COMMENTARY

Methodological Issues in Behavioral Immunology Research with Humans1

JANICE K. KIECOLT-GLASER2
Department of Psychiatry, The Ohio State University College of Medicine, Columbus, Ohio 43210

AND

RONALD GLASER
Department of Medical Microbiology and Immunology and Comprehensive Cancer Center, The Ohio State University College of Medicine, Columbus, Ohio 43210

This paper summarizes important methodological issues that are particularly relevant for behavioral immunology research with humans. The assessment of such salient parameters as nutrition, drug/alcohol use, physical activity, and health are discussed. In addition, a number of logistical issues are addressed.

This paper summarizes important methodological issues that are particularly relevant for behavioral immunology research with humans. Other experimental issues that are not unique to this area will not be discussed here, e.g., design and statistical issues.

The basis for the discussion that follows is the assumption that distressing psychological responses may be a common denominator through which psychologically related variables have an impact on the immune system. This assumption is based on data from such diverse groups as bereaved spouses, separated and divorced men and women, psychiatric patients with a major depression diagnosis, and family caregivers of Alzheimer's disease victims (Kiecolt-Glaser, Glaser, Dyer, Shuttleworth, Ogrocki, & Speicher, 1987b; Kiecolt-Glaser, Fisher, Ogrocki, Stout, Speicher, & Glaser, 1987a; Kiecolt-Glaser, Kennedy, Malkoff, Fisher, Speicher, & Glaser, 1988; Schleifer, Keller, Camerino, Thornton, & Stein, 1983; Stein, Keller, & Schleifer, 1985); these and related studies have shown that subjects from such groups are more distressed and show relatively poorer immune function than their well-matched community counterparts. Similarly, prospective longitudinal studies have suggested that there are downward alterations in immune function in medical students during examinations, compared to similar measures collected 1 month previously when subjects were less distressed (Glaser, Kiecolt-Glaser, Speicher, & Holliday, 1985a;

1 Work on this paper was supported in part by Grants ROI MH42096 and RO1 MH40787 from the National Institute of Mental Health.

2 Address correspondence to Janice K. Kiecolt-Glaser, Ph.D., Department of Psychiatry, 473 W. 12th Avenue, Columbus, OH 43210.

It is commonly assumed that these immunological alterations are at least in part a consequence of distress-related changes in endocrine function (Ader, 1981). However, the literature suggests a range of psychological responses in each of the groups noted above, as well as differences between populations. Bereaved individuals, for example, have been reported to have depressed mood, major depressive disorder, or anxiety states or disorders. The psychological "distress" response(s) may vary, and, in turn, the neurobiological concomitants may vary and have differential effects on the immune system.

This conceptualization has clear relevance for certain methodological issues, since distressed individuals tend to behave in ways that may both potentiate distress and concurrently affect both their immunological functioning and their health. For example, it is well known that more distressed individuals are more likely to abuse alcohol and other drugs, to have appetite and sleep disturbances, and to alter the frequency and duration of customary exercise compared to similar nondistressed individuals (Gregory & Smeltzer, 1983). These and other related issues are addressed below.

ALCOHOL AND OTHER DRUGS

Individuals who are depressed and/or anxious are more likely to self-medicate (ineffectively) with alcohol and other drugs (Grunberg & Baum, 1985). Substance abuse has direct adverse effects on immune function (Jaffe, 1980), as well as indirect effects via alterations in nutrition (Chandra & Newberne, 1977). Moreover, alcohol abuse can potentiate or enhance distress (Grunberg & Baum, 1985). Thus, if an individual is abusing alcohol and/or other drugs, it is probably not possible to separate the immunological consequences of substance abuse from any psychological mediation of immune function. It is not clear what level of alcohol and/or other drug use will have significant immunological consequences. However, it is safer to make a more conservative estimate of possible overuse, and exclude subjects on that basis.

There are several different methods for evaluating alcohol and drug consumption. The simplest method involves obtaining subjects' self-reports of alcohol/drug consumption (Babor, Stephens, & Marlatt, 1987). In order to provide a fairly conservative estimate of alcohol use, we routinely exclude data from individuals who report drinking 10 or more drinks a week, or who reports that they have had 10 or more drinks in the past week using the standard equivalents for alcohol, e.g., the alcohol content of five ounces of wine equals that in 12 ounces of beer or one ounce of whiskey. A relatively conservative exclusionary standard is preferable because of some evidence that substance abusers are likely to significantly underestimate intake (Grunberg & Baum, 1985).

