Research in Endorphins and Schizophrenia*

Glenn C. Davis, Monte S. Buchabaum, and William E. Bunney, Jr.

Abstract

It has been suggested that the newly discovered endogenous opiate peptides (called endorphins) might play a role in the symptoms of schizophrenia. The administration of narcotic antagonists provides both a test of the hypothesis and a potential treatment. In this article, we review the methods by which data have been gathered to test endorphin involvement in schizophrenia. Alternative strategies, which hold greater promise of producing conclusive positive or negative evidence, include exploitation of individual differences, use of psychophysiological measures, genetic strategies, and multivariate statistical techniques with larger sample sizes.

The simplest endorphin hypothesis of schizophrenia suggests that the symptoms directly result from an excess of these opiate-like peptides and implies that opiate (and hence endorphin) antagonists could not only be used as a test of the hypothesis but also provide effective treatment. While Volavka, Davis, and Ehrlich (1979) implicate a relationship to the dopamine system as the potential mechanism, a number of methodological problems apply to all biochemical variations on this theme.

In seeking biological markers for an etiology of schizophrenia, investigators face a number of difficulties: Schizophrenia is likely an illness with a variety of etiologies; a genetic factor may be widely prevalent and only expressed clinically in a proportion of those liable; environmental influences have a variable impact; and both the genetic vulnerability and environmental influences probably interact with the biology of the developmental process, especially during adolescence. Much of the study of schizophrenia has of necessity proceeded by careful identification and classification of symptoms. Complex behaviors such as the symptoms of schizophrenia are no doubt the expression of a multiplicity of neuronal events. Symptoms may be the final common expression of various etiologies. Further, classical diagnostic symptom clusters have been developed on theoretical bases not related, for example, to the cluster of behavioral effects of the opiates. Thus our careful attention to symptoms may be misleading and not yield biologically homogeneous subgroups. In this article we discuss many of these recent studies from a methodological viewpoint, including problems of the pharmacology (duration of action, concurrent neuroleptics and dose-response issues), behavioral assessment (rating scales), individual differences in response, and statistical analysis of small samples. Finally, we suggest alternative methodologies for testing the various endorphin hypotheses of schizophrenia.

Review of Studies

If an excess of endorphins were responsible for some of the symptoms of schizophrenia, then narcotic antagonists might be expected to bring about prompt relief in these symptoms. Narcotic antagonists such as naloxone or naltrexone act by blocking opiate receptors in the brain. Naloxone and naltrexone have been shown in vivo and in vitro to block and reverse the effects of endorphins (beta-endorphin and leucine- and methionine-enkephalin). This pharmacological strategy has...
been used in clinical studies in an attempt to link schizophrenia to endorphin abnormalities.

At least seven groups have used the short-acting parenterally administered antagonist naloxone in schizophrenics and three groups have given the longer acting naltrexone in clinical trials lasting from 2 to 8 weeks (table 1). In each of these studies, only a small number of schizophrenic patients have been used and several diagnostic criteria applied. Behavioral rating instruments such as the Brief Psychiatric Rating Scale (BPRS) or symptom specific scales (100mm global hallucination self-rating lines) have been commonly used as outcome measures. A minority of the studies were double-blind and counterbalanced with similar placebo procedures. Only 4 of 10 reported statistical tests, while 6 of 10 actually reported group data. The failure to report data or statistical tests is perhaps the rule rather than the exception in publication of negative results, a point we discuss later.

Acute vs. Chronic Drug Administration. Failure to demonstrate improvement with naloxone does not rule out a role for endorphins in schizophrenia. The behavioral effects of such an excess may be the consequences of subsequent actions of endorphins and not easily modified in the acute experiment. Possible therapeutic effects might depend upon changes, homeostatic compensations, or drug actions that take time to develop (Davis, Buchsbaum, and Bunney, in press). Most of the therapeutic agents used in psychiatry (tricyclic antidepressants, lithium, neuroleptics) require more than a single dose and more than 1 day for their target effects. This point is further complicated by the fact that drugs such as tricyclic antidepressants are thought to act by blockade of aminergic uptake, which occurs immediately, although clinical improvement may not be apparent for 3 to 4 weeks. Thus the failure of an acute antipsychotic efficacy is a weak test of a compound’s therapeutic value. A few studies have used the orally administered narcotic antagonist naltrexone (table 1), which provides continuous receptor blockade with single day dosing. All three studies were single-blind and, appropriate to the small sample size, no statistical analysis was reported. Longer trials of naltrexone with larger samples will be necessary to address the endorphin hypothesis.

