Design and Analysis Considerations in Evaluating the Chemiluminescence Response of Mouse Spleen Cells

DAVID M. GILTINAN,1* THOMAS P. CAPIZZI,1 GEORGE K. ABRUZZO,2 AND ROBERT A. FROMTLING2

Departments of Biometrics Research1 and Basic Microbiology,2 Merck Sharp & Dohme Research Laboratories, Rahway, New Jersey 07065

Received 24 July 1985/Accepted 6 December 1985

Chemiluminescence is the result of the respiratory burst generated by phagocytic cells after stimulation by antigen. The measurement of chemiluminescence represents a sensitive means for detecting antigenic stimulation and immune cell function. Although the kinetics of chemiluminescence reactions have been described, appropriate statistical methods for the evaluation of data from chemiluminescence assays have not been reported. Based on examination of data from several studies in which the chemiluminescence response of spleen cells was investigated after stimulation with the particulate antigen zymosan, recommendations are made for the design and statistical evaluation of such studies. Three parameters were used in assessing the chemiluminescence response; peak intensity of the emitted light, time to peak, and the area under the intensity-time curve. The data indicated that peak intensity alone provides an adequate characterization of the chemiluminescence response. Since percent change in response upon treatment is of interest, analysis on the log scale is appropriate, and the statistical procedure of choice in evaluating data of this type is a trend analysis. The need for a balanced allocation of treatments to avoid potential bias is demonstrated. The methods proposed are illustrated with data from two studies in which the effect of preincubation with low concentrations of ketoconazole, an antifungal agent, on the chemiluminescence response of BALB/cBY spleen cells was examined.

Normal function of phagocytic cells is critical for proper host defense against infectious agents. In particular, normal immune cell function is essential for successful defense against fungal agents of disease, since many antifungal agents are fungistatic and not fungicidal. An important activity of phagocytic cells is their ability to respond to appropriate stimuli by activation of the respiratory burst which comprises increased oxygen uptake; production of superoxide anions, hydrogen peroxide, hydroxyl radicals, and singlet oxygen; and the stimulation of the hexose monophosphate shunt. These active oxygen compounds participate in the destruction of microorganisms (1).

In the resting state, phagocytic cells consume little oxygen. The respiratory burst begins when, in response to a suitable stimulus, an oxidase which is dormant in the resting cell is converted to an active state. This oxidase catalyzes the one-electron reduction of oxygen to superoxide anion (1, 2). The stimulus may be a fungal or bacterial cell, a particulate antigen such as zymosan, or soluble agents such as plant lectins. This phenomenon has only been investigated extensively during the last 20 years. One result of these investigations has been the discovery that the respiratory burst is a complex metabolic series of pathways which serves to produce the highly reactive antimicrobial oxidants by the partial reduction of oxygen (1, 2).

A sensitive way to measure the respiratory burst of phagocytic cells is based on the finding that stimulated phagocytes generate photons of light which can be detected as chemical luminescence or chemiluminescence (CL) (3, 6, 8, 9, 11, 14). This light emission has been shown to depend both upon superoxide and myeloperoxidase-catalyzed reactions and is a measure of phagocytic as well as metabolic activity in the stimulated cells (12).

Several antifungal agents have been determined to have an inhibitory effect on the respiratory burst of phagocytic cells, as measured by CL response (G. K. Abruzzo, D. Giltinan, T. Capizzi, and R. A. Fromling, submitted for publication). A series of experiments was conducted in which a number of antifungal drugs were assayed to determine their effect on the generation of CL in mouse spleen cells in vitro. In the present paper, several issues pertaining to the interpretation of data from such experiments are discussed. Questions addressed include meaningful characterization of the CL response, choice of appropriate statistical methodology in analyzing the data, sensitivity of the assay, and design considerations in the planning of such experiments. The issues involved are illustrated with data from two studies performed to investigate the effect of ketoconazole, an antifungal agent, on the CL response of mouse spleen cells in vitro. Recommendations are made for conducting future studies.