There are a number of scales available that assess alcohol abuse within a clinical range, focusing on the extent to which alcohol and/or other drug use interferes
with work or home life (e.g., Selzer, 1971). Individuals with drug-related performance interference should certainly be excluded from most behavioral immunology studies. However, some individuals can drink a fair amount of alcohol without actually showing impairments that are sufficient to meet clinical criteria; for this reason it is important to ask how much an individual drinks in an average week, as well as how much an individual reports drinking in the past week.

To increase the investigator's level of certainty about alcohol use, the use of a brief questionnaire on alcoholism (Selzer, 1971) can be valuable. Moreover, use of the alcohol clinical index developed by Skinner, Holt, Shen, and Israel (1986) in combination with laboratory tests determining mean corpuscular volume and \( \gamma \)-glutamyltransferase activity (Skinner, Holt, Schuller, Ray, & Israel, 1984) can discriminate 88-97% of alcohol abusers (Skinner et al., 1986).

To detect relatively recent drug and alcohol use, individuals can be asked to provide urine for use in a drug screen. Clinical laboratories have such tests readily available, although the cost is variable and often high when multiple tests are requested. However, use of such tests can provide a very high level of certainty for exclusion of subjects whose recent drug/alcohol consumption may have consequences for immunity. Moreover, informing subjects that a urine screen will be used can itself have significant consequences for self-report data. In one study with pregnant women, a relatively high-risk group for alcohol use, subjects reported about twice as much alcohol consumption when informed that their urine would be tested (Lowe, Windsor, Adams, Morris, & Reese, 1986).

**NUTRITION**

It is important to assess the adequacy of subjects' nutrition, since there is excellent evidence that poor nutrition is associated with a variety of immunological impairments including cell-mediated immunity, phagocyte function, complement system integrity, and mucosal immunity (Chandra & Newberne, 1977). In addition, more distressed individuals often have appetite disturbances that can affect nutrition (Gregory & Smeltzer, 1983), so that any relationships between distress and immune function for these individuals could potentially be a function of underlying nutritional deficits.

The easiest and least expensive method for assessing recent nutritional adequacy of recent nutrition is to ask about recent weight changes. At a minimum, this information should be included in any study with human subjects in this area. However, while some individuals characteristically overeat during times when they are distressed (Gregory & Smeltzer, 1983), the nutritional adequacy of these subjects' diets may still be a problem. Thus, a secondary self-report step may be the inclusion of a food diary, reviewed by a qualified nutritionist. The diary should span at least 2 weeks (and preferably the prior month) because of the relatively slow rate for immunological changes. The diary should be prospective, because of the known problems with retrospective memory (Bradburn, Rips, & Shevell, 1987).

However, perhaps the simplest assessment method that provides the greatest assurance of nutritional adequacy is the inclusion of one or more biochemical nutritional assays for each subject. Our laboratory routinely uses one or two
plasma protein assays, generally with a shorter and longer half-life, so that we have information for both recent and somewhat longer term changes in nutrition. Protein assays provide better information and global nutritional status than those for carbohydrates and fats, since the former have varied nutritional building blocks, as well as very complex pathways for synthesis. Moreover, moderate to severe protein-caloric malnutrition is clearly associated with increased frequency and severity of infection (Chandra & Newberne, 1977). Different protein markers can provide evidence for more recent or less recent dietary insufficiency; e.g., the half-life of albumin is 2 to 3 weeks, in comparison to 8 days for total iron binding protein and transferrin. In particular, serum albumin levels are widely used as indicators of nutritional status; while earlier studies suggested age-related changes associated with serum albumin, the age-related decline in healthy subjects is not substantial (Campion, deLabry, & Glynn, 1988). When individuals fall out of normal range on one or more of these nutritional markers, their immunological data are discarded. This method provides some assurance that any immunological alterations are not simply reflections of nutritional inadequacy.