Drug Free vs. Stabilized on Neuroleptics. In the earliest report by Gunne, Lindstrom, and Terenius (1977) patients were stabilized on various neuroleptics and then concurrently administered the narcotic antagonist naloxone. Some of the subsequent replication attempts were done with the patients drug-free and others with the patients on maintenance neuroleptics (table 1). This situation presents two problems, clinical and experimental. Frequently clinical investigators presented with a new pharmacological approach are not prepared to mount a full double-blind placebo-controlled trial. Such a design requires weeks of inpatient care of medication-free patients for acute tests (e.g., single dose naloxone) and months for chronic trials (e.g., naltrexone), as well as possible delay in returning the patient to the community. A compromise is the result; the patient is administered the test drug in an additive trial, superimposing the new treatment on the ongoing regime. Experimentally, we are faced with a different problem; the interaction between the endorphin and dopamine systems might be responsible for any observed effects, since the initial positive studies were all neuroleptic-opiate antagonist combinations. An inevitable disadvantage of the additive trial is the difficulty in assessing the significance of results in the absence of a trial of naloxone alone, placebo, neuroleptic alone, and neuroleptic plus naloxone in a crossover design. In addition patients of investigators using the additive trial design are characterized on a variety of types and doses of neuroleptics, which further complicates the interpretation of results.

Behavior Rating Measures. One of the major stumbling blocks to effective study of psychiatric symptoms and psychiatric illnesses in general has been the difficulty of behavioral measurement. Even the best measures have problems with reliability, reproducibility, and lack of specificity. Measures tend to be of two types: global measures of severity of psychosis (Bunney and Hamburg 1963) or dimensional, semi-quantitative measures such as used in the Brief Psychiatric Rating Scale (BPRS) (Overall and Gorham 1962). These measures are useful in longitudinal studies in which random changes in ratings and variables influenced by rater variability may be minimized by averaging over days and weeks to more accurately reflect symptom improvement or deterioration. However, these measures have less reliability for repeated use in acute studies (e.g., BPRS every ½ hour for 3 hours). Items such as somatic concern, guilt feelings, grandiosity, and hostility that are based solely on patient reports are particularly difficult to rate frequently since even very ill patients may not spontaneously report these feelings every 30 minutes. A nonspecific decrease in verbalization may spuriously produce the
Table 1. Trials of naloxone and naltrexone in schizophrenic patients

<table>
<thead>
<tr>
<th>Study</th>
<th>Dose (mg)</th>
<th>Test period</th>
<th>Additive trial(^1)</th>
<th>Design(^2)</th>
<th>Outcome measures(^3)</th>
<th>Statistics</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Naloxone</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gunne, Lindstrom, and Terenius 1977</td>
<td>6</td>
<td>.4</td>
<td>8 hr</td>
<td>Yes</td>
<td>SB,PC</td>
<td>Hallucination self-rating</td>
<td>None</td>
</tr>
<tr>
<td>Volavka et al. 1977</td>
<td>7</td>
<td>.4-1.2</td>
<td>24 hr</td>
<td>Yes</td>
<td>DB,PC,RO</td>
<td>BPRS</td>
<td>None</td>
</tr>
<tr>
<td>Davis et al. 1977</td>
<td>14</td>
<td>.4-10</td>
<td>1 hr</td>
<td>Mixed</td>
<td>DB,PC,RO</td>
<td>BPRS,NIMH</td>
<td>Sign test, t-test</td>
</tr>
<tr>
<td>Janowsky et al. 1977</td>
<td>8</td>
<td>1.2</td>
<td>1 hr</td>
<td>Yes</td>
<td>DB,PC,RO</td>
<td>BPRS</td>
<td>ANOVA(^4)</td>
</tr>
<tr>
<td>Kurland et al. 1977</td>
<td>12</td>
<td>1.8</td>
<td>—</td>
<td>Yes</td>
<td>DB,PC,RO</td>
<td>Hallucination rating</td>
<td>None</td>
</tr>
<tr>
<td>Emrich et al. 1977</td>
<td>20(^5)</td>
<td>1.2-4</td>
<td>6 hr</td>
<td>Mixed</td>
<td>DB,PC,RO</td>
<td>IMPS, VBS</td>
<td>t-test</td>
</tr>
<tr>
<td>Watson et al. 1978</td>
<td>11</td>
<td>10</td>
<td>2 hr</td>
<td>Mixed</td>
<td>DB,PC,RO</td>
<td>BPRS,NIMH</td>
<td>ANOVA(^4)</td>
</tr>
<tr>
<td><strong>Naltrexone</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gitlin and Rosenblatt 1978</td>
<td>3</td>
<td>50-100</td>
<td>2 wk</td>
<td>Mixed</td>
<td>SB</td>
<td>BPRS</td>
<td>None</td>
</tr>
<tr>
<td>Mielke and Gallant 1977</td>
<td>5</td>
<td>50-250</td>
<td>6 wk</td>
<td>No</td>
<td>NB</td>
<td>CGI,BPRS,NOSIE</td>
<td>None</td>
</tr>
<tr>
<td>Simpson, Branchley, and Lee 1977</td>
<td>6</td>
<td>800</td>
<td>8 wk</td>
<td>No</td>
<td>SB</td>
<td>CGI,BPRS,NOSIE</td>
<td>None</td>
</tr>
</tbody>
</table>