MATERIALS AND METHODS

Chemiluminescence assay. Spleens from 6- to 8-week-old BALB/cBY female mice (Jackson Laboratories, Bar Harbor, Maine) were aseptically removed, gently massaged between two sterile ground glass slides, and washed three times in Hanks balanced salt solution. Washed spleen cells then were suspended in Dulbecco modified Eagle medium without phenol red and supplemented with 10% fetal bovine serum. Cells were counted by using a hemacytometer, adjusted to a concentration of 10⁷ cells per ml and incubated at 37°C. Viability was determined by using trypan blue exclusion. Chemiluminescence was measured in a Biolumat LB9505

* Corresponding author.
six-channel photoanalyzer (Berthold, Wildbad, Federal Republic of Germany [3]. This apparatus permits the measurement of six samples simultaneously. Before each experimental series, the Biolumat was calibrated with a standardized light source. Cells (0.47 µl containing 10^7 cells per ml) were placed in small plastic test tubes, and each tube was inserted into one of the six channels. The compound assayed (ketoconazole) was solubilized in dimethyl sulfoxide (DMSO). A preliminary study indicated that concentrations of DMSO up to and including 0.5% did not have a significant impact on the CL response. Accordingly, a DMSO concentration of 0.5% was used for solubilization purposes.

Spleen cells were preincubated with various concentrations of the test drug (in a volume of 10 µl) for 30 min. After this preincubation, 10 µl of zymosan (50 mg/ml) was injected into the assay tube to induce chemiluminescence. Luminol (5- amino-2,3-dihydro-1,4-phthalazinedione) also was used, in a standard concentration (2.0 µg/ml), to enhance the levels of chemiluminescence observed. During each experimental assay, the CL response (intensity of emitted light in counts per minute) was measured for each of the six treated cell groups every 20 s for 30 min after stimulation with zymosan. Measurements from the Biolumat were transmitted to an Apple IIE computer and Epson printer for purposes of data analysis and printing.

Statistical methods. For each assay, the six treatments in a study corresponded to a control (preincubation with 0.5% DMSO only) and five graded doses of the test compound. In a typical day of experimentation, four to six assays were completed. Preliminary results indicated that systematic differences existed among the measurement channels of the Biolumat. Accordingly, it was considered desirable to balance the allocation of treatments across channels of the machine as much as possible to avoid potential bias in the results. In the usual case, in which six experimental assays were performed in a day, complete balance was achieved by implementing a six by six Latin square design (5) in which each treatment appeared exactly once in each measurement channel of the Biolumat and once in each of the six experimental assays.

The Biolumat generated a considerable amount of data; measurements of the intensity of emitted light every 20 s for 30 min for each of the six measurement channels. This information was summarized by the Apple IIE in the form of a graph of intensity against time for each channel (Fig. 1). The peak intensity reached, the time at which this occurred and the area under the intensity-time curve over the 30-min period were also recorded by the computer. For the purposes of data analysis, characterization of the CL response in terms of a small number of parameters is desirable. In the evaluation of similar data, Welch et al. (14) reported that "the patterns of CL responses were reviewed for initial slopes, peaks, slopes of declining response, duration of response, and areas under the curve." We chose initially to work with three response variables; peak intensity (PEAK), time to peak (TIME), and area under the intensity-time curve (AREA). For similar data occurring in pharmacokinetics, these three parameters are important; this choice also had the virtue of practical convenience, these three quantities being among the summary information delivered by the computer. The variables peak intensity and area under the curve were evaluated on the log-transformed scale since it was felt that the percent change in these variables on treatment, as opposed to the absolute change, was of interest.

In assessing the effect on the CL response of the drug being tested, standard variance analysis techniques were used. The main focus of such an analysis is to test for the existence of systematic differences among the treatment mean responses. Consistency of treatment effects over several days of experimentation also may be investigated in the analysis of variance framework. Broadly speaking, this amounts to checking whether differences among response means remain the same from day to day, within the limits of random variation. In statistical terms this corresponds to testing for an interaction between the factors treatment and day.

Assuming that the overall analysis of variance indicates that significant differences among treatment mean responses exist, one will usually want to obtain a more detailed picture of the way in which treatment affects the response. This was done in two ways. Each dose tested was compared with the control by using the multiple comparison procedure of Dunnett (7) to account for the multiplicity of testing. In addition, since the six treatments typically corresponded to a control and five graded doses of the drug being tested, it was of interest to test for progressiveness of response with increasing concentrations of the drug. This was done by implementing the sequential test procedure of Tukey et al. (13). A description of this procedure, written for a nonstatistical audience, may be found in Capizzi et al. (4).