It may be important to assess components of diet in individuals as well. Ideally, one would control for monamines prior to phlebotomy, given the potentially important catecholamine effects that could mediate behavioral influences on immunity.

**MEDICATION AND CURRENT HEALTH STATUS**

In order to maximize confidence in immunological data, subjects should be unmedicated, and should have no chronic or acute health problems that might have any immunological or endocrinological consequences. Obvious exclusions include individuals who have had recent surgeries, who are pregnant, or who have recently given birth, since the anesthetics that are used and/or surgical trauma are known to adversely affect immune function (Lukomska, Waldemar, Engeset, & Kolstad, 1983); moreover, for the latter, there are pervasive immunological and endocrinological changes (e.g., Adams, 1982). In addition, exclusion of subjects with a cancer history is reasonable, since some cancers have immunosuppressive qualities (Doldi, Leroux, Augustin, Kirchner, & Kalden, 1985). Individuals who currently have (or have recently had) an episode of infectious disease are also potentially excludable, since there are alterations in natural killer cell activity, blastogenesis, and other parameters following viral infections (Lumio, Welen, Hirvonen, & Weber, 1983). At the very least, baseline measures should not be taken when individuals have reported an infectious illness within the last 2 weeks. There are some fluctuations in natural killer cell activity across the stages of the menstrual cycle (Lukomska et al., 1983). Changes in other immunological markers have not been systematically assessed, although it is reasonable to assume that the marked endocrinological fluctuations accompanying different phases of the menstrual cycle could have immunological correlates, given the interrelationships between the endocrine and immune systems (O'Dorisio, Wood, & O'Dorisio, 1985). Ideally, one would only use subjects during the same stage of their menstrual cycles; alternatively, if one is willing to tolerate the error variance, it is reasonable to ask female subjects about their current menstrual cycle stage, and
METHODOLOGICAL ISSUES

The use of subjects who are medicated is particularly problematic when studying the elderly, one of the groups of greatest interest. Some data suggest that as many as 86% of the elderly have one or more chronic health conditions (Krauss, 1980). To exclude such individuals leaves the possibility of systematically biased data.

However, there is evidence that some medications that are common among the elderly have immunological consequences. For example, beta blockers, used to treat hypertension, have been shown to have adverse effects on blastogenesis *in vitro* (Goodwin, Messner, & Williams, 1979), and estrogen is also known to have immunomodulatory effects (O'Dorisio et al., 1985). In our research with the elderly we exclude any individuals who are taking any medications with obvious immunological or endocrinological consequences, with the exception of beta blockers and estrogen supplements; instead, in the case of beta blockers and estrogen supplements, we match experimental and comparison subjects on the presence or absence of each. This method provides some assurance that any differences between groups of subjects are not simply a function of the medications they are taking; however, there may be dose and time effects that this method does not assess, and the effects on immunity could be due to either the illness or the medication, and/or the interaction of the variables.

Goodwin, Shearles, and Tung (1982) examined various immunological parameters in 279 healthy and 24 chronically ill individuals over 65, and in young controls. While they found the expected differences between the young and old subjects on certain aspects of the immune function, there were not differences between two groups of older adults. They suggested that the data supported age-related relationships in immune function, rather than age-associated diseases.

If an investigator simply selects individuals who have no medications, no chronic medical problems, and who are over 70 years of age, the sample is likely to be biased in a systematic fashion. Individuals over 70 years of age who have no medications or chronic illnesses are unlikely to be representative of individuals within their age group (Krauss, 1980). Thus, while they provide a relatively clean group for experimental purposes, the extent to which one can generalize from their data is limited.

SMOKING AND CAFFEINE INTAKE

Cigarette smoking can significantly affect a number of different neuroendocrine systems, and also appears to have adverse effects on some immune functions (Gatchel, Baum, & Krantz, in press). Many neuroendocrine studies eliminate subjects who smoke, in order to decrease error variance among subjects, particularly when assessing catecholamines (e.g., Chang, Richards, Kim, & Malarky, 1984).

