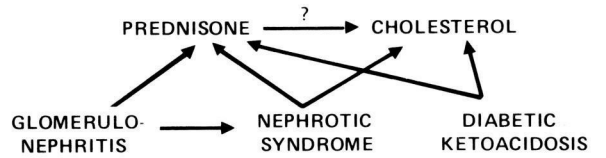

Figure 2. Causal relationships that may confound the effect of prednisone on cholesterol. These relationships have previously been stored in the medical knowledge base. For example, it is known that the nephrotic syndrome will elevate serum cholesterol levels. It is also known that the nephrotic syndrome may cause the physician to administer prednisone.

steroids. The actual links in the knowledge base contain detailed quantitative and probabilistic information on hundreds of relationships like these (11, 13).

Using this information, the Study Module searches for all diagnoses like the nephrotic syndrome that can affect both the independent variable (the drug) and the dependent variable (the side effect), however indirectly. These potentially confounding variables will then be individually controlled. For this study, the Study Module suggested that *glomerulonephritis, diabetic ketoacidosis, hepatitis,* and the *nephrotic syndrome* be controlled variables. The investigator may add or delete diagnoses from this list.

CONTROLLING CONFOUNDING VARIABLES

After determining that these clinical variables should be controlled, the Study Module proceeds to determine how each variable should be controlled, given the characteristics of the data in the database. The Study Module uses three methods to control confounding variables: elimination of entire patient records; elimination of time intervals containing confounding events; and statistical control for the presence of the confounder. To select a method to control each confounder, RX uses decision criteria stored in the knowledge base in the form of if/then rules. A typical decision rule is shown below, paraphrased in English:

> If: a particular confounding variable is limited to brief time intervals in most patient records in the study
> Then: eliminate the affected time intervals from those patient records.

If the decision is made to use some form of statistical control, then other rules may be invoked that choose among several strategies, including contingency table analysis and modeling by multivariate regression.

In the prednisone and cholesterol study the program makes the following decisions: control for hepatitis and ketoacidosis by eliminating the affected time intervals from patient records and control for the nephrotic syndrome statistically. The study further determines that no control is needed for glomerulonephritis, because control of the nephrotic syndrome should entirely remove its effect.

To control statistically for the nephrotic syndrome, albumin is selected as a proxy variable by the following means. The knowledge base contains a machine-readable definition of the nephrotic syndrome in terms of the attributes in the database (serum cholesterol level, serum albumin level, 24-hour urine protein level, and protein level by dipstick). For each attribute in the definition, the program samples patient records from the database and

compares the robust correlations of each variable with the proxy variable of interest. In this example, albumin was picked as a proxy, although in another database, the 24-hour urine protein level might be selected.

CHOOSING A STATISTICAL METHOD

To choose an appropriate statistical method the Study Module uses a statistical knowledge base containing information on a few dozen statistical methods. Their *procedures, prerequisites, assumptions,* and *objectives* are recorded in the knowledge base in machine-readable form. The *procedure* is the method for invoking the computer program that does the actual statistical analysis. The *prerequisites* are the conditions that must hold for the method to be mechanically applied. The *assumptions* are a list of conditions that must hold for the result to be valid. The *objectives* are the goals of the method. For the current study, the standard regression method was chosen based on the information that appears below. The computer stores not only the English text but the equivalent machine-executable code. Note that in the current version of the system no check is made on the validity of the *assumptions* of the model.

Standard Regression
procedure: Call program Regression with option = standard
prerequisites:
One or more independent variables
One dependent variable
Measurement-level of dependent variable = real valued
Measurement-level of independent variables = real valued
Number of observations > 1 + number of independent variables
assumptions:
Unknown causal ordering among the independent variables
Independent, normally distributed errors
Errors have constant variance (homoscedasticity)
Additive and linear effects
objectives:
Quantification of effects of independent variables

Having decided to use a regression method to fit the data still leaves open many issues as to when the data are sampled from patient records, how the regression model is formulated within each patient record, and how the data from separate patients are combined.

TIMING AND ELIGIBILITY CRITERIA

First, the data must be sampled from patient records at appropriate time lags. For example, time must elapse before a drug's effects become apparent. Furthermore, to control for intervals during which the patient has alterations in plasma lipids due to hepatitis or diabetes, their relevant durations, onset-delays, and carry-over periods must be taken into account.

The relevant time parameters may be accessed by the Study Module from the medical knowledge base, or, as in the present example, the time delays were manually entered into the program based on estimates from the medical literature. The Study Module next assembles a set of database queries, which when executed by the clinical database system will result in an appropriate data set for each patient.

All 1787 patient records in the database are not suitable for this study. Many patients never received prednisone, or had too few cholesterol levels recorded, or the timing of the cholesterol measurements with respect to prednisone may have been unsuitable. To create a set of patients who are eligible for the study, the Study Module creates a set of machine-readable eligibility criteria. These criteria, in combination with the timing information discussed above, give rise to the data set that is analyzed by the statistical package. For this study, patients were required to have had at least five recorded cholesterol levels of which at least two were preceded by periods of constant, daily-dose prednisone administration.