RESULTS

The general conclusions and recommendations concerning methodology reported here are based on our experience with approximately 20 studies. For purposes of exposition, results are illustrated with data from two specific experiments in which the effect of ketoconazole, an antifungal agent, on the CL response of mouse spleen cells in vitro was investigated.

Choice of response variable. In working with the three response variables PEAK, AREA, and TIME, our experience was that the first of these provided the best characterization of the CL response. This is in agreement with the findings of Welch et al. (14). We do recommend, however, that PEAK be analyzed on the log-transformed scale, since
treatment usually effected consistent changes in the magnitude of this response relative to control, rather than in the absolute size. The log-transformed variables LPEAK and LAREA were typically very highly correlated (in the studies considered, the sample correlation coefficient between the two ranged from 0.92 to 0.99) so that little extra information would be gained by analyzing LAREA as well as LPEAK.

The third variable considered, TIME, proved less useful as a characterization of the CL response. Many of the compounds assayed did not have a consistent effect on this variable from one experiment to the next; an agent which resulted in a statistically significant delay in the time to peak on one day might appear to cause an acceleration in subsequent experiments. This was explained in part by the relatively low variation in the measurement of TIME; because of this, differences in means of the order of a minute were often declared statistically significant. It is not clear, however, that such differences should be considered of practical importance, particularly in view of the fact that the starting time of a run is well-defined only to within about 30 s. Thus, TIME to peak intensity may not be a meaningful parameter for assessing the effect of treatment on the CL response.

Analysis of variance. The results of an analysis of variance performed on data from two studies investigating the effect of ketoconazole on the CL response of mouse spleen cells are listed in Table 1. In these studies, treatments corresponded to a control (preincubation with DMSO alone) and five concentrations of ketoconazole (0.08, 0.16, 0.31, 0.63, and 1.25 μg/ml) in DMSO. Experimentation was carried out over 2 days, with six experimental assays being performed on each day. Different batches of spleen cells were used on both days. Results for these two studies were representative of our experience with data of this nature and provide a useful frame of reference in illustrating the general issues that arise. Inspection of the analysis of variance table reveals a number of interesting features.

The treatment by day interaction was negligible for these data ($P = 0.823$). This was reassuring, since a significant interaction would mean that treatment had a different effect on the variable LPEAK on the experimentation of the 2 days. The fact that this interaction was not significant, (neither in these, nor the other studies considered) indicates that the effect of the antifungal agent tested was consistent over time and over the different cell batches used in the experimentation.

The amount of variation in response from assay to assay, as measured by the mean square for assay in Table 1, was considerable. This was not too surprising, since the average background response for assays later in the day tended to be systematically lower than for early assays. Similar remarks apply to channel-to-channel variability. This is measured by the mean square for channel in Table 1; the size of this quantity indicates that there was considerable systematic variation among the mean responses observed for the six channels of the Biolumat. Table 2 (response means by channel) shows the nature of this systematic effect. The numbers presented there are reasonably typical of our experience in these studies: on average, channel 1 always gave considerably higher readings than did channel 5 or channel 6, no matter what compound was being tested.

Evaluation of treatment effects. The absence of a treatment by day interaction indicated that it was reasonable to summarize the effect of the test compound by presenting the average response pooled over the experiments from both days for each dose level (Table 3). For the ketoconazole data, the overall $F$ test for differences among treatment means gave a $P$ value of 0.13. The interpretation is that there was some, though not overwhelming, evidence for differences among the treatment means. Since the $F$ test is nonspecific in nature and, thus, not very sensitive, more directed tests, which take the specific nature of the treat-

| TABLE 3. Ketoconazole chemiluminescence data; response means by dose level of ketoconazole |
|---------------------------------|---------------------------------|
| Treatment                      | Peak intensity$^*$              |
| Control                        | 6,323                           |
| Ketoconazole (μg/ml)           |                                 |
| 0.08                           | 6,529                           |
| 0.16                           | 6,204                           |
| 0.31                           | 6,248                           |
| 0.63                           | 6,057                           |
| 1.25                           | 6,143                           |