\(^1\)Concurrent neuroleptic treatment.

\(^2\)SB = single blind; DB = double blind; PC = placebo controlled; RO = random order; NB = nonblind.

\(^3\)BPRS = Brief Psychiatric Rating Scale; NIMH = National Institute of Mental Health Scales; CGI = Clinical Global Impression Scale; NOISE = Nurses Observation Scale of Inpatient Evaluation; VBS = Verlauf's-Beurteilungs-Skala; IMPS = Inpatient Multidimensional Psychiatric Scale.

\(^4\)Analysis of variance.

\(^5\)Not all patients schizophrenic.
appearance of improvement. With a structured interview technique in which specific probe questions are asked, repeated questioning over short intervals may artificially elevate positive responses. Such high rating levels may not produce a systematic advantage for drug or placebo, but the higher variances created may diminish the power of statistical comparison.

The absolute values of ratings often have little meaning in themselves. The worsening of "unusual thought content" as rated by the observer (e.g., Davis et al. 1977) might reflect an increasing opportunity to observe such abnormal content and thus only an apparent deterioration of symptoms. If the data are examined carefully, we discover that "unusual thought content" did not show a significant rating improvement when the naloxone treatment day is considered alone. The statistical effect is due to slight improvement on naloxone paired with deterioration of ratings on placebo. Does this mean (if replicated) that "unusual thought content" gets worse as the day goes on, or that placebo injections, for whatever reason, bring about deterioration? In acute studies many rating measures may suffer from this difficulty—namely the probability of observing a behavior increases with the duration of time it is measured, especially if investigators use unstructured interview paradigms. Nevertheless, although the degree of improvement is not interpretable, the double-blind placebo-controlled design still allows us to conclude that a salutary naloxone effect occurs.

The end-points of self-rating behavioral scales and what they suggest to the patient present other problems. Gunne, Lindstrom, and Terenius (1977) were the first to suggest that naloxone improves auditory hallucinations. This group used a 4-point intensity scale for the patients to rate their hallucinations, which ranged from "no voices" to "usual amount and intensity." The maximum hallucination rating of "usual amount and intensity" presents the patient with the examiner's bias—namely that it is not possible for the drug to increase hallucinations—and suggests that the examiner is looking for a reduction in this symptom.

Ethological measurement techniques have had only limited application in psychiatric illness. For example, psychomotor behavior has been measured by an activity monitor attached to extremities (Colburn et al. 1976). Some investigators use videotape or audiotape segments to rate speech production and movement disorders. In spite of the most thorough attempts to "objectify" the measurement of symptoms, some symptoms have relied on subjective reports that have no known correlates with measured biological variables, e.g., hallucinations. Nevertheless, strategies should be sought to assess objectively such subjective reports. One possible psychophysiological model of hallucinations was suggested to us by recent reports of obtaining auditory evoked potentials to a missing tone in a sequence (Simpson, Vaughan, and Ritter 1976; Sutton et al. 1967). Pharmacological manipulations using this model have not yet been tested.