Statistical Analysis: Two-Stage Regression

The application of the eligibility criteria and the timing criteria to each patient's time-oriented record results in a data set of values for cholesterol, prednisone, and the other covariates. Given that some form of regression is appropriate for the analysis, the program must still determine what form of regression analysis to use on the individual data sets.

Either a one-stage or two-stage method may be used. In the one-stage method the data from all patients are combined into one large data set and simultaneously fitted to a model. This action produces an overall correlation between prednisone and cholesterol. This is the traditional and most widely used regression method, but there are several problems with it when used for the analysis of routine clinical databases.

Foremost among these difficulties is that patients have greatly varying amounts of data in their records, both on the variables of interest and on the confounding variables. The one-stage method does not allow adjustment for the quantity of relevant data on each patient or for the range of variation in the independent variable. Furthermore, spurious or extreme values in one patient, for example, due to an unknown covariate, may markedly influence the overall result without becoming known to the investigator. Empirical studies have shown that the one-stage method frequently results in an inaccurate estimate of the overall effect (30, 31). (For readers unfamiliar with regression, the books by Draper and Smith [32] and by Brown and Hollander [33] are excellent.)

In the two-stage regression method each patient's data set is separately analyzed and the results are then combined (14, 31). For example, we first regress each patient's set of cholesterol levels on the prednisone dosages received. Second, the regression coefficients from the separate analyses are combined to test the overall hypothesis. The Study Module uses two-stage regression when the data in the patient records are adequate, and there are several advantages to this method.

First, an estimate of the size of the effect and its variance for each patient is obtained. Because the patients are analyzed individually, the unknown covariates and outliers from one record cannot influence the analyses of other records. This means that the analysis is easier to interpret than the model with all data analyzed concurrently. Second, the regression coefficients of all the patients may be plotted and analyzed as a function of other covariates to see, for example, how the effect varies across clinical settings or diseases. For example, one may plot the

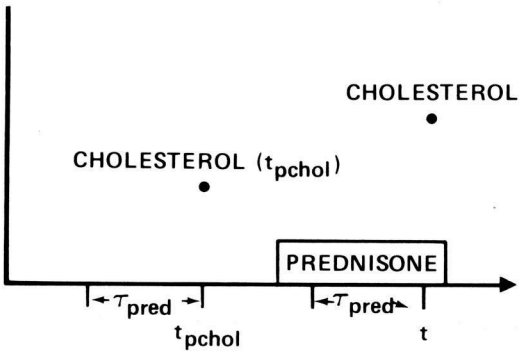

Figure 3. Time relationships in the study of the effect of prednisone on cholesterol. The regression model (equation 1 in the text) uses successive differences of cholesterol and prednisone. Each change in cholesterol level in the patient's record is associated with the preceding change in the prednisone dosage.

strength of the effect against the frequency of data collection to see if the effect is stronger in more severely ill patients, who typically have data recorded more frequently. Finally, the median and percentiles of the effect across patients may be calculated directly from the set of individual patient estimates to determine the frequency distribution of the effect in a sample of patients.

FITTING THE MODEL TO INDIVIDUAL PATIENT RECORDS

The regression model that was formulated by the Study Module and that it fitted to each patient record in the prednisone and cholesterol study is shown in EQUATION 1:

$$\Delta\text{cholesterol} = \beta_0 + \beta_1 \Delta\text{albumin} + \beta_2 \Delta\log(\text{prednisone} + 1)$$

In this equation the meaning of the variables are as follows:

$$\Delta\text{cholesterol} = \text{cholesterol}(t) - \text{cholesterol}(t_{pchol})$$

$$\Delta\text{albumin} = \text{albumin}(t - \tau_{ns}) - \text{albumin}(t_{pchol} - \tau_{ns})$$

$$\Delta\log(\text{prednisone}) = \log[\text{prednisone}(t - \tau_{pred})] - \log[\text{prednisone}(t_{pchol} - \tau_{pred})]$$

The time t_{pchol} represents the time of measurement of the previous cholesterol level, and τ_{ns} denotes the delay from the onset of the nephrotic syndrome to the establishment of a steady state for cholesterol. The time interval τ_{pred} is the analogous onset-delay for prednisone. No values are sampled during episodes of hepatitis or diabetic ketoacidosis. Figure 3 shows some of the time relationships that are used in the model. The parameter τ_{pred} is the estimated time required for cholesterol to reach a steady state after the administration of a constant daily dose of prednisone. We set τ_{pred} to 7 days based on estimates from the literature (18), although the actual time delay to steady state is probably dose-dependent and may be longer than a week (22). In most instances in the database, prednisone was begun several weeks before the cholesterol level was determined.