$^*$ Average peak intensity in counts per minute, backtransformed from log scale. By the Dunnett multiple comparisons procedure, testing at the 5% level of significance, none of the test results were significantly different from that of the control (0.5% DMSO in Dulbecco modified Eagle medium). The geometric standard deviation, a measure of relative error defined as the antilog of the standard deviation on the log scale, was 1.069 for this analysis. Twelve assays were performed for every treatment.
ments into account, are also appropriate. The results of the Dunnett test, comparing the individual dose means to that of the control are shown in Table 3. For these data, none of the individual doses of ketoconazole resulted in a mean response which was significantly different from that for control, by using the Dunnett test. The trend test of Tukey et al. (13), however, did indicate the existence of a statistically significant \( (P = 0.029) \) negative trend in response with increasing concentration of ketoconazole. Inspection of the treatment means indicates that this trend may be attributed to the lower values of the average peak intensity at the 0.63- and 1.25-\( \mu \)g/ml doses of ketoconazole. The significant negative trend persisted upon deletion of the highest dose from the analysis \((P = 0.031)\). However, deletion of results for the two highest doses from the analysis eliminated the evidence for a trend \((P = 0.324)\).

**DISCUSSION**

The results of the analyses indicate a number of general points worth noting. (i) The lack of a treatment by day interaction indicates a consistent effect of treatment over time and over the different cell batches used in experimentation. This reproducibility of treatment effects was evident when peak intensity was assessed on the log scale. The implication is that a particular treatment resulted in a fairly consistent percent change in response. While the background response level did not vary much for the 2 days of experimentation reported here, in general, the use of different cell batches resulted in quite disparate levels of background response. In this situation, percent change rather than absolute change in response is the pertinent measure of the effect of treatment. Accordingly, analysis on the log scale is appropriate. (ii) The considerable systematic variation in response from assay to assay and from channel to channel made it imperative to use a design in these studies in which the allocation of treatments was balanced across assays and across channels. That is, whenever possible, each treatment should appear the same number of times on each assay and in each channel. In our experimental framework of six treatments and six assays per day, a Latin square design provided a convenient means of achieving this balance; other balanced or partially balanced designs are possible and may be more convenient in different experimental settings. The use of a balanced design has two advantages. First, it allows the systematic assay-to-assay and channel-to-channel variability to be subtracted out of the residual (error) variation in the analysis. This is important, since the test for treatment differences essentially compares the variation among treatment mean responses to the residual, or random, variation. Failure to subtract out existing systematic variation would result in severe overestimation of the amount of random noise in the data, thereby making any treatment effects harder to detect. Thus, a balanced design allows a more sensitive test for treatment differences. The second advantage of balance is that it helps to avoid bias in the results. Given the magnitude of the observed systematic channel and assay effects, failure to implement a balanced design in these experiments would have serious consequences for the valid estimation of treatment effects. For example, if a single treatment were consistently assigned to the same measurement channel of the Biolumat, any observed lowering in the response could not be attributed solely to the effect of that treatment; the possibility of it being due to that particular channel yielding consistently lower readings could not be discounted. Similarly, running all replicates of a particular treatment in the same experimental assay would be inadvisable; to do so would confuse the effect of that treatment with the systematic assay effect. Balancing over assays and channels helps to avoid such bias, as well as enhancing precision, when assessing treatment effects. (iii) The fact that the trend analysis gave a significant result for the ketoconazole data, while the Dunnett test did not, reflects the greater sensitivity of the trend test. This increased sensitivity is a quite general phenomenon (4, 10, 13) and stems from the fact that the trend test addresses a more specific question than does the Dunnett procedure, and the trend test uses the data in a more directed and sensitive way to answer that question. In situations in which several doses of a drug are being tested, we strongly recommend that a trend analysis be performed, as it is a more powerful means of addressing the question of interest. For instance, the significant negative trend observed in the ketoconazole data indicates that, whereas comparison of each of the individual concentrations tested in these experiments with control were not statistically significant, testing at higher concentrations might well result in a more dramatic, and significant, inhibition of the CL response. Subsequent experiments with higher concentrations of ketoconazole have demonstrated that this was indeed the case (Abruzzo et al., submitted).

In conclusion, if a reduction in the CL response is to be used as an indicator of possible inhibition of the immune response, then the use of a statistically sound experimental design, as well as an appropriate analysis of the resulting data, is an absolute necessity. We believe that the methods proposed here provide a useful approach to the design and evaluation of studies to investigate the CL response.

**LITERATURE CITED**

11. Müller, S., S. Falkenberg, T. M. Ernst, U. Thalmann, M. Ruhl,