As opiates and endorphins have been shown to play an important role in pain appreciation, the measurement of pain sensation with psychophysical, psychophysiological, and behavioral tests could provide a useful index of endorphin effects. The finding of pain insensitivity in schizophrenia for clinical conditions (Marchand 1955; Marchand et al. 1969) and experimental pain (Hall and Stride 1954; Malmo and Shagass 1949; Malmo, Shagass, and Smith 1951; Sappington 1973) not only supports the use of this approach but supports the endorphin hypothesis as well (Davis, Buchsbaum, and Bunney in press). Several major advances in the assessment of pain appreciation have improved our ability to study possible abnormalities. Electrical stimuli may be used in the experimental situation to test for pain sensitivity. These stimuli are safe, brief, easily controlled both for delivery and intensity over a wide range, and highly reproducible from session to session. The concentric-ring electrode, developed by Turskey and Watson (1964), provides a standardized means for delivering shocks, increases control over skin current flow, and is well tolerated by both normal volunteers and psychiatric patients. Despite this advance, the ascending threshold paradigm commonly used to measure pain is contaminated with a variety of sensory, cognitive, and social variables. When a subject is asked to report at what point a series of shocks of increasing intensity becomes painful or unbearable, in certain experiment/patient pairings, the subject’s response may be influenced by the social context of the experiment, while in others, sensory factors may prove crucial. Thus the single threshold value contains variance from many complex sources. This problem has been somewhat alleviated by the introduction of signal detection analysis (Clark 1969), a putative technique for separating sensory factors from more complex behavioral mechanisms. The use of electrical stimuli is also especially suitable for the recording of cortical evoked potentials. Somatosensory evoked potentials are reduced in amplitude in schizophrenia (see review by Buchsbaum 1977), a finding that has been associ-
Statistical Analysis and Individual Differences. Many researchers currently believe, like Bleuler (1950), that schizophrenia is a group of illnesses and therefore have attempted to use strategies to search for subgroups. Both Volavka et al. (1977) and Watson et al. (1978) used individual difference strategies to isolate nalamoxone-responsive schizophrenic patients. Volavka et al., following their placebo-naloxone crossover trials, studied the responders in a second series of injections. This is an extremely valuable method for enhancing the proportion of responders in a clinical trial in which there is no clearly known clinical indicator or response other than that to the test drug. Furthermore this screening strategy may be useful in identifying individuals who will respond. However, if the proportion of responders (schizophrenics with “endorphin excess disease”) is small, then the chance of finding one in a group of seven patients becomes remote. The small sample size problem is even worse if the drug response measure (e.g., BPRS as discussed earlier) is highly variable and a significant number of false-positive and false-negative responders are identified.

Watson et al. (1978) screened over 1,000 clinical cases to find 11 patients with very frequent auditory hallucinations, suggested by Gunne, Lindstrom, and Terenius (1977) as a key target symptom. This strategy yielded a significant improvement on nalamoxone. Such an experimental approach is risky for the investigator, however, since a small sample size is vulnerable to a false-negative result (termed “high beta-error”). In fact, the efforts of this group nearly failed. They used a design in which pre- and post-injection hallucination ratings from drug and placebo days were compared, using a two-way analysis of variance (A:IOVA). They calculated a significant F value of 5.62, barely passing the F = 5.59, p< .05 level. A paired t-test comparing the placebo and naloxone pre-post difference scores did not reach significance (t = 2.29, but p< .05 for 2.31). This near-miss by t-test is not surprising. Their mean placebo to naloxone pre-post difference was only 1.88 units of hallucination improvement with a standard deviation of 2.41. Using only the t-test, how confident should these investigators be in concluding either that naloxone does or does not influence hallucinations? Dividing the change by the standard deviation (SD) yields a standard change of .78 SD units. From tables (Beyer 1968, p. 287) of the number of observations for a test on a single mean, the nine-patient sample size in the Watson et al. study had exactly a 50 percent chance of yielding a false-negative result. This means that of 100 investigative teams using nine patients each, only 50 would find a p < .05 t-test result. Watson et al. needed 22 patients to be 95 percent sure of obtaining a statistically significant result (p< .05, two-tailed) given a true effect of .78 SD units. No sample in table 1 is this large; in fact, 6 out of 10 have fewer than 9 subjects and thus less than a 50 percent chance of obtaining a statistically significant result if the magnitude of the true naloxone effect and variability are equal to that reported by Watson et al. Thus, quite apart from the question of whether naloxone or naltrexone are really effective treatments for psychotic symptomatology, from a statistical standpoint it is not at all surprising that the majority of the 10 studies conducted are negative. Publication of raw data in small N studies is helpful since it allows subsequent investigators to combine studies and ask additional questions.