Autocorrelation is another potential complication of studies involving time series variables. Autocorrelation refers to a variable's being correlated with itself. For example, if a cholesterol level is very high in a given patient, it is likely that cholesterol levels over the next few weeks will also be very high. Autocorrelation may cause the regression estimates to be overestimated unless corrective measures are taken. The Study Module uses regression models based on successive differences as in equation 1. In this model, successive differences in the cholesterol levels are regressed on successive differences in the prednisones, as shown in Figure 3. Using differences usually eliminates most of the problem with autocorrelation.

COMBINING PATIENT RESULTS

When the statistical package has fitted the regression model to each patient's record, a matrix of results is produced as shown in Table 2. The table shows the results in 34 patients who had adequate data and did not have the nephrotic syndrome at any clinic visit. Although some of these patients had occasional mild proteinuria, none had a 24-hour urine protein level greater than 3 g at any time. Patients were also excluded if at any time they had a spot urine protein level greater than +3 by dipstick or a serum albumin level less than 3 g/dL.

Table 2 shows the maximum recorded dose of prednisone (pred_{max}), the range of recorded cholesterols ($\text{chol}_{min,\,max}$), and the primary rheumatologic diagnosis. The column labeled n_i shows the number of paired prednisone-to-cholesterol measurements considered for each patient. Actually, the total number is $n_i + 1$, because successive differences were analyzed. The values for pred_{max} and for $\text{chol}_{min,\,max}$ are based on these pairs. (For example, Patient 11 received 100 mg/d of prednisone at one time, but no cholesterol was recorded at an appropriate time lag.)

Table 2 also shows the estimate (b_i) of that patient's true regression coefficient (β_i) of Δcholesterol on $\Delta\log_e$(prednisone). The coefficient b_i is the principal focus of our analysis; it quantifies the magnitude of the effect of prednisone on cholesterol (assuming this is a causal relationship). Simply stated, b_i is the slope of the regression line in a plot of the ith patient's successive changes in cholesterol against the preceding successive changes in prednisone. Patient 1 has a moderately positive slope of 16.8; increases in prednisone were associated with increases in cholesterol after a time delay of at least 1 week. Patient 34 has a strongly negative slope of −42.2. Each regression analysis also yields an estimate of the standard error of b_i around the ith patient's true β_i. The statistical significance (the p value) of the regression estimate is also shown.

To test the prednisone-to-cholesterol hypothesis in the overall group of patients, some method of weighting the slopes in individual patients is clearly needed, because the sizes of the records differ substantially. Patient 1 had 17 cholesterol values recorded as opposed to Patient 34 with only 6 recorded values. Similarly, some patients took prednisone throughout a broad range of dosages as compared to some whose prednisone varied within a narrow range. Intuitively, it would seem that the most reliable estimates of the effect of prednisone would be obtained from those patients who received it many times and in

Table 3. Summary of Statistics in Two Studies on the Effects of Prednisone on Cholesterol

	Forty-nine Patients with Chronic Rheumatologic Disorders*		Thirty-four Patients without the Nepthrotic Syndrome†
	Albumin	Prednisone	Prednisone
Mean b_i	−30.8	20.1	20.2
Median b_i	−0.54	12.1	16.8
Weighted mean b_w	−14.7	19.7	18.4
Standard deviation b_w	5.35	2.45	3.57
$t_{\text{two-sided}}$	−2.75	8.1	5.15
$p_{\text{two-sided}}$	0.008	$< 10^{-6}$	10^{-5}

* Patients analyzed using the model shown in equation 1; 15 patients had the nephrotic syndrome, which resulted in a significant correlation of albumin with cholesterol in this group.

† Full data on these patients are shown in Table 2. Albumin was no longer significant in this study. The prednisone effect was highly significant in both studies.

widely differing dosages. We would like to weight those estimates most heavily.

According to statistical theory, it can be shown that the mean effect over all the patients may be obtained by weighting each patient's regression estimate by the inverse of its variance (34). That is, if the precision of each patient's estimate is measured by the inverse of its variance, then we can weight our estimates by their respective precisions. It can further be shown that the precision of each estimate b_i is proportional to *nvarx,* the variation in prednisone dosage, also shown in Table 2. The actual weighting method used by the Study Module is somewhat more complicated and involves a random effects model that accounts for the variance across patients as well as the variance within an individual record (14, 31).

Although the details involved in calculating the weights are beyond the scope of this paper, it is worth noting that the weighting accomplishes just what is dictated by clinical intuition. If each patient record contains numerous divergent values for prednisone each followed by a cholesterol measurement, then the method assigns an equal weight to each record, because each patient's regression estimate is of equal precision. On the other hand, if some patient records contain large amounts of relevant data and others contain almost none, then the weights diverge, and the patients with the most data are assigned the heaviest weights.