Recent research suggests that caffeine can potentiate or intensify physiological effects of psychosocial stress (Lane & Williams, 1985). Moreover, contrary to popular wisdom, regular caffeine use does not necessarily lead to tolerance for interactions between caffeine and stressors. Caffeine has clear effects on cate-
cholamines (Dews, 1984), and also appears to have cardiovascular consequences, including increased forearm blood flow and increased forearm vascular resistance responses to stress (Lane & Williams, 1985). Thus, it is important to attempt to assess recent caffeine intake in subjects. However, this can be a very difficult task, given the wide variability in the amount of caffeine produced by a variety of different coffee-brewing procedures, as well as the variable amounts found in many other foods and over-the-counter medications (Dews, 1984).

**EVALUATION OF ILLNESS VERSUS ILLNESS BEHAVIOR**

One of the major issues of interest within the human behavioral immunology area is the extent to which behaviorally related immunological changes actually lead to alterations in health. There are only a handful of studies with human subjects that have found a confluence among distressing psychological responses, immune function, and actual health changes (Glaser et al., 1987; Kasl et al., 1979; Kiecolt-Glaser et al., 1988; Pennebaker et al. 1988). Infectious disease episodes occur with low frequency among most adults, so that longer study periods are better suited to the identification of more vulnerable individuals.

Another problem is the objective evaluation of illness. It is well known that certain individuals are prone to interpret distress in terms of somatic symptoms, and thus may bias the data (Mechanic, 1974).

There are several ways to evaluate the relationships. Perhaps the most promising work is being done by Sheldon Cohen and colleagues in collaboration with the Common Cold Unit in England. In that study they are systematically infecting subjects with a cold virus (Rhinovirus) or with placebo. They measure specific responses related to severity of infections in subjects, in addition to immune function and psychological variables. Such a design is an ideal way to evaluate the relationship, since they control timing and intensity of exposure to the pathogen.

A secondary way of evaluating the relationship is to obtain physicians' records of illness episodes and the assigned diagnosis with subjects' permission. While the study was not structured to do so, the Kasl et al. (1979) study does provide a more objective documentation of illness related to the variables of interest, as does the Pennebaker et al. (1988) study. Similarly, Baum and his students are using this method in their studies of more chronic stressors (Baum, Schaeffer, Lake, Fleming, & Collins, 1985). However, many individuals do not visit physicians for minor illnesses.

Alternative methods for providing some assurance of actual illness have been provided by other researchers. Rose, Jenkins, and Hurst (1979) have shown that the monthly questionnaire they developed for their air traffic controller study can be used by medical personnel to establish a diagnosis. The virtues of their method include specificity (actual illness can be separated somewhat more easily from illness behavior) and the frequency of assessment (self-reports are not diminished by forgetting). Moreover, they simply counted number of episodes, not the length of the episode, which they felt provided further control for illness behavior.

A broader related issue in evaluating self-reports of health or other behaviors concerns the accuracy of self-reports that span periods of several months or more. Studies on recall of events suggest that recall is poor past a period of a month or
two (Bradburn et al., 1987). However, there are a number of things that can be
done to prompt subjects' recall of such important events. An outstanding sum-
mmary of relevant interview issues is provided in a recent article (Bradburn et al.,
1987).

SLEEP

Distressed/depressed individuals can show a number of different kinds of sleep
disturbances, including early morning awakening, sleep onset insomnia, and
middle insomnia (wakening in the middle of the night for an hour or more)
(Gregory & Smeltzer, 1983). While there is some limited evidence that very
gross alterations in sleep (e.g., sleep deprivation for 2 to 3 days) can modulate
immune function (Palmblad, 1981), it is not known whether the less severe sleep
disturbances associated with distress can lead to immunological alterations.
However, it is likely that sleep disturbances could have a number of different
effects; dramatic increases in IL-1 activity as well as other changes in immune
functions have been shown to be related to the onset of slow wave sleep
(Moldofsky, Lue, Eisen, Keystone, & Gorczynski, 1986).

In order to evaluate the relationship of distress-related sleep disturbances to
immunological alterations, it may be helpful for investigators to begin to system-
atically collect data. For example, in our medical student studies we routinely
ask how much sleep students have had in the prior 3 nights, so we can compare
total sleep for the 3 nights preceding lower-stress baseline periods with sleep
before and during their examination periods. While we always find reliable
differences, the absolute amounts are not generally large and do not reliably
correlate with immunological alterations.