New Strategies

Based on our comments above, we suggest a series of strategies that may be useful individually or in conjunction with each other.

Single Trial Screening. If a small proportion of symptom responders to naloxone is expected, the use of a double-blind placebo-controlled trial to identify them is inefficient because the sample is heavily weighted with nonresponders. Psychophysiological response to a single dose of a rapid acting agent (e.g., I.V. naloxone) may function as a screening test for subsequent clinical effects. Responders may then be admitted to a double-blind, placebo-controlled crossover trial.

Biological Indicator Screening. Drug response (as used by Volavka et al. 1977) and symptom pattern (as used by Watson et al. 1978) are not the only screening variables possible. Biological and psychophysiological indicators of endorphin hyperactivity may be potent markers. Possible candidates for such an approach include individual measures of pain insensitivity, altered respiratory rates, diminished pupillary diameter, evoked potential measures of somatosensory function, and plasma and CSF endorphin levels as such assays become available. These objective measures may be used in conjunction with the single trial screening strategy to assess funct-
tional endorphin status and, perhaps, to generate a sample of schizophrenics with "endorphin excess disease." Such a strategy will also tend to mitigate the problem of lack of placebo.

Psychophysiological Models. The abnormal arousal and disordered perception seen in schizophrenia lend themselves to laboratory investigation using psychophysiological techniques. Much information about arousal and perceptual processes has been obtained from recordings of autonomic and electrophysiological variables in schizophrenics, but these techniques have been little used in psychopharmacological studies. Psychophysiological measurements can be made unobtrusively, repeated at brief intervals, and scored uniformly and automatically, making them especially suitable for acute drug infusion studies.

Genetic Strategies. Once a biological or behavioral indicator of endorphin dysfunction has been found, the question remains whether it is an etiologically significant trait variable, a homeostatic response to psychosis, or an artifact of neuroleptic treatment, hospitalization, or some other concomitant of chronic psychiatric illness. If patients can be identified using the screening strategies described above, their relatives can be examined and compared with relatives of other patient populations. Several alternative results are informative. If only the psychiatrically ill relatives of "endorphin dysfunction" probands also have the biological abnormality, credence is produced for a genetic and etiological role for endorphins in this subgroup of schizophrenics. The implications of this and other genetic patterns are discussed elsewhere (Buchsbaum and Haier 1978; Rieder and Gershon 1978).

Statistics. Because of the small sample size of studies done to date, powerful statistical tools for coping with highly variable behavioral measures have not been used. Multivariate techniques, such as multiple analysis of variance, Hotelling's $T^2$, and factor analysis, can handle a host of behavioral measures. Variables such as age, sex, and race, which might affect the magnitude of drug effect but are irrelevant to the effect under study, may be removed by analysis of covariance.

Conclusion

In this review we have not directly addressed the question of the validity of the endorphin hypothesis, but rather focused on the methods by which data have been gathered. A series of small sample size studies, each failing to reach statistical significance, increases the danger of prematurely rejecting the endorphin hypothesis. This premature rejection is further exacerbated by the likelihood that a multiplicity of biological factors are responsible for the group of schizophrenics. Burgeoning research on endorphins and other peptides clearly holds promise for understanding a variety of important behavioral mechanisms.

The evidence for endorphin involvement in schizophrenia has so far brought in only a scotch verdict; if endorphins are indeed at fault in this syndrome, creative prosecution is needed. This creativity may arise from the excitement that the discovery of the opiate receptor and opiate peptides and the associated new technology have brought to the neurosciences.

References


The Authors

Glenn C. Davis, M.D., is Research Psychiatrist, Biological Psychiatry Branch, Monte S. Buchsbaum, M.D., is Chief, Section on Clinical Psychopharmacology, Biological Psychiatry Branch, and William E. Bunney, Jr., M.D., is Chief, Biological Psychiatry Branch, Intramural Research Program, National Institute of Mental Health, Bethesda, Md.