Results

Of the 1787 patients in the overall Stanford ARAMIS database, only 49 patients had adequate data to test the hypothesis. The remaining patients were either not given prednisone, or there was no variation in their prednisone dosage, or they had too few cholesterol measurements to be included. The results for this sample of patients are shown in Table 3. Prednisone's correlation with cholesterol was highly significant ($p < 10^{-6}$) after the effect of the nephrotic syndrome (as represented by serum albumin) was removed statistically. The overall (weighted) regression coefficient b_w was 19.7, which when substituted into equation 1 means that, on average, a dose of 30 mg/d of prednisone was associated with an increase in cholesterol of 68 mg/dL.

In Table 3, the statistics for albumin show that the nephrotic syndrome was also significantly correlated with cholesterol ($p = 0.008$), as expected (35). The relative lack of significance of albumin compared to prednisone is due to the fact that only 15 of the 49 patients had the nephrotic syndrome. In a separate study of the nephrotic patients alone it was found that the effect of the nephrotic syndrome on cholesterol was three times as large ($b_w = -48.0$) and more highly significant ($p = 0.0003$). This means that, on average, a decrease in serum albumin of 1 g/dL was associated with an increase in cholesterol of 48 mg/dL. The relatively low regression coefficient in the study with all patients ($b_w = -14.7$) was evidently due to a dilution of the true effect of the nephrotic syndrome in a study that included 34 patients without the nephrotic syndrome.

Because our primary focus was on the association of prednisone with changes in cholesterol, we did a new study in which we overrode the Study Module's decision to control the nephrotic syndrome statistically, and instead we required that it eliminate those patients from the study. This procedure resulted in a new data set of 34 patients without the nephrotic syndrome. This subset formed the basis for all our subsequent studies. These 34 patients had various rheumatologic diagnoses as shown in Table 2. There were 23 women and 11 men followed for a median of 6.8 years (quartiles at 4.5 and 9.7 years) and with a median of 40 clinic visits (quartiles at 26 and 53 visits).

These data were first analyzed using the model of equation 1, which includes the serum albumin term, to determine whether the effect of the nephrotic syndrome on cholesterol had been fully removed. In that study the overall regression coefficient for serum albumin (b_w) was equal to +11.4 with a p value of 0.051. Although the p value is in the equivocal range, it is noteworthy that the sign of the regression coefficient was now positive. That is, decreases in serum albumin were associated with decreases in cholesterol. Based on this reversal of sign and the equivocal p value, the albumin term was dropped from the model.

The final model, then, included only Δlog(prednisone) as the independent variable. The results for the individual patients were shown in Table 2, and the overall results appear in Table 3. Again, the association between prednisone and cholesterol was highly significant ($p = 10^{-5}$). The weighted mean of the effect b_w was 18.4. Further-

more, it is reassuring that the mean, the median, and weighted mean are not statistically different. That is, the distribution of the magnitudes of the effect in our sample of patient was not grossly skewed.

Interestingly, the mean effect was not significantly different from the effect of prednisone in the group of 15 patients with the nephrotic syndrome ($b_w = 22.4$; standard error (SE) = 4.76). The effects of prednisone and the nephrotic syndrome on cholesterol appear to be linearly additive. The steroid effect occurs in the absence of the nephrotic syndrome, and, in its presence, can result in extremely high serum cholesterol levels. Values in excess of 350 mg/dL were frequently seen in this setting in patients who had no intrinsic hyperlipoproteinemia.

UNIFORMITY OF THE EFFECT ACROSS PATIENTS

Even though the overall p value is highly significant, the existence of a true prednisone-to-cholesterol association becomes far more convincing when the data on the individual patients are examined. Table 2 shows the results of the regression analyses for individual non-nephrotic patients. If those individual effects are plotted on a scatter diagram as in Figure 4 against their respective p values (Table 2), we see that far more patients than expected have statistically significant positive coefficients. (In both Figures 4 and 5 each data point corresponds to one patient.) If there was no effect, we would expect a symmetric distribution of the points about the horizontal line $b_i = 0$.

A scatter plot of the regression coefficients against $n_i \text{var}(x_i)$ (Table 2) as shown in Figure 5 is equally persuasive. The plot shows that the range of the b_i s tends to converge toward a value of approximately 20 as more information is gathered on the patients, either due to an increase in the amount of data n_i or due to an increase in the range of values throughout which prednisone is administered $\text{var}(x_i)$. It also appears that no *surveillance bias* is present, because the mean of the values for patients with little information (left-hand side of the abscissa) is about the same as the mean of the values for patients with much information (right-hand side of the abscissa).