The best objective assessment of sleep quantity and quality can be obtained in
a sleep laboratory, a procedure that is too expensive and time consuming for
most studies such as these. However, there are several brief scales developed to
assess sleep disturbances that may be useful for investigators in this area (Hoch,
Reynolds, Kupper, Berman, Houck, & Stack, 1987; Jenkins, Stanton, Niemcryk,
& Rose, in press).

PHYSICAL ACTIVITY

Relatively strong and consistent associations have been found between health
and physical activity (LaPorte, Montoye, & Caspersen, 1985). While some of the
strongest evidence is within the cardiovascular arena, there is also growing evi-
dence that physical activity may have both immunological and endocrinological
consequences (Simon, 1984). For example, Temoshok, Zich, Solomon, Stites, and
O'Leary (1987) showed a relationship between certain aspects of immunity and
exercise/fitness in HIV-infectioned men. Similarly, women who run 10 miles a
week or more have substantially different 24-h endocrine profiles than women
who exercise less vigorously (Chang et al., 1984).

A recent study suggests that some of the differences seen in endocrine and
immune function related to aerobid fitness may be associated with the rapidity of
recovery from psychological and physiological stressors. Subjects with high aer-
obic fitness showed lower heart rates, a faster recovery from subjective anxiety,
and a more efficient recovery for relative plasma epinephrine measures in response to a psychological stressor, compared to low fit subjects (Brooke & Long, 1987).

While these and other data suggest that physical activity and aerobic fitness have immunological and endocrinological correlates, there has been little systematic assessment. This is hardly surprising, because there is no instrument that assesses physical activity accurately and reliably, while not affecting the behavior being measured (Laporte et al., 1985). The instruments that are the most accurate tend to be quite impractical on a population basis (Laporte et al., 1985).

If there are wide variations in level of physical activity within subject samples, it may be fruitful to use one of the questionnaires described by Washburn and Montoye (1986) that were developed for this purpose. At the very least, such information might be useful in partialling out some of the potential error variance attributable to differences in exercise.

**CHOICE OF CONTROL/COMPARISON GROUPS AND THEIR PSYCHOMETRIC ASSESSMENT**

There are a number of studies in which psychiatric patients' values on immunological assays have been compared with those of nonpsychiatric subjects, particularly studies addressing herpesvirus antibody titers. Certain psychiatric patient subgroups, especially those who are more depressed, have significantly higher herpesvirus antibody titers than nonpsychiatric controls; no differences between the patient and control groups have been found when other viral antigens such as measles or rubella were used (e.g., Halonen, Rimon, Arohonka, & Jantti, 1974; Lycke, Norrby, & Roos, 1974). Based on these findings, some researchers have suggested that the herpesviruses may have an etiologic role for certain psychiatric disorders. However, we and others have been able to show stress-related changes in antibody titers to various herpesviruses, including herpes simplex virus, cytomegalovirus, and Epstein-Barr virus in medical students and other populations (Glaser et al., 1985x; Glaser et al., 1987; Kasl et al., 1979; Kiecolt-Glaser et al., 1987x, b). Thus, a more parsimonious explanation for the psychiatric patient data may be related to the higher levels of distress that are characteristic of most hospitalized psychiatric patient populations compared to their nonpsychiatric community counterparts.

However, several other laboratories have done similar studies, but have failed to find significant differences (DeLisi, Smith, Hamovit, Maxwell, Goldin, Dingman, & Gershon, 1986). In such studies, one issue that has not received much attention is the level of distress in comparison subjects. Since normal (i.e., no psychiatric diagnosis) individuals can show elevated herpesvirus antibody titers as well as other immunological changes following both minor and major stressful life events (Glaser et al., 1985x; Glaser et al., 1987; Kasl et al., 1979; Kiecolt-Glaser et al., 1987x, b), it becomes important for investigators who are interested in differences between psychiatric patients and nonpsychiatric comparison subjects to make sure that the nonpsychiatric comparison subjects are not distressed. None of the above studies using psychiatric patient samples (DeLisi et al., 1986; Halonen et al., 1974; Lycke et al., 1974) evaluated level of distress in the non-
psychiatric portion of the sample. In fact, in the DeLisi et al. (1986) study, the comparison subjects were people who worked in the psychiatric unit. Such a job can be reasonably stressful at times, and may not provide a fair evaluation of actual differences between psychiatric patients and comparison subjects.