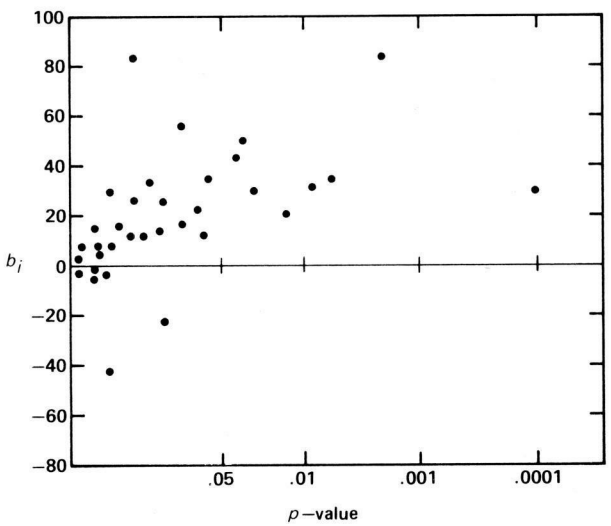

Figure 4. Magnitude of the effect versus statistical significance. Each point represents data from 1 of the 34 patients in Table 2. To the right of $p = 0.05$ we see that all patients had positive regression coefficients. That is, an increase in cholesterol was associated with an increase in prednisone at least 7 days earlier.

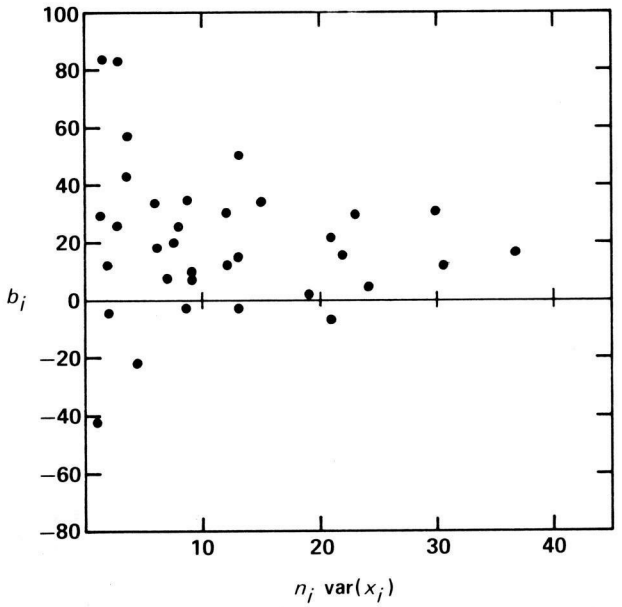

Figure 5. Magnitude of the effect versus amount of information as given for the 34 patients in Table 2. As the amount of information in each record increases, as measured by the total dispersion of the prednisone dosages, the magnitudes of the effects of prednisone on cholesterol tend to converge.

MAGNITUDE OF THE EFFECT

The fitted regression model for cholesterol on prednisone in the 34 patients without the nephrotic syndrome shown in Table 2 is shown in EQUATION 2:

$$\Delta\text{cholesterol} = 18.4 \log_e (\text{prednisone} + 1)$$

Substituting into this equation, we find, for example, that continued prednisone dosages of 20, 30, and 60 mg/d would be associated on average with increases in cholesterol of 56, 63, and 76 mg/dL after a delay of at least 1 week. (The actual delay to steady state may be several weeks.)

Previous studies of the steroid-to-cholesterol effect have mainly used fixed-dose protocols; hence, dose/response relationships have not been explicitly quantitated. Nonetheless, in a study of renal transplant patients, Curtis and associates (19) saw an increase in cholesterol from 177 mg/dL before transplant to 241 mg/dL 4 months after transplant in patients on 30 mg/d of prednisone. (The cholesterol level gradually dropped to baseline if patients were then switched to alternate-day steroids.) Ibels and coworkers (22) reported an average increase in cholesterol of 40 mg/dL in patients after transplant on an average of 15 mg/d. Stern and associates (23) reported a mean rise in cholesterol of 88 mg/dL in 12 patients with various rheumatologic conditions on a wide range of prednisone dosages (median, 40 mg/d; range, 15 to 60 mg/d). Zimmerman and colleagues (18), studying 12 patients who were free of metabolic disor-

Table 4. Frequency Distribution of the Effect of Prednisone on Cholesterol*

Range of Cholesterol	Patients
mg/dL	*%*
100-150	1
150-195	2
195-210	0
210-225	12
225-230	0
230-235	0
235-250	9
250-280	18
280-360	50
360-700	8

* Distribution of cholesterol (baseline, 230 mg/dL) after at least 1 week of daily oral prednisone, 30 mg/d.

ders, noted an average increase in cholesterol of 30 mg/dL over baseline after 30 days on a tapering prednisone regimen with an asymptote of 20 mg/d. It appears that values from previous reports accord well with the fitted model shown as equation 2.

FREQUENCY DISTRIBUTION OF THE EFFECT

With a fitted model for each patient, it is possible to calculate the frequency distribution of the effect across patients. For example, if we arbitrarily assume a baseline serum cholesterol level of 230 mg/dL and take a reference level of prednisone at 30 mg/d, we can then calculate the expected level of cholesterol for each patient by substituting each of the 34 b_i s from Table 2 into equation 2. The resulting set of Δcholesterol values, weighted by their corresponding w_i s, may then be displayed as a frequency distribution as in Table 4. As shown in the table, after 7 days 85% of patients had some increase in their serum cholesterol and 58% had an increase of at least 50 mg/dL over their baseline. We suspect that those patients whose cholesterol values appeared to drop had covariates that were not adjusted for by the study design. (Because this frequency distribution is based on a small sample of patients, it should be taken only as a rough guide.)

CONTROL FOR THE UNDERLYING DISEASE

The possibility must be addressed that the strong, time-lagged association that was seen could have been due to the underlying diseases for which prednisone was administered. One may conjecture that in patients with systemic lupus erythematosus, a flare of the condition caused the physician to administer prednisone and some weeks later also caused an increase in the serum cholesterol level. If this sequence of events repeatedly and uniformly occurred in many patients over time, the statistical results would be identical to what was seen.

To check the possibility that the prednisone-to-cholesterol association may have been due to the patients' underlying disease, we regressed cholesterol on various clinical variables that may, under certain circumstances, be markers of or proxies for lupus erythematosus activity. These variables included C3 (the third component of complement), erythrocyte sedimentation rate, anti-DNA antibody titer, and blood urea nitrogen. Each variable was substituted for albumin in equation 1 and the resulting model was tested on the 34 non-nephrotic patients shown in Table 2. In every case, the proxy variable was not significantly correlated with cholesterol, whereas the statistical significance of prednisone was not diminished. In general, if a confounding variable is only occasionally associated with prednisone and cholesterol, or its time lag is inconsistent, or it does not occur in most records, then it will not be statistically significant. Nonetheless, the association may still be explained by some factor associated with the underlying disease that we have failed to record.

In an additional effort to determine whether the association was due to systemic lupus erythematosus or one of its sequelae, we divided the set of 34 patients shown in Table 2 into two groups: the 17 patients with systemic lupus erythematosus and the 17 patients with other rheumatologic diagnoses. In the patients with systemic lupus erythematosus the value of b_w was 20.1 (SE = 4.5). In the patients without systemic lupus erythematosus, the value of b_w was 16.9 (SE = 5.78). The difference between the two values is not statistically significant. Because the association appears to be equal in both groups of patients, it seems reasonable to assume that it may appear with equal magnitude in nonrheumatologic settings in which constant daily glucocorticoids are administered: for example, cancer chemotherapy, chronic asthma, and inflammatory dermatoses.

Discussion

In a recent review, Barnett (1) summarized the reasons for the rapid growth in the development of computerized medical record systems. Expanded use of medical records comprises the first set of reasons: for communication among health care providers and for meeting the reporting requirements of insurers, governmental agencies, and quality assurance programs. Further promoting the rapid development of these systems have been advantages shared by all computer data processing systems: concurrent accessibility of records at multiple sites, ease of reformatting data for multiple objectives, automated error checking, legibility, and elimination of redundant data entry. For these reasons and because of the continually increasing performance-to-price ratio of computing equipment, it is likely that computerized medical record systems will soon be ubiquitous in hospitals and clinics.

We have shown the feasibility of using routinely collected data for the performance of biomedical research. Because of the immense amount of data that will be rapidly accessible from ambulatory systems, there are exciting prospects for using these data to do clinical studies rapidly and perhaps automatically, discovering new medical hypotheses with some data sets and confirming the results on others.

Although the prospect of discovering and studying medical hypotheses using routine clinical data is tantalizing, the use of routinely collected data for the performance of clinical research carries a number of potential hazards (26, 36, 37). These hazards include problems due to inadequate and improperly sampled data, biases due to non-random selection of patients, and failure to

adjust for recorded and unrecorded confounding variables.

Databases on ambulatory patients generally contain sporadically collected data, and patient follow-up may be haphazard. Furthermore, all clinical variables will not be relevant for every patient. For these reasons, the amount of data relevant to a given study may be a small fraction of the database. In the present study only 49 patients from a database of 1787 had data adequate for our study design, and of these, 15 patients had the nephrotic syndrome and were subsequently eliminated. Under these circumstances, appropriate methods should be used to increase confidence in the ability to generalize the result.

In the current study, a measure of generality was obtained by doing, in effect, a separate study of each of the 49 eligible patients. Using the two-stage regression method, a separate measure of the prednisone-to-cholesterol association was obtained for each patient. Figures 4 and 5 showed that these individual associations were consistent in direction and magnitude, corroborating the existence of a uniform underlying effect.

Even though a separate regression analysis was done on each patient, the results could still have been explained by a third variable, such as the underlying disease for which prednisone was administered. This explanation is impossible to rule out, but the prednisone-to-cholesterol association remained intact despite controlling for the underlying diseases and despite separate analyses of patients with lupus erythematosus and with other rheumatologic ailments.