TIMING OF BLOOD SAMPLES AND LABORATORY ANALYSES

In order to control for possible diurnal variations in immunological parameters, it is important to take all blood samples from a given population within the same time period. The exact time of day when blood samples are drawn is probably not critical, as long as blood samples for all subjects in a study are drawn within the same 1- to 2-h period. Thus, at the beginning of the study, the time period needs to be set for the remainder of the study.

A related issue is the amount of time that blood samples are allowed to sit before the assays are run in the laboratory. Fletcher, Baron, Ashman, Fischl, and Klimas (1987) have shown that sample storage time can have significant effects on some immunological parameters. This becomes a particularly important issue when one is sending samples to a commercial laboratory, since laboratories may run samples immediately on some days, while waiting up to 24 h (or more) before running them on other days. Although an investigator may have no other alternative than the use of a commercial laboratory, the error variance resulting from differences in storage time is likely to be much greater in commercial laboratories.

In the ideal study, blood would be collected from all experimental and control or comparison subjects at precisely the same time. The realities of research normally make such a plan impossible. When samples are collected on multiple days from groups of subjects who are hypothesized to differ on some characteristic, it is important that blood samples from the different subject cohorts be run simultaneously (e.g., rather than assaying samples from depressed subjects on one day and comparison subjects on the next day). Schleifer, Keller, Bond, Cohen, and Stein (1987) have shown that the day on which samples are gathered and analyses done can account for as much as 85% of between-group differences. They show that there are high correlations due to what they term the "measurement occasion," the day on which samples are obtained. Thus, in order to avoid systematic bias, it is important to intermingle subjects from various groups when samples are collected.

LABORATORY SUPPLIES

In the beginning of a particular study, it is wise to buy sufficient quantities of laboratory supplies for the entire study if at all possible (e.g., mitogens, fetal bovine serum, media, plasticware, etc.). While this suggestion may seem rather trivial, it can have enormous consequences for laboratory data in a given study. For example, we found a 10-fold difference in the relative values obtained for γ-interferon using different lots of concanavalin A (Con A) in two studies in which lymphocytes were stimulated with Con A to produce interferon (Glaser et al., 1987; Glaser, Rice, Speicher, Stout, & Kiecolt-Glaser, 1986). Thus, if an inves-
tigator were to buy different batches of mitogen without attempting to examine their relative potencies, one could easily show remarkable artifactual changes over time.

**BEHAVIORAL IMMUNOLOGY RESEARCH IN THE BEST OF ALL POSSIBLE WORLDS ...**

With unlimited laboratory personnel, unconstrained subject access, and limitless funding, an ideal study would include multiple immunological samples from the same subjects over a period of several days for each data point. It would also include multiple psychological measures over a 2-week period or more, since most immunological functions change over days or weeks, so that some of the relevant psychological changes may precede immunological changes by several days or more. The subjects in such studies would be taking no medications, they would be in perfect health, they would all have a uniform physical activity level, and they would, of course, be remarkably compliant with all of these procedures.

However, the reality for researchers who are working in this area is normally quite different from this ideal scenario. In fact, because the most interesting and potentially most important differences are related to differences in affect, all the other issues described related to affect in this paper come into play. The most difficult subject population in this respect is HIV-seropositive individuals, who provide a very important area for inquiry. The multiple problems that such subjects pose can include self-medication with a variety of street drugs, megadoses of vitamins, multiple drug regimens, disordered sleep and nutrition due to both the distress associated with the illness as well as changes in those parameters that are often associated with simply being ill, and wide variance in physical activity levels. For these and other difficult subject populations, it becomes especially important to evaluate the parameters described in this paper that have relevance for immune function. By making such assessments a routine part of the protocol, it may be possible to statistically control some of the error variance due to these factors, thus providing a better understanding of CNS-immune system interactions.

**REFERENCES**


Kiecolt-Glaser, J. K., Glaser, R., Williger, D., Stout, J., Messick, G., Sheppard, S., Ricker, D.,


Received November 23, 1987