Because of the statistical complexity and the immense computational requirements of studies like the present one, it is clear that strong computational assistance is required. With the RX Study Module we incorporated as much of this expertise as possible into the software itself. The medical and statistical knowledge bases assist with the selection of confounding variables and with the construction of the statistical model. The program further assists with the elaboration of eligibility criteria, the accession of an appropriate data set from time-oriented patient records, and the execution of the study design.

What, then, will be the role of clinical database studies in the formation of biomedical knowledge? Without randomizing patients to treatments, it is never possible to control for unknown or unrecorded influences. Therefore, it is never possible to claim that a correlation is due to a causal relationship when using non-randomized, nonprotocol data. Furthermore, it is even unrealistic to expect to control fully for known covariates. There is no doubt that a well-conducted randomized trial will generally produce a more convincing result than will a nonrandomized study.

Nonetheless, results based on routine clinical databases are useful for generating new hypotheses and for strengthening belief in hypotheses that have been partially confirmed. That is, studies like the present one are useful for moving a hypothesis along the spectrum from preliminary hunch to established fact. Because nonprotocol, routine health care databases can be expected to become vastly more widespread, the availability of methods for rapidly and reliably exploring them can be expected to yield valuable biomedical findings.

ACKNOWLEDGMENTS: The authors thank Robert Guy Kraines and Byron William Brown for statistical analysis and adaptation of the two-stage regression method; Michael Walker for assistance with program maintenance and performance of the data analyses; Gio Wiederhold for assistance with project management; Jack Kahoun for assistance with the medical literature review; Ronald Kaplan and Beau Shiel, Xerox Palo Alto Research Center, for their assistance with IDL; James Fries and Dennis McShane for allowing us to analyze the Stanford Immunology database; to ARAMIS group members including Katherine Williams, Alison Harlow, and James Standish for assistance in transferring the database to SUMEX; and Peter Rudd, Mark Musen, Michael Kahn, Stephen Fortmann, and Allen Cooper.

The RX Study Module is in the public domain. Because it is a research prototype, capability in INTERLISP programming is required to install it at other laboratories. Interested researchers with the requisite computing facilities to maintain the system independently may contact the authors.

Grant support: In part by grant LM-04334 from the National Library of Medicine and grant IST-8317858 from the National Science Foundation. Previous funding included grant HS-04389 from the National Center for Health Services Research, grant LM-03370 from the National Library of Medicine, and by the Pharmaceutical Manufacturers Association Foundation. The ARAMIS database is sponsored by grants AM-21393 and HS-03802 from the National Institutes of Health; and SUMEX-AIM by grant RR-00785 from the Biotechnology Resources Program, National Institutes of Health.

▶ Requests for reprints should be addressed to Robert L. Blum, M.D., Ph.D.; Department of Computer Science, Margaret Jacks Hall, Stanford University, Stanford, CA 94305.

References

1. BARNETT GO. The application of computer-based medical-record systems in ambulatory practice. *N Engl J Med.* 1984;**310:**1643-50.
2. BARNETT GO, JUSTINCE NS, SOMAND ME, et al. COSTAR—a computer-based medical information system for ambulatory care. *Proc. IEEE.* 1979;**67:**1226-37.
3. HAMMOND WE. Ambulatory Care Systems: TMR In: *Proceedings of the Sixth Annual Symposium on Computer Applications in Medical Care.* Washington, D.C.: IEEE Computer Society Press; 1982:75-97.
4. MCDONALD CJ, BLEVINS L, GLAZENER T, HAAS J, LEMMON L, MEEKS-JOHNSON J. Data Base Management, Feedback Control and the Regenstrief Medical Record. In: *Proceedings of the Sixth Annual Symposium on Computer Applications in Medical Care.* Washington, D.C.: IEEE Computer Society Press; 1982:52-60.
5. MCDONALD CJ, HUI SL, SMITH DM, et al. Reminders to physicians from an introspective computer medical record: a two-year randomized trial. *Ann Intern Med.* 1984;**100:**130-8.
6. SIMBORG DW, WHITING-O'KEEFE QE. Summary time oriented record (STOR)—a progress report. In: *Proceedings of the Fifth Annual Symposium on Computer Applications in Medical Care.* Washington, D.C.: IEEE Computer Society Press; 1981:100-10.
7. FRIES JF, MCSHANE DJ. ARAMIS: a national chronic disease data bank system. In: *Proceedings of the Third Annual Symposium on Computer Applications in Medical Care.* Washington, D.C.: IEEE Computer Society Press; 1979:798-801.
8. LAYARD MW, MCSHANE DJ. Applications of MEDLOG, a microcomputer-based system for time-oriented clinical data. In: *Proceedings of the Seventh Annual Symposium on Computer Applications in Medical Care.* Washington, D.C.: IEEE Computer Society Press; 1983:731-4.
9. BOLT BERANEK AND NEWMAN, INC. *CLINFO: An Introduction to the CLINFO Data Management and Analysis System.* Cambridge, Massachusetts: Bolt Beranek and Newman, Inc.; 1981.
10. BLUM RL. Discovery, confirmation, and incorporation of causal relationships from a large time-oriented database: the RX project. *Comput Biomed Res.* 1982;**15:**164-87.
11. BLUM RL. *Discovery and Representation of Causal Relationships from a Large Time-oriented Clinical Database: The RX Project.* New York: Springer-Verlag; 1982.
12. BLUM RL. Displaying clinical data from a time-oriented database. *Comput Biol Med.* 1981;**11:**197-210.
13. BLUM RL. *Modeling and encoding clinical causal relationships in a medical knowledge base.* In: *Proceedings of the Seventh Annual Symposium on Computer Applications in Medical Care.* Washington, D.C.: IEEE Computer Society Press; 1983:837-41.
14. BLUM RL. *Two-Stage Regression: Application to a Time-Oriented Clinical Database; Knowledge Systems Laboratory Report KSL-85-43.* Stanford, California: Computer Science Department, Stanford University; 1985.
15. ADLERSBERG D, et al. Adrenal cortex and lipid metabolism: effects of

cortisone and adrenocorticotrophin (ACTH) on serum lipids in man. *Proc Soc Exp Biol Med.* 1950;**74:**877-9.
16. Moran TJ. Cortisone-induced alterations in lipid metabolism. *Arch Pathol.* 1962;**73:**52-64.
17. Reaven EP, Kolterman OG, Reaven GM. Ultrastructural and physiological evidence for corticosteroid-induced alterations in hepatic production of very low density lipoprotein particles. *J Lipid Res.* 1974;**15:**74-83.
18. Zimmerman J, Fainaru M, Eisenberg S. The effects of prednisone therapy on plasma lipoproteins and apolipoproteins: a prospective study. *Metabolism.* 1984;**33:**521-6.
19. Curtis JJ, Galla JH, Woodford SY, Lucas BA, Luke RG. Effects of alternate-day prednisone on plasma lipids in renal transplant recipients. *Kidney Int.* 1982;**22:**42-7.
20. Cattran DC, Steiner G, Wilson MD, Fenton MD. Hyperlipidemia after renal transplantation: natural history and pathophysiology. *Ann Intern Med.* 1979;**91:**554-9.
21. Ibels LS, Stewart JH, Mahony JF, Sheil AG. Deaths from occlusive arterial disease in renal allograft recipients. *Br Med J.* 1974;**3:**552-4.
22. Ibels LS, Alfrey AC, Weil R III. Hyperlipidemia in adult, pediatric and diabetic renal transplant recipients. *Am J Med.* 1978;**64:**634-42.
23. Stern MP, Kolterman OG, Fries JF, McDevitt HO, Reaven GM. Adrenocortical steroid treatment of rheumatic diseases: effects on lipid metabolism. *Arch Intern Med.* 1973;**132:**97-101.
24. McShane DJ, Harlow A, Kraines RG, Fries JF. TOD: a software system for the ARAMIS data bank. *Computer.* 1979;**12:**34-40.
25. Hess EV. A uniform database for rheumatic diseases: prepared by the Computer Committee of the American Rheumatism Association. *Arthritis Rheum.* 1976;**19:**645-8.
26. Feinstein AR. An additional basic science for clinical medicine: III. the challenges of comparison and measurement. *Ann Intern Med.* 1983;**99:**705-12.
27. XEROX Corporation. *Interlisp Reference Manual.* Pasadena; XEROX Special Information Systems; 1983.
28. Walker MG, Blum RL. A Lisp Tutorial. *MD Computing.* 1985;**2:**56-63.
29. Kaplan RM, Sheil BA, Smith ER. *The interactive Data-Analysis Language Reference Manual.* Palo Alto: Xerox Palo Alto Research Center; 1978.
30. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters: I. Michaelis-Menten model: routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm.* 1980;**8:**553-71.
31. Hyde J. Determining an Average Slope. In: Miller RG, Efron B, Brown BW, Moses LE, eds. *Biostatistics Casebook.* New York: John Wiley & Sons; 1980:171.
32. Draper NR, Smith H. *Applied Regression Analysis.* New York: John Wiley & Sons; 1981.
33. Brown BW, Hollander M. *Statistics: A Biomedical Introduction.* New York: John Wiley & Sons; 1977.
34. Norwood TE Jr, Hinkelmann K. Estimating the common mean of several normal populations. *Ann Stat.* 1977;**5:**1047-50.
35. Baxter JH, Goodman HC, Havel RJ. Serum lipid and lipoprotein alterations in nephrosis. *J Clin Invest.* 1960;**39:**455.
36. Byar DP. Why databases should not replace randomized clinical trials. *Biometrics.* 1980;**36:**337-42.
37. Dambrosia JM, Ellenberg JH. Statistical considerations for a medical data base. *Biometrics.* 1980;**36:**323-32.